



SAPIENZA
UNIVERSITÀ DI ROMA

Laboratorio di Ricerca Operativa

Modelli continui di ottimizzazione
per l'Ingegneria Gestionale

Dispensa didattica con modelli, esempi numerici svolti, casi di studio, implementazioni complete in Python/Gurobi e analisi di sensitività. Programmazione lineare, quadratica e non lineare; decisioni sotto incertezza (Newsvector, VaR e CVaR); ottimizzazione per il machine learning (SVM, regressione robusta e quantile).

Fabio Furini

sites.google.com/view/fabiofurini

Dipartimento di Ingegneria informatica, automatica e gestionale — Sapienza Università di Roma

Versione online e materiale scaricabile:

fabiofurini.github.io/laboratorio-ricerca-operativa/
github.com/fabiofurini/laboratorio-ricerca-operativa

Anno accademico 2026–2027

Indice

I	Strumenti	5
1	Introduzione al laboratorio	6
1.1	Che cosa faremo	6
1.2	Le tre classi di modelli e le sigle usate	6
1.3	Notazione	6
1.4	Una scelta di campo: solo variabili continue	7
1.5	Mappa dei modelli	8
1.6	Struttura di ogni capitolo	8
1.7	Organizzazione del corso e valutazione	8
2	Richiami: LP, QP, NLP e analisi di sensitività	9
2.1	Teoria della programmazione lineare	9
2.1.1	La dualità	9
2.1.2	Le condizioni di ottimalità: gli scarti complementari	12
2.1.3	L'analisi di sensitività e i prezzi ombra	15
2.2	Teoria dell'ottimizzazione non lineare	16
2.2.1	Convessità e programmazione quadratica	16
2.2.2	Le condizioni di ottimalità: le condizioni KKT	17
2.3	Il protocollo di analisi di sensitività	19
3	Il solver: costruire, risolvere, interpretare	20
3.1	I comandi generali	20
3.1.1	Installazione e licenza	20
3.1.2	I cinque passi per costruire un modello	20
3.1.3	Far girare il modello e leggere il log	21
3.1.4	Recuperare la soluzione	21
3.2	I comandi per i modelli lineari	22
3.2.1	Prezzi ombra, costi ridotti e intervalli: gli attributi	22
3.2.2	Interpretare l'output: la checklist	22
3.2.3	Due letture complete	23
3.3	I comandi per i modelli non lineari	24
3.3.1	Vincoli funzionali e termini bilineari	24
3.3.2	Tolleranze e scalatura	25
3.3.3	Quando i duali non ci sono	25

II	Modelli deterministici continui	26
4	Produzione e scorte multiperiodali	27
4.1	Motivazione gestionale	27
4.2	Costruzione guidata del modello	27
4.3	Esempio svolto a mano	29
4.4	Caso di studio	29
4.5	Implementazione	30
4.6	Risultati	30
4.7	Analisi di sensitività	31
4.8	Esercizi	33
5	Supply chain con congestione e sostenibilità	34
5.1	Motivazione gestionale	34
5.2	Costruzione guidata del modello	34
5.3	Esempio svolto a mano	36
5.4	Caso di studio	36
5.5	Risultati	36
5.6	Analisi di sensitività: il prezzo della CO ₂	38
5.7	Esercizi	38
6	Portafoglio di Markowitz	40
6.1	Motivazione gestionale	40
6.2	Costruzione guidata del modello	40
6.3	Esempio svolto a mano	42
6.4	Caso di studio	42
6.5	Risultati	42
6.6	Analisi di sensitività	44
6.7	Esercizi	44
7	Pricing e revenue management	46
7.1	Motivazione gestionale	46
7.2	Costruzione guidata del modello	46
7.3	Esempio svolto a mano	48
7.4	Caso di studio e implementazione	48
7.5	Versione multiprodotto	50
7.6	Esercizi	50
8	Allocazione del budget pubblicitario	51
8.1	Motivazione gestionale	51
8.2	Costruzione guidata del modello	51
8.3	Esempio svolto a mano	53
8.4	Caso di studio e implementazione	53
8.5	Risultati e verifica delle KKT	53
8.6	Analisi di sensitività	54
8.7	Esercizi	55

9	Localizzazione continua di un servizio	56
9.1	Motivazione gestionale	56
9.2	Costruzione guidata del modello	56
9.3	Esempio svolto a mano	58
9.4	Caso di studio	58
9.5	Risultati	59
9.6	Analisi di sensitività: la frontiera efficienza-equità	59
9.7	Esercizi	60
10	Ricarica intelligente di veicoli elettrici	62
10.1	Motivazione gestionale	62
10.2	Costruzione guidata del modello	62
10.3	Esempio svolto a mano	64
10.4	Caso di studio	64
10.5	Risultati	64
10.6	Analisi di sensitività	65
10.7	Esercizi	66
11	Capacità di servizio e tempi di attesa	68
11.1	Motivazione gestionale	68
11.2	Costruzione guidata del modello	68
11.3	Esempio svolto a mano	69
11.4	Caso di studio e risultati	70
11.5	Analisi di sensitività: il prezzo di una promessa	71
11.6	Esercizi	72
III	Decisioni sotto incertezza	73
12	Il Newsvendor e le sue varianti	74
12.1	Motivazione gestionale	74
12.2	Il modello base	74
12.3	Esempio svolto a mano (domanda discreta)	75
12.4	Caso di studio: la regola del quantile in continuo	76
12.5	La formulazione lineare a scenari	76
12.6	Livelli di servizio	78
12.7	Avversione al rischio: media-CVaR	78
12.8	Multiprodotto con budget condiviso	79
12.9	Esercizi	79
13	VaR e CVaR	81
13.1	Motivazione gestionale	81
13.2	Definizioni e proprietà	81
13.3	Esempio svolto a mano	82
13.4	Caso di studio 1: portafoglio mean-CVaR contro Markowitz	83
13.5	Caso di studio 2: supply chain a due stadi	84
13.6	Analisi di sensitività	85
13.7	Esercizi	86

14 Arbitraggio e prezzatura	87
14.1 Motivazione gestionale	87
14.2 Costruzione guidata del modello	87
14.3 La dualità è la prezzatura: le probabilità neutrali al rischio	89
14.4 Esempio svolto a mano	90
14.5 Caso di studio	90
14.6 Risultati	90
14.7 Analisi di sensitività: l'intervallo dei prezzi coerenti	91
14.8 Esercizi	92
 IV Ottimizzazione e machine learning	 93
15 Support Vector Machine	94
15.1 Motivazione gestionale	94
15.2 Hard margin: la geometria	94
15.3 Soft margin	96
15.4 La formulazione duale e i support vector	98
15.5 Kernel: frontiere non lineari da un QP convesso	98
15.6 Classi sbilanciate e costi asimmetrici	99
15.7 Support Vector Regression	100
15.8 Esercizi	100
 16 Regressione robusta e quantile	 102
16.1 Motivazione gestionale	102
16.2 Il modello	102
16.3 Caso di studio: lo storico della panetteria	104
16.4 La lettura duale: i punti di appoggio	105
16.5 Regressione quantile e scorta di sicurezza	106
16.6 Selezione delle caratteristiche con un budget	107
16.7 Esercizi	109
 V Organizzazione del laboratorio	 110
17 Organizzazione del laboratorio	111
17.1 Percorso essenziale in quattro laboratori	111
17.2 Struttura della consegna	111
17.3 Criteri di valutazione	112
17.4 Domande tipiche di discussione	112
17.5 Errori ricorrenti da evitare	112
17.6 Materiali del corso	113

Parte I

Strumenti



Introduzione al laboratorio

1.1 Che cosa faremo

Questo laboratorio insegna a usare la programmazione matematica come strumento di *decisione*, non solo di calcolo. Ogni capitolo parte da un problema gestionale concreto — quanto produrre, dove localizzare un servizio, quale prezzo fissare, quanto rischio accettare — e lo trasforma in un modello di ottimizzazione che viene poi implementato in Python, risolto con un solver e, soprattutto, *interrogato*: quanto vale un'ora di capacità in più? La soluzione resiste se i dati cambiano del 5%? Quale compromesso consiglieremmo a un decisore?

La domanda giusta

Alla fine di ogni esercitazione la domanda non è soltanto “*qual è l'ottimo?*”, ma “*quale decisione suggeriamo e quanto è robusta?*”. Il valore di un modello sta nella qualità della raccomandazione che permette di formulare.

1.2 Le tre classi di modelli e le sigle usate

In tutta la dispensa i modelli appartengono a tre classi, indicate con le sigle standard internazionali:

- **LP** (*Linear Programming*, programmazione lineare): obiettivo e vincoli sono funzioni lineari delle variabili;
- **QP** (*Quadratic Programming*, programmazione quadratica): obiettivo quadratico, vincoli lineari;
- **NLP** (*Nonlinear Programming*, programmazione non lineare): obiettivo o vincoli non lineari generali.

Un problema è *convesso* quando ogni minimo locale è anche globale (definizione precisa nel capitolo 2): per LP è sempre vero, per QP e NLP dipende dalle funzioni, e lo segnaleremo ogni volta. Le altre sigle ricorrenti sono definite al primo uso nel rispettivo capitolo.

1.3 Notazione

In tutta la dispensa vale la stessa convenzione:

- **scalari e indici**: lettere minuscole in corsivo (x_{it} , b_t , λ , i , t);

- **insiemi di oggetti indicizzati dai numeri:** gli oggetti di un modello (prodotti, canali, titoli, veicoli, quartieri, scenari) sono numerati e gli indici corrono su insiemi enumerati esplicitamente, $i \in \{1, 2, \dots, n\}$ — così la notazione del testo è già quella del modello costruito; i *conteggi* sono interi positivi ($n \in \mathbb{Z}_{\geq 1}$);
- **dati numerici:** valori razionali, $\mathbb{Q}_{>0}$ o $\mathbb{Q}_{\geq 0}$ (i dati di un'istanza reale sono sempre numeri di macchina);
- **vettori:** lettere minuscole in **grassetto** ($\mathbf{x}$, $\mathbf{w}$, $\boldsymbol{\mu}$); tutti i vettori sono colonna e $\mathbf{a}'\mathbf{b}$ indica il prodotto scalare;
- **matrici:** lettere maiuscole in **grassetto** ($\mathbf{Q}$, $\mathbf{A}$); $\mathbf{Q} \succeq 0$ significa semidefinita positiva;
- **grafi:** $G = (N, A)$ con nodi N e archi A (gli unici insiemi non enumerati, per loro natura);
- **variabili duali:** π_i , una per vincolo; **costi ridotti:** $\bar{c}_j$; **scarti:** $\bar{s}_i$;
- **soluzioni:** le variabili incognite si scrivono lisce (x_j , π_i); la *barra* indica i valori di una soluzione *ammissibile* ($\bar{x}_j$, $\bar{\pi}_i$, $\bar{s}_i$, $\bar{c}_j$), la *tilde* quelli di una soluzione *ottima* ($\tilde{x}_j$, $\tilde{\pi}_i$, $\tilde{z}$).

Ogni coefficiente e ogni insieme è *definito prima di essere usato*. In ogni modello: (i) le variabili sono introdotte esplicitamente prima della formulazione; (ii) obiettivo e ciascuna famiglia di vincoli hanno un numero d'equazione; (iii) i vincoli di non negatività, che definiscono le variabili, chiudono il modello; (iv) dopo il modello, un elenco descrive obiettivo e vincoli famiglia per famiglia, con il numero di vincoli di ciascuna.

1.4 Una scelta di campo: solo variabili continue

Tutti i modelli di questa dispensa usano **esclusivamente variabili continue**: niente variabili binarie o intere. È una scelta deliberata, con tre motivazioni.

1. **Didattica:** con variabili continue valgono la dualità, i prezzi ombra, le condizioni KKT — gli strumenti che trasformano una soluzione numerica in una spiegazione economica. Con variabili binarie questi strumenti si perdono o si indeboliscono.
2. **Copertura:** una quantità sorprendente di decisioni reali è naturalmente continua: quantità, flussi, quote di portafoglio, prezzi, potenze, coordinate.
3. **Progressione:** la stessa applicazione può passare dalla programmazione lineare (LP) a quella quadratica (QP) e non lineare (NLP), mostrando come cresce l'espressività del modello senza cambiare strumento.

1.5 Mappa dei modelli

Cap.	Applicazione	Classe	Concetto centrale
4	Produzione e scorte multiperiodali	LP / QP	dualità, capacità, smoothing
5	Supply chain e sostenibilità	LP / NLP	flussi, congestione, CO ₂
6	Portafoglio di Markowitz	QP	rischio-rendimento
7	Pricing e revenue management	NLP	elasticità e capacità
8	Budget pubblicitario	NLP conv.	rendimenti decrescenti
9	Localizzazione continua	NLP conv.	efficienza-equità
10	Ricarica di veicoli elettrici	LP / QP	energia, picchi
11	Capacità e tempi di attesa	NLP conv.	capacità-attesa
12	Newsvendor e varianti	LP stoc.	decidere prima di sapere
13	VaR e CVaR	LP	rischio di coda
14	Arbitraggio e prezzatura	LP	la dualità che prezza
15	Support Vector Machine	QP	ottimizzazione per il ML

1.6 Struttura di ogni capitolo

Ogni capitolo applicativo segue la stessa scaletta in otto passi, pensata per essere replicata nelle relazioni di laboratorio:

1. **Motivazione gestionale** — il problema raccontato come decisione aziendale;
2. **Costruzione guidata del modello** — insiemi, parametri, variabili, vincoli e obiettivo, spiegati riga per riga;
3. **Esempio numerico svolto a mano** — un'istanza minima risolta con tutti i passaggi, per capire il meccanismo prima di accendere il solver;
4. **Caso di studio** — un'istanza realistica con dati riproducibili (CSV);
5. **Implementazione Python/Gurobi** — il codice completo, commentato;
6. **Risultati** — l'output del solver tradotto in linguaggio manageriale;
7. **Analisi di sensitività** — prezzi ombra, scenari, trade-off;
8. **Esercizi** — dalla verifica alla piccola estensione di ricerca.

I numeri riportati in ogni capitolo sono generati dagli script della cartella `python/`: eseguendo `python3 esegui_tutti.py` si rigenerano dati, tabelle e figure. Se modificate i dati, i numeri della dispensa vanno rigenerati.

1.7 Organizzazione del corso e valutazione

Il percorso essenziale prevede quattro laboratori (capitolo 17 per il dettaglio): (1) produzione e scorte con dualità e prezzi ombra; (2) Markowitz e frontiera efficiente; (3) modellazione non lineare (pricing o budget); (4) progetto a scelta con presentazione manageriale. La valutazione suggerita pesa: correttezza della formulazione 30%, implementazione e verifica numerica 25%, analisi di sensitività 25%, interpretazione manageriale e qualità delle visualizzazioni 20%.

Richiami: programmazione lineare, quadratica e non lineare

Questo capitolo raccoglie gli strumenti teorici usati in tutta la dispensa, in due sottocapitoli: la *teoria della programmazione lineare* (dualità, condizioni di ottimalità, analisi di sensitività) e la sua *estensione non lineare* (convessità, condizioni KKT). Chiude il protocollo di sensitività usato in ogni laboratorio. Serve anche da riferimento rapido.

2.1 Teoria della programmazione lineare

2.1.1 La dualità

Chiamiamo $M = \{1, 2, \dots, m\}$ l'insieme degli indici dei vincoli e $N = \{1, 2, \dots, n\}$ quello degli indici delle variabili. I *dati* del problema sono i costi $c_j \in \mathbb{R}$ per $j \in N$, i coefficienti $a_{ij} \in \mathbb{R}$ e i termini noti $b_i \in \mathbb{R}$ per $i \in M$. Gli insiemi $M_{\leq}$, $M_{=}$ e $M_{\geq}$ ripartiscono M secondo il verso del vincolo; gli insiemi $N_{\geq 0}$, $N_{\leq 0}$ e $N_{=0}$ ripartiscono N secondo il segno della variabile (non negative, libere, non positive). Le *variabili* sono le x_j del primale e, nel duale, una π_i per ciascuno degli m vincoli del primale.

Il primale (P), forma generale

$$\begin{aligned}
 \text{(P)} \quad & \max \sum_{j \in N} c_j x_j \\
 \text{soggetto a} \quad & \sum_{j \in N} a_{ij} x_j \leq b_i, & \forall i \in M_{\leq}, \\
 & \sum_{j \in N} a_{ij} x_j = b_i, & \forall i \in M_{=}, \\
 & \sum_{j \in N} a_{ij} x_j \geq b_i, & \forall i \in M_{\geq}, \\
 & x_j \geq 0, & \forall j \in N_{\geq 0}, \\
 & x_j \leq 0, & \forall j \in N_{\leq 0}, \\
 & x_j \text{ libero}, & \forall j \in N_{=0}.
 \end{aligned}$$

Il duale (D), forma generale

$$\begin{aligned}
 \text{(D)} \quad & \min \sum_{i \in M} b_i \pi_i \\
 \text{soggetto a} \quad & \sum_{i \in M} a_{ij} \pi_i \geq c_j, & \forall j \in N_{\geq 0}, \\
 & \sum_{i \in M} a_{ij} \pi_i = c_j, & \forall j \in N_{=0}, \\
 & \sum_{i \in M} a_{ij} \pi_i \leq c_j, & \forall j \in N_{\leq 0}, \\
 & \pi_i \geq 0, & \forall i \in M_{\leq}, \\
 & \pi_i \geq 0, & \forall i \in M_{=}, \\
 & \pi_i \leq 0, & \forall i \in M_{\geq}.
 \end{aligned}$$

Le regole della dualità

Ogni vincolo del primale genera una variabile duale, ogni variabile del primale un vincolo duale, con le corrispondenze:

Primale (max)	Duale (min)
vincolo $\leq b_i$ ($i \in M_{\leq}$)	variabile $\pi_i \geq 0$
vincolo $= b_i$ ($i \in M_{=}$)	variabile $\pi_i \geq 0$
vincolo $\geq b_i$ ($i \in M_{\geq}$)	variabile $\pi_i \leq 0$
variabile $x_j \geq 0$ ($j \in N_{\geq 0}$)	vincolo $\geq c_j$
variabile $x_j \leq 0$ ($j \in N_{\leq 0}$)	vincolo $\leq c_j$
variabile x_j libero ($j \in N_{=0}$)	vincolo $= c_j$

Due letture veloci: le *uguaglianze* del primale danno variabili duali *libere* (e viceversa); in un massimo un vincolo $\leq$ di risorsa ha $\pi_i \geq 0$ (più risorsa, più valore). Per un primale di *minimo* la tabella si legge da destra a sinistra.

La coppia è legata dai due teoremi della dualità.

Teorema della dualità debole

Per ogni soluzione ammissibile $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$ del primale (P) e ogni soluzione ammissibile $(\bar{\pi}_1, \bar{\pi}_2, \dots, \bar{\pi}_m)$ del duale (D) vale

$$\sum_{j \in N} c_j \bar{x}_j \leq \sum_{i \in M} b_i \bar{\pi}_i.$$

Teorema della dualità forte

Il primale (P) ha una soluzione ottima $(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ se e solo se il duale (D) ha una soluzione ottima $(\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_m)$, e in tal caso i due valori ottimi coincidono:

$$\sum_{j \in N} c_j \tilde{x}_j = \sum_{i \in M} b_i \tilde{\pi}_i.$$

Spiegazione riga per riga

La dualità debole dice che ogni soluzione ammissibile del duale è un *limite superiore* per il valore del primale di massimo (e ogni soluzione ammissibile del primale un limite inferiore per il duale): se due soluzioni ammissibili hanno lo stesso valore, sono entrambe ottime. La dualità forte garantisce che all'ottimo questo accade sempre: il divario si chiude. Esistenza non significa unicità: gli ottimi possono essere più d'uno, ma i valori ottimi coincidono. Le variabili duali π_i e i costi ridotti che presto definiremo sono i numeri che useremo di continuo.

Costruire il duale con le regole: l'LP 2×2

Applichiamo le regole all'LP 2×2 , che è proprio nella forma canonica di massimo (vincoli $\leq$, variabili non negative):

$$\begin{array}{ll} \max & 30x_1 + 50x_2 \\ \text{soggetto a} & x_1 + 3x_2 \leq 90, \\ & 2x_1 + x_2 \leq 80, \\ & x_1, x_2 \geq 0. \end{array}$$

- **Passo 1 — ogni vincolo genera una variabile duale.** Due vincoli $\leq$ in un massimo $\Rightarrow \pi_1, \pi_2 \geq 0$.
- **Passo 2 — ogni variabile genera un vincolo duale.** I coefficienti sono le *colonne* del primale, i termini noti i costi: da $x_1 \geq 0$ nasce $\pi_1 + 2\pi_2 \geq 30$; da $x_2 \geq 0$ nasce $3\pi_1 + \pi_2 \geq 50$.
- **Passo 3 — l'obiettivo scambia i ruoli:** i termini noti (90, 80) diventano i costi del duale, che è un problema di *minimo*.

$$\begin{array}{ll} \min & 90\pi_1 + 80\pi_2 \\ \text{soggetto a} & \pi_1 + 2\pi_2 \geq 30, \\ & 3\pi_1 + \pi_2 \geq 50, \\ & \pi_1, \pi_2 \geq 0. \end{array}$$

Costruire il duale con le regole: tutti i casi in un solo LP

Costruiamo ora un LP di *minimo* che contiene tutti i casi della tabella — tre variabili con i tre segni e tre vincoli generici con i tre versi; essendo un minimo, la tabella si legge da

destra a sinistra:

$$\begin{array}{llll}
 \min & 5x_1 + 8x_2 - 9x_3 \\
 \text{soggetto a} & x_1 + x_2 & \geq & 30, \\
 & x_1 + x_2 - x_3 & = & 100, \\
 & x_1 - 2x_2 & \leq & -20, \\
 & x_1 & \geq & 0, \\
 & x_2 & \geq & 0, \\
 & x_3 & \leq & 0.
 \end{array}$$

- **Passo 1 — le variabili duali**, una per vincolo, con il segno dato dal verso: $\pi_1 \geq 0$ (vincolo $\geq$ in un minimo), $\pi_2 \geq 0$ (uguaglianza), $\pi_3 \leq 0$ (vincolo $\leq$ in un minimo).
- **Passo 2 — i vincoli duali**, uno per variabile, con il verso dato dal segno: da $x_1 \geq 0$ nasce $\pi_1 + \pi_2 + \pi_3 \leq 5$; da $x_2 \geq 0$ nasce $\pi_1 + \pi_2 - 2\pi_3 = 8$; da $x_3 \leq 0$ nasce $-\pi_2 \geq -9$.
- **Passo 3 — l'obiettivo**: il duale è un massimo, $\max 30\pi_1 + 100\pi_2 - 20\pi_3$.

$$\begin{array}{llll}
 \max & 30\pi_1 + 100\pi_2 - 20\pi_3 \\
 \text{soggetto a} & \pi_1 + \pi_2 + \pi_3 & \leq & 5, \\
 & \pi_1 + \pi_2 - 2\pi_3 & = & 8, \\
 & -\pi_2 & \geq & -9, \\
 & \pi_1 & \geq & 0, \\
 & \pi_2 & \geq & 0, \\
 & \pi_3 & \leq & 0.
 \end{array}$$

2.1.2 Le condizioni di ottimalità: gli scarti complementari

A ogni vincolo del primale associamo il suo *scarto* $\bar{s}_i$, e a ogni variabile il suo *costo ridotto* $\bar{c}_j$, cioè lo scarto del corrispondente vincolo *duale*:

$$\bar{s}_i = \sum_{j \in N} a_{ij} \bar{x}_j - b_i, \quad \forall i \in M, \quad \bar{c}_j = c_j - \sum_{i \in M} a_{ij} \bar{\pi}_i, \quad \forall j \in N.$$

Lo scarto misura quanto il vincolo i è lontano dall'essere attivo; il costo ridotto misura il coefficiente c_j al netto del valore delle risorse consumate da x_j , valutate con le variabili duali.

Condizioni di ottimalità dell'LP (scarti complementari)

Una coppia di soluzioni ammissibili, $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$ per il primale (P) e $(\bar{\pi}_1, \bar{\pi}_2, \dots, \bar{\pi}_m)$ per il duale (D), è ottima per entrambi i problemi *se e solo se*

$$\begin{array}{ll}
 \bar{\pi}_i \cdot \bar{s}_i = 0, & \forall i \in M, \\
 \bar{x}_j \cdot \bar{c}_j = 0, & \forall j \in N.
 \end{array}$$

Spiegazione riga per riga

In parole: **vincolo lasco** $\Rightarrow$ **duale a zero** ($\bar{s}_i \neq 0 \Rightarrow \bar{\pi}_i = 0$) e **variabile diversa da zero** $\Rightarrow$ **costo ridotto nullo** ($\bar{x}_j \neq 0 \Rightarrow \bar{c}_j = 0$). All'ottimo sono *implicazioni*, non equivalenze: per la degenerazione un vincolo attivo può avere duale nullo, e una variabile a zero costo ridotto nullo. Gli esempi di verifica più avanti controllano proprio queste condizioni.

Spiegazione riga per riga

Due letture da ricordare per il laboratorio:

- **Prezzo ombra** $\tilde{\pi}_i$: il valore marginale del termine noto b_i ,

$$\tilde{\pi}_i = \frac{\partial \tilde{z}}{\partial b_i};$$

vale per b_i nell'*intervallo di validità del termine noto* $[b_i^{\min}, b_i^{\max}]$: il più grande intervallo in cui, variando solo b_i , il prezzo ombra resta esattamente $\tilde{\pi}_i$ (oltre, cambia e va ricalcolato). Caso particolare: vincolo non attivo $\Rightarrow \tilde{\pi}_i = 0$.

- **Costo ridotto** $\bar{c}_j$ (lo scarto del vincolo duale, definito sopra): per una variabile a zero dice di quanto deve migliorare c_j perché convenga attivarla; vale per c_j nell'*intervallo di validità del coefficiente* $[c_j^{\min}, c_j^{\max}]$: il più grande intervallo in cui, variando solo c_j , la soluzione ottima non cambia (per una variabile a zero la soglia di convenienza è proprio uno dei due estremi). Caso particolare: variabile diversa da zero $\Rightarrow \bar{c}_j = 0$.
- **Attenzione alla degenerazione**: può esistere una variabile a zero con costo ridotto nullo, anche se non viene usata. Vale quindi solo l'implicazione " $\bar{c}_j \neq 0 \Rightarrow$ variabile su un suo bound", non il viceversa.

I segni dei prezzi ombra, spiegati bene

Il prezzo ombra è sempre, con un'unica convenzione,

$$\tilde{\pi}_i = \frac{\partial \tilde{z}}{\partial b_i},$$

la derivata del valore ottimo rispetto al termine noto del vincolo i : di quanto cambia $\tilde{z}$ se b_i aumenta di una (piccola) unità. Il segno si deduce sempre rispondendo a due domande:

- **aumentare b_i allarga o restringe la regione ammissibile?** Con un vincolo $\leq$ la allarga (il tetto si alza); con un vincolo $\geq$ la restringe (la richiesta cresce); con un'uguaglianza la sposta soltanto;
- **una regione più ampia come cambia l'ottimo?** Non può mai peggiorarlo: in un minimo il valore ottimo scende o resta uguale, in un massimo sale o resta uguale.

Combinando le due risposte:

Verso del vincolo	problema di minimo	problema di massimo
$\leq$ (b_i cresce $\Rightarrow$ regione più ampia)	$\tilde{\pi}_i \leq 0$	$\tilde{\pi}_i \geq 0$
$\geq$ (b_i cresce $\Rightarrow$ regione più stretta)	$\tilde{\pi}_i \geq 0$	$\tilde{\pi}_i \leq 0$
$=$	segno qualunque (duale libera)	

Due letture tipiche. In un *minimo*: un vincolo $\leq$ (una capacità) ha $\tilde{\pi}_i \leq 0$ — allentarlo riduce il costo, e $|\tilde{\pi}_i|$ è il risparmio per ogni unità di capacità in più; un vincolo $\geq$ (una richiesta da soddisfare) ha $\tilde{\pi}_i \geq 0$ — alzare la richiesta costa, e $\tilde{\pi}_i$ è il costo marginale. In un *massimo* le letture si ribaltano. In ogni caso un vincolo non attivo ha $\tilde{\pi}_i = 0$ per complementarità.

Dualità su un LP 2×2 , svolta per intero

Riprendiamo l'LP 2×2 e il suo duale, costruiti sopra, e verifichiamo le condizioni di ottimalità.

- **Passo 1 — risolvendo si ottiene:**

$$(\tilde{x}_1, \tilde{x}_2) = (30, 20), \quad \tilde{z} = 1900, \quad (\tilde{\pi}_1, \tilde{\pi}_2) = (14, 8), \quad (\bar{c}_1, \bar{c}_2) = (0, 0),$$

con intervalli di validità $[b_1^{\min}, b_1^{\max}] = [40, 240]$, $[b_2^{\min}, b_2^{\max}] = [30, 180]$, $[c_1^{\min}, c_1^{\max}] = [50/3, 100]$ e $[c_2^{\min}, c_2^{\max}] = [15, 90]$. Entrambi i vincoli sono attivi.

- **Passo 2 — verifica dei prezzi ombra.** Una unità in più di b_1 vale 14, una unità in più di b_2 vale 8. Verifichiamo la proprietà che li caratterizza: $\tilde{x}_1$ e $\tilde{x}_2$ sono diverse da zero, quindi per gli scarti complementari i loro vincoli duali devono essere attivi ($\bar{c}_1 = \bar{c}_2 = 0$):

$$\tilde{\pi}_1 + 2\tilde{\pi}_2 = 14 + 16 = 30 = c_1, \quad 3\tilde{\pi}_1 + \tilde{\pi}_2 = 42 + 8 = 50 = c_2. \quad \checkmark$$

- **Passo 3 — verifica della dualità forte.**

$$\sum_{i \in M} b_i \tilde{\pi}_i = 90 \cdot 14 + 80 \cdot 8 = 1260 + 640 = 1900 = \tilde{z}. \quad \checkmark$$

- **Passo 4 — verifica dei costi ridotti.** Con la definizione $\bar{c}_j = c_j - \sum_{i \in M} a_{ij} \tilde{\pi}_i$:

$$\bar{c}_1 = 30 - (1 \cdot 14 + 2 \cdot 8) = 0, \quad \bar{c}_2 = 50 - (3 \cdot 14 + 1 \cdot 8) = 0 :$$

entrambi nulli, come per ogni variabile positiva all'ottimo: le due variabili ripagano esattamente le risorse che consumano. Un costo ridotto diverso da zero può averlo solo una variabile che sta su un proprio bound.

Tutti i casi in un solo LP: verifica delle condizioni

Riprendiamo l'LP “tutti i casi” e il suo duale (esempio 2.1.1) — i casi sono gli stessi che torneranno, uno alla volta, nei capitoli applicativi.

- **Passo 1 — risolvendo si ottiene:**

$$(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) = (60, 40, 0), \quad \tilde{z} = 620,$$

$$(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3) = (0, 6, -1), \quad (\bar{c}_1, \bar{c}_2, \bar{c}_3) = (0, 0, -3),$$

con intervalli di validità dei termini noti $(-\infty, 100]$, $[30, +\infty)$ e $[-200, +\infty)$, e dei coefficienti $(-\infty, 8]$, $[5, 17]$ e $(-\infty, -6]$. All'ottimo sono attivi il secondo e il terzo vincolo, con $\tilde{x}_3 = 0$.

- **Passo 2 — verifica dei prezzi ombra con le regole della tabella.** $\tilde{\pi}_1 = 0$, coerente con la complementarità: il primo vincolo è lasco ($\tilde{x}_1 + \tilde{x}_2 = 100 > 30$). $\tilde{\pi}_2 = 6$ (variabile duale libera, qui positiva). $\tilde{\pi}_3 = -1 \leq 0$, come dev'essere per un vincolo $\leq$ in un minimo. Verifichiamo che, come impongono gli scarti complementari, i vincoli duali di x_1 e x_2 (diverse da zero) siano attivi:

$$\underbrace{0 + 6 - 1}_{x_1} = 5 = c_1, \quad \underbrace{0 + 6 - 2 \cdot (-1)}_{x_2} = 8 = c_2, \quad \checkmark$$

(x_2 è libera: il suo vincolo duale è di uguaglianza). Verifica per perturbazione:

$$b_2 = 101 \Rightarrow \tilde{z} = 626 (+6), \quad b_3 = -19 \Rightarrow \tilde{z} = 619 (-1). \quad \checkmark$$

- **Passo 3 — verifica della dualità forte.**

$$\sum_{i \in M} b_i \tilde{\pi}_i = 30 \cdot 0 + 100 \cdot 6 + (-20) \cdot (-1) = 620 = \tilde{z}. \quad \checkmark$$

- **Passo 4 — verifica dei costi ridotti, uno per segno.** Con la definizione $\bar{c}_j = c_j - \sum_{i \in M} a_{ij} \tilde{\pi}_i$:

$$\bar{c}_1 = 5 - (0 + 6 - 1) = 0, \quad \bar{c}_2 = 8 - (0 + 6 + 2) = 0:$$

coerente con $\tilde{x}_1, \tilde{x}_2 \neq 0$. x_3 vale zero, che per una variabile ≤ 0 è il suo *bound superiore*:

$$\bar{c}_3 = -9 - (-1) \cdot 6 = -3,$$

cioè forzare $x_3 = -1$ peggiorerebbe l'ottimo di 3; x_3 si attiverebbe solo se il suo coefficiente salisse almeno a -6 (proprio l'estremo $c_3^{\max}$ dell'intervallo di validità). Verifica:

$$x_3 = -1 \Rightarrow \tilde{z} = 623 (+3). \quad \checkmark$$

2.1.3 L'analisi di sensitività e i prezzi ombra

Risolvere un LP dà una soluzione ottima; l'**analisi di sensitività** risponde alla domanda successiva, quella che interessa davvero un decisore: *come cambia l'ottimo se cambiano i dati?* Negli LP gran parte della risposta arriva gratis insieme alla soluzione, sotto forma di quattro numeri per ogni vincolo e variabile:

- il **prezzo ombra** $\tilde{\pi}_i$ di un vincolo: di quanto migliora il valore ottimo se il termine noto del vincolo aumenta di una unità — “quanto pagherei per un'unità in più di quella risorsa”;
- l'**intervallo di validità del termine noto** $[b_i^{\min}, b_i^{\max}]$: dove quel prezzo ombra resta esatto (fuori, va rifatta l'ottimizzazione);
- il **costo ridotto** $\bar{c}_j$ di una variabile a zero: di quanto deve migliorare il suo coefficiente in obiettivo perché convenga attivarla;
- l'**intervallo di validità del coefficiente** in obiettivo $[c_j^{\min}, c_j^{\max}]$: dove il coefficiente può variare senza che la soluzione ottima cambi. Nell'LP 2×2 : c_1 può variare in $[50/3 \approx 16,7, 100]$ e c_2 in $[15, 90]$ senza spostare la soluzione $(30, 20)$.

Prezzi ombra in pratica, sull'LP 2×2

Riprendiamo l'LP 2×2 : ottimo 1900 con duali $\tilde{\pi}_1 = 14$ e $\tilde{\pi}_2 = 8$.

Lettura. Una unità in più di b_1 fa crescere l'ottimo di 14; se procurarla costasse 12, converrebbe ($14 > 12$).

Verifica per perturbazione. Ririsolvendo:

$$b_1 = 91 \Rightarrow 1914 = 1900 + 14, \quad b_1 = 100 \Rightarrow 2040 = 1900 + 10 \cdot 14 :$$

il prezzo ombra è ancora esatto perché 100 è dentro l'intervallo di validità $[40, 240]$.

Dove smette di valere. A $b_1 = 240$ il primo vincolo smette di essere la risorsa scarsa: oltre, ogni unità aggiuntiva vale meno di 14 e va ricalcolato tutto. Ecco perché ogni prezzo ombra si cita *insieme* al suo intervallo di validità.

Vincoli non attivi. Se fosse $b_1 = 300$ (vincolo non attivo all'ottimo), il suo prezzo ombra sarebbe 0: una risorsa che avanza non vale nulla al margine.

Costi ridotti in pratica: conviene una terza variabile?

Nell'LP 2×2 entrambe le variabili sono positive all'ottimo e i loro costi ridotti valgono 0: è sempre così per le variabili diverse da zero. Per vedere un costo ridotto all'opera serve una variabile a zero: aggiungiamo al modello una variabile $x_3 \geq 0$ con coefficiente $c_3 = 20$ in obiettivo e consumi $a_{13} = a_{23} = 1$ nei due vincoli.

- **Passo 1 — valutazione con i prezzi ombra.** Le risorse assorbite da un'unità di x_3 valgono, ai prezzi ombra,

$$1 \cdot 14 + 1 \cdot 8 = 22 > 20 = c_3 :$$

ogni unità di x_3 sottrarrebbe risorse a impieghi migliori.

- **Passo 2 — il costo ridotto.** La differenza è proprio il costo ridotto:

$$\bar{c}_3 = 20 - 22 = -2.$$

Risolvendo si conferma: la soluzione ottima resta $(30, 20, 0)$ con $\tilde{z} = 1900$, la variabile x_3 è a zero con $\bar{c}_3 = -2$ e $c_3^{\max} = 22$: il coefficiente c_3 dovrebbe salire almeno a 22 perché convenga attivarla.

- **Passo 3 — controprova.** Con $c_3 = 23 > 22$ la soluzione cambia davvero, e non di poco: la nuova soluzione ottima è $(0, 5, 75)$ con valore 1975. Il costo ridotto dice *quando* una variabile a zero entra in soluzione, non *come* cambia la soluzione: quella va ricalcolata.

2.2 Teoria dell'ottimizzazione non lineare

2.2.1 Convessità e programmazione quadratica

Un problema $\min f(x_1, x_2, \dots, x_n)$ su un insieme ammissibile X è **convesso** se f è convessa e X è convesso. La proprietà che useremo di continuo:

Perché la convessità conta

In un problema convesso ogni minimo locale è globale: l'ottimalità si può *certificare*. Nei problemi non convessi (ad esempio un problema di pricing con termine bilineare $p \cdot q$) un metodo locale garantisce solo un ottimo locale; i QP non convessi si risolvono comunque all'ottimo globale con tecniche apposite, ma il costo computazionale cresce.

Un **QP** ha obiettivo $\frac{1}{2} \sum_i \sum_j q_{ij} x_i x_j + \sum_j c_j x_j$ e vincoli lineari; è convesso se e solo se la matrice $Q = (q_{ij})$ è semidefinita positiva ($Q \succeq 0$). L'esempio che segue ne mostra uno in piccolo; nella dispensa i QP compaiono ogni volta che un costo o un rischio cresce più che linearmente con la decisione.

Un QP 2×2 , svolto per intero

Consideriamo un QP di minimo con obiettivo quadratico separabile (la seconda variabile “costa” il doppio della prima, al margine) e un vincolo di soglia sulla somma:

$$\begin{aligned} \min \quad & x_1^2 + 2x_2^2 \\ \text{soggetto a} \quad & x_1 + x_2 \geq 6, \\ & x_1, x_2 \geq 0. \end{aligned}$$

Qui $Q = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \succ 0$: il QP è convesso e l'ottimo trovato è globale.

- **Passo 1 — risolvendo si ottiene:**

$$\tilde{x}_1 = 4, \quad \tilde{x}_2 = 2, \quad \tilde{f} = 24, \quad \lambda = 8.$$

- **Passo 2 — verifica del prezzo ombra.** Il vincolo è attivo ($4 + 2 = 6$: l'obiettivo cresce con le variabili, quindi la somma non supera il necessario) e le derivate parziali coincidono:

$$2\tilde{x}_1 = 8 = 4\tilde{x}_2 = \lambda. \quad \checkmark$$

Se fossero diverse converrebbe spostare quantità dall'una all'altra variabile; la derivata comune è proprio il moltiplicatore.

- **Passo 3 — verifica per perturbazione.** Con termine noto 7 si ottiene $(\tilde{x}_1, \tilde{x}_2) = (14/3, 7/3)$ con

$$\tilde{f} = 32,67 \approx 24 + 8$$

(in più c'è il termine di curvatura: qui il prezzo ombra è esatto solo al margine, non per incrementi finiti).

2.2.2 Le condizioni di ottimalità: le condizioni KKT

Le condizioni di ottimalità dei problemi non lineari vincolati prendono il nome di **condizioni KKT** (da Karush, Kuhn e Tucker, che le formularono). Per il problema $\min f(x_1, \dots, x_n)$ soggetto a $g_i(x_1, \dots, x_n) \leq 0$ per ogni $i \in \{1, 2, \dots, m\}$ e $h_j(x_1, \dots, x_n) = 0$ per ogni $j \in \{1, 2, \dots, q\}$, la Lagrangiana è

$$L(x_1, \dots, x_n) = f(x_1, \dots, x_n) + \sum_{i=1}^m \lambda_i g_i(x_1, \dots, x_n) + \sum_{j=1}^q \nu_j h_j(x_1, \dots, x_n), \quad \lambda_i \geq 0.$$

Condizioni KKT

In un punto di ottimo regolare $(\tilde{x}_1, \dots, \tilde{x}_n)$ esistono moltiplicatori $\lambda_i \geq 0$ e ν_j tali che

$$\frac{\partial L}{\partial x_k}(\tilde{x}_1, \dots, \tilde{x}_n) = 0 \quad \forall k \in \{1, 2, \dots, n\} \quad (\text{stazionarietà}),$$

$$\lambda_i g_i(\tilde{x}_1, \dots, \tilde{x}_n) = 0 \quad \forall i \in \{1, 2, \dots, m\} \quad (\text{complementarietà}).$$

Se il problema è convesso, le KKT sono anche *sufficienti*: un punto che le soddisfa è un ottimo globale.

Le KKT sono gli scarti complementari, generalizzati

Le due parti del capitolo sono speculari. Per un LP le KKT si riducono *esattamente* alle condizioni di ottimalità del paragrafo 2.1.2: la stazionarietà della Lagrangiana equivale all'ammissibilità del duale (i moltiplicatori λ_i sono le variabili duali π_i , e i moltiplicatori dei vincoli di segno sono i costi ridotti), e la complementarietà $\lambda_i g_i(\tilde{x}_1, \dots, \tilde{x}_n) = 0$ è lo scarto complementare $\bar{\pi}_i \cdot \bar{s}_i = 0$. Gli scarti complementari sono le KKT nel caso lineare.

Le KKT sul QP dell'esempio, svolte per intero

Riprendiamo il QP appena risolto: $\min x_1^2 + 2x_2^2$ soggetto a $x_1 + x_2 \geq 6$, $x_1, x_2 \geq 0$.

- **Passo 1.** Riscriviamo il vincolo in forma standard: $g(x_1, x_2) = 6 - x_1 - x_2 \leq 0$.
- **Passo 2.** Stazionarietà della Lagrangiana $L = x_1^2 + 2x_2^2 + \lambda(6 - x_1 - x_2)$:

$$\frac{\partial L}{\partial x_1} = 2x_1 - \lambda = 0, \quad \frac{\partial L}{\partial x_2} = 4x_2 - \lambda = 0 \Rightarrow x_1 = \frac{\lambda}{2}, \quad x_2 = \frac{\lambda}{4}$$

(i vincoli $x_1, x_2 \geq 0$ risulteranno non attivi: i loro moltiplicatori sono nulli per complementarietà).

- **Passo 3.** Complementarietà: se il vincolo di soglia fosse inattivo ($\lambda = 0$) avremmo $x_1 = x_2 = 0$, che viola $x_1 + x_2 \geq 6$: impossibile. Quindi il vincolo è attivo: $\lambda/2 + \lambda/4 = 6 \Rightarrow \lambda = 8$, $\tilde{x}_1 = 4$, $\tilde{x}_2 = 2$, $\tilde{f} = 24$ — gli stessi numeri trovati per via elementare nell'esempio del QP, ora ottenuti con la procedura generale.
- **Passo 4.** Lettura del moltiplicatore: con termine noto $6 + \varepsilon$ il valore ottimo cresce di circa $\lambda\varepsilon = 8\varepsilon$. Verifica: con termine noto d la soluzione è $\tilde{x}_1 = 2d/3$, $\tilde{x}_2 = d/3$, quindi

$$\tilde{f}(d) = \frac{2}{3}d^2 \Rightarrow \tilde{f}(6 + \varepsilon) = 24 + 8\varepsilon + \frac{2}{3}\varepsilon^2. \quad \checkmark$$

Spiegazione riga per riga

I moltiplicatori KKT generalizzano i prezzi ombra dell'LP: misurano la variazione marginale del valore ottimo rispetto al termine noto del vincolo. Nel laboratorio, quando i duali non sono disponibili (modelli non lineari generali), stimiamo il moltiplicatore *per perturbazione*: si aumenta il termine noto di ε , si riottimizza e si calcola il rapporto incrementale. Questa verifica numerica è essa stessa un esercizio ricorrente.

2.3 Il protocollo di analisi di sensitività

L'analisi di sensitività è parte del modello, non un'appendice. In ogni esercitazione seguiamo lo stesso protocollo in sei passi:

1. Scenario base	risolvere, verificare l'ammissibilità, identificare i vincoli attivi.
2. One-at-a-time	variare un parametro chiave su una griglia; grafico valore/-decisioni.
3. Prezzi ombra e costi ridotti	confrontare la previsione marginale del duale con una riottimizzazione dopo una piccola perturbazione; leggere il costo ridotto $\bar{c}_j$ di ogni variabile a zero: la soglia oltre la quale conviene attivarla.
4. Scenari	costruire almeno tre scenari coerenti (pessimistico, centrale, ottimistico).
5. Trade-off	variare un peso o una soglia per tracciare una frontiera (costo-servizio, rischio-rendimento, costo-emissioni).
6. Stabilità	segnalare cambi bruschi del mix ottimo e parametri stimati fragili.

Esercizio 2.1 — dualità

Nell'LP 2×2 dell'esempio è possibile aumentare b_1 di 10 unità pagando 12 per unità. Conviene? Fino a che prezzo unitario conviene? E fino a quale valore di b_1 vale il prezzo ombra di 14? (Suggerimento: ricalcolare il vertice ottimo quando il primo vincolo si sposta a $90 + \Delta$ e trovare il Δ oltre il quale la soluzione cambia struttura.)

Esercizio 2.2 — KKT

Risolvere con le KKT: $\min (x - 3)^2 + (y - 1)^2$ soggetto a $x + 2y \leq 2$, $x, y \geq 0$. Dire quali vincoli sono attivi all'ottimo e interpretare i moltiplicatori.

Esercizio 2.3 — convessità

Per quali valori del parametro γ la funzione $f(x_1, x_2) = x_1^2 + \gamma x_1 x_2 + x_2^2$ è convessa? (Scrivere Q e imporre $Q \succeq 0$.)

Il solver: costruire, risolvere, interpretare

Questo capitolo è la cassetta degli attrezzi del laboratorio: come si costruisce un modello con `gurobipy`, come si fa girare, come si recupera la soluzione e — la parte più importante — come si interpreta l'output. Gli esempi sono generici e usano gli stessi LP del capitolo 2: insieme, i due capitoli danno le basi per capire tutti i modelli e gli esempi dei capitoli successivi. Una versione autonoma di questa guida è nel file `GUIDA_GUROBI.md` nella cartella del corso. Tre sottocapitoli: i *comandi generali* (costruire, risolvere, leggere la soluzione), poi i comandi specifici per i *modelli lineari* (duali, costi ridotti, intervalli di validità) e per i *modelli non lineari* (vincoli funzionali e bilineari).

3.1 I comandi generali

3.1.1 Installazione e licenza

```
python3 -m pip install gurobipy
```

Il pacchetto include una licenza dimostrativa (fino a 2000 variabili e 2000 vincoli), sufficiente per quasi tutti i modelli della dispensa. Per la licenza accademica completa e gratuita: registrarsi su <https://portal.gurobi.com> con l'email istituzionale, richiedere una *Named-User Academic License* ed eseguire il comando `grbgetkey` indicato dal portale (sotto rete di ateneo o VPN).

3.1.2 I cinque passi per costruire un modello

Riprendiamo l'LP 2×2 del capitolo 2: $\max 30x_1 + 50x_2$ con $x_1 + 3x_2 \leq 90$ (vincolo 1), $2x_1 + x_2 \leq 80$ (vincolo 2) e $x_1, x_2 \geq 0$.

```
1 import gurobipy as gp
2 from gurobipy import GRB
3
4 # Passo 1: il contenitore del modello
5 m = gp.Model("lp_2x2")
6
7 # Passo 2: le variabili decisionali (default: continue, >= 0)
8 x1 = m.addVar(name="x1")
9 x2 = m.addVar(name="x2")
10
11 # Passo 3: i vincoli (conservare l'oggetto per leggere poi il duale)
12 v1 = m.addConstr(x1 + 3*x2 <= 90, name="vincolo1")
13 v2 = m.addConstr(2*x1 + x2 <= 80, name="vincolo2")
```

```

14
15 # Passo 4: la funzione obiettivo
16 m.setObjective(30*x1 + 50*x2, GRB.MAXIMIZE)
17
18 # Passo 5 (debug): scrivere il modello in formato leggibile
19 m.write("lp_2x2.lp")

```

Le tre regole d'oro di gurobipy

1. **Variabili indicizzate** con `addVars`:
`x = m.addVars(I, T, name="x")` crea la famiglia `x[i,t]` sugli insiemi di indici `I` e `T`; le somme si scrivono `gp.quicksum(...)` o `x.sum(i, "*")`.
2. **Variabili libere** (che possono essere negative, come l'intercetta b della SVM o la soglia η del CVaR): `m.addVar(lb=-GRB.INFINITY)`. Dimenticare `lb` è l'errore più frequente del laboratorio: il default è ≥ 0 !
3. **Famiglie di vincoli** con generatore: `m.addConstrs((x.sum("*", t) <= b[t] for t in T), name="cap")`.

3.1.3 Far girare il modello e leggere il log

```

1 m.Params.OutputFlag = 0      # 0 = silenzioso, 1 = log a video
2 m.Params.TimeLimit   = 60    # secondi
3 m.Params.NonConvex   = 2     # serve solo per QP non convessi (es.
   pricing p*q)
4 m.optimize()

```

Il log del solver, quando attivo, va letto in tre punti:

```

Optimize a model with 2 rows, 2 columns and 4 nonzeros  <- dimensioni attese?
Coefficient statistics: Matrix range [1e+00, 3e+00]     <- dati ben scalati?
Optimal objective   1.900000000e+03                     <- esito e valore

```

Se i range dei coefficienti spaziano da 10^{-6} a 10^9 il modello è mal scalato: cambiare unità di misura (migliaia di euro, MWh, ...) prima di cercare altri errori.

3.1.4 Recuperare la soluzione

Controllare sempre lo stato prima di leggere qualunque numero:

```

1 if m.Status == GRB.OPTIMAL:
2     print(m.ObjVal, x1.X, x2.X)
3 elif m.Status == GRB.INFEASIBLE:
4     m.computeIIS(); m.write("conflitto.ilp") # quali vincoli si
   contraddicono
5 elif m.Status == GRB.UNBOUNDED:
6     pass # quasi sempre manca un vincolo o un bound

```

Attributo	Oggetto	Significato
Status	modello	OPTIMAL=2, INFEASIBLE=3, UNBOUNDED=5, TIME_LIMIT=9
ObjVal	modello	valore ottimo dell'obiettivo
X	variabile	valore della variabile all'ottimo
Slack	vincolo	scarto: $0 \Rightarrow$ vincolo attivo

Gli attributi specifici degli LP (prezzi ombra, costi ridotti, intervalli di validità) sono nel sottocapitolo che segue.

3.2 I comandi per i modelli lineari

3.2.1 Prezzi ombra, costi ridotti e intervalli: gli attributi

Il ponte con la notazione del capitolo 2

Gli attributi del solver sono esattamente le quantità teoriche dei richiami:

Teoria (cap. 2)	Attributo Gurobi	Oggetto
prezzo ombra $\tilde{\pi}_i$	Pi	vincolo
costo ridotto $\bar{c}_j$	RC	variabile
scarto $\bar{s}_i$	Slack	vincolo
intervallo $[b_i^{\min}, b_i^{\max}]$	SARHSLow-SARHSUp	vincolo
intervallo $[c_j^{\min}, c_j^{\max}]$	SAObjLow-SAObjUp	variabile

Se un vincolo contiene una costante a sinistra, come `quicksum(x[v,t] ...) + base[t] <= C`, Gurobi la sposta nel termine noto: il RHS memorizzato è $C - \text{base}[t]$. Chi in un'analisi di sensitività scrive `constr.RHS = nuovaC` sta in realtà imponendo `quicksum(x) <= nuovaC`, cioè un'altra cosa. La forma corretta è `constr.RHS = nuovaC - base[t]`. Questo errore ci è capitato davvero preparando il caso di studio della ricarica dei veicoli elettrici, che lo discute.

3.2.2 Interpretare l'output: la checklist

1. Lo stato è OPTIMAL? Se no, fermarsi e diagnosticare (`computeIIS` per l'inammissibilità, file `.lp` per l'illimitatezza).
2. Il valore ottimo ha l'ordine di grandezza atteso?
3. Quali vincoli sono attivi (`Slack = 0`)? Sono quelli previsti?
4. Quanto valgono i prezzi ombra? Quale risorsa conviene potenziare per prima? (Ricordare i segni: in un problema di minimo, un vincolo $\leq$ ha $Pi \leq 0$.)
5. Ci sono variabili a zero? Leggere il loro costo ridotto (RC): dice di quanto deve migliorare il coefficiente in obiettivo perché convenga attivarle (paragrafo 2.1.3).
6. La soluzione è stabile? Rieseguire con dati perturbati del 5%.
7. Tradurre tutto in una raccomandazione manageriale di tre righe.

3.2.3 Due letture complete

Lettura completa dell'esempio 2×2

Eseguendo il codice della sezione precedente:

```
Status: 2 (OPTIMAL)      ObjVal: 1900.0
x1 = 30.0    RC = 0.0    SAObj = [16.7, 100.0]
x2 = 20.0    RC = 0.0    SAObj = [15.0, 90.0]
vincolo1: Slack = 0.0    Pi = 14.0    SARHS = [40.0, 240.0]
vincolo2: Slack = 0.0    Pi = 8.0     SARHS = [30.0, 180.0]
```

Lettura: entrambi i vincoli sono attivi (saturi); una unità in più di b_1 vale 14, e la valutazione resta valida finché b_1 resta in $[40, 240]$; oltre, va ricalcolato tutto. Se si potesse aumentare un solo termine noto, meglio b_1 ($14 > 8$ per unità, a parità di costo). I costi ridotti (attributo $x1.RC$) sono nulli perché entrambe le variabili sono in base — attenzione: può esistere anche una variabile *in base a valore zero* (degenerazione), quindi $RC = 0$ da solo non dice che la variabile è usata. Aggiungendo la terza variabile del capitolo 2 il solver darebbe $x_3 = 0$ con $RC = -2$: il suo coefficiente deve salire di almeno 2 (cioè a 22) perché entri in soluzione, con range di validità $SAObj = (-\infty, 22]$.

Nel caso generale: che cosa restituisce il solver e con che segno

Prezzi ombra (attributo Pi , uno per vincolo): sempre la “derivata del valore ottimo rispetto al termine noto”. Il segno si deduce in due mosse: aumentare b_i allarga la regione ammissibile (vincolo $\leq$) o la restringe (vincolo $\geq$); e una regione più ampia non può mai peggiorare l’ottimo. Quindi:

- in un problema di *massimo*: vincolo $\leq$ di risorsa $\Rightarrow Pi \geq 0$ (più risorsa, più profitto); vincolo $\geq \Rightarrow Pi \leq 0$;
- in un problema di *minimo*: vincolo $\geq$ di domanda $\Rightarrow Pi \geq 0$ (più domanda, più costo); vincolo $\leq$ di capacità $\Rightarrow Pi \leq 0$ (più capacità, meno costo);
- vincolo di *uguaglianza* $\Rightarrow$ segno qualunque, come da regole della dualità. Vincolo non attivo $\Rightarrow Pi = 0$; validità nel range $SARHS_{Low} - SARHS_{Up}$.

Costi ridotti (attributo RC , uno per variabile):

- variabile *in base*: $RC = 0$ — di norma è strettamente tra i suoi bound, ma può anche trovarsi su un bound con $RC = 0$ (base *degenere*);
- variabile al *bound inferiore* (tipicamente a zero): $RC \geq 0$ in un minimo (di quanto deve scendere il suo costo perché convenga usarla), $RC \leq 0$ in un massimo (di quanto deve salire il suo margine); validità nel range $SAObj_{Low} - SAObj_{Up}$;
- variabile al *bound superiore*: RC è il prezzo ombra del bound — in un modello di flusso di minimo costo un arco saturo ha $RC < 0$: un’unità di capacità in più su quell’arco fa risparmiare $|RC|$.

Il ciclo completo, per *tutti* i vincoli e le variabili (solo LP):

```
1 for v in m.getConstrs():
2     print(v.ConstrName, v.Pi, v.SARHSLow, v.SARHSUp)
3 for v in m.getVars():
4     print(v.VarName, v.X, v.RC, v.SAObjLow, v.SAObjUp)
```

L'esempio completo: tutti i casi, dal modello alla lettura

Implementiamo l'LP "tutti i casi" del capitolo 2 (tre versi di vincolo, tre segni di variabile) e leggiamo *tutto* quello che il solver restituisce:

```

1 m = gp.Model("tutti_i_casi")
2 x1 = m.addVar(name="x1") # x1 >= 0 (default)
3 x2 = m.addVar(lb=-GRB.INFINITY, name="x2") # x2 libera
4 x3 = m.addVar(lb=-GRB.INFINITY, ub=0, name="x3") # x3 <= 0
5 v1 = m.addConstr(x1 + x2 >= 30, name="vincolo1") # verso >=
6 v2 = m.addConstr(x1 + x2 - x3 == 100, name="vincolo2") # verso =
7 v3 = m.addConstr(x1 - 2*x2 <= -20, name="vincolo3") # verso <=
8 m.setObjective(5*x1 + 8*x2 - 9*x3, GRB.MINIMIZE)
9 m.optimize()
10 for v in m.getVars():
11     print(v.VarName, v.X, v.RC, v.SAObjLow, v.SAObjUp)
12 for c in m.getConstrs():
13     print(c.ConstrName, c.Slack, c.Pi, c.SARHSLow, c.SARHSUp)

```

```

Status: 2 (OPTIMAL)    ObjVal: 620.0
x1: X = 60.0 RC = 0.0 SAObj = [-inf, 8.0]
x2: X = 40.0 RC = 0.0 SAObj = [5.0, 17.0]
x3: X = 0.0 RC = -3.0 SAObj = [-inf, -6.0]
vincolo1: Slack = -70.0 Pi = 0.0 SARHS = [-inf, 100.0]
vincolo2: Slack = 0.0 Pi = 6.0 SARHS = [30.0, inf]
vincolo3: Slack = 0.0 Pi = -1.0 SARHS = [-200.0, inf]

```

Qui compaiono tutti i casi della spiegazione precedente: il vincolo non attivo con $Pi = 0$ (vincolo1, $Slack \neq 0$); l'uguaglianza con duale positivo (+6); il $\leq$ in un minimo con duale negativo (-1); due variabili in base con $RC = 0$; e la variabile x_3 ferma al suo bound superiore (0) con $RC = -3$ e soglia di convenienza $SAObjUp = -6$. La verifica a mano di ognuno di questi numeri, con le regole della dualità e per perturbazione, è nell'esempio "tutti i casi" del capitolo 2.

3.3 I comandi per i modelli non lineari

Un solver solo per tutto il laboratorio: Gurobi copre nativamente LP e QP (anche non convessi, con `m.Params.NonConvex = 2`) e, dalla versione 12, risolve *globalmente* anche i modelli con funzioni non lineari generali.

3.3.1 Vincoli funzionali e termini bilineari

Le funzioni non lineari — logaritmi, esponenziali, potenze — entrano nel modello tramite i *vincoli funzionali*: si introduce una variabile ausiliaria y e si impone $y = f(x)$. I rapporti come $1/(\mu - \lambda)$ si riformulano invece con una variabile ausiliaria e un vincolo bilineare ($w(\mu - \lambda) = 1$).

```

1 m.Params.FuncNonlinear = 1 # funzioni trattate in modo esatto (
   globale)
2 z = m.addVar(lb=-GRB.INFINITY)
3 m.addGenConstrLog(g, z) # z = log(g); anche Exp, Pow, Sin, ...
4 m.addQConstr(w * v == 1) # termini bilineari: servono NonConvex =
   2

```

3.3.2 Tolleranze e scalatura

Due accorgimenti pratici, entrambi usati negli script della dispensa:

- **tolleranze:** per analisi marginali accurate (differenze tra due ottimi vicini) stringere `MIPGap`, `FeasibilityTol` e `OptimalityTol` a 10^{-9} ;
- **scalatura:** riformulare per evitare quantità minuscole — ad esempio $q \leq A p^{-\varepsilon}$ si scrive $q \cdot r \leq A$ con $r = p^\varepsilon$, che tiene i numeri in un range sano.

3.3.3 Quando i duali non ci sono

Per i modelli con vincoli funzionali o bilineari il solver non fornisce i duali: prezzi ombra e moltiplicatori si stimano *per perturbazione* (aumentare il termine noto di ε , riottimizzare, rapporto incrementale), come nel protocollo del capitolo 2.

Perché un solver solo

Stessa sintassi, stessa checklist di interpretazione e — soprattutto — ottimo *globale certificato* anche nei problemi non convessi: con un solver locale ogni risultato andrebbe accompagnato dalla domanda “è solo un ottimo locale?”. I problemi convessi con sola struttura quadratica o conica (vincoli come $d_x^2 + d_y^2 \leq d^2$) non richiedono nulla di speciale.

Esercizio 3.1

Implementare l'LP 2×2 , stampare `Pi`, `Slack`, `SARHSLow/Up` di entrambi i vincoli e verificare il prezzo ombra del primo vincolo ri-ottimizzando con $b_1 : 90 \rightarrow 91$.

Esercizio 3.2

Togliere il secondo vincolo e osservare l'esito. Poi cambiare l'obiettivo in minimizzazione senza altri cambiamenti: perché il risultato è $x = 0$?

Esercizio 3.3

Costruire apposta un modello inammissibile (due vincoli incompatibili, ad esempio $x_1 + x_2 \geq 10$ e $x_1 + x_2 \leq 8$) e usare `computeIIS` per individuare il conflitto.

Esercizio 3.4

Risolvere con i vincoli funzionali di Gurobi il problema in una variabile $\max 5 \log(1+x) - x$ con $x \in [0, 10]$ e confrontare con la soluzione analitica ($5/(1+x) - 1 = 0 \Rightarrow x = 4$).

Parte II

Modelli deterministici continui

Produzione e scorte multiperiodali

Classe: LP / QP convesso

Script: [python/lab04_produzione.py](#)

[Pagina online](#)

4.1 Motivazione gestionale

Un'azienda decide quanto produrre oggi e quanto conservare a scorta per la domanda futura. Produrre in anticipo costa giacenza; produrre all'ultimo momento rischia di sbattere contro il limite di capacità proprio nei mesi di picco. Il modello bilancia margini, capacità, scorte e stabilità del piano — ed è il punto di partenza ideale del laboratorio perché ogni suo numero ha un'interpretazione economica immediata.

In pratica questo problema si presenta ogni volta che un'azienda pianifica la produzione su più periodi con risorse condivise: piani principali di produzione (MPS), pianificazione aggregata, contratti di fornitura con profili mensili. Il nucleo della difficoltà è il *legame intertemporale*: la decisione di oggi (produrre in anticipo o no) cambia i vincoli di domani.

In questo capitolo impareremo a: (1) formulare un LP multiperiodale con vincoli di bilancio delle scorte; (2) leggere e verificare i prezzi ombra della capacità e i loro intervalli di validità; (3) usare un termine quadratico per stabilizzare il piano; (4) costruire la curva del valore della capacità.

4.2 Costruzione guidata del modello

Il problema a parole. *Decidiamo* quante unità di ogni prodotto produrre in ogni mese e quante tenerne a scorta. *L'obiettivo* è minimizzare la somma dei costi di produzione e di giacenza sull'intero orizzonte. *I vincoli* sono due: la contabilità di magazzino deve tornare in ogni mese (ciò che entra uguaglia ciò che esce, e tutta la domanda va servita) e le ore macchina usate in un mese non possono superare quelle disponibili.

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di prodotti; i prodotti sono indicizzati da $i \in \{1, 2, \dots, n\}$
m	$\in \mathbb{Z}_{\geq 1}$	numero di mesi dell'orizzonte; i mesi sono indicizzati da $t \in \{1, 2, \dots, m\}$
d_{it}	$\in \mathbb{Q}_{\geq 0}$	domanda del prodotto i nel mese t (unità)
c_{it}	$\in \mathbb{Q}_{> 0}$	costo di produzione di un'unità di i nel mese t (€)
h_{it}	$\in \mathbb{Q}_{> 0}$	costo di giacenza di un'unità di i per il mese t (€)
a_i	$\in \mathbb{Q}_{> 0}$	ore macchina per produrre un'unità del prodotto i
b_t	$\in \mathbb{Q}_{\geq 0}$	ore macchina disponibili nel mese t
s_{i0}	$\in \mathbb{Q}_{\geq 0}$	scorta iniziale del prodotto i (unità)

Variabili decisionali

Introduciamo le seguenti $2nm$ variabili non negative:

$$\begin{cases} x_{it} = \text{unità del prodotto } i \text{ prodotte nel mese } t \\ s_{it} = \text{unità del prodotto } i \text{ a scorta alla fine del mese } t \end{cases} \quad \forall i \in \{1, 2, \dots, n\}, \forall t \in \{1, 2, \dots, m\}.$$

Usando queste variabili, un modello LP per il problema è il seguente:

Modello LP di produzione e scorte a costo minimo

$$\min \sum_{i=1}^n \sum_{t=1}^m (c_{it} x_{it} + h_{it} s_{it}) \quad (4.1a)$$

$$\text{soggetto a } s_{i,t-1} + x_{it} = d_{it} + s_{it}, \quad \forall i \in \{1, 2, \dots, n\}, \forall t \in \{1, 2, \dots, m\}, \quad (4.1b)$$

$$\sum_{i=1}^n a_i x_{it} \leq b_t, \quad \forall t \in \{1, 2, \dots, m\}, \quad (4.1c)$$

$$x_{it} \geq 0, \quad \forall i \in \{1, 2, \dots, n\}, \forall t \in \{1, 2, \dots, m\}, \quad (4.1d)$$

$$s_{it} \geq 0, \quad \forall i \in \{1, 2, \dots, n\}, \forall t \in \{1, 2, \dots, m\}. \quad (4.1e)$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (4.1a) minimizza il costo totale del piano, somma dei costi di produzione $c_{it} x_{it}$ e di giacenza $h_{it} s_{it}$ su tutti i prodotti e i mesi;
- i vincoli lineari (4.1b) sono i bilanci delle scorte: per ogni prodotto e ogni mese, ciò che entra (scorta precedente più produzione) è uguale a ciò che esce (domanda servita più scorta successiva); per $t = 1$ si usa la scorta iniziale s_{i0} (nm vincoli lineari);
- i vincoli lineari (4.1c) assicurano che, in ogni mese, le ore macchina usate da tutti i prodotti non superino quelle disponibili (m vincoli lineari);
- i vincoli (4.1d) e (4.1e) definiscono le variabili del modello.

Spiegazione riga per riga

Perché il bilancio (4.1b) è il cuore del modello. È il vincolo che *lega i mesi tra loro*: senza di esso il problema si spezzerebbe in m problemi indipendenti e le scorte non avrebbero senso.

Cosa succede se... togliamo il bilancio? Il modello produce zero (costo minimo!) perché nulla lo obbliga a servire la domanda. Se la domanda totale supera la capacità cumulata, il modello diventa inammissibile: in tal caso o si consente la domanda persa con una variabile $u_{it} \geq 0$ e una penalità π_{it} in obiettivo, oppure si rivede la capacità.

4.3 Esempio svolto a mano

Un prodotto, due mesi

Domanda $d_1 = 60$, $d_2 = 100$; capacità $b_1 = b_2 = 80$ unità/mese ($a = 1$); costo di produzione $c = 10$ €, giacenza $h = 1$ €/unità/mese, scorta iniziale nulla.

- **Passo 1 — è ammissibile?** Domanda totale $160 \leq$ capacità totale 160: sì, ma solo producendo al massimo in entrambi i mesi.
- **Passo 2 — bilanci.** $x_1 = 60 + s_1$ e $s_1 + x_2 = 100$. Con $x_2 \leq 80$ serve $s_1 \geq 20$, quindi $x_1 \geq 80$: ma $x_1 \leq 80$, dunque $x_1 = 80$, $s_1 = 20$, $x_2 = 80$.
- **Passo 3 — costo.** $10 \cdot (80 + 80) + 1 \cdot 20 = 1620$ €. La soluzione è *forzata*: entrambi i vincoli di capacità sono attivi.
- **Passo 4 — prezzo ombra a mano.** Con un'ora in più nel mese 2 ($b_2 = 81$): $x_2 = 81$, $s_1 = 19$, $x_1 = 79$: il costo scende di 1 € (un'unità non sta più a scorta un mese). Il duale della capacità del mese 2 è dunque -1 €/unità. Con un'ora in più nel mese 1 invece non cambia nulla (il mese 1 produce già il necessario): duale 0. Questo meccanismo — *la capacità di un mese vale quanto giacenza fa risparmiare* — riappare identico nel caso di studio.

4.4 Caso di studio

Tre prodotti ($n = 3$) su sei mesi ($m = 6$), con domanda stagionale crescente (picco nei mesi 4–5), un'unica risorsa condivisa e questi parametri (file CSV `produzione_domanda` e `produzione_capacita` nella cartella `dati`):

Prodotto	c_i (€/u)	h_i (€/u/mese)	a_i (ore/u)	s_{i0}
1	12,00	0,80	0,9	30
2	18,00	1,20	1,4	20
3	25,00	1,60	2,1	10

mese	1	2	3	4	5	6
b_t (ore)	420	420	460	460	460	420
$d_{1,t}$	110	130	150	190	210	160
$d_{2,t}$	70	80	110	140	150	100
$d_{3,t}$	50	55	60	80	95	70

4.5 Implementazione

Il cuore dello script (versione integrale in lab04_produzione.py):

```

1 m = gp.Model("produzione_scorde")
2 x = m.addVars(prodotti, mesi, name="x")      # quantità prodotta
3 s = m.addVars(prodotti, mesi, name="s")      # scorta a fine mese
4
5 # bilancio: scorta entrante + produzione
6 #           = domanda + scorta uscente
7 m.addConstrs(
8     ((s0[i] if t == 1 else s[i, t-1]) + x[i, t]
9      == domanda[i, t] + s[i, t]
10      for i in prodotti for t in mesi), name="bilancio")
11
12 # capacità condivisa (conservo i vincoli per i duali)
13 v_cap = m.addConstrs(
14     (gp.quicksum(a[i] * x[i, t] for i in prodotti)
15      <= capacita[t] for t in mesi), name="capacita")
16
17 m.setObjective(
18     gp.quicksum(c[i]*x[i, t] + h[i]*s[i, t]
19                for i in prodotti for t in mesi),
20     GRB.MINIMIZE)
21 m.optimize()

```

4.6 Risultati

Costo totale ottimo: 32.889,33 EUR

Piano di produzione (unita'/mese):

mese	1	2	3	4	5	6
prod.1	80.0	130.0	150.0	190.0	210.0	160.0
prod.2	50.0	80.0	110.0	140.0	150.0	100.0
prod.3	89.5	91.0	81.4	44.3	29.0	64.8

Scorte a fine mese:

mese	1	2	3	4	5	6
prod.1	0.0	0.0	0.0	0.0	0.0	0.0
prod.2	0.0	0.0	0.0	0.0	0.0	0.0
prod.3	49.5	85.5	106.9	71.2	5.2	0.0

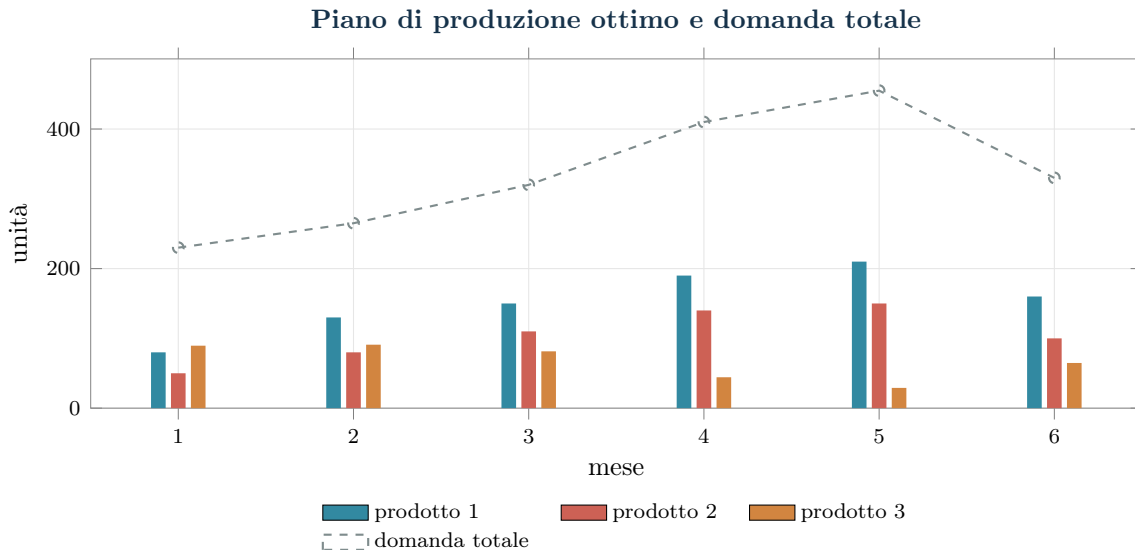


Figura 4.1: Piano di produzione ottimo dell'LP: barre = unità prodotte dei tre prodotti in ciascun mese; linea tratteggiata = domanda totale del mese. Nei mesi di picco (4–5) la produzione totale resta sotto la domanda: la differenza è coperta dalle scorte del prodotto 3 accumulate nei primi mesi.

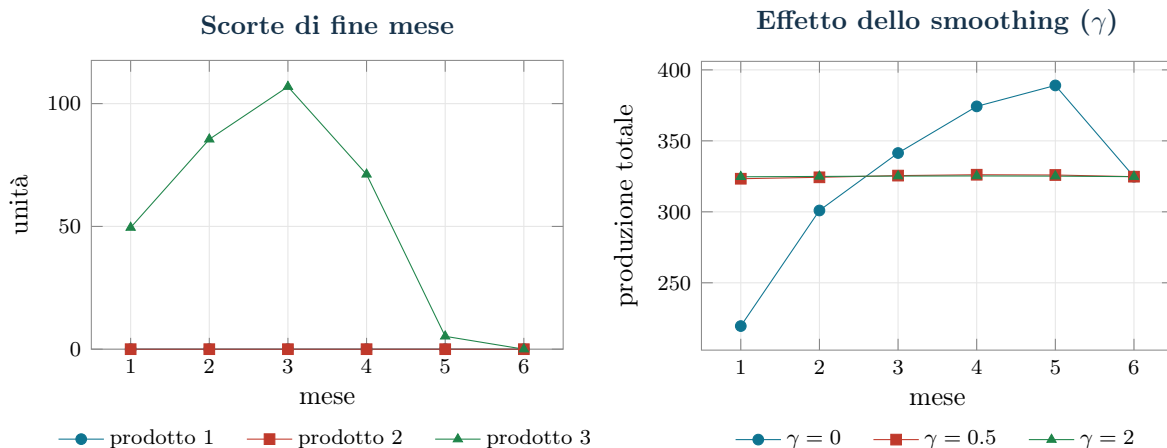


Figura 4.2: A sinistra: scorte di fine mese per prodotto — solo il prodotto 3 usa il magazzino (strategia di pre-build), e le scorte si azzerano nel mese finale. A destra: produzione totale mensile per tre valori del peso di smoothing γ : già con $\gamma = 0,5$ il profilo diventa quasi piatto.

Insight manageriale

Il solver ha scoperto da solo la strategia del *pre-build*: i prodotti 1 e 2 (margin di giacenza alti rispetto alle ore assorbite) vengono prodotti su domanda, mentre il prodotto 3 — il più oneroso in ore macchina — viene prodotto in anticipo nei mesi 1–3 e accumulato (fino a 107 unità a scorta) per liberare capacità nei mesi di picco. Le scorte tornano a zero nel mese finale: una scorta terminale positiva sarebbe denaro sprecato, e il modello lo sa.

4.7 Analisi di sensitività

Prezzi ombra della capacità

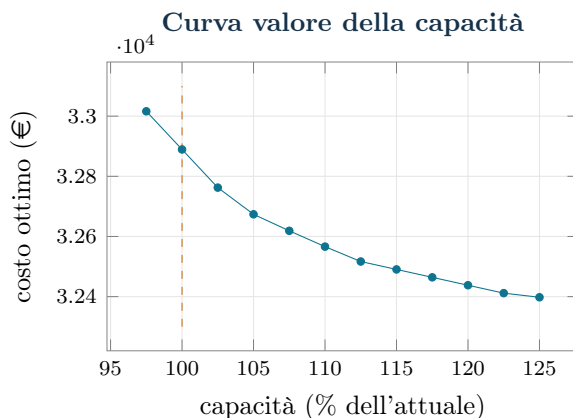
mese 1: uso	330.0/420	Pi =	0.000 EUR/ora	valido per b_t in	[330.0, +inf]
mese 2: uso	420.0/420	Pi =	-0.762 EUR/ora	valido per b_t in	[330.0, 524.0]
mese 3: uso	460.0/460	Pi =	-1.524 EUR/ora	valido per b_t in	[370.0, 564.0]
mese 4: uso	460.0/460	Pi =	-2.286 EUR/ora	valido per b_t in	[370.0, 564.0]
mese 5: uso	460.0/460	Pi =	-3.048 EUR/ora	valido per b_t in	[399.0, 564.0]
mese 6: uso	420.0/420	Pi =	-3.810 EUR/ora	valido per b_t in	[330.0, 431.0]

Verifica: +1 ora nel mese 6 -> costo da 32889.33 a 32885.52 (delta -3.810)

Da dove viene il numero 0,762

I duali crescono di esattamente 0,762 € al mese. Non è un caso: $0,762 = h_3/a_3 = 1,6/2,1$. Un'ora in più nel mese t permette di produrre $1/a_3$ unità del prodotto 3 *un mese più tardi*, risparmiando un mese di giacenza: $h_3 \cdot (1/a_3)$. Il duale del mese 6 vale $5 \times 0,762 = 3,81$ perché un'ora lì evita cinque mesi di giacenza a un'unità prodotta nel mese 1. Quando il solver e un ragionamento economico elementare danno lo stesso numero, il modello è probabilmente giusto.

Il valore della capacità e il costo della regolarità



La curva è **convessa e decrescente**: le prime ore extra valgono molto, poi sempre meno. Sotto il 97% dell'attuale capacità il modello diventa *inammissibile*: la domanda non è più servibile per intero — informazione gestionale preziosa quanto l'ottimo stesso.

La variante QP aggiunge in obiettivo $\gamma \sum_{t=2}^m (v_t - v_{t-1})^2$ con $v_t = \sum_{i=1}^n x_{it}$: con $\gamma = 0,5$ il profilo produttivo diventa quasi piatto (variazione massima da 81 a 1 unità/mese) al costo di appena $33.003 - 32.889 = 114$ € (+0,35%).

Figura 4.3: Costo totale ottimo in funzione della capacità disponibile (in % dell'attuale; la linea tratteggiata arancione segna la capacità corrente): curva convessa e decrescente, interrotta a sinistra dove il problema diventa inammissibile.

Insight manageriale

Raccomandazione tipo da relazione di laboratorio: “Conviene acquistare ore straordinarie nei mesi 5–6 fino a 3–3,8 €/ora sotto il prezzo interno; un piano livellato costa solo lo 0,35% in più e riduce drasticamente i cambi di ritmo; sotto il 97% della capacità attuale il servizio completo non è garantibile.”

4.8 Esercizi

Esercizio 4.1 — verifica

Verificare per perturbazione i duali dei mesi 3 e 4 (aggiungere un'ora e riottimizzare). Perché il duale del mese 1 è nullo?

Esercizio 4.2 — domanda persa

Aggiungere la variabile u_{it} (domanda non servita) con penalità $\pi = 40$ €/unità e ridurre la capacità al 90%: quanta domanda si perde, di quale prodotto, in quali mesi? Interpretare la scelta del modello.

Esercizio 4.3 — massimo profitto

Trasformare il modello in massimizzazione del profitto con ricavi $p = (20, 30, 42)$ € e servizio facoltativo ($y_{it} \leq d_{it}$). Quale domanda conviene *non* servire, e perché?

Esercizio 4.4 — emissioni

Ogni unità del prodotto 3 emette 2,5 kg di CO₂, i prodotti 1 e 2 un kg. Aggiungere il termine $\tau \sum_{i=1}^n \sum_{t=1}^m e_i x_{it}$ all'obiettivo e tracciare la frontiera costo-emissioni per $\tau \in [0, 10]$ €/kg.

Esercizio 4.5 — scorta di sicurezza

Imporre $s_{it} \geq 0,1 d_{i,t+1}$ (scorta di sicurezza del 10%). Quanto costa questa politica? Quale vincolo di capacità diventa più critico?

Supply chain con congestione e sostenibilità

Classe: LP / NLP convesso

Script: [python/lab05_supplychain.py](#)

[Pagina online](#)

5.1 Motivazione gestionale

I prodotti attraversano una rete di stabilimenti, hub e mercati. Ogni tratta ha un costo, una capacità e un'impronta di CO₂; il costo effettivo cresce rapidamente quando una tratta si avvicina alla saturazione. Tre domande manageriali: come instradare i flussi al minimo costo? cosa cambia se penalizziamo la congestione? quanto deve valere la CO₂ perché le rotte “pulite” diventino convenienti?

È il problema quotidiano di chi gestisce distribuzione e logistica: scegliere le rotte, decidere quanto caricare ciascuna tratta, valutare investimenti in capacità e, sempre più spesso, rendere conto dell'impronta carbonica delle spedizioni. La struttura matematica — il flusso su rete — è tra le più riutilizzabili di tutta la ricerca operativa.

In questo capitolo impareremo a: (1) formulare un problema di flusso a costo minimo con vincoli di conservazione; (2) modellare la congestione con termini quadratici convessi; (3) tracciare una frontiera costo-emissioni con un prezzo interno della CO₂; (4) capire perché le soluzioni degli LP “saltano” tra vertici.

5.2 Costruzione guidata del modello

Il problema a parole. *Decidiamo* quante unità far viaggiare su ciascuna tratta della rete. *L'obiettivo* è minimizzare il costo totale di trasporto (nelle varianti: più congestione o CO₂). *I vincoli* impongono che in ogni nodo i flussi tornino — gli stabilimenti non spediscono più di quanto producono, gli hub non trattengono nulla, i mercati ricevono esattamente la loro domanda — e che nessuna tratta superi la propria capacità.

Dati (input del modello)

Simbolo	Tipo	Significato
$G = (N, A)$	grafo diretto	nodi N e archi $A \subseteq N \times N$; scriviamo $ N $ e $ A $ per il numero di nodi e di archi
b_i	$\in \mathbb{Q}$	saldo del nodo $i \in N$: offerta se > 0 , domanda se < 0 , transito se $= 0$
u_{ij}	$\in \mathbb{Q}_{>0}$	capacità dell'arco $(i, j) \in A$ (unità)
c_{ij}	$\in \mathbb{Q}_{\geq 0}$	costo di trasporto di un'unità sull'arco (i, j) (€)
e_{ij}	$\in \mathbb{Q}_{\geq 0}$	emissioni per unità trasportata su (i, j) (kg CO ₂)

Variabili decisionali

Introduciamo una variabile non negativa per ciascuno degli $|A|$ archi:

$$x_{ij} = \text{unità di prodotto inviate sull'arco } (i, j), \quad \forall (i, j) \in A.$$

Usando queste variabili, un modello LP per il problema è il seguente:

Modello LP di flusso a costo minimo

$$\min \sum_{(i,j) \in A} c_{ij} x_{ij} \quad (5.1a)$$

$$\text{soggetto a } \sum_{j: (i,j) \in A} x_{ij} - \sum_{j: (j,i) \in A} x_{ji} = b_i, \quad \forall i \in N, \quad (5.1b)$$

$$x_{ij} \leq u_{ij}, \quad \forall (i, j) \in A, \quad (5.1c)$$

$$x_{ij} \geq 0, \quad \forall (i, j) \in A. \quad (5.1d)$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (5.1a) minimizza il costo totale di trasporto: ogni unità che attraversa l'arco (i, j) paga il costo unitario c_{ij} ;
- i vincoli lineari (5.1b) sono le conservazioni del flusso: in ogni nodo, ciò che esce meno ciò che entra è pari al saldo b_i — positivo negli stabilimenti, negativo nei mercati, nullo negli hub ($|N|$ vincoli lineari); la somma dei saldi su tutti i nodi deve essere ≥ 0 , altrimenti il modello è inammissibile;
- i vincoli lineari (5.1c) assicurano che nessuna tratta trasporti più della propria capacità ($|A|$ vincoli lineari);
- i vincoli (5.1d) definiscono le variabili del modello.

Varianti sull'obiettivo (i vincoli non cambiano):

$$\underbrace{\sum_{(i,j) \in A} (c_{ij} x_{ij} + \alpha c_{ij} x_{ij}^2 / u_{ij})}_{\text{congestione (convessa)}} \quad \underbrace{\sum_{(i,j) \in A} (c_{ij} + \tau e_{ij}) x_{ij}}_{\text{prezzo interno CO}_2} \quad \underbrace{\min z : x_{ij} / u_{ij} \leq z, \quad \forall (i, j) \in A}_{\text{minimax utilizzo}}$$

dove $\alpha \in \mathbb{Q}_{\geq 0}$ pesa la congestione e $\tau \in \mathbb{Q}_{\geq 0}$ (€/kg) è il prezzo interno della CO₂.

Spiegazione riga per riga

Congestione. Il termine $\alpha c_{ij} x_{ij}^2 / u_{ij}$ è convesso e cresce con l'utilizzo dell'arco: rappresenta straordinari, code ai terminali, premi assicurativi. Il suo effetto è *ripartire* i flussi su più rotte invece di saturare la più economica.

Prezzo della CO₂. Non serve un vincolo sulle emissioni: basta aggiungerle al costo con il prezzo interno τ . Variando τ si traccia l'intera frontiera costo-emissioni (approccio "a peso"; l'alternativa "a vincolo" impone $\sum_{(i,j) \in A} e_{ij} x_{ij} \leq \bar{e}$ e legge il duale).

5.3 Esempio svolto a mano

Due rotte in concorrenza

Un mercato chiede 100 unità servibili da due rotte: rotta 1 con $c_1 = 2 \text{ €}$ e $u_1 = 80$; rotta 2 con $c_2 = 5 \text{ €}$ e capacità abbondante.

LP puro. Si riempie la rotta economica e si trabocca sull'altra: $x_1 = 80$, $x_2 = 20$, costo $2 \cdot 80 + 5 \cdot 20 = 260 \text{ €}$. Prezzo ombra della domanda: servire la 101-esima unità costerebbe 5 € (la rotta 1 è saturata); prezzo ombra della capacità u_1 : +1 unità di capacità fa risparmiare $5 - 2 = 3 \text{ €}$.

Con congestione ($\alpha = 1$): costo rotta 1 = $2x_1 + 2x_1^2/80$. Il costo marginale della rotta 1, $2 + 4x_1/80$, va eguagliato a quello della rotta 2 (altrimenti conviene spostare flusso): $2 + x_1/20 = 5 \Rightarrow x_1 = 60$, $x_2 = 40$. La rotta economica *non viene più saturata*: il modello lascia margine di sicurezza, proprio come farebbe un pianificatore prudente.

5.4 Caso di studio

Rete con 2 stabilimenti (S1, S2), 2 hub (H1, H2) e 4 mercati (M1–M4); 12 archi con capacità, costi ed emissioni *non* proporzionali ai costi: le tratte economiche sono "su strada" (inquinanti), quelle costose "su ferro" (pulite). Dati in [dati/supplychain_archi.csv](#); offerta (260, 240), domanda totale 450.

```

1 m = gp.Model("supply_chain")
2 x = m.addVars(A, name="x", ub=U) # ub direttamente sull'arco
3 m.addConstrs((x.sum(s, "*") <= offerta[s] for s in offerta))
4 m.addConstrs((x.sum("h", "*") == x.sum(h, "*") for h in hub))
5 v_dom = m.addConstrs((x.sum("k", "*") == domanda[k]
6                       for k in domanda), name="domanda")
7 obj = gp.quicksum((c[a] + tau * e[a]) * x[a] for a in A)
8 if congestione > 0: # termine quadratico convesso
9     obj += gp.quicksum(congestione * c[a] * x[a] * x[a] / U[a]
10                      for a in A)
11 m.setObjective(obj, GRB.MINIMIZE)

```

5.5 Risultati

LP costo minimo : costo 3.385,00 EUR emissioni 2.730 kg utilizzo max 100%
 archi saturi: S1->H1, S2->H2, H2->M3
 prezzi ombra della domanda: M1 8,00 M2 9,50 M3 10,00 M4 10,50 EUR/unità'

```

costi ridotti archi a zero (soglia = SAObjLow):
S2->H1 +2,00 (soglia 5,00)   H1->M4 +0,50 (soglia 5,50)
H2->M1 +4,50 (soglia 1,50)   H2->M2 +1,00 (soglia 3,00)   EUR/unità
Congestione a=1: costo 3.702,81 EUR   emissioni 2.530 kg   utilizzo max 90%
Minimax      : utilizzo massimo minimo possibile 58,4%

```

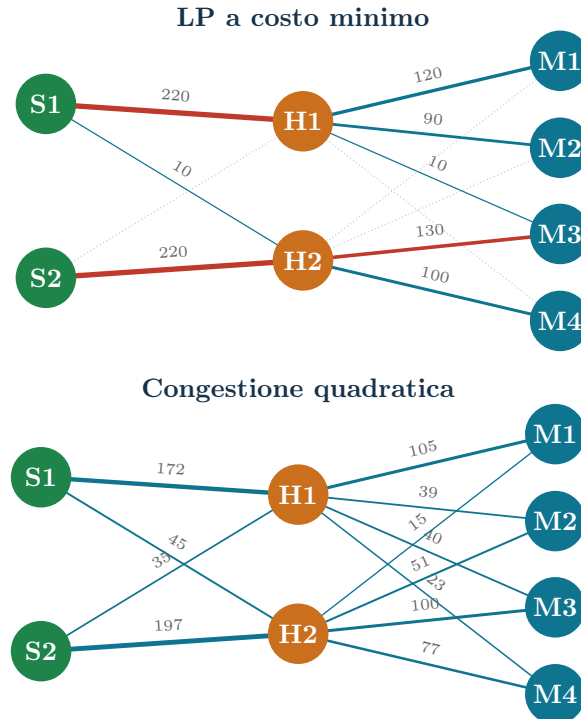


Figura 5.1: I flussi ottimi sulla rete (stabilimenti S verdi, hub H arancioni, mercati M teal; lo spessore degli archi è proporzionale al flusso, il rosso indica un arco saturo, il numero è il flusso in unità). Sopra: LP a costo minimo, che satura tre archi. Sotto: modello con congestione quadratica, che ripartisce i flussi su più rotte e non satura alcun arco.

Insight manageriale

I prezzi ombra della domanda (8–10,50 €/unità) sono i *costi marginali di servizio* dei quattro mercati: è il numero da confrontare con il margine commerciale prima di accettare nuovi ordini, e la base corretta per un prezzo di trasferimento interno. M4 è il mercato più caro da servire: +31% rispetto a M1.

I costi ridotti degli archi non usati, con il loro range

Quattro archi restano a zero e il loro costo ridotto (attributo $x[a].RC$) è la loro *soglia di convenienza*: la rotta $S2 \rightarrow H1$ (costo 7 €/unità, $RC = +2$) entrerebbe nella soluzione ottima solo se il suo costo scendesse sotto $7 - 2 = 5$ €/unità — il numero da portare al tavolo in una trattativa con il vettore. Come sempre, accanto al costo ridotto citiamo il suo range di validità $SAObjLow-SAObjUp$: per una variabile a zero è $[soglia, +\infty)$, cioè il costo dell'arco può crescere quanto vuole (resta inutilizzato) e la soglia stessa è $SAObjLow$. La rotta $H1 \rightarrow M4$ è a un passo dall'essere usata ($RC = +0,50$): basta uno sconto di 50 centesimi. Le variabili positive hanno tutte costo ridotto nullo.

5.6 Analisi di sensitività: il prezzo della CO₂

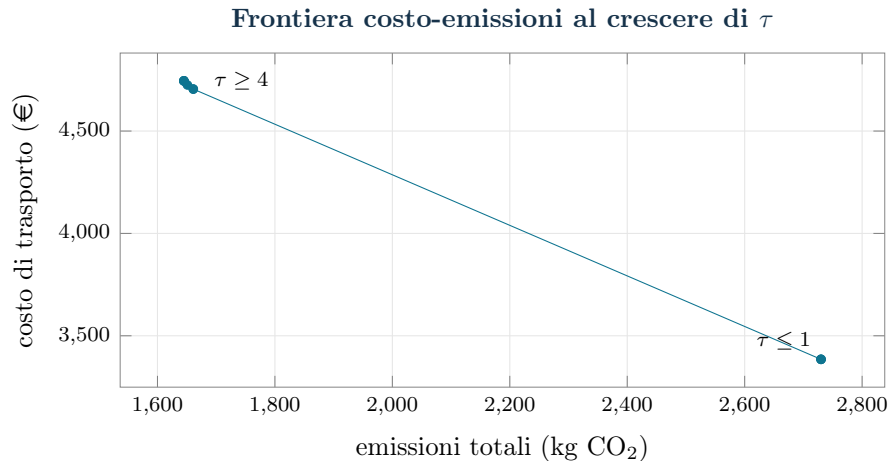


Figura 5.2: Frontiera costo-emissioni: ogni punto è la soluzione ottima per un prezzo interno della CO₂ (τ); in basso a destra l'assetto "su strada" (economico, 2730 kg), in alto a sinistra l'assetto "su ferro" (caro, 1645 kg). La frontiera procede per salti tra vertici dell'LP.

$\tau = 0.0 \dots 1.0$	: costo 3385	emissioni 2730	(rotte su strada)
$\tau = 1.24$	: costo 4025	emissioni 2173	(vertice intermedio)
$\tau = 1.26$	: costo 4265	emissioni 1981	
$\tau = 1.5$	: costo 4705	emissioni 1661	(assetto "ferro")
$\tau = 4.0$ e oltre	: costo 4745	emissioni 1645	(stabile)

Perché la frontiera procede a salti

Le soluzioni di un LP stanno nei *vertici* del poliedro ammissibile: al crescere di τ l'ottimo resta fermo su un vertice, poi *salta* al successivo. Una bisezione sul modello mostra che fino a $\tau = 1$ non cambia nulla, che tra $\tau = 1,0$ e $\tau = 1,5$ la rete attraversa vertici intermedi (a $\tau = 1,24$: 2173 kg) e che il salto principale avviene esattamente a $\tau = 1,25$ €/kg. Managerialmente: un prezzo interno della CO₂ sotto la soglia non cambia le rotte; appena sopra, ristruttura la rete in blocco. Il modello con congestione, strettamente convesso, si sposterebbe invece con continuità.

Il minimax "puro" ($\min z$) è degenere: qualunque instradamento con utilizzo $\leq z^*$ è ottimo, anche costosissimo. In pratica si usa sempre una combinazione (costo $+\rho \cdot z$) o si vincola z e si minimizza il costo.

5.7 Esercizi

Esercizio 5.1 — verifica

Verificare per perturbazione il prezzo ombra della domanda di M4 (portarla a 101) e quello della capacità dell'arco H2→M3.

Esercizio 5.2 — guasto di un hub

Ridurre del 50% le capacità degli archi entranti e uscenti da H2: quanto costa il guasto? quali mercati soffrono di più? (Confrontare i prezzi ombra prima/dopo.)

Esercizio 5.3 — soglia del ferro

Riprodurre con una bisezione su τ la soglia $\tau = 1,25$ €/kg trovata nel testo, ed elencare quali archi cambiano flusso attraversandola. Confrontare la soglia con il prezzo ETS corrente della CO₂.

Esercizio 5.4 — vincolo sulle emissioni

Riformulare con vincolo $\sum_{(i,j) \in A} e_{ij} x_{ij} \leq \bar{e}$ per $\bar{e} = 2000$ kg e verificare che il duale del vincolo coincide con il τ “a peso” che produce le stesse emissioni.

Esercizio 5.5 — barriera

Sostituire il termine di congestione con la barriera $\alpha x_{ij}/(u_{ij} - x_{ij})$ e confrontare: quale formulazione tiene i flussi più lontani dalla saturazione?

Portafoglio di Markowitz

Classe: QP convesso

Script: [python/lab06_markowitz.py](#)

[Pagina online](#)

6.1 Motivazione gestionale

Quanto investire in ciascun titolo per bilanciare rendimento atteso e rischio? Il modello di Markowitz è il QP più famoso della storia: minimizza la varianza del portafoglio imponendo un rendimento minimo. Il risultato non è un numero ma una *frontiera efficiente*: il menu completo dei compromessi rischio-rendimento tra cui il decisore sceglie in base alla propria avversione al rischio.

Il problema nasce in finanza (Markowitz, 1952, premio Nobel) ma la sua struttura — allocare un budget tra alternative rischiose e correlate — riappare ovunque: mix di progetti di R&S, portafogli di clienti, perfino l’allocazione di capacità produttiva tra mercati volatili. È anche il ponte naturale tra l’ottimizzazione e la statistica: i dati del modello sono *stime*, con tutte le conseguenze del caso.

In questo capitolo impareremo a: (1) formulare e risolvere un QP convesso; (2) costruire e leggere la frontiera efficiente; (3) capire l’effetto dei vincoli di mandato; (4) diffidare delle stime dei rendimenti attesi.

6.2 Costruzione guidata del modello

Il problema a parole. *Decidiamo* la quota di capitale da investire in ciascun titolo. *L’obiettivo* è minimizzare il rischio del portafoglio, misurato dalla varianza del suo rendimento. *I vincoli*: investire tutto il capitale (le quote sommano a 1), garantire un rendimento atteso minimo e rispettare le quote minime e massime per titolo.

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di titoli; i titoli sono indicizzati da $i \in \{1, 2, \dots, n\}$
μ_i	$\in \mathbb{Q}$	rendimento atteso del titolo i (per anno)
q_{ij}	$\in \mathbb{Q}$	covarianza tra i rendimenti dei titoli i e j ; la matrice $\mathbf{Q} = (q_{ij})$ è semidefinita positiva ($\mathbf{Q} \succeq 0$)
$\bar{r}$	$\in \mathbb{Q}$	rendimento atteso minimo richiesto
ℓ_i, u_i	$\in \mathbb{Q}, 0 \leq \ell_i \leq u_i \leq 1$	quota minima e massima investibile nel titolo i

Variabili decisionali

Introduciamo le seguenti n variabili non negative:

$$x_i = \text{quota di capitale investita nel titolo } i, \quad \forall i \in \{1, 2, \dots, n\}.$$

Usando queste variabili, un modello QP per il problema è il seguente:

Modello QP di minima varianza con rendimento minimo

$$\min \sum_{i=1}^n \sum_{j=1}^n q_{ij} x_i x_j \quad (6.1a)$$

$$\text{soggetto a } \sum_{i=1}^n \mu_i x_i \geq \bar{r}, \quad (6.1b)$$

$$\sum_{i=1}^n x_i = 1, \quad (6.1c)$$

$$x_i \leq u_i, \quad \forall i \in \{1, 2, \dots, n\}, \quad (6.1d)$$

$$x_i \geq \ell_i, \quad \forall i \in \{1, 2, \dots, n\}. \quad (6.1e)$$

Descrizione della funzione obiettivo e dei vincoli del modello QP:

- la funzione obiettivo quadratica (6.1a) minimizza la varianza del rendimento del portafoglio; contiene le covarianze q_{ij} , ed è qui che nasce la diversificazione: combinare titoli poco correlati abbassa il rischio complessivo sotto quello dei singoli componenti;
- il vincolo lineare (6.1b) impone che il rendimento atteso del portafoglio raggiunga la soglia $\bar{r}$: è la “manopola” con cui si percorre la frontiera efficiente, e il suo moltiplicatore dice quanta varianza costa ogni punto di rendimento in più (un vincolo lineare);
- il vincolo lineare (6.1c) impone di investire tutto il capitale (un vincolo lineare);
- i vincoli lineari (6.1d) sono i tetti per titolo, tipici vincoli regolamentari o di mandato (n vincoli lineari);
- i vincoli (6.1e) definiscono le variabili del modello (con $\ell_i = 0$: divieto di vendita allo scoperto).

Formulazioni equivalenti: $\max \sum_{i=1}^n \mu_i x_i - \lambda \sum_{i=1}^n \sum_{j=1}^n q_{ij} x_i x_j$ (media-varianza) e $\max \sum_{i=1}^n \mu_i x_i$ con varianza $\leq \bar{\sigma}^2$ (massimo rendimento a rischio limitato). Variando $\bar{r}$ (o λ , o $\bar{\sigma}$) si percorre la stessa frontiera.

Spiegazione riga per riga

Convessità. Ogni matrice di covarianza è semidefinita positiva, quindi il QP è convesso e l'ottimo trovato è *globale certificato*.

6.3 Esempio svolto a mano

Due titoli non correlati

Volatilità $\sigma_1 = 20\%$, $\sigma_2 = 30\%$, correlazione nulla. Minimizzare la varianza con $x_1 + x_2 = 1$ (senza vincolo di rendimento).

- **Passo 1.** Sostituendo $x_2 = 1 - x_1$: $f(x_1) = 0,04 x_1^2 + 0,09 (1 - x_1)^2$.
- **Passo 2.** $f'(x_1) = 0,08 x_1 - 0,18 (1 - x_1) = 0 \Rightarrow x_1 = \frac{0,09}{0,04 + 0,09} = \frac{9}{13} \approx 69,2\%$.
- **Passo 3.** Varianza del portafoglio: $f(9/13) = 0,04 \cdot 0,479 + 0,09 \cdot 0,095 = 0,0277$, cioè volatilità 16,6%: **meno rischioso di entrambi i titoli**. È l'effetto diversificazione nella sua forma più pura: i pesi ottimi sono inversamente proporzionali alle varianze.

6.4 Caso di studio

Otto ETF settoriali (*Exchange Traded Fund*: fondi quotati che replicano un paniere di titoli di un settore), 60 rendimenti mensili simulati con un modello a un fattore di mercato (file [dati/markowitz_rendimenti.csv](#)); μ e Q sono *stimati* da questi dati, come nella pratica. Statistiche annualizzate:

ENE: mu = 19,12%	vol = 18,65%	SAN: mu = 2,80%	vol = 11,93%
FIN: mu = 8,27%	vol = 19,59%	CON: mu = 11,25%	vol = 11,67%
TEC: mu = 1,58%	vol = 29,27%	UTL: mu = 11,99%	vol = 7,53%
IND: mu = 9,97%	vol = 16,60%	MAT: mu = 20,93%	vol = 23,93%

```

1 m = gp.Model("markowitz")
2 x = m.addVars(n, ub=u, name="x")          # quote, 0 <= x_i <= u
3 m.addConstr(x.sum() == 1, name="budget")
4 m.addConstr(gp.quicksum(mu[i]*x[i] for i in range(n)) >= r_min)
5 m.setObjective(gp.quicksum(Q[i, j]*x[i]*x[j]
6                   for i in range(n) for j in range(n)),
7                 GRB.MINIMIZE)
8 m.optimize()
```

6.5 Risultati

Minima varianza globale : rendimento 11,06%, volatilità 6,03%
 Equipesato (1/n) : rendimento 10,74%, volatilità 10,13%
 Composizione min varianza: ENE 5,2% IND 1,9% SAN 11,5% CON 22,3% UTL 58,3%

Frontiera efficiente: la diversificazione domina i titoli singoli

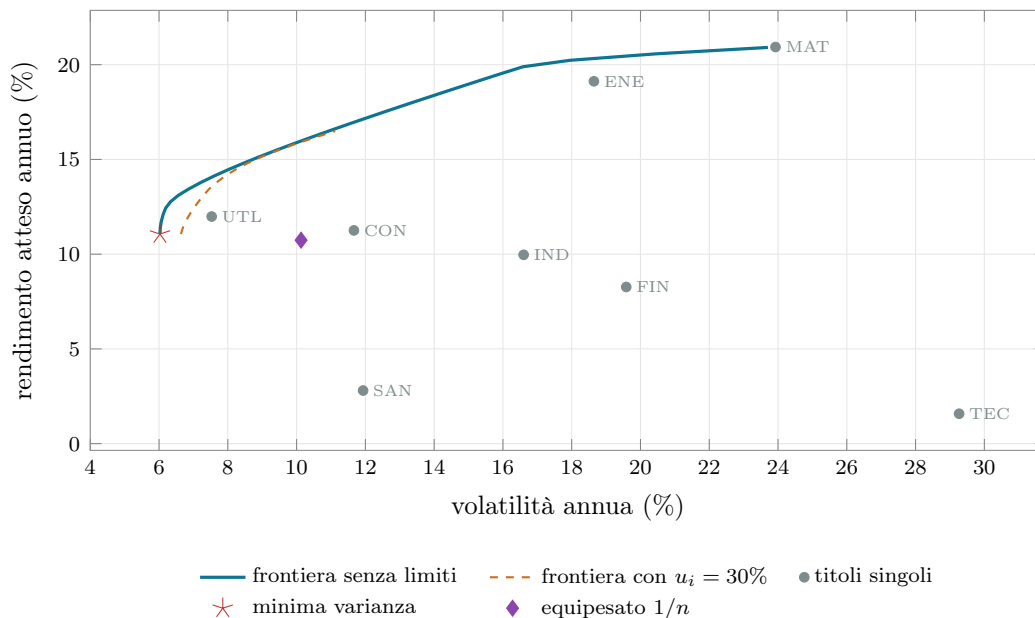


Figura 6.1: Piano rischio-rendimento: gli otto titoli (punti grigi), la frontiera efficiente senza limiti (linea teal), la frontiera con tetto del 30% per titolo (tratteggiata arancione), il portafoglio di minima varianza (stella rossa) e l'equipesato 1/n (rombo viola), che risulta dominato.

Composizione ottima lungo la frontiera

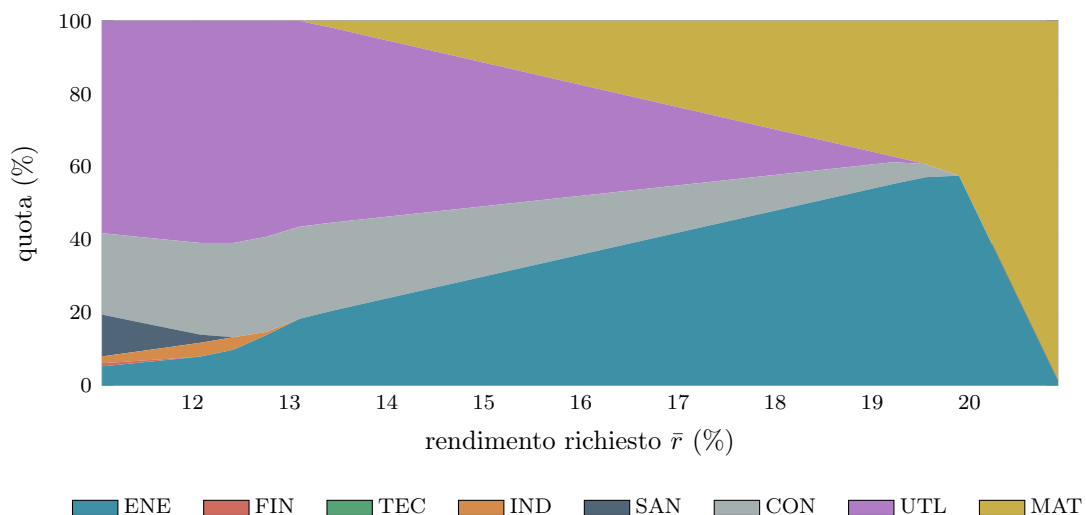


Figura 6.2: Composizione del portafoglio ottimo (aree impilate, in %) al crescere del rendimento richiesto $\bar{r}$: a sinistra domina il mix difensivo (UTL, CON, SAN), verso destra il portafoglio si concentra progressivamente sui titoli ad alto rendimento (MAT, ENE) fino al portafoglio monotitolo all'estremo.

Insight manageriale

Tre messaggi dalla figura. (1) Il portafoglio di minima varianza (volatilità 6%) è molto meno rischioso del miglior titolo singolo (UTL, 7,5%) pur rendendo l'11%. (2) L'equipesato $1/n$ — la strategia “non so nulla” — è nettamente dominato: stessa area di rendimento ma quasi il doppio della volatilità. (3) Il tetto $u_i = 30\%$ taglia la parte alta della frontiera: il costo dei vincoli di mandato si *vede* come distanza tra le due curve.

6.6 Analisi di sensitività

```
r_min = 6%: vol 6,03% (vincolo NON attivo: coincide col min varianza)
r_min = 8%: vol 6,03% (idem)
r_min = 10%: vol 6,03% (idem)
r_min = 12%: vol 6,10% d(varianza)/d(r_min) ~ 0,0221
```

Un vincolo che non costa nulla

Fino a $\bar{r} = 11\%$ il vincolo di rendimento è *inattivo*: il portafoglio di minima varianza rende già l'11,06%, quindi chiedere “almeno l'8%” non costa nulla e il moltiplicatore è zero. Solo oltre l'11,06% il vincolo morde e ogni punto di rendimento in più si paga in varianza (moltiplicatore $\approx 0,022$ a $\bar{r} = 12\%$). Riconoscere i vincoli inattivi evita di “pagare” vincoli gratis.

Nei dati simulati il titolo TEC ha α vero dell'11% annuo, ma su 60 mesi il rendimento *stimato* è 1,6%: il rumore domina. È la lezione pratica più importante di Markowitz: **le stime dei rendimenti attesi sono molto più instabili di quelle delle covarianze**, e i portafogli ottimizzati inseguono gli errori di stima. Rimedi usati in pratica: vincoli u_i , shrinkage delle stime, oppure ottimizzare solo il rischio (minima varianza).

6.7 Esercizi

Esercizio 6.1 — verifica

Verificare a mano l'esempio dei due titoli con $\rho = 0,5$: la formula diventa $x_1 = (\sigma_2^2 - \rho\sigma_1\sigma_2)/(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)$. La diversificazione conviene ancora?

Esercizio 6.2 — frontiera

Ricalcolare la frontiera con $u_i = 20\%$: quanto rendimento massimo si perde? Rappresentare le tre frontiere ($u = 1, 0,3, 0,2$) nello stesso grafico.

Esercizio 6.3 — stabilità delle stime

Rieseguire lo script con un seme casuale diverso (`default_rng(1)`) e confrontare le composizioni ottime a parità di $\bar{r}$: quanto cambiano? Quale implicazione operativa ne trae?

Esercizio 6.4 — tracking error

Il benchmark è l'equipesato $x_i^B = 1/8$. Minimizzare il tracking error $\sum_{i=1}^n \sum_{j=1}^n q_{ij}(x_i - x_i^B)(x_j - x_j^B)$ con rendimento atteso $\geq 12\%$: quanto ci si deve scostare dal benchmark?

Esercizio 6.5 — costi di transazione

Partendo dal portafoglio equipesato, aggiungere il costo quadratico $\kappa \sum_{i=1}^n (x_i - 1/8)^2$ con $\kappa = 0,5$: come cambia la composizione rispetto all'ottimo "pulito"?

Pricing e revenue management

Classe: NLP concavo / non convesso

Script: [python/lab07_pricing.py](#)

[Pagina online](#)

7.1 Motivazione gestionale

Quale prezzo massimizza il profitto di un concerto, di un hotel, di un abbonamento, quando la domanda diminuisce al crescere del prezzo? Qui il prezzo non è un dato ma una *variabile*: la domanda diventa endogena e il profitto $p \cdot q$ introduce un termine bilineare. È il capitolo in cui il laboratorio incontra per la prima volta la non convessità — e impara a maneggiarla.

Fissare il prezzo è forse la decisione aziendale a più alto impatto sul margine: un punto percentuale di prezzo vale spesso più di molti punti di volume. Il revenue management — nato nelle compagnie aeree e oggi standard in hotel, eventi, noleggi — aggiunge il vincolo di capacità: i posti invenduti stasera non si recuperano domani.

In questo capitolo impareremo a: (1) modellare la domanda come funzione del prezzo e capire quando il profitto è concavo; (2) trattare un obiettivo bilineare non convesso; (3) valutare correttamente la capacità (il posto in più *non* vale il margine pieno); (4) riconoscere gli ottimi di spigolo, dove nessuna derivata si annulla.

7.2 Costruzione guidata del modello

Il problema a parole. *Decidiamo* il prezzo di vendita e la quantità da vendere. *L'obiettivo* è massimizzare il profitto, cioè ricavi meno costi variabili. *I vincoli*: non si può vendere più di quanto il mercato chiede a quel prezzo, né più della capacità disponibile.

Dati (input del modello)

Simbolo	Tipo	Significato
$d(\cdot)$	funzione $\mathbb{Q}_{\geq 0} \rightarrow \mathbb{Q}_{\geq 0}$	domanda attesa al prezzo p , decrescente in p
c	$\in \mathbb{Q}_{\geq 0}$	costo marginale unitario (€)
k	$\in \mathbb{Q}_{> 0}$	capacità disponibile (posti, camere, unità)

Variabili decisionali

Introduciamo le seguenti 2 variabili non negative:

$$\begin{cases} p = \text{prezzo di vendita deciso dall'impresa (€)} \\ q = \text{quantità venduta (unità)} \end{cases}$$

Usando queste variabili, un modello di pricing per il problema è il seguente:

Modello di pricing con vincolo di capacità

$$\max \quad pq - cq \quad (7.1a)$$

$$\text{soggetto a} \quad q \leq d(p), \quad (7.1b)$$

$$q \leq k, \quad (7.1c)$$

$$p \geq 0, \quad (7.1d)$$

$$q \geq 0. \quad (7.1e)$$

Descrizione della funzione obiettivo e dei vincoli del modello:

- la funzione obiettivo (7.1a) massimizza il profitto $pq - cq = (p - c)q$, ricavo meno costo variabile; contiene il prodotto tra le due variabili p e q (termine *bilineare*, non convesso), ed è qui che nasce tutta la difficoltà del capitolo;
- il vincolo (7.1b) impone di non vendere più di quanto il mercato chiede al prezzo p (un vincolo, lineare se d è lineare);
- il vincolo lineare (7.1c) impone di non vendere più della capacità disponibile; quando è attivo, è lui a determinare il prezzo ottimo (un vincolo lineare);
- i vincoli (7.1d) e (7.1e) definiscono le variabili del modello.

Funzioni di domanda usate nel caso di studio, con i rispettivi dati (tutti razionali positivi): lineare $d(p) = a - bp$; a elasticità costante $d(p) = \theta p^{-\varepsilon}$ con $\varepsilon > 1$; logistica $d(p) = m/(1 + e^{-\alpha + \beta p})$.

Spiegazione riga per riga

Perché $q \leq d(p)$ e non $q = d(p)$? All'ottimo la disuguaglianza (7.1b) è attiva da sola (vendere meno del possibile a quel prezzo non conviene mai se $p > c$), ma scriverla come $\leq$ mantiene la regione ammissibile più semplice e il modello ammissibile anche quando $d(p) > k$.

Dove sta la non convessità? Nel prodotto $p \cdot q$ dell'obiettivo. Con domanda lineare, sostituendo $q = a - bp$, il profitto $(p - c)(a - bp)$ è una parabola concava: il problema *ridotto* è facile. Gurobi risolve comunque la versione bilineare all'ottimo globale (`NonConvex=2`); per le domande non lineari generali (elasticità costante, logistica) bastano i vincoli funzionali di Gurobi (`addGenConstrPow`, `addGenConstrExp`): stesso solver, ottimo ancora globale.

7.3 Esempio svolto a mano

Domanda lineare, con e senza capacità

Concerto: $d(p) = 1200 - 5p$ (quindi $a = 1200$, $b = 5$), costo marginale $c = 20$ €, teatro da $k = 400$ posti.

- **Passo 1 — senza capacità.** Profitto $\Pi(p) = (p - 20)(1200 - 5p)$; $\Pi'(p) = 1200 - 10p + 100 = 0$ dà $p^\circ = \frac{a/b + c}{2} = \frac{240 + 20}{2} = 130$ €, con $q^\circ = 1200 - 650 = 550$ biglietti.
- **Passo 2 — la capacità morde.** $q^\circ = 550 > 400$: il vincolo è attivo, quindi si vende esattamente k e il prezzo che svuota il teatro è $p^* = (a - k)/b = (1200 - 400)/5 = 160$ €. Profitto: $(160 - 20) \cdot 400 = 56.000$ €.
- **Passo 3 — quanto vale un posto in più?** Aumentando k , il prezzo ottimo scende di $1/b$: $\frac{d\Pi}{dk} = \underbrace{(p^* - c)}_{140} + k \underbrace{\frac{dp^*}{dk}}_{-1/5} = 140 - 80 = 60$ €. Un posto in più *non* vale il margine pieno (140 €): per riempirlo bisogna abbassare il prezzo a tutti. Errore classico di valutazione degli investimenti di capacità.

7.4 Caso di studio e implementazione

Stessi dati dell'esempio (è il caso raro in cui l'istanza realistica è già piccola), più le varianti a elasticità costante ($\theta = 6 \cdot 10^6$, $\varepsilon = 2,2$), logistica ($m = 900$, $\alpha = 6$, $\beta = 0,045$) e il multiprodotto.

```

1 m = gp.Model("pricing_lineare")
2 m.Params.NonConvex = 2 # obiettivo bilineare p*q
3 p = m.addVar(ub=a/b, name="p")
4 q = m.addVar(name="q")
5 m.addConstr(q <= a - b*p, name="domanda")
6 v_cap = m.addConstr(q <= K, name="capienza")
7 m.setObjective(p*q - c*q, GRB.MAXIMIZE)
8 m.optimize()
9 # niente duali nei QP non convessi: sensitivita' per perturbazione
10 v_cap.RHS = K + 1; m.optimize()

```

Gurobi: $p^* = 160,00$ EUR, $q^* = 400$, profitto = 56.000,00 EUR
 Valore marginale di un posto: 59,80 EUR (teoria: 60,00)
 Elasticità costante: $p^* = 79,11$ EUR (ottimo nel punto in cui $d(p) = k$)
 Logistica : $p^* = 138,29$ EUR, profitto = 47.316,83 EUR

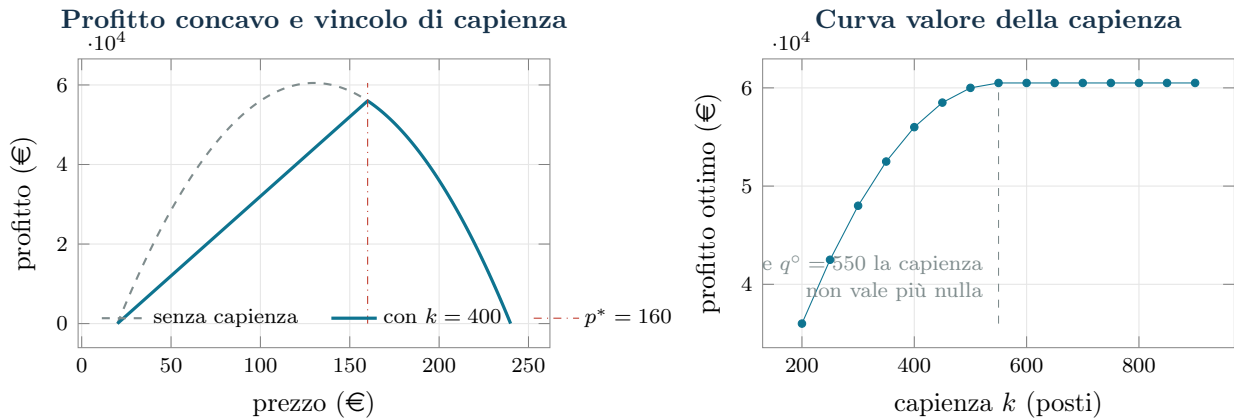


Figura 7.1: A sinistra: profitto in funzione del prezzo per la domanda lineare, senza vincolo (parabola grigia tratteggiata) e con capienza $k = 400$ (linea teal): il vincolo taglia la parabola e sposta l'ottimo a $p^* = 160$ €. A destra: profitto ottimo al crescere della capienza; oltre $q^o = 550$ posti la curva diventa piatta e la capacità aggiuntiva non vale più nulla.

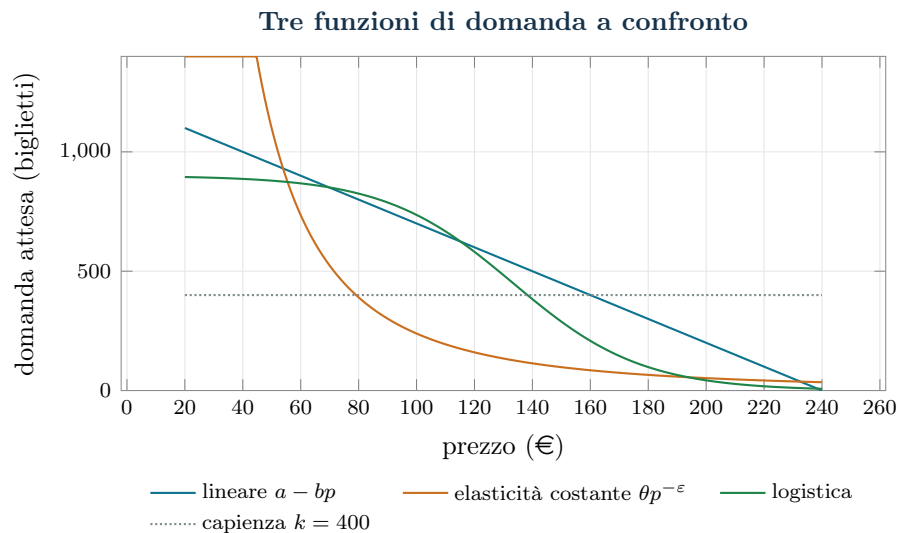


Figura 7.2: Le tre funzioni di domanda del caso di studio a confronto (biglietti attesi in funzione del prezzo) con la capienza $k = 400$ (linea punteggiata): la lineare taglia l'asse, l'elasticità costante ha coda lunga, la logistica descrive la transizione graduale tra domanda alta e bassa.

L'ottimo nel punto di spigolo

Con elasticità costante l'ottimo *non vincolato* sarebbe $p = c\varepsilon/(\varepsilon - 1) = 36,67$ €, ma a quel prezzo la domanda supererebbe di gran lunga la capienza. Per p sotto $79,11$ € si vende comunque k (conviene quindi alzare il prezzo); sopra, il profitto $(p - c)\theta p^{-\varepsilon}$ decresce. L'ottimo $p^* = 79,11 = (\theta/k)^{1/\varepsilon}$ sta nel *punto di spigolo* dove $d(p) = k$: non annulla nessuna derivata. Morale: mai cercare l'ottimo solo tra i punti stazionari quando ci sono vincoli.

7.5 Versione multiprodotto

Due categorie di biglietti con sostituzione (alzare il prezzo della platea spinge parte della domanda in galleria): $d_1(p_1, p_2) = 500 - 2p_1 + 0,6p_2$, $d_2(p_1, p_2) = 900 + 0,8p_1 - 4p_2$, costi (30, 15) €, capacità (150, 300) posti.

```
platea    : p1* = 234,04 EUR    q1* = 150/150 (piena)
galleria  : p2* = 196,81 EUR    q2* = 300/300 (piena)
profitto totale: 85.148,94 EUR
```

Insight manageriale

Entrambe le categorie si riempiono, ma i prezzi non sono “quelli che svuotano” ciascuna sala presa da sola: il modello sfrutta la sostituzione, tenendo la platea cara per spingere domanda verso la galleria. Con prodotti sostituiti i prezzi vanno decisi *congiuntamente*: ottimizzarli uno alla volta lascia soldi sul tavolo.

Con funzioni di domanda generali il profitto può non essere concavo: un solver locale può fermarsi su un ottimo locale. Prassi del laboratorio: studiare dominio e derivate della funzione *prima* di fidarsi del solver, provare più punti iniziali, e tracciare sempre il grafico del profitto sull'intervallo rilevante.

7.6 Esercizi

Esercizio 7.1 — verifica

Verificare con Gurobi il valore marginale del posto per $k = 300$ (dalla tabella: 100 €) e ricavarlo anche con la formula $p^* - c - k/b$.

Esercizio 7.2 — costo marginale

Come cambiano p^* , q^* e il valore del posto se il costo marginale sale a $c = 60$ €? Rifare i tre passi dell'esempio a mano.

Esercizio 7.3 — prezzo di riferimento

Aggiungere la penalità reputazionale $-\gamma(p-120)^2$ con $\gamma = 2$: quale prezzo bilancia profitto e reputazione? Il problema resta concavo?

Esercizio 7.4 — elasticità

Per la domanda a elasticità costante, tracciare p^* in funzione di $\varepsilon \in [1,3; 3]$ con $k = 400$: per quali ε il vincolo di capienza resta attivo?

Esercizio 7.5 — multiprodotto

Nel modello a due categorie, di quanto aumenta il profitto con 50 posti di platea in più? E con 50 di galleria? Quale ampliamento consigliereste, a parità di costo?

Allocazione del budget pubblicitario

Classe: NLP convesso

Script: [python/lab08_budget.py](#)[Pagina online](#)

8.1 Motivazione gestionale

Come ripartire una campagna da 100.000 € tra social, search, TV e influencer, quando ogni canale mostra rendimenti marginali decrescenti? I primi euro su un canale rendono molto, i successivi sempre meno: la risposta è una funzione concava della spesa, e l'ottimo segue una regola economica elegante che questo capitolo dimostra e verifica numericamente.

L'allocazione di un budget tra impieghi con rendimenti decrescenti è uno schema universale: marketing su più canali, fondi tra progetti, tempo tra attività, fertilizzante tra campi. La teoria dice qualcosa di forte e verificabile: all'ottimo i rendimenti marginali si eguagliano — ed è esattamente ciò che il solver restituirà, cifra per cifra.

In questo capitolo impareremo a: (1) modellare rendimenti decrescenti con funzioni concave; (2) risolvere NLP convessi con i vincoli funzionali di Gurobi; (3) verificare numericamente le condizioni KKT; (4) costruire la curva valore-budget e usare λ per negoziare il budget.

8.2 Costruzione guidata del modello

Il problema a parole. Decidiamo quanto spendere in ciascun canale pubblicitario. *L'obiettivo* è massimizzare la risposta totale attesa della campagna. *I vincoli*: la spesa complessiva non può superare il budget, e ciascun canale ha un investimento massimo utile.

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di canali pubblicitari; i canali sono indicizzati da $i \in \{1, 2, \dots, n\}$
b	$\in \mathbb{Q}_{>0}$	budget totale (migliaia di €)
$r_i(\cdot)$	funzione concava crescente	risposta attesa del canale i in funzione della spesa; nel caso di studio $r_i(x) = a_i \log(1 + k_i x)$ con $a_i, k_i \in \mathbb{Q}_{>0}$ dati
u_i	$\in \mathbb{Q}_{>0}$	massimo investimento utile o consentito nel canale i

Variabili decisionali

Introduciamo le seguenti n variabili non negative:

$$x_i = \text{spesa nel canale } i \text{ (migliaia di €)}, \quad \forall i \in \{1, 2, \dots, n\}.$$

Usando queste variabili, un modello NLP convesso per il problema è il seguente:

Modello di allocazione con risposta concava

$$\max \sum_{i=1}^n r_i(x_i) \quad (8.1a)$$

$$\text{soggetto a } \sum_{i=1}^n x_i \leq b, \quad (8.1b)$$

$$x_i \leq u_i, \quad \forall i \in \{1, 2, \dots, n\}, \quad (8.1c)$$

$$x_i \geq 0, \quad \forall i \in \{1, 2, \dots, n\}. \quad (8.1d)$$

Descrizione della funzione obiettivo e dei vincoli del modello:

- la funzione obiettivo concava (8.1a) massimizza la risposta totale della campagna; è separabile per canale e somma di funzioni concave, quindi il problema ha ottimo globale certificabile;
- il vincolo lineare (8.1b) impone che la spesa complessiva non superi il budget: è l'unico vincolo che lega i canali tra loro — senza di esso ogni canale andrebbe al proprio tetto (un vincolo lineare);
- i vincoli lineari (8.1c) sono i tetti per canale: il limite oltre cui il canale non assorbe altra spesa utile (n vincoli lineari);
- i vincoli (8.1d) definiscono le variabili del modello.

La condizione economica all'ottimo

Per le KKT, all'ottimo esiste $\lambda \geq 0$ (prezzo ombra del budget) tale che, per ogni canale non bloccato ai limiti 0 o u_i :

$$r'_i(x_i^*) = \lambda.$$

In parole: **l'ultimo euro investito rende lo stesso in tutti i canali attivi**. Se così non fosse, spostare un euro dal canale con marginale basso a quello con marginale alto migliorerebbe la risposta: non saremmo all'ottimo. I canali al tetto u_i possono avere marginale $> \lambda$ (vorremmo investirne di più ma non si può); quelli a zero hanno marginale iniziale $< \lambda$ (non valgono nemmeno il primo euro).

8.3 Esempio svolto a mano

Due canali logaritmici

$r_1(x) = 100 \log(1 + 0,2x)$, $r_2(x) = 60 \log(1 + 0,1x)$, budget $b = 50$ (migliaia di €), senza tetti.

- **Passo 1 — condizione di uguaglianza dei marginali.**

$$\frac{100 \cdot 0,2}{1 + 0,2x_1} = \frac{60 \cdot 0,1}{1 + 0,1x_2} \implies \frac{20}{1 + 0,2x_1} = \frac{6}{1 + 0,1x_2}.$$

- **Passo 2 — budget.** $x_2 = 50 - x_1$. Sostituendo: $20(1 + 0,1(50 - x_1)) = 6(1 + 0,2x_1)$,
cioè $20(6 - 0,1x_1) = 6 + 1,2x_1 \Rightarrow 120 - 2x_1 = 6 + 1,2x_1 \Rightarrow x_1 = \frac{114}{3,2} = 35,6$, $x_2 = 14,4$.
- **Passo 3 — prezzo ombra.** $\lambda = 20/(1 + 0,2 \cdot 35,6) = 2,46$: un euro di budget in più renderebbe $\approx 2,46$ unità di risposta — lo stesso numero da qualunque canale lo si faccia entrare.

8.4 Caso di studio e implementazione

Quattro canali, $r_i(x) = a_i \log(1 + k_i x)$, budget $b = 100$ (migliaia di €), dati in [dati/budget_canali.csv](#). Il problema è concavo e lo risolviamo con i vincoli funzionali di Gurobi: una variabile ausiliaria $z_i = \log(g_i)$ con $g_i = 1 + k_i x_i$, e l'obiettivo diventa lineare nelle z_i (ottimo globale certificato):

```
1 m.Params.FuncNonlinear = 1 # log trattato in modo esatto
2 x = m.addVars(4, ub=u)
3 g = m.addVars(4, lb=1.0) # g_i = 1 + k_i x_i
4 z = m.addVars(4, lb=-GRB.INFINITY)
5 for i in range(4):
6     m.addConstr(g[i] == 1 + k[i] * x[i])
7     m.addGenConstrLog(g[i], z[i])
8 m.addConstr(x.sum() <= B)
9 m.setObjective(gp.quicksum(a[i] * z[i] for i in range(4)),
10                GRB.MAXIMIZE)
```

8.5 Risultati e verifica delle KKT

Risposta totale: 1.449,8 (migliaia di contatti utili)

canale	spesa	tetto	risposta	marginale
social	24,3	60	385,3	8,2693
search	34,8	80	539,5	8,2693
TV	22,9	120	235,2	8,2693
influencer	18,0	35	289,8	8,2693

Verifica: +1000 EUR di budget -> risposta +8,244 ~ lambda

Insight manageriale

La condizione teorica si vede nei numeri: i quattro marginali coincidono alla quarta cifra ($\lambda = 8,2693$), e l'aumento reale di risposta con mille euro in più (8,244) conferma la lettura del moltiplicatore. Nota manageriale: la TV riceve *meno* dei social nonostante il tetto più alto — non conta la dimensione del canale ma la velocità con cui satura (k_i).

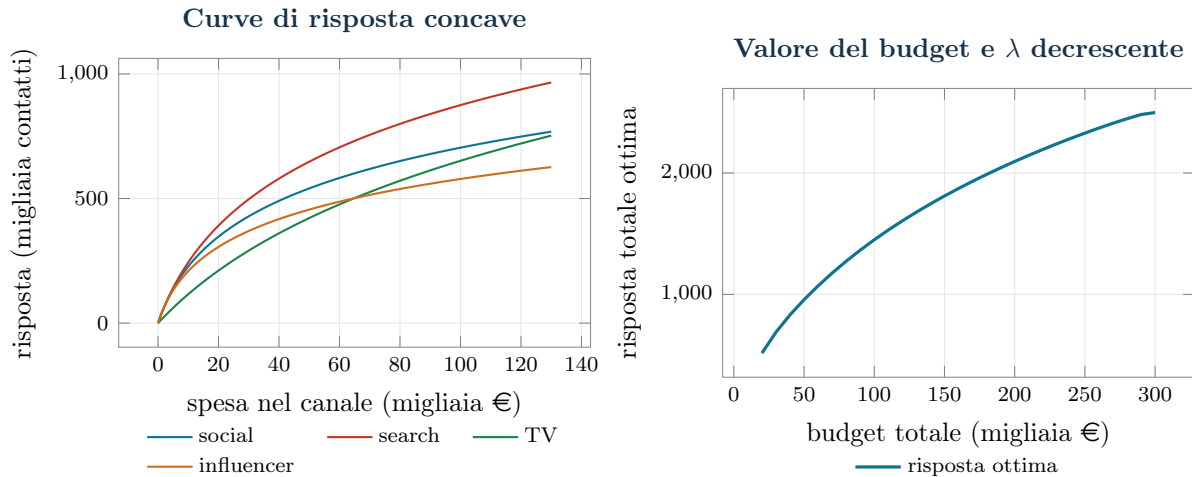


Figura 8.1: A sinistra: le curve di risposta $r_i(x) = a_i \log(1 + k_i x)$ dei quattro canali — concave, con velocità di saturazione diverse. A destra: la risposta totale ottima al crescere del budget (curva valore), concava: ogni trancina di budget rende meno della precedente.

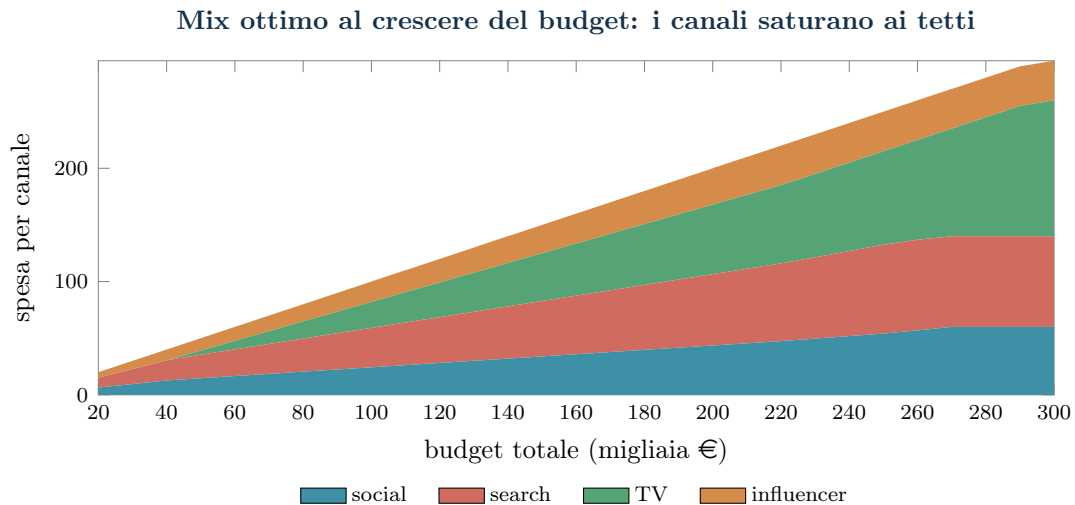


Figura 8.2: Mix ottimo di spesa (aree impilate, migliaia di €) al crescere del budget totale: i canali entrano e crescono insieme mantenendo uguali i ritorni marginali; l'influencer satura per primo al suo tetto (35), seguito dagli altri.

8.6 Analisi di sensitività

b = 20:	risposta	515,9	lambda = 18,971
b = 60:	risposta	1.070,4	lambda = 10,909
b = 100:	risposta	1.449,8	lambda = 8,244

```

b = 180: risposta 1.988,2    lambda = 5,538
b = 260: risposta 2.370,2    lambda = 4,049
b = 300: risposta 2.498,1    lambda = 0,000 (tutti i canali ai tetti)

```

Spiegazione riga per riga

La curva valore-budget è concava: λ scende da 19 a 4 man mano che il budget cresce. A $b = 300$ tutti i canali sono ai tetti ($60 + 80 + 120 + 35 = 295$): il budget smette di essere la risorsa scarsa e λ crolla a zero. Per un CFO: la *curva* di λ è l'argomento quantitativo per decidere quanto budget concedere al marketing — si finanzia finché λ supera il valore di un euro investito altrove.

8.7 Esercizi

Esercizio 8.1 — verifica

Risolvere l'esempio a mano con $b = 80$: dove finisce il budget aggiuntivo? Come cambia λ ?

Esercizio 8.2 — risposta esponenziale

Sostituire la risposta della TV con $r(x) = 520(1 - e^{-0,025x})$ e ripetere l'ottimizzazione: cambia il mix? La funzione ha un tetto naturale: il bound u_i serve ancora?

Esercizio 8.3 — canale escluso

Aggiungere un quinto canale con marginale iniziale $a_5 k_5 = 5 < \lambda$: verificare che l'ottimo lo lascia a zero e interpretare il suo “costo ridotto” $a_5 k_5 - \lambda$.

Esercizio 8.4 — copertura equa

Il budget serve tre segmenti di clientela con coperture diverse per canale. Massimizzare la copertura del segmento peggiore ($\max z$ con $z \leq \text{copertura}_g$): quanto costa l'equità in risposta totale?

Localizzazione continua di un servizio

Classe: NLP convesso

Script: [python/lab09_localizzazione.py](#)

[Pagina online](#)

9.1 Motivazione gestionale

Dove collocare una stazione di ricarica rapida, un micro-hub logistico, un presidio sanitario, per essere “vicini” alla domanda? La risposta dipende da cosa significa vicini: distanza *media* (efficienza), distanza *massima* (equità verso l’utente più sfortunato) o distanza *quadratica* (che porta al baricentro). Tre obiettivi, tre punti diversi sulla mappa: la geometria rende il trade-off visibile come in nessun altro modello.

Le decisioni di localizzazione sono tra le più costose da correggere: un hub nel posto sbagliato resta nel posto sbagliato per anni. La versione continua del problema (coordinate libere sul piano) è il laboratorio ideale per un tema che riappare in tutte le decisioni pubbliche e private: il conflitto tra *efficienza* (servire bene in media) ed *equità* (non abbandonare nessuno).

In questo capitolo impareremo a: (1) formulare tre obiettivi di localizzazione e capire quando coincidono; (2) riformulare le distanze euclidee con vincoli conici e risolverle con Gurobi da buoni punti iniziali; (3) gestire vincoli geografici convessi; (4) costruire una frontiera efficienza-equità con il metodo ε -constraint.

9.2 Costruzione guidata del modello

Il problema a parole. Decidiamo le coordinate in cui collocare la nuova struttura. *L’obiettivo* — ed è la vera scelta manageriale — può essere minimizzare la distanza media pesata dagli utenti, oppure la distanza dell’utente più lontano, oppure la distanza quadratica. *I vincoli*, quando presenti, delimitano l’area ammissibile (una zona rettangolare, un raggio massimo da un’infrastruttura).

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di punti di domanda (quartieri), indicizzati da $i \in \{1, 2, \dots, n\}$
(a_i, b_i)	$\in \mathbb{Q}^2$	coordinate del quartiere i (km)
w_i	$\in \mathbb{Q}_{\geq 0}$	peso del quartiere i : popolazione, domanda o priorità
$(a_0, b_0), r$	$\in \mathbb{Q}^2, \in \mathbb{Q}_{>0}$	centro e raggio dell'eventuale vincolo geografico (distanza massima da un'infrastruttura)

Variabili decisionali

Introduciamo le seguenti 2 variabili libere (possono assumere qualunque valore reale) e, per la versione minimax, una variabile ausiliaria non negativa:

$$\begin{cases} x = \text{coordinata est della nuova struttura (km)} \\ y = \text{coordinata nord della nuova struttura (km)} \\ z = \text{distanza massima dal quartiere più lontano (solo minimax)} \end{cases}$$

Usando queste variabili, i modelli per i tre obiettivi classici sono i seguenti:

Modello di Weber (distanza media pesata)

$$\min \sum_{i=1}^n w_i \sqrt{(x - a_i)^2 + (y - b_i)^2} \quad (9.1a)$$

$$\text{soggetto a } x \geq 0, \quad (9.1b)$$

$$y \geq 0. \quad (9.1c)$$

Modello minimax (distanza massima)

$$\min z \quad (9.2a)$$

$$\text{soggetto a } \sqrt{(x - a_i)^2 + (y - b_i)^2} \leq z, \quad \forall i \in \{1, 2, \dots, n\}, \quad (9.2b)$$

$$x \geq 0, \quad (9.2c)$$

$$y \geq 0, \quad (9.2d)$$

$$z \geq 0. \quad (9.2e)$$

Descrizione delle funzioni obiettivo e dei vincoli dei due modelli:

- la funzione obiettivo convessa (9.1a) minimizza la distanza totale pesata dai quartieri (efficienza); i vincoli (9.1b) e (9.1c) definiscono le variabili libere;
- la funzione obiettivo (9.2a) minimizza la distanza del quartiere più lontano (equità); i vincoli convessi (9.2b) impongono che ogni quartiere disti al più z dalla struttura (n vincoli); i vincoli (9.2c)–(9.2e) definiscono le variabili;

- terza variante, **quadratica**: minimizzare $\sum_{i=1}^n w_i [(x - a_i)^2 + (y - b_i)^2]$, che ha soluzione in forma chiusa (il baricentro pesato).

Vincoli geografici convessi tipici, da aggiungere a qualunque variante: zona rettangolare $x^L \leq x \leq x^U$, $y^L \leq y \leq y^U$; distanza massima da un'infrastruttura $(x - a_0)^2 + (y - b_0)^2 \leq r^2$.

Spiegazione riga per riga

Quadratico $\Rightarrow$ **baricentro**. Annullando il gradiente: $x^* = \sum_{i=1}^n w_i a_i / \sum_{i=1}^n w_i$ (e analogo per y): l'unico caso con formula chiusa, utile come punto di partenza per gli altri.

Weber. Non ha forma chiusa; la funzione non è differenziabile nei punti (a_i, b_i) . All'ottimo i "tiri" unitari pesati dei quartieri si annullano: $\sum_{i=1}^n w_i \mathbf{u}_i = \mathbf{0}$, dove $\mathbf{u}_i$ è il versore dalla struttura verso il quartiere i .

Minimax. L'ottimo è il centro del *cerchio minimo* che racchiude tutti i punti: dipende solo dai quartieri estremi e ignora i pesi — per questo protegge le periferie.

9.3 Esempio svolto a mano

Tre clienti su una strada

Clienti nei km 0, 4 e 10 con pesi 1, 1 e 3.

Weber 1D = mediana pesata. Il costo è $f(x) = |x| + |x - 4| + 3|x - 10|$. Spostandosi di δ a sinistra da $x = 10$, il costo dei primi due clienti scende di 2δ ma quello del terzo sale di 3δ : peggiora. Da destra, simmetricamente, peggiora ancora. Ottimo: $x^* = 10$, il punto che lascia *almeno metà del peso totale* ($5/2 = 2,5$) da entrambi i lati — la mediana pesata.

Baricentro. $x = (1 \cdot 0 + 1 \cdot 4 + 3 \cdot 10) / 5 = 6,8$: sensibile a *quanto* sono lontani i punti, non solo da che parte stanno.

Minimax. Punto medio degli estremi: $(0 + 10) / 2 = 5$, pesi ignorati.

Tre obiettivi ragionevoli, tre risposte diverse: 10, 6,8, 5. La scelta dell'obiettivo È la decisione manageriale.

9.4 Caso di studio

Dodici quartieri con pesi tra 5 e 25 mila abitanti ([dati/localizzazione_quartieri.csv](#)); vincolo tecnico: la stazione deve stare entro $r = 2$ km dalla cabina elettrica in (7,6).

Il trucco di modellazione: una variabile $d_k \geq$ distanza euclidea dal quartiere k , imposta con il vincolo conico (convesso) $d_x^2 + d_y^2 \leq d_k^2$; Weber minimizza $\sum_k w_k d_k$, il minimax una z con $d_k \leq z$.

```

1 d = m.addVars(n)                                # d_k >= distanza dal quartiere
   k
2 for k in range(n):
3     m.addConstr(dx[k] == px - coord[k, 0])
4     m.addConstr(dy[k] == py - coord[k, 1])
5     m.addQConstr(dx[k]**2 + dy[k]**2 <= d[k]*d[k])    # cono
6 # Weber:    min sum w_k d_k
7 # minimax: min z con d_k <= z
8 # vincolo geografico: (px-7)^2 + (py-6)^2 <= R^2

```

9.5 Risultati

Baricentro pesato	:	(4,897, 5,397)		
Weber	:	(4,886, 5,310)	dist. media	3,344 km max 6,314 km
Minimax	:	(5,496, 5,029)	dist. media	3,391 km max 5,680 km
Weber vincolato	:	(5,083, 5,429)	costo	+0,2% (vincolo attivo, ottimo sul bordo)

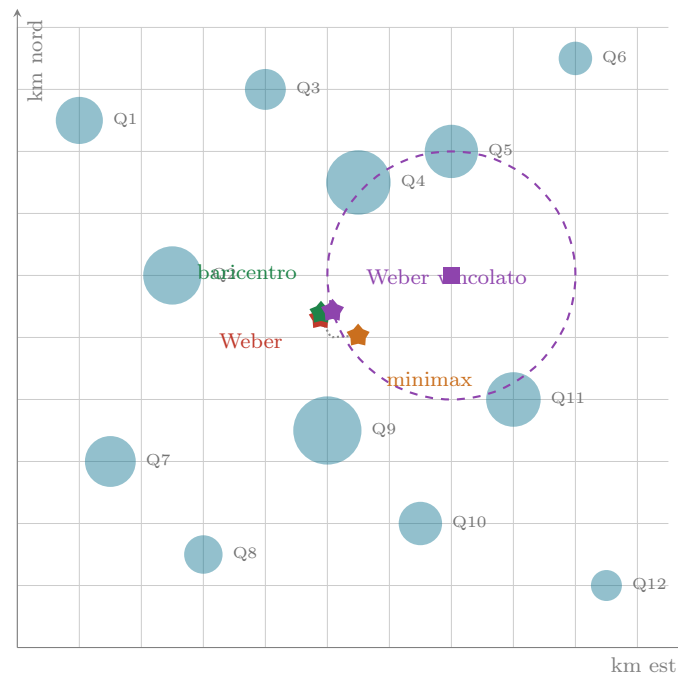


Figura 9.1: Mappa della città: i cerchi teal sono i 12 quartieri (area proporzionale alla popolazione); le stelle indicano le quattro localizzazioni ottime (Weber in rosso, minimax in arancione, baricentro in verde, Weber vincolato in viola); il cerchio viola tratteggiato è il vincolo dei 2 km dalla cabina elettrica; la linea punteggiata è la traiettoria delle soluzioni di compromesso tra minimax e Weber.

Insight manageriale

Weber e baricentro quasi coincidono qui (i pesi sono distribuiti in modo bilanciato), ma il minimax si sposta di oltre mezzo km verso sud-est per proteggere i quartieri periferici. Il vincolo della cabina costa pochissimo (+0,2% di distanza totale): scoprire che un vincolo temuto è quasi gratis è un risultato manageriale importante quanto l'ottimo stesso.

9.6 Analisi di sensitività: la frontiera efficienza-equità

Minimizziamo la distanza media imponendo un tetto D alla distanza massima (metodo ε -constraint), e variamo D tra i due estremi:

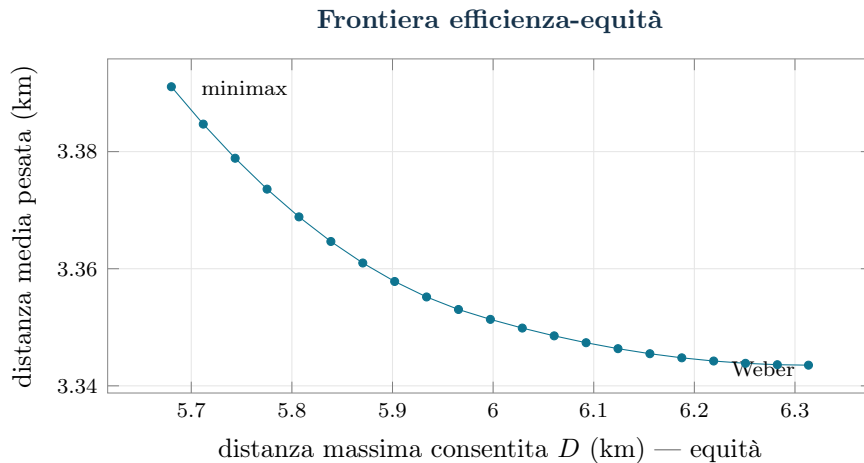


Figura 9.2: Frontiera efficienza-equità (ε -constraint): distanza media pesata minima ottenibile per ogni tetto D imposto alla distanza massima. Gli estremi sono la soluzione minimax (a sinistra) e il punto di Weber (a destra); la curva quasi piatta indica che l’equità costa pochissimo in efficienza.

tetto $D = 5,680$ km:	media 3,391	(soluzione minimax)
tetto $D = 5,997$ km:	media 3,351	
tetto $D = 6,314$ km:	media 3,344	(soluzione Weber)

Spiegazione riga per riga

La frontiera è quasi piatta: garantire al quartiere più lontano 0,63 km in meno costa solo 47 metri di distanza media. Quando la frontiera è così piatta, la soluzione equa è “quasi gratis” — un argomento fortissimo in una discussione pubblica. Il caso opposto (frontiera ripida) segnalerebbe un vero conflitto efficienza-equità.

9.7 Esercizi

Esercizio 9.1 — verifica

Calcolare a mano il baricentro pesato dei 12 quartieri dai dati CSV e confrontarlo con l’output dello script.

Esercizio 9.2 — peso che cambia tutto

Portare il peso di Q12 (il quartiere in basso a destra) da 5 a 40: di quanto si spostano Weber e minimax? Quale dei due è più sensibile ai pesi, e perché?

Esercizio 9.3 — due strutture

Con *due* stazioni e l’assegnazione di ogni quartiere alla più vicina, il problema resta convesso? Provare un approccio alternato (assegna–ottimizza) e discutere perché trova solo ottimi locali.

Esercizio 9.4 — distanza Manhattan

Sostituire la distanza euclidea con $|x - a_i| + |y - b_i|$ e mostrare che il problema di Weber diventa un LP (introdurre variabili ausiliarie per i valori assoluti). Risolverlo con Gurobi e confrontare l'ottimo.

Esercizio 9.5 — moltiplicatore del raggio

Stimare per perturbazione il moltiplicatore del vincolo $r = 2$ km (risolvere con $r = 2,1$): quanto vale, in km di distanza totale, ogni 100 m di raggio in più?

Ricarica intelligente di veicoli elettrici

Classe: LP / QP convesso

Script: [python/lab10_ricarica_ev.py](#)

[Pagina online](#)

10.1 Motivazione gestionale

Una flotta aziendale deve arrivare carica al mattino sfruttando le ore notturne a prezzo basso — ma se tutti i veicoli caricano insieme nelle ore economiche, il prelievo di punta esplode e con esso i costi di potenza e il rischio di superare la capacità del contatore. Il modello decide *quanta potenza* dare a ciascun veicolo in ciascuna ora: una decisione naturalmente continua, senza alcun bisogno di variabili di accensione/spegnimento.

La ricarica intelligente (*smart charging*) è uno dei problemi operativi nuovi creati dalla transizione energetica: flotte elettriche, prezzi orari dell'energia, limiti di potenza dei contatori. È anche un esempio perfetto di problema che *sembra* richiedere variabili di accensione/spegnimento e invece è un LP puro, perché la decisione vera — quanta potenza — è continua.

In questo capitolo impareremo a: (1) modellare decisioni su un orizzonte orario con finestre di disponibilità; (2) leggere i duali come “prezzi delle ore marginali”; (3) gestire il trade-off costo-picco con un approccio multiobiettivo; (4) diagnosticare un'insidia reale dell'analisi di sensitività (il RHS con costanti).

10.2 Costruzione guidata del modello

Il problema a parole. *Decidiamo* quanta potenza erogare a ogni veicolo in ogni ora. *L'obiettivo* è minimizzare la spesa energetica della notte. *I vincoli*: ogni veicolo deve ricevere tutta l'energia che gli serve entro la partenza, può caricare solo quando è collegato e mai oltre la potenza del suo caricatore, e il prelievo complessivo (ricarica più edificio) non può mai superare la potenza del contatore.

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di veicoli, indicizzati da $v \in \{1, 2, \dots, n\}$
m	$\in \mathbb{Z}_{\geq 1}$	numero di intervalli orari, indicizzati da $t \in \{1, 2, \dots, m\}$, di durata Δt (qui $\Delta t = 1$ h)
π_t	$\in \mathbb{Q}_{\geq 0}$	prezzo dell'energia nell'ora t (€/kWh)
a_{vt}	$\in \{0, 1\}$	<i>dato</i> osservato: 1 se il veicolo v è collegato nell'ora t , 0 altrimenti
e_v	$\in \mathbb{Q}_{>0}$	energia richiesta dal veicolo v entro la partenza (kWh)
$\bar{p}_v$	$\in \mathbb{Q}_{>0}$	potenza massima di ricarica del veicolo v (kW)
b_t	$\in \mathbb{Q}_{\geq 0}$	carico di base dell'edificio nell'ora t (kW)
k	$\in \mathbb{Q}_{>0}$	potenza massima prelevabile dal contatore (kW)
η	$\in \mathbb{Q}, 0 < \eta \leq 1$	rendimento di carica

Variabili decisionali

Introduciamo le seguenti nm variabili non negative:

x_{vt} = potenza assegnata al veicolo v nell'ora t (kW), $\forall v \in \{1, 2, \dots, n\}, \forall t \in \{1, 2, \dots, m\}$.

Usando queste variabili, un modello LP per il problema è il seguente:

Modello LP di ricarica a costo minimo

$$\min \sum_{v=1}^n \sum_{t=1}^m \pi_t \Delta t x_{vt} \quad (10.1a)$$

$$\text{soggetto a } \eta \sum_{t=1}^m \Delta t x_{vt} \geq e_v, \quad \forall v \in \{1, 2, \dots, n\}, \quad (10.1b)$$

$$\sum_{v=1}^n x_{vt} + b_t \leq k, \quad \forall t \in \{1, 2, \dots, m\}, \quad (10.1c)$$

$$x_{vt} \leq a_{vt} \bar{p}_v, \quad \forall v \in \{1, 2, \dots, n\}, \forall t \in \{1, 2, \dots, m\}, \quad (10.1d)$$

$$x_{vt} \geq 0, \quad \forall v \in \{1, 2, \dots, n\}, \forall t \in \{1, 2, \dots, m\}. \quad (10.1e)$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (10.1a) minimizza la spesa energetica della notte: la potenza x_{vt} erogata per la durata Δt costa $\pi_t \Delta t x_{vt}$ euro;
- i vincoli lineari (10.1b) assicurano che ogni veicolo riceva tutta l'energia richiesta entro la partenza; il rendimento η sta a sinistra: dell'energia prelevata solo la frazione η finisce in batteria (n vincoli lineari);
- i vincoli lineari (10.1c) impongono che il prelievo complessivo (ricarica più carico di base) non superi mai la potenza del contatore: sono i vincoli che accoppiano i veicoli tra loro (m vincoli lineari);

- i vincoli lineari (10.1d) limitano la potenza di ogni veicolo al suo caricatore e la azzerano nelle ore in cui il veicolo è assente ($a_{vt} = 0$): la disponibilità è un *dato*, non una decisione, ed è per questo che non servono variabili binarie (nm vincoli lineari);
- i vincoli (10.1e) definiscono le variabili del modello.

Varianti: peak shaving ($\min z$ con $\sum_{v=1}^n x_{vt} + b_t \leq z$ per ogni $t \in \{1, 2, \dots, m\}$); profilo regolare (QP con $\gamma \sum_{t=2}^m (z_t - z_{t-1})^2$); multiobiettivo costo $+ \rho z$.

Spiegazione riga per riga

Vincolo di energia scritto come $\geq$. Caricare più del necessario è ammesso ma mai conveniente (il prezzo è positivo), quindi all'ottimo vale con uguaglianza.

10.3 Esempio svolto a mano

Un veicolo, due ore

Servono $e = 10$ kWh ($\eta = 1$); prezzi $\pi_1 = 0,10$ e $\pi_2 = 0,20$ €/kWh; potenza massima $\bar{p} = 8$ kW.

Caso $\bar{p} = 8$. Nell'ora economica si carica al massimo (8 kWh); i restanti 2 kWh vanno nell'ora cara. Costo: $8 \cdot 0,10 + 2 \cdot 0,20 = 1,20$ €. *Prezzo ombra del fabbisogno*: l'undicesimo kWh finirebbe nell'ora cara $\Rightarrow 0,20$ €/kWh. *Prezzo ombra del limite di potenza* nell'ora 1: un kW in più permetterebbe di spostare un kWh dall'ora cara a quella economica $\Rightarrow 0,10$ €/kW.

Caso $\bar{p} = 12$. Tutto nell'ora economica: costo 1,00 €, duale del fabbisogno 0,10, duale della potenza 0 (vincolo non attivo). Il pattern "il duale è il prezzo dell'ora marginale" è la chiave di lettura di tutto il capitolo.

10.4 Caso di studio

Sei furgoni con finestre di ricarica notturne diverse (dati in [dati/ev_flotta.csv](#)), prezzi orari stilizzati con picco serale, carico di base dell'edificio, contatore da $k = 120$ kW, $\eta = 0,95$.

```

1 x = m.addVars(veicoli, ore, name="x")
2 for v in veicoli:
3     for t in ore:
4         x[v, t].UB = P[v] * disp[v, t]      # 0 se non collegato
5 v_ene = m.addConstrs((eta * gp.quicksum(x[v, t] for t in ore)
6                     >= E[v] for v in veicoli), name="energia")
7 v_rete = m.addConstrs((gp.quicksum(x[v, t] for v in veicoli)
8                     + base[t] <= C for t in ore), name="rete")
9 m.setObjective(gp.quicksum(prezzo[t] * x[v, t]
10                        for v in veicoli for t in ore), GRB.MINIMIZE
)
```

10.5 Risultati

Costo minimo : 20,36 EUR/notte picco di prelievo 103,4 kW (limite 120)
 Peak shaving : picco minimo possibile 68,0 kW
 Prezzi ombra del fabbisogno (EUR/kWh):
 V1 0,0947 V2 0,0842 V3 0,0842 V4 0,0947 V5 0,0947 V6 0,0842

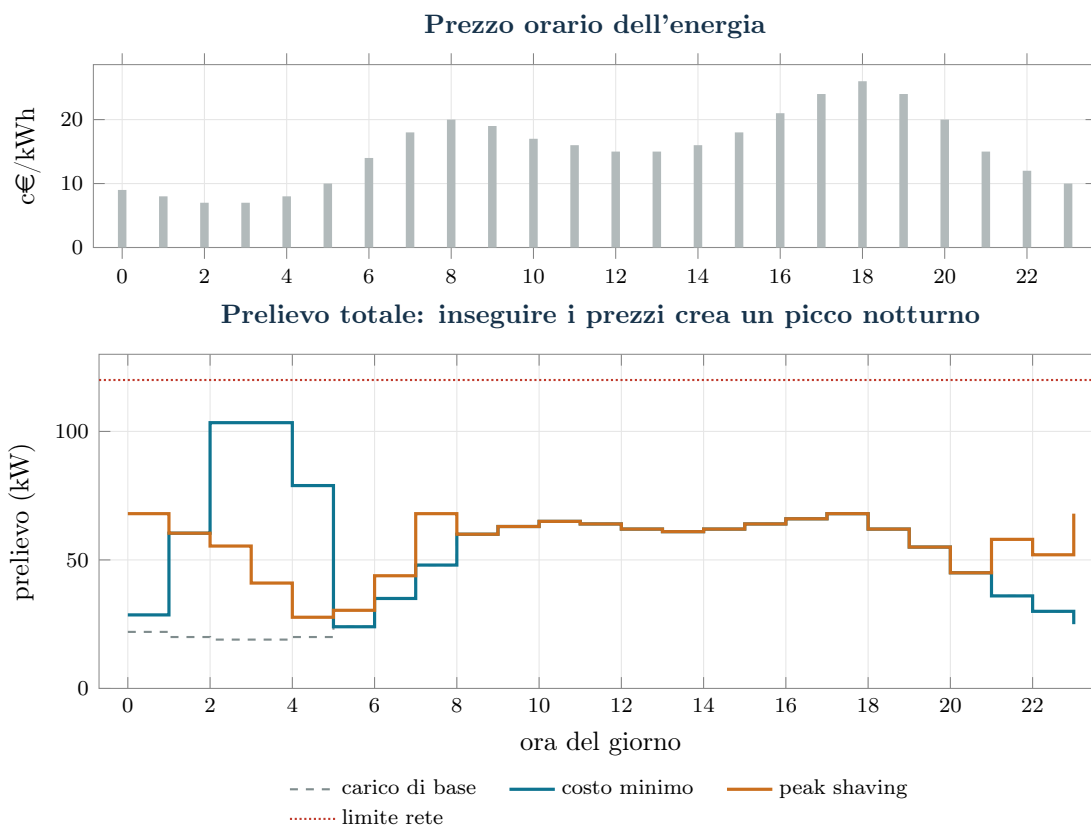


Figura 10.1: Sopra: prezzo orario dell'energia (barre grigie), basso di notte e con picco serale. Sotto: prelievo totale (base + ricarica) ora per ora nelle due strategie: il costo minimo (teal) concentra la ricarica nelle ore economiche e crea un picco notturno di 103 kW; il peak shaving (arancione) distribuisce la ricarica e tiene il prelievo a 68 kW, sotto il limite di rete (linea rossa).

I duali raccontano le ore marginali

I prezzi ombra del fabbisogno valgono $0,0842 = 0,08/\eta$ oppure $0,0947 = 0,09/\eta$: sono i prezzi delle *ore marginali* di ciascun veicolo (l'ora più economica ancora libera nella sua finestra), corretti per il rendimento. V2, V3 e V6 hanno finestre che coprono le ore a 0,08; V1, V4 e V5 arrivano tardi o partono presto e la loro ora marginale costa 0,09. Un kWh di fabbisogno in più costa quindi 8,4–9,5 centesimi a seconda del veicolo.

10.6 Analisi di sensitività

Compromesso costo + $\rho \cdot \text{picco}$:

$\rho = 0,00$:	costo 20,36	picco 103,4	$\rho = 0,20$:	costo 22,15	picco 68,0
$\rho = 0,05$:	costo 20,69	picco 86,5	$\rho = 0,50$:	costo 22,15	picco 68,0
$\rho = 0,10$:	costo 21,39	picco 74,9			

Capacità del contatore:

k = 60, 65 kW: INAMMISSIBILE
 k = 70 kW : 21,93

k = 80: 21,09
 k = 90: 20,63

k = 120: 20,36

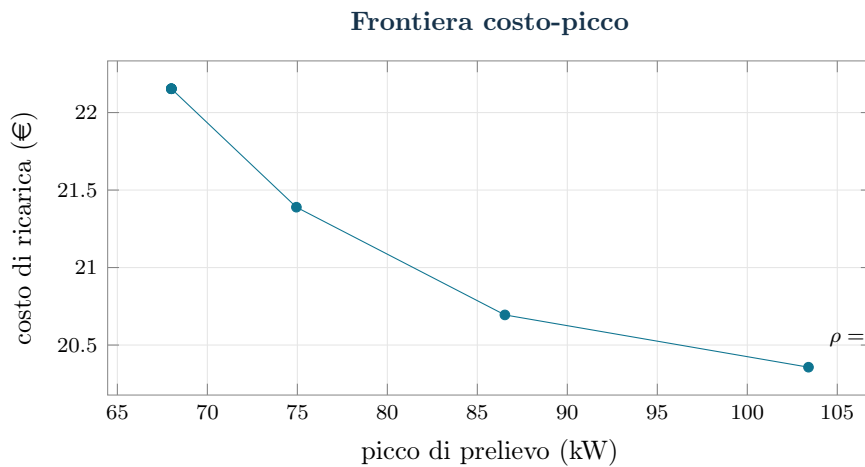


Figura 10.2: Frontiera costo-picco ottenuta variando il peso ρ del picco nell'obiettivo: ogni punto è una strategia di ricarica ottima. Ridurre il picco da 103 a 75 kW costa circa un euro a notte; il punto $\rho \geq 0,2$ raggiunge il picco minimo possibile (68 kW) al costo più basso che lo consente.

Insight manageriale

Tagliare il picco da 103 a 75 kW (-28%) costa solo un euro a notte; arrivare a 68 kW ne costa meno di due. Se la tariffa di potenza o l'upgrade del contatore costano più di così, il peak shaving si ripaga da solo. Nota: il minimax puro (picco 68) senza termine di costo spenderebbe 27,79 € — il compromesso con $\rho = 0,2$ ottiene lo stesso picco a 22,15 €: mai ottimizzare un solo obiettivo quando ce ne sono due.

Nel vincolo $\text{quicksum}(x) + \text{base}[t] \leq C$, Gurobi sposta la costante **base[t]** nel termine noto: il RHS memorizzato è $C - \text{base}[t]$. La prima versione della nostra analisi di sensitività scriveva $v.\text{RHS} = \text{nuovaC}$ e otteneva costi identici per ogni capacità — stava in realtà allentando il vincolo di **base[t]**. La forma corretta è $v.\text{RHS} = \text{nuovaC} - \text{base}[t]$. Quando una sensitività non cambia nulla, sospettare prima del proprio codice che del modello.

10.7 Esercizi

Esercizio 10.1 — verifica

Verificare per perturbazione il prezzo ombra del fabbisogno di V1 (portare e_1 da 46 a 47 kWh) e spiegare perché vale 0,09/0,95.

Esercizio 10.2 — arrivo in ritardo

V3 arriva alle 23 invece che alle 20: come cambiano costo e picco? Quale prezzo ombra aumenta?

Esercizio 10.3 — profilo regolare

Implementare la variante QP con $\gamma \sum_{t=2}^m (z_t - z_{t-1})^2$ dove $z_t = b_t + \sum_{v \in V} x_{vt}$, per $\gamma \in \{0,01; 0,1; 1\}$: confrontare costo, picco e “nervosismo” del profilo.

Esercizio 10.4 — emissioni

Con intensità carbonica oraria g_t (alta la sera, bassa di notte), tracciare la frontiera costo-emissioni. Le ore economiche sono anche le più pulite in questo caso?

Esercizio 10.5 — capacità minima

Trovare per bisezione la capacità minima k che mantiene il problema ammissibile, e spiegare quale combinazione di finestre e $\bar{p}_v$ la determina.

Capacità di servizio e tempi di attesa

Classe: NLP convesso (coda M/M/1)

Script: [python/lab11_code.py](#)

[Pagina online](#)

11.1 Motivazione gestionale

Quanta capacità assegnare a un servizio clienti, a uno sportello, a un servizio cloud? Più capacità costa; poca capacità fa esplodere le attese. La teoria delle code fornisce la relazione quantitativa tra capacità e attesa, e l'ottimizzazione la trasforma in una decisione. Il messaggio centrale è contro-intuitivo per chi ragiona “a efficienza”: **l'utilizzazione ottima non è il 100%**.

Ogni servizio con arrivi aleatori — call center, pronto soccorso, help desk, API di un servizio cloud — vive lo stesso dilemma: la capacità costa, l'attesa pure. La teoria delle code dà la relazione esatta tra le due; l'ottimizzazione sceglie il punto giusto. Il capitolo è breve ma il suo messaggio (l'utilizzazione ottima non è il 100%) vale da solo una discussione in consiglio di amministrazione.

In questo capitolo impareremo a: (1) usare le formule M/M/1 dentro un problema di ottimizzazione; (2) ricavare l'ottimo analiticamente e confermarlo numericamente; (3) prezzare una promessa di servizio; (4) proteggere il dimensionamento dall'incertezza sulla domanda.

11.2 Costruzione guidata del modello

Il problema a parole. *Decidiamo* quanta capacità di servizio acquistare. *L'obiettivo* è minimizzare il costo totale: capacità comprata più il valore del tempo che i clienti passano nel sistema. *Il vincolo* strutturale è uno solo, la stabilità della coda ($\mu > \lambda$); nelle varianti si aggiunge la promessa di servizio $w \leq w_{\max}$.

Usiamo il modello di coda più semplice, indicato con la notazione standard **M/M/1**: arrivi *markoviani* (processo di Poisson), tempi di servizio *markoviani* (esponenziali) e *un* servente. Per questo sistema valgono le formule chiuse riportate in tabella; ciò che qui è nuovo è trattare la capacità μ come una variabile di ottimizzazione.

Dati (input del modello)

Simbolo	Tipo	Significato
λ	$\in \mathbb{Q}_{>0}$	tasso medio di arrivo dei clienti (richieste/ora)
c	$\in \mathbb{Q}_{>0}$	costo di un'unità di capacità (€ per unità di μ all'ora)
h	$\in \mathbb{Q}_{>0}$	valore di un'ora di permanenza di un cliente nel sistema (€)

Variabile decisionale e grandezze derivate

Introduciamo una variabile continua:

$$\mu = \text{capacità del servizio (tasso di servizio, richieste/ora),}$$

con il requisito di stabilità $\mu > \lambda$. Dalla teoria delle code M/M/1 derivano il tempo medio nel sistema $w(\mu) = 1/(\mu - \lambda)$ (in letteratura: W) e il numero medio di clienti nel sistema $l(\mu) = \lambda/(\mu - \lambda)$ (legge di Little; in letteratura: L).

Usando questa variabile, il modello per il problema è il seguente:

Modello di costo totale capacità + attesa

$$\min \quad c\mu + h \frac{\lambda}{\mu - \lambda} \quad (11.1a)$$

$$\text{soggetto a } \mu > \lambda. \quad (11.1b)$$

Descrizione della funzione obiettivo e del vincolo del modello:

- la funzione obiettivo (11.1a) è la somma di due termini: il primo, $c\mu$, compra la capacità; il secondo, $h \cdot l(\mu)$, monetizza il tempo complessivo che i clienti passano nel sistema; è convessa su $(\lambda, +\infty)$;
- il vincolo (11.1b) definisce la variabile e garantisce la stabilità della coda: con $\mu \leq \lambda$ la fila cresce senza limite.

Annullando la derivata dell'obiettivo si ottiene la soluzione in forma chiusa:

$$C'(\mu) = c - \frac{h\lambda}{(\mu - \lambda)^2} = 0 \quad \Rightarrow \quad \boxed{\mu^* = \lambda + \sqrt{h\lambda/c}}$$

Spiegazione riga per riga

La formula ha la struttura classica “capacità = domanda + scorta di sicurezza”: si copre la domanda media λ e si aggiunge un margine $\sqrt{h\lambda/c}$ che cresce con il valore del tempo dei clienti e decresce con il costo della capacità. Il margine cresce come $\sqrt{\lambda}$: i sistemi grandi godono di economie di scala nel “cuscinetto” (pooling).

11.3 Esempio svolto a mano

Sportello con 8 clienti l'ora

$\lambda = 8/h$, $c = 2 \text{ €}$ per unità di capacità, $h = 4 \text{ €/h}$.

- **Passo 1.** $\mu^* = 8 + \sqrt{4 \cdot 8/2} = 8 + 4 = 12$ clienti/ora.
- **Passo 2.** Utilizzazione $\rho = 8/12 = 66,7\%$; tempo medio $w = 1/(12 - 8) = 15$ minuti; costo $C = 2 \cdot 12 + 4 \cdot 8/4 = 32 \text{ €/h}$.
- **Passo 3 — perché non tagliare la capacità?** Con $\mu = 9$ ($\rho = 89\%$, “più efficiente”): $C = 18 + 32/1 = 50 \text{ €/h}$, molto peggio. L'apparente spreco (33% di capacità inutilizzata) è ciò che tiene corte le code.

11.4 Caso di studio e risultati

Servizio clienti: $\lambda = 42$ richieste/ora, $c = 3 \text{ €}$, $h = 1,5 \text{ €}$.

```
Analitico : mu* = 42 + sqrt(1,5*42/3) = 46,583    costo 153,495 EUR/h
Gurobi    : mu* = 46,582 (bilineare w(mu-lam) = 1, NonConvex=2: coincide)
All'ottimo: rho = 90,2%    w = 13,1 minuti
```

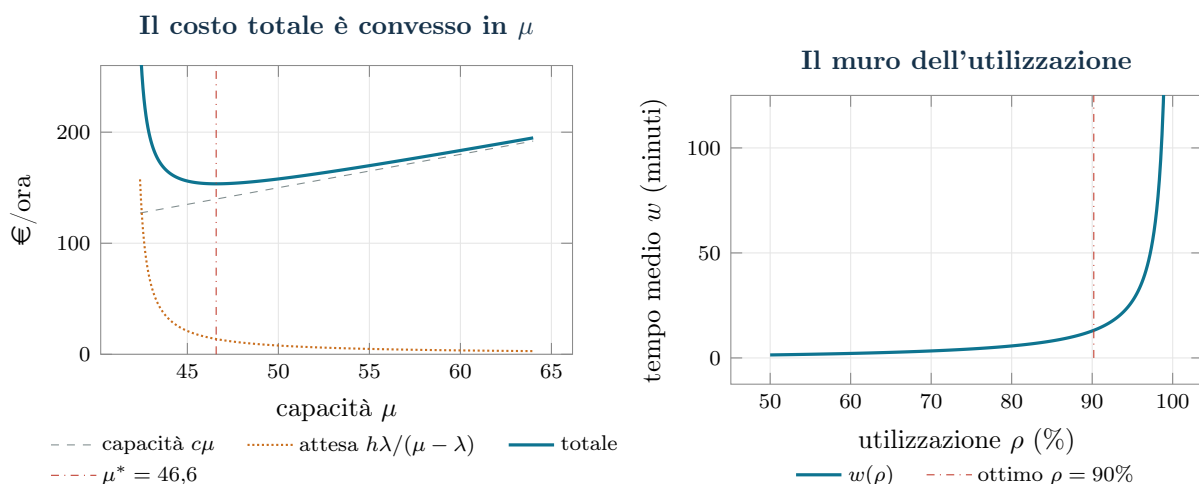


Figura 11.1: A sinistra: le due componenti del costo (capacità, crescente; attesa, decrescente) e il costo totale, convesso, con il minimo in $\mu^* = 46,6$. A destra: il “muro dell'utilizzazione” — il tempo medio nel sistema w in funzione dell'utilizzazione ρ : oltre il 90% l'attesa esplode in modo iperbolico.

Insight manageriale

Il grafico di destra è forse il più importante della dispensa da mostrare a un manager: tra $\rho = 90\%$ e $\rho = 99\%$ il tempo di attesa si moltiplica per dieci. “Saturare le risorse” e “dare un buon servizio” sono obiettivi in conflitto matematico, non organizzativo: nessuna riorganizzazione elimina l'iperbole $w = 1/(\mu - \lambda)$.

11.5 Analisi di sensitività: il prezzo di una promessa

Se il marketing promette “tempo medio sotto $w_{\max}$ ”, serve $\mu \geq \lambda + 1/w_{\max}$: il vincolo diventa attivo quando è più stringente dell’ottimo economico (13,1 minuti).

$w_{\max} = 12$	min: $\mu = 47,00$	costo 153,60	(quasi gratis)
$w_{\max} = 9$	min: $\mu = 48,67$	costo 155,45	
$w_{\max} = 6$	min: $\mu = 52,00$	costo 162,30	
$w_{\max} = 4$	min: $\mu = 57,00$	costo 175,20	
$w_{\max} = 3$	min: $\mu = 62,00$	costo 189,15	
$w_{\max} = 2$	min: $\mu = 72,00$	costo 218,10	(+42% sul costo base)

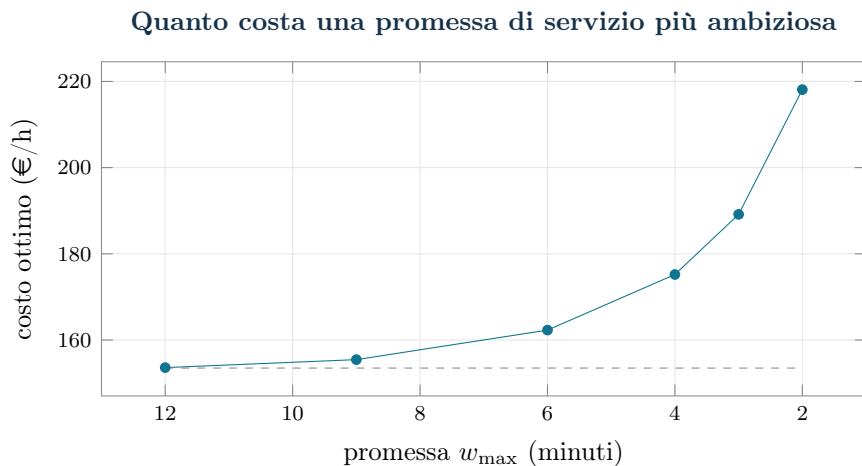


Figura 11.2: Costo ottimo in funzione della promessa di servizio $w_{\max}$ (asse invertito: verso destra la promessa diventa più ambiziosa). La linea tratteggiata è il costo senza promessa: fino a 12 minuti la promessa è quasi gratuita, sotto i 4 minuti ogni minuto in meno costa in modo esplosivo.

Il costo marginale della promessa

Il prezzo ombra del vincolo cresce in modo esplosivo: passare da 12 a 9 minuti costa 1,85 €/h, da 3 a 2 minuti costa 28,95 €/h. Quando il vincolo è attivo, $\mu = \lambda + 1/w_{\max}$ e $C = c\lambda + c/w_{\max} + h\lambda w_{\max}$, da cui $dC/dw_{\max} = -c/w_{\max}^2 + h\lambda$ (con $w_{\max}$ in ore): a $w_{\max} = 6$ minuti vale $-300 + 63 = -237$ €/h per ora di promessa — esattamente la stima per differenze finite dello script. È il numero da dare al marketing *prima* che firmi lo SLA (*Service Level Agreement*, l’impegno contrattuale sul livello di servizio).

Robustezza

Se λ è incerto in $[36, 48]$, dimensionare sull’ottimo nominale $\mu^* = 46,58$ è **instabile**: con $\lambda = 48 > \mu^*$ la coda diverge (costo infinito). Minimizzare il costo del caso peggiore dà $\mu_{\text{rob}} = 52,90$ con costo worst-case 173,39 €/h: si paga un premio assicurativo di $\approx 13\%$ per essere protetti da un disastro. Le decisioni di capacità vanno sempre verificate contro il *massimo* plausibile della domanda, non solo contro la media.

11.6 Esercizi

Esercizio 11.1 — verifica

Ricavare per conto proprio la formula $dC/dw_{\max} = -c/w_{\max}^2 + h\lambda$ della spiegazione precedente e verificarla numericamente a $w_{\max} = 4$ e 9 minuti con differenze finite.

Esercizio 11.2 — penalità quadratica

Sostituire il costo d'attesa lineare con $h [1/(\mu - \lambda)]^2$: ricavare μ^* analiticamente e confrontare l'utilizzazione ottima con il caso lineare.

Esercizio 11.3 — due code vs una

Due sportelli separati ricevono $\lambda = 21$ ciascuno; in alternativa un unico sistema riceve $\lambda = 42$ con capacità aggregata. A parità di costo di capacità, quale configurazione dà attese minori? (È l'argomento quantitativo per il pooling.)

Esercizio 11.4 — frontiera capacità-servizio

Tracciare la frontiera (costo, w) variando $w_{\max}$ con continuità e individuare il punto oltre il quale ogni minuto promesso in meno costa più di 10 €/h.

Parte III

Decisioni sotto incertezza

Il Newsvendor e le sue varianti

Classe: convesso 1D / LP stocastico a scenari

Script: [python/lab12_newsvendor.py](#)

[Pagina online](#)

12.1 Motivazione gestionale

Il Newsvendor (“venditore di giornali”) descrive una decisione comunissima: scegliere una quantità *prima* di osservare la domanda. Ordinare troppo genera invenduto; troppo poco, vendite perse. Si applica a moda e stagionali, freschi e deperibili, farmaci, capacità alberghiera, componenti di ricambio. È la porta d’ingresso dell’ottimizzazione stocastica, e il suo punto di forza didattico è un risultato netto: **la quantità ottima non è la domanda media**.

Il modello ha più di un secolo e resta il fondamento della gestione delle scorte per prodotti a ciclo singolo. La sua importanza didattica va oltre: è il primo incontro con l’idea che *si decide prima e si scopre poi*, e con gli strumenti per farlo bene (scenari, valore della soluzione stocastica, misure di rischio).

In questo capitolo impareremo a: (1) ricavare e applicare la regola del quantile; (2) trasformare un problema stocastico in un LP a scenari con il trucco delle parti positive; (3) quantificare il valore di modellare l’incertezza (VSS); (4) incorporare livelli di servizio e avversione al rischio.

12.2 Il modello base

Il problema a parole. *Decidiamo* una sola quantità q , prima di conoscere la domanda. *L’obiettivo* è minimizzare il costo atteso complessivo dell’errore: quello di aver ordinato troppo (invenduto) più quello di aver ordinato troppo poco (vendite perse). *Vincoli* non ce ne sono nel modello base ($q \geq 0$); compaiono nelle varianti: budget condiviso, livelli di servizio, limiti di capacità.

Dati (input del modello)

Simbolo	Tipo	Significato
D	variabile aleatoria ≥ 0	domanda, con funzione di ripartizione F ; eccezione alla convenzione sulle maiuscole, per aderenza alla letteratura: variabile aleatoria maiuscola, realizzazioni minuscole (d_s)
c_u	$\in \mathbb{Q}_{>0}$	costo unitario di <i>underage</i> (domanda non servita): margine perso più eventuale penalità (€)
c_o	$\in \mathbb{Q}_{>0}$	costo unitario di <i>overage</i> (unità invenduta): costo meno valore di recupero (€)

Con prezzo di vendita p , costo d'acquisto c e valore di recupero $v \leq c$ (tutti in $\mathbb{Q}_{>0}$):
 $c_u = p - c$ e $c_o = c - v$.

Variabile decisionale

Introduciamo una variabile non negativa:

q = quantità ordinata *prima* di osservare la domanda D .

Usando questa variabile, il modello per il problema è il seguente:

Modello Newsvendor in forma di costo

$$\min c_o \mathbb{E}[(q - D)^+] + c_u \mathbb{E}[(D - q)^+] \quad (12.1a)$$

$$\text{soggetto a } q \geq 0. \quad (12.1b)$$

Descrizione della funzione obiettivo e del vincolo del modello:

- la funzione obiettivo convessa (12.1a) minimizza il costo atteso dell'errore: eccedenza $(q - D)^+$ pagata c_o e carenza $(D - q)^+$ pagata c_u (il simbolo $(x)^+ = \max\{x, 0\}$ indica la parte positiva);
- il vincolo (12.1b) definisce la variabile del modello.

Se F è continua, l'ottimo soddisfa la **regola del quantile**:

$$F(q^*) = \alpha^* = \frac{c_u}{c_u + c_o} \quad \implies \quad q^* = F^{-1}(\alpha^*).$$

Da dove viene la regola del quantile

Si ragiona sull'unità marginale: ordinare la q -esima unità in più rende c_u se la domanda la assorbe (probabilità $1 - F(q)$) e costa c_o se resta invenduta (probabilità $F(q)$). Conviene finché $c_u(1 - F(q)) \geq c_o F(q)$, cioè finché $F(q) \leq c_u/(c_u + c_o)$. Se la carenza costa molto più dell'eccedenza, α^* è alto e si ordina un quantile alto; se l'invenduto è costoso (deperibili), si ordina poco.

12.3 Esempio svolto a mano (domanda discreta)

Tre scenari equiprobabili

$D \in \{80, 100, 120\}$ con probabilità $1/3$; $c_u = 9$, $c_o = 4$ (quindi $\alpha^* = 9/13 = 0,6923$).

- **Passo 1 — costo atteso per i tre candidati naturali.**

$$\begin{aligned} C(80) &= \frac{1}{3} [0 + 9 \cdot 20 + 9 \cdot 40] = 180 \\ C(100) &= \frac{1}{3} [4 \cdot 20 + 0 + 9 \cdot 20] = \frac{260}{3} = 86,7 \\ C(120) &= \frac{1}{3} [4 \cdot 40 + 4 \cdot 20 + 0] = 80 \end{aligned}$$

- **Passo 2 — regola del quantile discreta.** Serve il più piccolo q con $F(q) \geq 0,6923$: $F(80) = 1/3$, $F(100) = 2/3 = 0,667 < 0,6923$, $F(120) = 1$. Quindi $q^* = 120$: coerente con il passo 1.
- **Passo 3 — lettura.** Ordinare il *massimo* storico nonostante la media sia 100: con $c_u = 9 \gg c_o = 4$, restare corti costa più del doppio che restare lunghi. Chi ordina “la media” paga 86,7 contro 80: l’incertezza si gestisce, non si ignora.

12.4 Caso di studio: la regola del quantile in continuo

Panetteria: $p = 15$, $c = 6$, $v = 2 \text{ €} \Rightarrow c_u = 9$, $c_o = 4$; domanda normale con $\mu = 100$, $\sigma = 20$.

```
alpha* = 9/13 = 0,6923
q* = F^-1(0,6923) = 110,05 unita'    (media = 100)
costo atteso in q*: 91,43 EUR        in q = media: 103,72 EUR
```

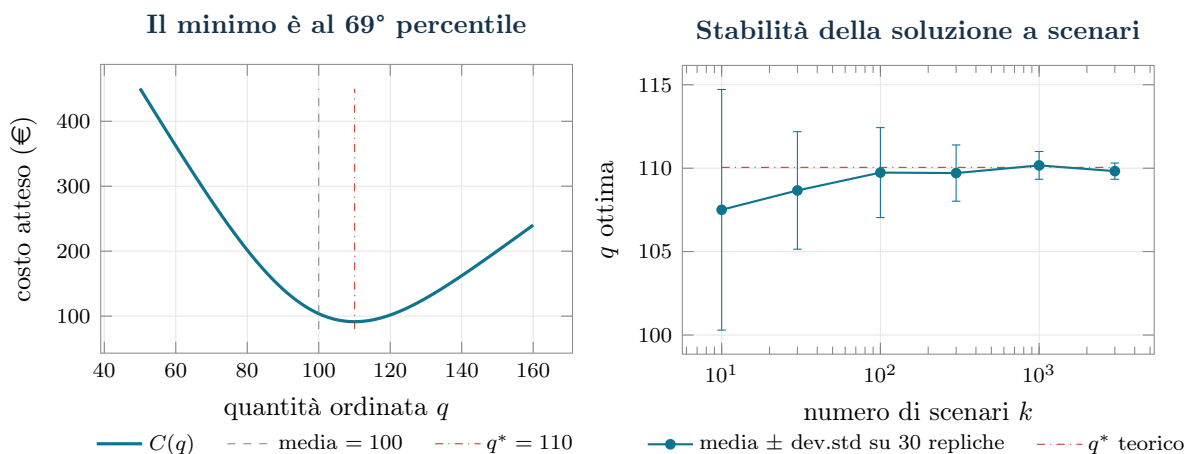


Figura 12.1: A sinistra: il costo atteso $C(q)$ è convesso e il minimo cade in $q^* = 110$ (69° percentile della domanda), non sulla media 100. A destra: la q ottima stimata da k scenari campionati, media e deviazione standard su 30 repliche — con pochi scenari la decisione è instabile.

12.5 La formulazione lineare a scenari

Quando F non è nota in forma chiusa — il caso normale in azienda — si usano gli scenari: dati storici, campioni da una distribuzione stimata, o scenari pesati. Nuovi dati del modello:

il numero di scenari $k \in \mathbb{Z}_{\geq 1}$, indicizzati da $s \in \{1, 2, \dots, k\}$; per ogni scenario, la domanda $d_s \in \mathbb{Q}_{\geq 0}$ e la probabilità $\pi_s \in \mathbb{Q}_{\geq 0}$, con $\sum_{s=1}^k \pi_s = 1$.

Oltre a q , introduciamo le seguenti $2k$ variabili non negative:

$$\begin{cases} o_s = \text{eccedenza (unità invendute) nello scenario } s \\ u_s = \text{carenza (domanda non servita) nello scenario } s \end{cases} \quad \forall s \in \{1, 2, \dots, k\}.$$

Usando queste variabili, un modello LP per il problema è il seguente:

Modello LP del Newsvendor a scenari

$$\min \sum_{s=1}^k \pi_s (c_o o_s + c_u u_s) \quad (12.2a)$$

$$\text{soggetto a } o_s \geq q - d_s, \quad \forall s \in \{1, 2, \dots, k\}, \quad (12.2b)$$

$$u_s \geq d_s - q, \quad \forall s \in \{1, 2, \dots, k\}, \quad (12.2c)$$

$$q \geq 0, \quad (12.2d)$$

$$o_s, u_s \geq 0, \quad \forall s \in \{1, 2, \dots, k\}. \quad (12.2e)$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (12.2a) è la media pesata, sugli scenari, dei costi di eccedenza e di carenza: la versione “a scenari” del costo atteso del modello base;
- i vincoli lineari (12.2b) e (12.2c) definiscono eccedenza e carenza di ogni scenario a partire dall’unica vera decisione q ; insieme alle non negatività realizzano le parti positive, come spiegato sotto ($2k$ vincoli lineari);
- i vincoli (12.2d) e (12.2e) definiscono le variabili del modello.

Perché è lineare (il trucco delle parti positive)

$(q - d_s)^+$ non è lineare, ma all’*ottimo* la variabile o_s , che compare nell’obiettivo con coefficiente positivo, si schiaccia sul massimo tra $q - d_s$ e 0: il vincolo (12.2b) insieme a $o_s \geq 0$ fa esattamente il lavoro di $\max\{q - d_s, 0\}$ senza variabili binarie. Stesso meccanismo per u_s . Questo trucco — linearizzare le parti positive che l’obiettivo spinge verso il basso — torna identico nel CVaR e nella SVM.

```
1 q = m.addVar(name="q")
2 o = m.addVars(S, name="o")      # eccedenza per scenario
3 u = m.addVars(S, name="u")      # carenza per scenario
4 m.addConstrs((o[s] >= q - dom[s] for s in range(S)))
5 m.addConstrs((u[s] >= dom[s] - q for s in range(S)))
6 m.setObjective(gp.quicksum((Co*o[s] + Cu*u[s]) / S
7                               for s in range(S)), GRB.MINIMIZE)
```

LP con $k = 600$ scenari: $q = 108,61$
 quantile empirico al 69,23%: 108,60 (coincidono: e’ un teorema)
 VSS: decisione $q = E[D] = 100$ costa 99,50; q stocastica costa 88,33
 valore della soluzione stocastica = 11,17 EUR per ciclo

Insight manageriale

Il *valore della soluzione stocastica* (VSS) risponde alla domanda del capo: “perché complicarsi la vita con gli scenari?”. Qui vale 11,17 € a ciclo, l’11% del costo: è il guadagno di *modellare* l’incertezza invece di sostituirla con la media. La figura sulla stabilità dà l’altra metà del messaggio: sotto i 100 scenari la q ottima balla di ± 3 unità tra una stima e l’altra — gli scenari sono anch’essi un campione, e la decisione eredita la loro varianza.

12.6 Livelli di servizio

```
cycle service level 90%: q = 125,63 (costo +23,41 EUR sull'ottimo economico)
cycle service level 95%: q = 132,90 (costo +45,59)
cycle service level 99%: q = 146,53 (costo +95,56)
fill rate in q*: 96,06%   probabilita' di non-stockout in q*: 69,23%
```

In $q^* = 110$ la probabilità di evitare lo stock-out è solo il 69%, ma il *fill rate* (quota di domanda servita in media) è il 96%: misure diverse che i contratti spesso confondono. Promettere “95% di probabilità di copertura totale” costa +45,59 € a ciclo rispetto all’ottimo economico; promettere “95% di fill rate” è quasi gratis. Leggere bene lo SLA prima di firmarlo.

12.7 Avversione al rischio: media-CVaR

Il costo atteso non distingue tra tante piccole perdite e rari disastri. Con la formulazione lineare del CVaR (qui solo anticipata) minimizziamo $(1 - \lambda) \mathbb{E}[\text{costo}] + \lambda \text{CVaR}_{0,90}(\text{costo})$:

lambda = 0,00:	q = 108,61	costo medio	88,33	CVaR90	245,93
lambda = 0,25:	q = 112,45	costo medio	90,37	CVaR90	229,29
lambda = 0,50:	q = 114,48	costo medio	92,52	CVaR90	225,49
lambda = 1,00:	q = 116,17	costo medio	94,87	CVaR90	224,56

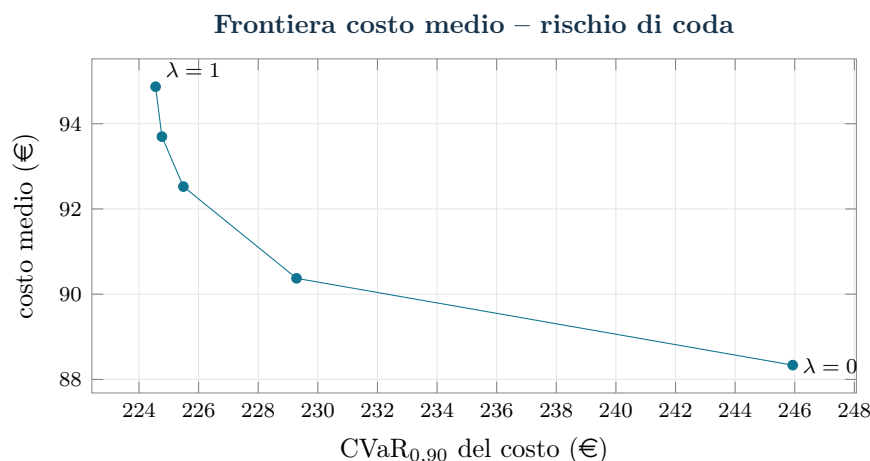


Figura 12.2: Frontiera tra costo medio e $\text{CVaR}_{0,90}$ del costo, ottenuta variando il peso λ dell’avversione al rischio: da $\lambda = 0$ (ottimo in media, coda pesante) a $\lambda = 1$ (protezione massima). Il primo tratto è quasi verticale: molta protezione per pochissimo costo medio.

Il decisore avverso al rischio ordina di più (da 108,6 a 116,2 unità): paga 6,5 € in più in media per tagliare di 21,4 € il costo medio dei peggiori 10% degli scenari. La gran parte della protezione si ottiene già con $\lambda = 0,25$: la frontiera è ripida all'inizio e piatta poi.

12.8 Multiprodotto con budget condiviso

Tre dolci con domande *correlate* ($\rho = 0,7$: le feste vanno bene o male per tutti insieme) e un budget di produzione di 1200 €:

prodotto	q senza budget	q con budget
panettone	111,1	99,3
pandoro	91,5	75,4
torrone	65,9	56,8
Spesa: 1200/1200 prezzo ombra del budget: -0,553		

Spiegazione riga per riga

Il budget vincolante comprime le tre quantità sotto i rispettivi quantili ottimi, ma non proporzionalmente: il taglio dipende dal rapporto c_u/c_o e dal costo unitario di ciascun prodotto. Il duale dice che un euro di budget in più ridurrebbe il costo atteso di 0,55 €: un rendimento del 55% che giustificerebbe quasi qualunque finanziamento — informazione che senza il modello nessuno avrebbe quantificato. La correlazione alta, inoltre, toglie il beneficio di diversificazione: i tre prodotti falliscono insieme, e trattarli come indipendenti sottostimerebbe il rischio.

12.9 Esercizi

Esercizio 12.1 — verifica

Ricalcolare a mano α^* e q^* se il Comune impone una penalità $b = 3$ € per cliente non servito ($c_u = p - c + b$). Verificare con lo script.

Esercizio 12.2 — dati storici

Usare come scenari le 24 osservazioni di domanda degli ultimi due anni (generarle con lo script) invece della distribuzione: quanto differisce q ? Quale approccio consigliereste a un'azienda che ha solo 24 osservazioni?

Esercizio 12.3 — deperibili

Il recupero diventa negativo ($v = -1$ €: smaltire costa): come cambiano α^* e q^* ? Tracciare q^* in funzione di $v \in [-2, 4]$.

Esercizio 12.4 — EVPI

Calcolare l'EVPI (*Expected Value of Perfect Information*, valore atteso dell'informazione perfetta): costo medio con q scelta *dopo* aver visto la domanda (cioè $q = d_s$, costo 0) contro costo della soluzione stocastica. Quanto vale al massimo una previsione perfetta della domanda?

Esercizio 12.5 — fill rate vincolato

Aggiungere all'LP il vincolo di fill rate $\sum_{s=1}^k \pi_s u_s \leq (1 - 0,98) \mathbb{E}[D]$ e confrontare quantità e costo con il cycle service level al 98%.

Esercizio 12.6 — multiprodotto

Rieseguire il multiprodotto con domande indipendenti ($\rho = 0$) a parità di tutto il resto: il costo atteso ottimo migliora o peggiora? Perché?

VaR e CVaR: misurare e ottimizzare il rischio

Classe: LP a scenari

Script: [python/lab13_var_cvar.py](#)

[Pagina online](#)

13.1 Motivazione gestionale

Molti modelli ottimizzano un valore medio. Scelta semplice, ma che può nascondere perdite rare e molto elevate. VaR e CVaR descrivono la *coda* della distribuzione delle perdite e trasformano l'avversione al rischio in decisioni quantitative. Il VaR risponde a “qual è una soglia di perdita elevata?”; il CVaR risponde anche a “quanto perdiamo *in media* quando quella soglia viene superata?”.

Dopo le crisi finanziarie, “quanto perdiamo negli scenari peggiori?” è diventata una domanda regolamentare oltre che gestionale. VaR e CVaR sono il linguaggio standard con cui banche, assicurazioni e, sempre più, supply chain manager parlano di rischio di coda. La sorpresa didattica è che il CVaR — concetto apparentemente sofisticato — si ottimizza con un semplice LP.

In questo capitolo impareremo a: (1) definire e calcolare VaR e CVaR; (2) usare la formulazione lineare di Rockafellar–Uryasev; (3) costruire portafogli e piani logistici protetti dagli scenari peggiori; (4) riconoscere i limiti statistici delle stime di coda.

13.2 Definizioni e proprietà

Il problema a parole. Nei modelli di questo capitolo *decidiamo* come prima (quote di portafoglio, quantità, capacità prenotate), ma *l'obiettivo* cambia: non minimizzare la perdita *media*, bensì una misura della sua *coda* — quanto si perde negli scenari peggiori. *I vincoli* restano quelli del problema sottostante, più i vincoli lineari ausiliari che definiscono il CVaR.

Sia L la perdita aleatoria (valori alti = esiti peggiori) e $\alpha \in (0, 1)$ un livello di confidenza (tipicamente 0,90–0,99). Come per la domanda del newsvendor, la variabile aleatoria è maiuscola e le sue realizzazioni minuscole (ℓ_s): è l'eccezione, per aderenza alla letteratura, alla nostra convenzione che riserva le maiuscole a insiemi e matrici.

VaR e CVaR

$$\text{VaR}_\alpha(L) = \inf\{\eta \in \mathbb{R} : \mathbb{P}(L \leq \eta) \geq \alpha\},$$

$$\text{CVaR}_\alpha(L) = \text{perdita media nella coda peggiore di massa } 1 - \alpha.$$

Convessità	il CVaR è convesso nelle decisioni (se L lo è); il VaR in generale no
Subadditività	il CVaR premia la diversificazione; il VaR può violarla
Coda	il VaR non distingue due code con lo stesso quantile ma gravità diversa
Ottimizzazione	il CVaR ammette una formulazione LP a scenari; il VaR no

Dati del modello a scenari: il numero di scenari $k \in \mathbb{Z}_{\geq 1}$, indicizzati da $s \in \{1, 2, \dots, k\}$; per ogni scenario, la probabilità $\pi_s \in \mathbb{Q}_{\geq 0}$ (con $\sum_{s=1}^k \pi_s = 1$) e la perdita $\ell_s(\mathbf{x})$, funzione lineare del vettore delle decisioni $\mathbf{x}$, vincolato in un poliedro X ; il livello di confidenza $\alpha \in (0, 1)$.

Oltre alle decisioni $\mathbf{x}$, introduciamo una variabile libera e k variabili non negative:

$$\begin{cases} \eta = \text{soglia candidata di perdita (all'ottimo: un VaR)} \\ \xi_s = \text{eccesso di perdita oltre } \eta \text{ nello scenario } s \end{cases} \quad \forall s \in \{1, 2, \dots, k\}.$$

Usando queste variabili, il modello LP di Rockafellar–Uryasev è il seguente:

Modello LP di Rockafellar–Uryasev per il CVaR

$$\min \quad \eta + \frac{1}{1-\alpha} \sum_{s=1}^k \pi_s \xi_s \quad (13.1a)$$

$$\text{soggetto a} \quad \xi_s \geq \ell_s(\mathbf{x}) - \eta, \quad \forall s \in \{1, 2, \dots, k\}, \quad (13.1b)$$

$$\mathbf{x} \in X, \quad (13.1c)$$

$$\eta \geq 0, \quad (13.1d)$$

$$\xi_s \geq 0, \quad \forall s \in \{1, 2, \dots, k\}. \quad (13.1e)$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (13.1a) è la soglia η più la media pesata degli eccessi di coda, amplificata di $1/(1-\alpha)$; *all'ottimo* η^* è un VaR e il valore obiettivo è il CVaR: un solo LP fornisce entrambe le misure;
- i vincoli lineari (13.1b) definiscono gli eccessi: solo gli scenari con perdita oltre la soglia contribuiscono (k vincoli lineari);
- il vincolo (13.1c) raccoglie i vincoli del problema decisionale sottostante (poliedrali);
- i vincoli (13.1d) e (13.1e) definiscono le variabili: η è libera (il simbolo $\geq$), ξ_s non negative.

Spiegazione riga per riga

La variabile $\xi_s = (\ell_s - \eta)^+$ è l'eccesso di perdita oltre la soglia candidata η (stesso trucco delle parti positive del Newsvendor). Minimizzare su η bilancia due spinte: alzare η riduce gli eccessi ma sposta in su la soglia; il punto di equilibrio è esattamente il quantile α . Se X è un poliedro e ℓ_s è lineare, tutto il modello è un LP.

13.3 Esempio svolto a mano

Sei scenari, $\alpha = 0,80$

Perdite equiprobabili $\{2, 4, 5, 7, 12, 20\}$.

VaR. La cumulata raggiunge 0,80 in corrispondenza della quinta perdita ordinata: $\text{VaR}_{0,80} = 12$.

CVaR con la formula. La coda peggiore ha massa 0,20; lo scenario da 20 pesa $1/6 = 0,167$, e la massa restante 0,033 sta *sul punto* 12:

$$\text{CVaR}_{0,80} = 12 + \frac{1}{0,20} \cdot \frac{1}{6} (20 - 12) = 12 + 6,67 = 18,67.$$

Verifica con l'LP (script): $\eta^* = 12$, $\xi = (0, 0, 0, 0, 0, 8)$, valore 18,6667. La formula tratta correttamente la massa "a cavallo" del quantile senza selezionare scenari con variabili binarie.

Confronto. Perdita media 8,33; VaR 12; CVaR 18,67: tre numeri, tre storie. La media rassicura, il VaR dà una soglia, il CVaR dice quanto fa male quando va male.

13.4 Caso di studio 1: portafoglio mean-CVaR contro Markowitz

Stessi otto titoli del capitolo 6, ma $k = 220$ scenari mensili generati con *code grasse* (fattore di mercato t di Student, gradi di libertà 4): il mondo in cui il CVaR mostra il suo valore. Perdita di scenario: $\ell_s(\mathbf{x}) = -\sum_{i=1}^n r_{is} x_i$, dove r_{is} è il rendimento del titolo i nello scenario s . Minimizziamo $\text{CVaR}_{0,90}$ con rendimento atteso $\geq 8\%$ annuo.

```

1 eta = m.addVar(lb=-GRB.INFINITY, name="eta")    # soglia: libera!
2 xi  = m.addVars(S, name="xi")
3 m.addConstr(x.sum() == 1)
4 m.addConstr(gp.quicksum(mu[i]*x[i] for i in range(n)) >= r_min)
5 m.addConstrs((xi[s] >= -gp.quicksum(R[s, i]*x[i] for i in range(n))
6               - eta for s in range(S)))
7 m.setObjective(eta + gp.quicksum(xi[s] for s in range(S))
8               / (S * (1 - alpha)), GRB.MINIMIZE)

```

	perdita media	VaR90	CVaR90	(perdite mensili)
mean-CVaR	-0,0067	0,0327	0,0489	
Markowitz	-0,0067	0,0329	0,0531	
Composizioni:	mean-CVaR: ENE 9,9 FIN 1,9 TEC 10,5 SAN 8,6 UTL 46,5 MAT 22,5			
	Markowitz: ENE 18,4 FIN 6,4 TEC 6,5 SAN 13,6 UTL 40,4 MAT 14,6			

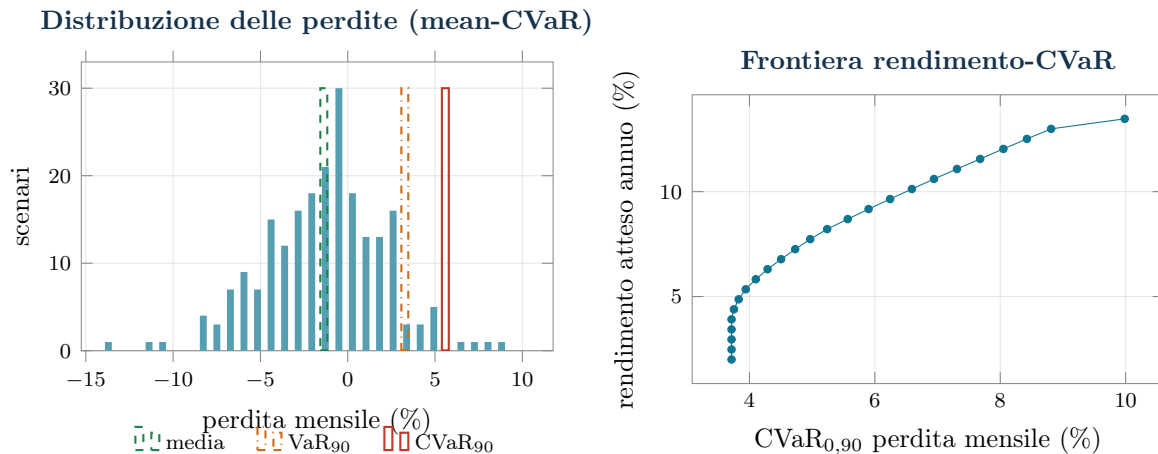


Figura 13.1: A sinistra: istogramma delle perdite mensili del portafoglio mean-CVaR sui 220 scenari, con media (verde), VaR_{90} (arancione) e $CVaR_{90}$ (rosso): il CVaR è la media della coda a destra del VaR. A destra: frontiera rendimento-CVaR ottenuta variando il rendimento minimo richiesto.

Insight manageriale

A parità di rendimento richiesto e di perdita media, il portafoglio mean-CVaR taglia la coda: $CVaR_{90}$ del 4,89% contro 5,31% di Markowitz (−8%). Lo fa riducendo i titoli ad alto beta (ENE dal 18% al 10%) a favore di UTL e — sorpresa — di una quota di TEC che, pur mediocre in media, si muove in controtendenza negli scenari peggiori di questo campione. La varianza penalizza simmetricamente sopra e sotto la media; il CVaR guarda solo dove fa male.

13.5 Caso di studio 2: supply chain a due stadi

Prima di osservare lo scenario si *prenota* capacità dai fornitori; poi, visto lo scenario, si decide il ricorso. F1 è economico ma fragile (nel 12% degli scenari la sua disponibilità crolla al 30%); F2 è caro ma affidabile. Penalità di shortage: 40 €/unità; $k = 400$ scenari.

Dati: due fornitori indicizzati da $a \in \{1, 2\}$ (1 economico ma fragile, 2 caro ma affidabile) con costi di prenotazione $c_a^{\text{pre}} \in \mathbb{Q}_{>0}$ e di acquisto $c_a^{\text{uso}} \in \mathbb{Q}_{>0}$; k scenari $s \in \{1, 2, \dots, k\}$ con domanda $d_s \in \mathbb{Q}_{\geq 0}$ e disponibilità $\delta_{as} \in (0, 1]$ (il dato che modella il guasto); penalità di shortage $g \in \mathbb{Q}_{>0}$; peso del rischio $\lambda \in [0, 1]$ e livello $\alpha \in (0, 1)$.

Introduciamo le variabili non negative x_a (capacità prenotata dal fornitore a , di *primo stadio*: uguale in tutti gli scenari), f_{as} (acquisto dal fornitore a nello scenario s) e u_s (domanda non servita nello scenario s), queste ultime di *ricorso*.

Modello a due stadi con CVaR

$$\min (1 - \lambda) \left[\sum_{a=1}^2 c_a^{\text{pre}} x_a + \sum_{s=1}^k \pi_s \phi_s \right] + \lambda \text{CVaR}_\alpha \quad (13.2a)$$

$$\text{soggetto a } f_{as} \leq \delta_{as} x_a, \quad \forall a \in \{1, 2\}, \forall s \in \{1, 2, \dots, k\}, \quad (13.2b)$$

$$f_{1s} + f_{2s} + u_s \geq d_s, \quad \forall s \in \{1, 2, \dots, k\}, \quad (13.2c)$$

$$x_a, f_{as}, u_s \geq 0, \quad \forall a \in \{1, 2\}, \forall s \in \{1, 2, \dots, k\}, \quad (13.2d)$$

dove $\phi_s = \sum_{a=1}^2 c_a^{\text{uso}} f_{as} + g u_s$ è il costo di ricorso dello scenario s e CVaR_α è il CVaR del costo totale, linearizzato con le variabili η, ξ_s di (13.1).

Descrizione della funzione obiettivo e dei vincoli del modello:

- la funzione obiettivo (13.2a) combina, con peso λ , il costo medio (prenotazioni più valore atteso di acquisti e penalità) e il CVaR del costo totale;
- i vincoli lineari (13.2b) impongono che nello scenario s si possa usare il fornitore a solo fino alla capacità prenotata x_a , ridotta dalla disponibilità δ_{as} ($2k$ vincoli lineari);
- i vincoli lineari (13.2c) impongono di coprire la domanda con i flussi o, in mancanza, di conteggiarla come shortage u_s , che in obiettivo paga la penalità: il modello non è mai inammissibile, “compra” l’inammissibilità al prezzo della penalità (k vincoli lineari);
- i vincoli (13.2d) definiscono le variabili del modello.

```
lambda = 0,0: prenoto forn.1 161,5   forn.2 134,2
              costo medio 1.724   CVaR90 2.902   servizio 98,7%
lambda = 0,5: prenoto forn.1  92,7   forn.2 212,7
              costo medio 1.803   CVaR90 2.217   servizio 99,9%
lambda = 1,0: prenoto forn.1  68,5   forn.2 236,9
              costo medio 1.906   CVaR90 2.139   servizio 99,0%
```

Insight manageriale

Il decisore neutrale compra dal fornitore economico e accetta che nel 12% degli scenari lo shortage faccia esplodere i costi (CVaR 2.902). Al crescere di λ la capacità migra verso il fornitore affidabile: +79 € di costo medio comprano -685 € di CVaR. Questo è il *costo della resilienza*, ed è esattamente il numero che serve in una discussione su rishoring e dual sourcing.

13.6 Analisi di sensitività

I parametri da variare, con l'effetto atteso:

α	più alto $\Rightarrow$ coda più estrema stimata da meno scenari: servono più scenari per stime stabili
λ	controlla il compromesso prestazione media / protezione
π_s	scenari ripesati o stress test (aggiungere scenari estremi)
$ S $	stabilità statistica: ripetere con campioni diversi, come nel capitolo 12

Con $\alpha = 0,99$ e $k = 220$ scenari la coda contiene solo 2–3 scenari: il CVaR stimato è quasi rumore. Regola pratica: servono almeno qualche decina di scenari *oltre* il quantile. Inoltre con distribuzioni discrete il VaR può non essere unico (η^* è un VaR): confrontare sempre η^* con il quantile empirico.

13.7 Esercizi

Esercizio 13.1 — verifica

Ricalcolare a mano VaR e CVaR dell'esempio a 6 scenari con $\alpha = 0,90$ e verificare con l'LP. Attenzione: la coda ha massa 0,10 e lo scenario da 20 pesa 0,167 > 0,10.

Esercizio 13.2 — subadditività

Costruire un esempio a scenari con due attività in cui $\text{VaR}(L_1 + L_2) > \text{VaR}(L_1) + \text{VaR}(L_2)$ e verificare che il CVaR invece non viola la subadditività.

Esercizio 13.3 — frontiera in α

Per il portafoglio, tracciare il CVaR ottimo per $\alpha \in \{0,80; 0,90; 0,95; 0,99\}$ a parità di rendimento: come cambia la composizione? Quando iniziano le instabilità numeriche?

Esercizio 13.4 — CVaR vincolato

Riformulare il portafoglio come $\max \sum_{i=1}^n \mu_i x_i$ soggetto a $\text{CVaR}_{0,90} \leq \bar{c}$: verificare che variando $\bar{c}$ si percorre la stessa frontiera e leggere il duale del vincolo.

Esercizio 13.5 — stress test

Aggiungere alla supply chain uno scenario estremo (domanda +50% e F1 al 10%) con probabilità 1%: quanto cambia la prenotazione ottima per $\lambda = 0,5$? E il CVaR?

Arbitraggio e prezzatura senza arbitraggio

Classe: LP

Script: [python/lab14_arbitraggio.py](#)[Pagina online](#)

14.1 Motivazione gestionale

Un *arbitraggio* è una strategia di compravendita che crea denaro dal nulla: incassa oggi senza alcun rischio di perdita domani, oppure non costa nulla oggi e può solo guadagnare domani. In un mercato ben funzionante queste occasioni non possono durare: chiunque le sfrutterebbe su scala illimitata, e i prezzi si aggiusterebbero subito. “Assenza di arbitraggio” è quindi l’ipotesi di lavoro di tutta la finanza dei prezzi: dei desk di trading che cercano incoerenze tra strumenti quotati, dei risk manager che validano i prezzi di listino, di chi deve dare un prezzo a un titolo nuovo coerente con quelli esistenti.

La sorpresa didattica è che tutto questo è *programmazione lineare*: rilevare un arbitraggio è risolvere un LP, e la dualità del capitolo 2 — prezzi ombra, scarti complementari, dualità forte — diventa la teoria della prezzatura. È l’applicazione più pulita della coppia primale-duale di tutta la dispensa.

In questo capitolo impareremo a: (1) formulare la ricerca di un arbitraggio come LP; (2) leggere nei duali dei vincoli le *probabilità neutrali al rischio*; (3) calcolare l’intervallo di prezzi coerenti di un titolo; (4) prezzare un titolo nuovo (un’opzione) in un mercato completo o incompleto.

14.2 Costruzione guidata del modello

Il problema a parole. Il mercato ha due istanti: oggi ($t = 0$), in cui compriamo o vendiamo titoli ai prezzi correnti, e domani ($t = 1$), in cui si realizza uno tra m *stati del mondo* e ogni titolo paga un importo noto che dipende dallo stato. *Decidiamo* le quantità di ciascun titolo in portafoglio (positive se comprate, negative se vendute allo scoperto). *L’obiettivo* è minimizzare il costo del portafoglio oggi. *I vincoli* impongono che domani, in *ogni* stato del mondo, il portafoglio non perda: se con queste regole si riesce a spendere meno di zero (cioè a incassare), quello è un arbitraggio.

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di titoli rischiosi; i titoli sono numerati $0, 1, \dots, n$ e il titolo 0 è privo di rischio
m	$\in \mathbb{Z}_{\geq 1}$	numero di stati del mondo a $t = 1$
r	$\in \mathbb{Q}_{\geq 0}$	tasso privo di rischio; $R = 1 + r$ è il rendimento lordo del titolo 0, uguale in ogni stato
s_i^0	$\in \mathbb{Q}_{> 0}$	prezzo del titolo i oggi ($s_0^0 = 1$: il titolo 0 costa 1 e paga R)
s_{ij}^1	$\in \mathbb{Q}_{\geq 0}$	payoff del titolo i nello stato j ($s_{0j}^1 = R$ per ogni stato j)

Variabili decisionali

Introduciamo una variabile libera per ciascuno degli $n + 1$ titoli:

x_i = quantità del titolo i in portafoglio (> 0 acquisto, < 0 vendita allo scoperto), $\forall i \in \{0, 1, \dots, n\}$.

Usando queste variabili, il modello LP per la ricerca di un arbitraggio è il seguente:

Modello LP di ricerca dell'arbitraggio

$$\min \sum_{i=0}^n s_i^0 x_i \quad (14.1a)$$

$$\text{soggetto a } \sum_{i=0}^n s_{ij}^1 x_i \geq 0, \quad \forall j \in \{1, 2, \dots, m\}, \quad (14.1b)$$

$$x_i \geq 0, \quad \forall i \in \{0, 1, \dots, n\}. \quad (14.1c)$$

Descrizione della funzione obiettivo e dei vincoli del modello:

- la funzione obiettivo lineare (14.1a) minimizza il costo del portafoglio oggi: un valore negativo è un incasso immediato;
- i vincoli lineari (14.1b) impongono che domani il portafoglio non perda in nessuno degli m stati del mondo (m vincoli lineari);
- i vincoli (14.1c) definiscono le variabili, libere: si può comprare e vendere allo scoperto ($n + 1$ variabili).

Leggere l'esito: tipo A, tipo B, e la normalizzazione

Il portafoglio nullo è sempre ammissibile, quindi l'ottimo è sempre ≤ 0 . Tre esiti possibili:

- **ottimo < 0 (arbitraggio di tipo A):** si incassa oggi senza rischi domani. I vincoli sono omogenei: la strategia si può scalare a piacere e il modello è *illimitato*. Per ottenere un certificato finito si aggiunge la *normalizzazione* $\sum_{i=0}^n s_i^0 x_i \geq -1$ ("incasso oggi al più 1"): l'ottimo diventa -1 e la soluzione è una strategia concreta;
- **ottimo $= 0$ con qualche vincolo di stato lasco in ogni soluzione ottima (arbitraggio di tipo B):** oggi non costa nulla, domani non si perde mai e in almeno uno stato si guadagna;

- **ottimo = 0 e all'ottimo tutti i vincoli attivi:** nessun arbitraggio, i prezzi sono coerenti.

14.3 La dualità è la prezzatura: le probabilità neutrali al rischio

Applichiamo al modello (14.1) le regole della dualità del capitolo 2 (primale di minimo: tabella da destra a sinistra). Ogni vincolo di stato (14.1b), di verso $\geq$, genera una variabile duale $p_j \geq 0$; ogni variabile libera x_i genera un vincolo duale di uguaglianza; i termini noti (tutti nulli) diventano l'obiettivo del duale:

Il duale: misure di prezzatura

$$\max \sum_{j=1}^m 0 \cdot p_j \quad (14.2a)$$

$$\text{soggetto a } \sum_{j=1}^m s_{ij}^1 p_j = s_i^0, \quad \forall i \in \{0, 1, \dots, n\}, \quad (14.2b)$$

$$p_j \geq 0, \quad \forall j \in \{1, 2, \dots, m\}. \quad (14.2c)$$

- l'obiettivo (14.2a) è identicamente nullo: del duale conta solo l'*ammissibilità* (ogni soluzione ammissibile è ottima);
- i vincoli (14.2b) dicono che il prezzo di oggi di ogni titolo è la somma dei suoi payoff pesati con i p_j : i duali *prezzano* i titoli ($n+1$ vincoli lineari);
- i vincoli (14.2c) definiscono le variabili (m variabili non negative).

Il vincolo duale del titolo 0 dà $\sum_{j=1}^m R p_j = 1$, cioè $\sum_{j=1}^m p_j = 1/R$: i numeri $q_j = R p_j$ sono non negativi e sommano a 1 — sono *probabilità*, dette **neutrali al rischio**. Con esse i vincoli (14.2b) si leggono:

$$s_i^0 = \frac{1}{R} \sum_{j=1}^m q_j s_{ij}^1, \quad \forall i \in \{0, 1, \dots, n\} :$$

il prezzo di oggi è il valore atteso scontato dei payoff di domani, sotto le probabilità q .

Teorema fondamentale della prezzatura

Non esiste alcun arbitraggio (di tipo A né di tipo B) *se e solo se* esiste una misura di probabilità neutrale al rischio con $q_j > 0$ per ogni $j \in \{1, 2, \dots, m\}$.

Perché è vero: dualità forte e scarti complementari

È il capitolo 2 al lavoro. Se non c'è arbitraggio di tipo A, l'ottimo del primale è 0: per la *dualità forte* il duale è ammissibile, cioè una misura di prezzatura $p \geq 0$ esiste. Se non c'è nemmeno arbitraggio di tipo B, tutti i vincoli di stato sono attivi all'ottimo, e per gli *scarti complementari* esiste una soluzione duale con $p_j > 0$ per ogni stato: moltiplicando per R si ottengono le $q_j > 0$. Viceversa, se esiste $q > 0$, ogni strategia con payoff non negativi ha

costo $\sum_i s_i^0 x_i = \frac{1}{R} \sum_j q_j (\sum_i s_{ij}^1 x_i) \geq 0$: nessun arbitraggio.

14.4 Esempio svolto a mano

Tre stati, un arbitraggio evidente

Mercato con $m = 3$ stati, $r = 4\%$ ($R = 1,04$) e due titoli rischiosi con payoff $s_1^1 = (10, 15, 13)$ e $s_2^1 = (30, 15, 25)$; prezzi di oggi $s_1^0 = 6$ e $s_2^0 = 20$.

- **Passo 1 — una strategia candidata.** Compriamo un'unità di ciascun titolo rischioso ($x_1 = x_2 = 1$, spesa 26) vendendo allo scoperto il titolo privo di rischio per 27 ($x_0 = -27$). Costo oggi:

$$1 \cdot (-27) + 6 \cdot 1 + 20 \cdot 1 = -1 :$$

incassiamo 1.

- **Passo 2 — nessun rischio domani.** Nei tre stati il portafoglio paga
 $-27 \cdot 1,04 + 10 + 30 = 11,92$, $-27 \cdot 1,04 + 15 + 15 = 1,92$, $-27 \cdot 1,04 + 13 + 25 = 9,92$,
 tutti ≥ 0 : è un arbitraggio di tipo A.
- **Passo 3 — perché il modello è illimitato.** Raddoppiando la strategia si incassano 2 oggi con payoff ancora ≥ 0 : senza la normalizzazione il costo può scendere quanto si vuole.

14.5 Caso di studio

Stesso mercato dell'esempio ($R = 1,04$, payoff in [dati/arbitraggio_payoff.csv](#)), analizzato con lo script in due configurazioni di prezzi: quella con arbitraggio ($s_1^0 = 6$, $s_2^0 = 20$) e una coerente ($s_1^0 = 13$, $s_2^0 = 18,6923$).

```
1 x = m.addVars(n1, lb=-GRB.INFINITY, name="x")      # posizioni libere
2 v_stato = m.addConstrs(
3     (gp.quicksum(s1[i, j] * x[i] for i in range(n1)) >= 0
4     for j in range(n_stati)), name="stato")
5 if normalizza: # senza: con arbitraggio il modello e' illimitato
6     m.addConstr(gp.quicksum(s0[i] * x[i] for i in range(n1)) >= -1)
7 m.setObjective(gp.quicksum(s0[i] * x[i] for i in range(n1)),
8                 GRB.MINIMIZE)
```

14.6 Risultati

```
Prezzi (6, 20):      senza normalizzazione: UNBOUNDED (tipo A)
                    normalizzato: ottimo -1 -> incasso 1 oggi, payoff (0; 0; 0,20) >= 0
Prezzi (13, 18.69): ottimo 0, nessun arbitraggio
duali dei vincoli di stato p* = (0,2846; 0,6769; 0)
q = R p*              = (0,296; 0,704; 0) somma 1
```

verifica prezzatura: $s_0 = (1; 13; 18,6923)$ riprodotti esattamente

I duali prezzano

Con i prezzi coerenti l'LP restituisce ottimo 0 e, nei duali dei vincoli di stato, la misura di prezzatura: $q = (0,296; 0,704; 0)$. Ogni titolo — e ogni titolo *nuovo* — si prezza come valore atteso scontato dei payoff sotto q . La $q_3 = 0$ segnala una degenerazione informativa: lo stato 3 non serve a prezzare questi titoli (il suo payoff è replicabile con gli altri due stati), ed è il motivo per cui qui il “se e solo se” del teorema richiede $q > 0$: con $q_3 = 0$ resta un arbitraggio di tipo B latente sullo stato 3.

14.7 Analisi di sensitività: l'intervallo dei prezzi coerenti

Se sono quotati solo il titolo 0 (prezzo 1) e il titolo 1 (prezzo 13), quali prezzi del titolo 2 *non* creano arbitraggio? Due LP sulle misure di prezzatura — minimizzare e massimizzare $\sum_j s_{2j}^1 p_j$ sui $p \geq 0$ coerenti con i titoli quotati — danno l'intervallo

$$s_2^0 \in [18,69, 21,54].$$

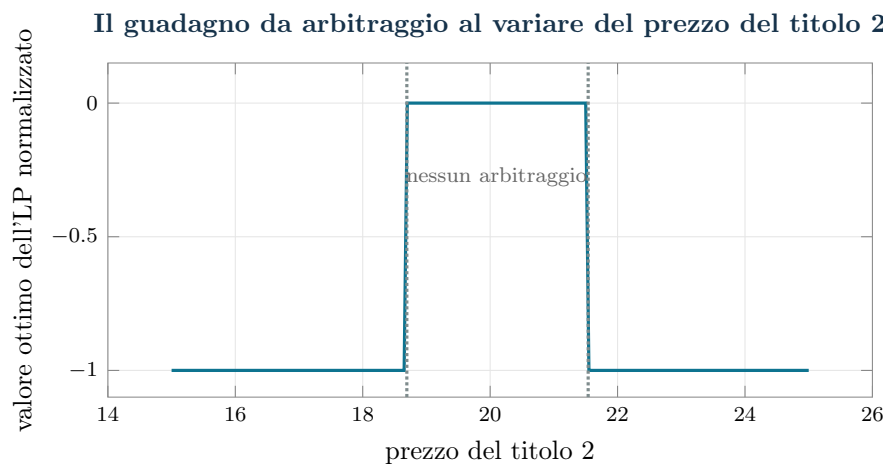


Figura 14.1: Valore ottimo dell'LP di arbitraggio normalizzato al variare del prezzo del titolo 2 (titoli 0 e 1 quotati a 1 e 13). Dentro l'intervallo $[18,69, 21,54]$ (linee tratteggiate) il valore è 0: prezzi coerenti; fuori, il valore scende a -1 : esiste un arbitraggio, in una direzione o nell'altra.

Prezzare un titolo nuovo: completo vs incompleto

La stessa coppia di LP prezza qualunque titolo nuovo. Una *call* sul titolo 1 con strike 12 paga $\max(s_{1j}^1 - 12, 0) = (0, 3, 1)$ nei tre stati.

- **Mercato completo** (titoli 0, 1, 2 quotati): la misura di prezzatura è unica e il prezzo pure: 2,0308.
- **Mercato incompleto** (solo titoli 0 e 1): le misure coerenti sono infinite e il prezzo è un *intervallo*, $[1,4615, 2,0308]$: sotto il minimo o sopra il massimo, chiunque potrebbe costruire un arbitraggio combinando call e titoli quotati.

14.8 Esercizi

Esercizio 14.1 — scala dell'arbitraggio

Con i prezzi $(6, 20)$, trovare una strategia che incassa 5 oggi senza rischi domani. (Suggerimento: le soluzioni dell'LP omogeneo si possono scalare.)

Esercizio 14.2 — estremo dell'intervallo

Porre il prezzo del titolo 2 all'estremo superiore 21,5385 e risolvere l'LP di arbitraggio: verificare che l'ottimo è 0 e calcolare la misura di prezzatura associata. Che cosa succede appena sopra, a 21,60?

Esercizio 14.3 — prezzatura di una call

Nel mercato completo (prezzi 1, 13, 18,6923), prezzare una call sul titolo 2 con strike 20 (payoff $\max(s_{2j}^1 - 20, 0)$) e verificare il prezzo con le probabilità neutrali al rischio del caso di studio.

Parte IV

Ottimizzazione e machine learning

Support Vector Machine: ottimizzazione per il machine learning

Classe: QP convesso

Script: [python/lab15_svm.py](#)

[Pagina online](#)

15.1 Motivazione gestionale

La Support Vector Machine collega direttamente l'ottimizzazione convessa al machine learning: il classificatore si ottiene risolvendo un QP, con una lettura geometrica chiara. Le applicazioni gestionali sono immediate: rischio di credito, churn, manutenzione predittiva, controllo qualità, frodi. In questo capitolo *non* usiamo librerie di ML: ogni modello è un QP scritto e risolto con Gurobi, perché l'obiettivo è capire *che cosa* ottimizza un classificatore.

Nota modellistica: le etichette $y_i \in \{-1, +1\}$ sono *dati osservati*, non variabili — le variabili (coefficienti, intercetta, scarti, moltiplicatori) sono tutte continue.

Questo capitolo chiude il cerchio: gli stessi strumenti usati per produzione, logistica e finanza — QP convessi, dualità, moltiplicatori — sono il motore di uno dei classificatori più eleganti del machine learning. Capire la SVM come problema di ottimizzazione, e non come scatola nera di libreria, dà il controllo completo su ciò che il modello sta davvero facendo.

In questo capitolo impareremo a: (1) formulare hard e soft margin come QP; (2) derivare e risolvere il duale, identificando i support vector; (3) ottenere frontiere non lineari con i kernel restando nel mondo convesso; (4) gestire classi sbilanciate con costi asimmetrici e valutare con le metriche giuste.

15.2 Hard margin: la geometria

Il problema a parole. *Decidiamo* i coefficienti di un iperpiano che separi due classi di osservazioni. *L'obiettivo* è massimizzare il margine, cioè la distanza tra l'iperpiano e i punti più vicini delle due classi (un classificatore con margine ampio generalizza meglio). *I vincoli* chiedono che ogni osservazione stia dal lato giusto dell'iperpiano — nel soft margin, a meno di uno scarto pagato in obiettivo.

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di osservazioni del training set, indicizzate da $i \in \{1, 2, \dots, n\}$
p	$\in \mathbb{Z}_{\geq 1}$	numero di caratteristiche, indicizzate da $j \in \{1, 2, \dots, p\}$
x_{ij}	$\in \mathbb{Q}$	valore della caratteristica j per l'osservazione i ; $\mathbf{x}_i \in \mathbb{Q}^p$ è il vettore dell'osservazione i
y_i	$\in \{-1, +1\}$	etichetta <i>osservata</i> della classe di i (un dato, non una variabile)
C	$\in \mathbb{Q}_{>0}$	peso delle violazioni del margine (solo soft margin); maiuscolo per aderenza alla letteratura di machine learning, eccezione alla convenzione che riserva le maiuscole a insiemi e matrici

Variabili decisionali

Introduciamo le seguenti $p + 1$ variabili libere:

$$\begin{cases} w_j = \text{coefficiente } j \text{ dell'iperpiano separatore} \\ b = \text{intercetta dell'iperpiano} \end{cases} \quad \forall j \in \{1, 2, \dots, p\},$$

che definiscono l'iperpiano $\sum_{j=1}^p w_j x_j + b = 0$.

Usando queste variabili, il modello QP hard margin è il seguente:

Modello QP hard margin

$$\min \quad \frac{1}{2} \sum_{j=1}^p w_j^2 \quad (15.1a)$$

$$\text{soggetto a} \quad y_i \left(\sum_{j=1}^p w_j x_{ij} + b \right) \geq 1, \quad \forall i \in \{1, 2, \dots, n\}, \quad (15.1b)$$

$$w_j \geq 0, \quad \forall j \in \{1, 2, \dots, p\}, \quad (15.1c)$$

$$b \geq 0. \quad (15.1d)$$

Descrizione della funzione obiettivo e dei vincoli del modello QP:

- la funzione obiettivo quadratica convessa (15.1a) minimizza $\frac{1}{2} \|\mathbf{w}\|^2$; poiché il margine geometrico tra gli iperpiani di supporto è $2/\|\mathbf{w}\|$, minimizzarla equivale a *massimizzare il margine*;
- i vincoli lineari (15.1b) impongono che ogni osservazione stia dal lato giusto dell'iperpiano, a distanza almeno unitaria in scala $\mathbf{w}$ (n vincoli lineari);
- i vincoli (15.1c) e (15.1d) definiscono le variabili del modello, tutte libere.

Due punti, tutto a mano

$\mathbf{x}_1 = (0, 0)$ con $y_1 = -1$ e $\mathbf{x}_2 = (2, 2)$ con $y_2 = +1$.

- **Passo 1 — vincoli.** $-(b) \geq 1$ e $2w_1 + 2w_2 + b \geq 1$, cioè $b \leq -1$ e $2(w_1 + w_2) \geq 1 - b \geq 2$.
- **Passo 2 — simmetria e ottimo.** Per minimizzare $w_1^2 + w_2^2$ con $w_1 + w_2 \geq 1$ conviene $w_1 = w_2 = t$ e prendere le disuguaglianze attive: $4t + b = 1$, $b = -1 \Rightarrow t = 1/2$. Quindi $\mathbf{w} = (0,5; 0,5)$, $b = -1$: l'iperpiano $0,5x_1 + 0,5x_2 = 1$ è l'asse del segmento tra i due punti.
- **Passo 3 — margine.** $2/\|\mathbf{w}\| = 2/\sqrt{0,5} = 2\sqrt{2} = 2,83$: esattamente la distanza tra i due punti. Entrambi i vincoli sono attivi: entrambi i punti sono *support vector*. Con dati reali sovrapposti, però, l'hard margin è inammissibile: serve il soft margin.

15.3 Soft margin

Con dati reali sovrapposti l'hard margin è inammissibile: si aggiungono n variabili non negative di scarto,

$$\xi_i = \text{violazione del margine dell'osservazione } i, \quad \forall i \in \{1, 2, \dots, n\},$$

con $\xi_i > 1$ che corrisponde a una classificazione errata.

Modello QP soft margin (primale)

$$\min \quad \frac{1}{2} \sum_{j=1}^p w_j^2 + C \sum_{i=1}^n \xi_i \quad (15.2a)$$

$$\text{soggetto a} \quad y_i \left(\sum_{j=1}^p w_j x_{ij} + b \right) \geq 1 - \xi_i, \quad \forall i \in \{1, 2, \dots, n\}, \quad (15.2b)$$

$$w_j \geq 0, \quad \forall j \in \{1, 2, \dots, p\}, \quad (15.2c)$$

$$b \geq 0, \quad (15.2d)$$

$$\xi_i \geq 0, \quad \forall i \in \{1, 2, \dots, n\}. \quad (15.2e)$$

Descrizione della funzione obiettivo e dei vincoli del modello QP:

- la funzione obiettivo quadratica convessa (15.2a) bilancia due termini: la semplicità del modello ($\frac{1}{2}\|\mathbf{w}\|^2$, margine ampio) e il costo delle violazioni sul training set ($C \sum_{i=1}^n \xi_i$); C è il prezzo assegnato alle violazioni;
- i vincoli lineari (15.2b) richiedono la classificazione corretta *a meno dello scarto* ξ_i , pagato in obiettivo (n vincoli lineari);
- i vincoli (15.2c), (15.2d) e (15.2e) definiscono le variabili del modello: w_j e b libere, ξ_i non negative.

Caso di studio: rischio di credito

90 clienti, due indicatori standardizzati (solidità patrimoniale, puntualità nei pagamenti), 55 affidabili e 35 insolventi, parzialmente sovrapposti ([dati/svm_clienti.csv](#)).

```
C = 0,05: w = (0,678, 0,474)  b = -0,148 | margine 2,418 | errori 2 | nel
      margine 22
C = 1,00: w = (1,433, 0,611)  b = -0,236 | margine 1,284 | errori 2 | nel
      margine 7
C = 20,00: w = (5,058, 2,338)  b = -2,936 | margine 0,359 | errori 0 | nel
      margine 3
```

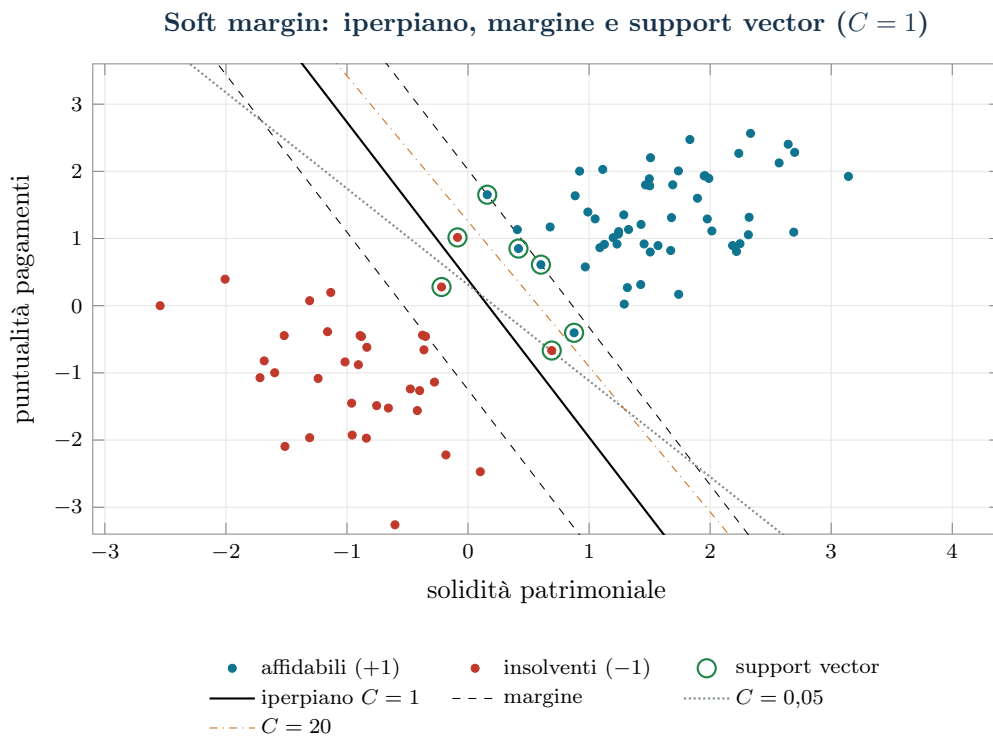


Figura 15.1: I 90 clienti nel piano dei due indicatori: affidabili (teal), insolventi (rosso), support vector cerchiati in verde. La retta nera è l'iperpiano ottimo per $C = 1$ con le sue rette di margine (tratteggiate); le rette grigia e arancione sono gli iperpiani per $C = 0,05$ (margine ampio) e $C = 20$ (margine stretto).

Insight manageriale

C piccolo compra un margine ampio accettando 22 punti dentro il margine; C grande insegue ogni punto del training set (zero errori, margine strettissimo): è il compromesso **regolarizzazione–aderenza**, e va scelto con dati di validazione, mai guardando il test set. Si noti che le tre rette sono simili ma non uguali: la pendenza cambia perché i punti “pagati” cambiano.

15.4 La formulazione duale e i support vector

Associando un moltiplicatore α_i a ciascun vincolo di classificazione (15.2b), si ottengono n nuove variabili:

α_i = moltiplicatore del vincolo di classificazione dell'osservazione i , $\forall i \in \{1, 2, \dots, n\}$.

Modello QP duale del soft margin

$$\max \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}'_i \mathbf{x}_j \quad (15.3a)$$

$$\text{soggetto a } \sum_{i=1}^n \alpha_i y_i = 0, \quad (15.3b)$$

$$\alpha_i \leq C, \quad \forall i \in \{1, 2, \dots, n\}, \quad (15.3c)$$

$$\alpha_i \geq 0, \quad \forall i \in \{1, 2, \dots, n\}. \quad (15.3d)$$

Descrizione della funzione obiettivo e dei vincoli del modello duale:

- la funzione obiettivo quadratica concava (15.3a) dipende dai dati *solo* tramite i prodotti scalari $\mathbf{x}'_i \mathbf{x}_j$: è l'osservazione che apre la porta al kernel;
- il vincolo lineare (15.3b) deriva dalla stazionarietà rispetto all'intercetta b (un vincolo lineare);
- i vincoli (15.3c) limitano ogni moltiplicatore al prezzo C delle violazioni (n vincoli lineari);
- i vincoli (15.3d) definiscono le variabili del modello.

Relazione primale-duale: $\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i$ e $f(\mathbf{x}) = \text{sign}(\mathbf{w}'\mathbf{x} + b)$.

```
w dal duale = (1,433, 0,611)    b = -0,236    (identici al primale)
valore duale 4,7868 = valore primale (dualita' forte)
alpha = 0      : 83 punti -> NON influenzano la frontiera
0 < alpha < C : 2 punti  -> support vector esattamente sul margine
alpha = C      : 5 punti  -> dentro il margine o mal classificati
```

Perché si chiamano support vector

La soluzione dipende *solo* dai 7 punti con $\alpha_i > 0$: sono loro a “sostenere” l'iperpiano; gli altri 83 potrebbero sparire senza cambiare nulla. La complementarietà KKT organizza tutto: $\alpha_i = 0 \Rightarrow$ punto oltre il margine; $0 < \alpha_i < C \Rightarrow$ esattamente sul margine (da questi si ricava b); $\alpha_i = C \Rightarrow$ il vincolo “paga” scarto $\xi_i > 0$. Il duale, inoltre, dipende dai dati solo tramite i prodotti scalari $\mathbf{x}'_i \mathbf{x}_j$: questa osservazione apre la porta al kernel.

15.5 Kernel: frontiere non lineari da un QP convesso

Sostituendo $\mathbf{x}'_i \mathbf{x}_j$ con un kernel $K(\mathbf{x}_i, \mathbf{x}_j)$ — ad esempio il kernel RBF (*Radial Basis Function*) gaussiano $K(\mathbf{x}, \mathbf{z}) = e^{-\gamma \|\mathbf{x} - \mathbf{z}\|^2}$ — si classifica in uno spazio trasformato senza mai costruirlo: $f(\mathbf{x}) = \text{sign}(\sum_{i=1}^n \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b)$. La frontiera nello spazio originale può essere arbitrariamente curva, ma il problema di training *resta un QP convesso* se il kernel è valido (matrice semidefinita positiva).

Caso di studio: anomalie “a corona”

90 osservazioni di processo: il funzionamento normale al centro, le anomalie disposte ad anello — nessuna retta può separarle.

```
gamma = 0,1: errori di training 1   support vector 20
gamma = 0,7: errori di training 0   support vector 15
gamma = 5,0: errori di training 0   support vector 61 <- "memorizza" i punti
```

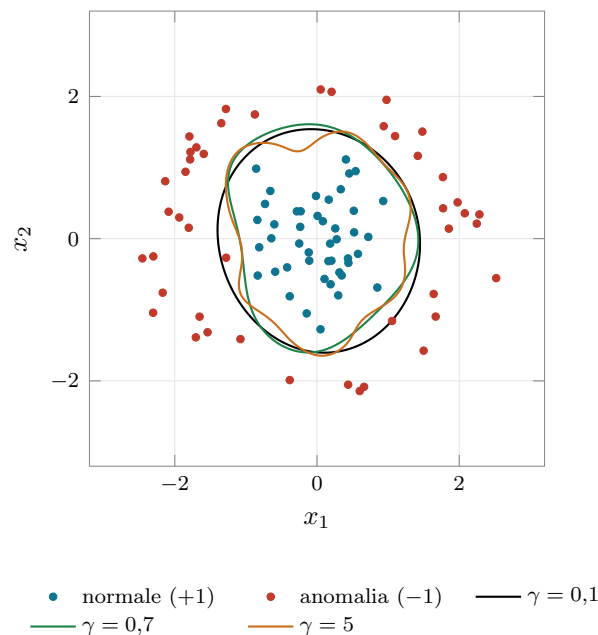
Kernel RBF: la frontiera di livello 0 per tre valori di γ 

Figura 15.2: Il dataset “a corona” (funzionamento normale al centro, anomalie ad anello) e le frontiere di decisione del kernel RBF per tre valori di γ : con $\gamma = 0,1$ la frontiera è un cerchio regolare, con $\gamma = 5$ si frammenta intorno ai singoli punti (overfitting).

Con $\gamma = 5$ la frontiera si frammenta in isole intorno ai singoli punti e i support vector salgono a 61 su 90: il modello sta memorizzando il training set (overfitting), pur restando un QP convesso — convessità del *training* e capacità di *generalizzare* sono proprietà diverse. C e γ vanno scelti insieme per cross-validation.

15.6 Classi sbilanciate e costi asimmetrici

Nelle applicazioni gestionali la classe importante è rara e l'errore sui rari costa di più. Basta pesare gli scarti: $C_- \sum_{i: y_i = -1} \xi_i + C_+ \sum_{i: y_i = +1} \xi_i$.

costi uguali ($C=1$)	: precision 1,00	recall insolventi 0,94
costo 10x sugli insolventi	: precision 0,97	recall insolventi 1,00

Insight manageriale

Con il costo $10\times$ la frontiera arretra verso la classe “sana”: nessun insolvente sfugge (recall 100%) al prezzo di qualche falso allarme in più (precision dal 100% al 97%). I pesi devono riflettere i *costi decisionali* (quanto costa un fido concesso a chi non ripagherà vs. un cliente buono respinto), non le frequenze delle classi.

15.7 Support Vector Regression

La SVR stima una funzione continua ignorando gli errori entro un tubo di ampiezza ε : $\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$ con $|y_i - (\mathbf{w}'\mathbf{x}_i + b)| \leq \varepsilon + \xi_i^{(*)}$ per ogni i .

Retta SVR: domanda = $-10,86 * \text{prezzo} + 220,32$ (vera: $-11p + 220$)
 Tubo eps = 8: 14 punti su 40 fuori dal tubo -> solo loro determinano la retta

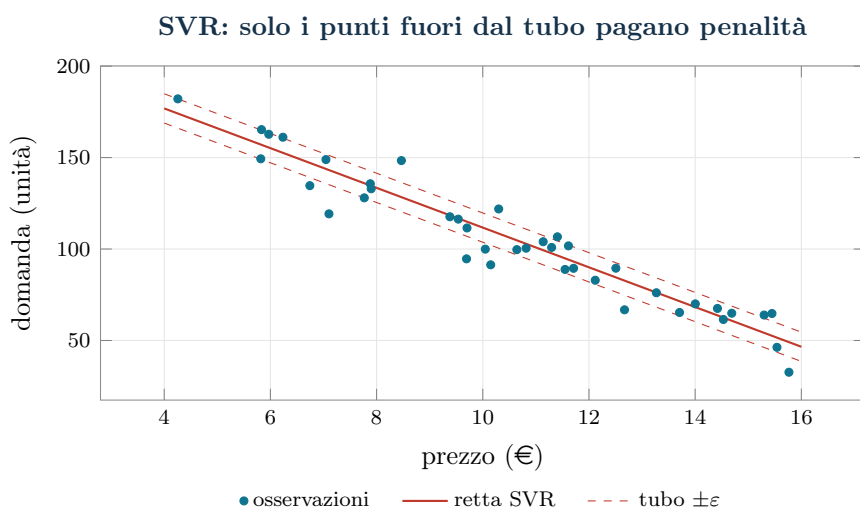


Figura 15.3: Support Vector Regression sulla relazione prezzo–domanda: le 40 osservazioni, la retta stimata e il tubo di tolleranza $\pm\varepsilon = 8$: solo i 14 punti fuori dal tubo pagano penalità e determinano la retta.

15.8 Esercizi**Esercizio 15.1 — verifica**

Risolvere a mano l’hard margin con $\mathbf{x}_1 = (0,0)$, $y_1 = -1$ e $\mathbf{x}_2 = (4,0)$, $y_2 = +1$, e verificare con Gurobi ($\mathbf{w} = (0,5; 0)$, $b = -1$, margine 4).

Esercizio 15.2 — complementarità

Per $C = 1$, elencare dallo script i 7 punti con $\alpha_i > 0$ e verificare numericamente le tre condizioni di complementarità ($\alpha_i = 0 \Rightarrow \xi_i = 0$ e margine ≥ 1 ; ecc.).

Esercizio 15.3 — curva di validazione

Dividere i 90 clienti in training (60) e validazione (30); tracciare l'errore di validazione per $C \in \{0,01; 0,1; 1; 10; 100\}$ e scegliere C . L'errore di training è monotono in C ? E quello di validazione?

Esercizio 15.4 — soglia decisionale

A parità di modello ($C = 1$), spostare la soglia: classificare insolvente se $\mathbf{w}'\mathbf{x} + b < \theta$ con $\theta \in [-1, 1]$. Tracciare precision e recall in funzione di θ : dove sta il punto che massimizza il profitto se un insolvente non individuato costa 10 e un falso allarme 1?

Esercizio 15.5 — one-class

Implementare la One-Class SVM sul dataset a corona usando solo i punti normali: $\min \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu n} \sum_{i=1}^n \xi_i - \rho$ soggetto a $\mathbf{w}'\phi(\mathbf{x}_i) \geq \rho - \xi_i$ (kernel RBF, $\gamma = 0,7$). Quale frazione di anomalie individua per $\nu \in \{0,05; 0,1; 0,2\}$?

Regressione robusta e quantile: stimare i parametri con un LP

Classe: LP (confronto con QP)

Script: [python/lab16_regressione.py](#)

[Pagina online](#)

16.1 Motivazione gestionale

Tutti i modelli visti finora partono da parametri noti: la curva di domanda del capitolo 7, i costi del capitolo 4, gli scenari del capitolo 12. Quei numeri, però, qualcuno li deve stimare — e la stima è essa stessa un problema di ottimizzazione, che si risolve con lo stesso solver del resto del corso.

Il capitolo 15 ha usato l'ottimizzazione per *classificare*. Qui la usiamo per *prevedere una quantità*, e la scelta della funzione obiettivo diventa una decisione gestionale: minimizzando la somma degli scarti *assoluti* il problema resta un LP, la stima non si lascia trascinare dai giorni anomali e — pesando in modo asimmetrico gli scarti in eccesso e in difetto — si può puntare a un *quantile* della domanda invece che al suo centro. È esattamente il numero che serve al newsvendor del capitolo 12.

In questo capitolo impareremo a: (1) formulare la regressione come LP; (2) leggerne il duale, riconoscendo i punti che sostengono la stima e la proprietà del quantile; (3) stimare la quantità da produrre condizionatamente alle condizioni del giorno; (4) selezionare le caratteristiche rilevanti con un budget sui coefficienti, leggendone il prezzo ombra.

16.2 Il modello

Il problema a parole. *Decidiamo* i coefficienti di una funzione lineare che, a partire dalle caratteristiche osservate di un giorno, prevede la domanda. *L'obiettivo* è rendere piccoli gli scostamenti fra domanda osservata e domanda prevista, pesando in modo asimmetrico quelli in eccesso e quelli in difetto. *I vincoli* definiscono, osservazione per osservazione, lo scostamento della previsione dal dato.

Dati (input del modello)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di osservazioni storiche, indicizzate da $i \in \{1, 2, \dots, n\}$
p	$\in \mathbb{Z}_{\geq 0}$	numero di caratteristiche osservate, indicizzate da $j \in \{1, 2, \dots, p\}$
x_{ij}	$\in \mathbb{Q}$	valore della caratteristica j nell'osservazione i
y_i	$\in \mathbb{Q}$	valore osservato da spiegare nell'osservazione i (la domanda)
τ	$\in \mathbb{Q}, 0 < \tau < 1$	quantile a cui si punta: $\tau = 1/2$ dà la mediana, τ grande una previsione prudente verso l'alto

Variabili decisionali

Introduciamo $p + 1$ variabili libere per la funzione da stimare,

$$\begin{cases} w_j = \text{coefficiente della caratteristica } j \\ b = \text{intercetta della funzione} \end{cases} \quad \forall j \in \{1, 2, \dots, p\},$$

e, per ogni osservazione, due variabili non negative che misurano lo scostamento della previsione dal dato:

$$\begin{cases} u_i = \text{scostamento in eccesso dell'osservazione } i \\ v_i = \text{scostamento in difetto dell'osservazione } i \end{cases} \quad \forall i \in \{1, 2, \dots, n\}.$$

Usando queste variabili, il modello LP di regressione quantile è il seguente:

Modello LP di regressione quantile

$$\min \sum_{i=1}^n [\tau u_i + (1 - \tau) v_i] \quad (16.1a)$$

$$\text{soggetto a } \sum_{j=1}^p w_j x_{ij} + b + u_i - v_i = y_i, \quad \forall i \in \{1, 2, \dots, n\}, \quad (16.1b)$$

$$w_j \geq 0, \quad \forall j \in \{1, 2, \dots, p\}, \quad (16.1c)$$

$$b \geq 0, \quad (16.1d)$$

$$u_i, v_i \geq 0, \quad \forall i \in \{1, 2, \dots, n\}. \quad (16.1e)$$

Descrizione della funzione obiettivo e dei vincoli del modello LP:

- la funzione obiettivo lineare (16.1a) paga τ per ogni unità di scostamento in eccesso e $1 - \tau$ per ogni unità di scostamento in difetto: con $\tau = 1/2$ è metà della somma degli scarti assoluti, con τ grande restare sotto il dato costa più che restare sopra;
- i vincoli lineari (16.1b) definiscono lo scostamento di ogni osservazione come differenza fra le sue due parti, $u_i - v_i = y_i - \text{previsione}$ (n vincoli lineari);
- i vincoli (16.1c), (16.1d) e (16.1e) definiscono le variabili del modello: coefficienti e intercetta liberi, scostamenti non negativi.

Perché due variabili per un solo scostamento

Lo scostamento $y_i - (\sum_j w_j x_{ij} + b)$ ha segno qualunque, ma l'obiettivo deve pagarne il *valore assoluto*: spezzandolo nella parte positiva u_i e nella parte negativa v_i , entrambe non negative, l'obiettivo resta lineare. All'ottimo almeno una delle due è nulla: se fossero entrambe positive, ridurle della stessa quantità lascerebbe soddisfatto il vincolo (16.1b) e farebbe diminuire l'obiettivo, perché $\tau > 0$ e $1 - \tau > 0$. È lo stesso trucco delle parti positive già usato per il newsvendor a scenari nel capitolo 12.

Senza caratteristiche: torna il quantile empirico

Con $p = 0$ resta la sola intercetta e il modello sceglie un unico numero b . Sette giorni di vendite osservate: 88, 96, 104, 112, 120, 132, 148 pezzi. La panetteria del capitolo 12 ha $c_u = 9$ e $c_o = 4$, quindi $\tau = \alpha^* = 9/13 = 0,6923$.

- **Passo 1 — che cosa costa muovere b .** Spostando b di una unità verso l'alto si pagano τ in più su ogni osservazione che sta *sopra* b e si risparmiano $1 - \tau$ su ogni osservazione che sta *sotto*. Conviene salire finché le osservazioni sopra sono tante da compensare.
- **Passo 2 — l'ottimo.** Il valore ottimo è $\tilde{b} = 120$, la quinta osservazione in ordine crescente: sotto ce ne sono 4, sopra 2, e vale $4/7 = 0,5714 \leq \tau \leq 5/7 = 0,7143$. Il costo ottimo è

$$\tilde{z} = \frac{4}{13} (32 + 24 + 16 + 8) + \frac{9}{13} (12 + 28) = \frac{320+360}{13} = \frac{680}{13} = 52,3077.$$

- **Passo 3 — i duali.** I duali dei sette vincoli valgono $\tilde{\pi} = (-\frac{4}{13}, -\frac{4}{13}, -\frac{4}{13}, -\frac{4}{13}, -\frac{2}{13}, \frac{9}{13}, \frac{9}{13})$: $\tau = 9/13$ sulle osservazioni sopra la stima, $-(1 - \tau) = -4/13$ su quelle sotto, un valore interno sull'osservazione interpolata. La loro somma è nulla e il valore duale $\sum_i \tilde{\pi}_i y_i = 680/13$ coincide con quello primale.

Il modello, senza caratteristiche, restituisce il **quantile empirico**: è la regola del quantile del capitolo 12, scritta come LP. Con le caratteristiche diventa un quantile *condizionato* alle condizioni del giorno.

16.3 Caso di studio: lo storico della panetteria

Sessanta giorni di vendite della panetteria del capitolo 12: per ogni giorno la domanda in pezzi, la temperatura massima, il prezzo praticato e l'indicazione di weekend ([dati/panetteria_storico.csv](#)). Cinque di quei giorni cadono durante due scioperi dei trasporti: il centro è deserto e le vendite crollano a una ventina di pezzi.

Stimiamo prima la relazione con la sola temperatura ($p = 1$), che si vede in un grafico, confrontando due obiettivi diversi sugli stessi dati: la somma degli scarti assoluti (LP, $\tau = 1/2$) e la somma degli scarti al quadrato (QP dei minimi quadrati, $\min \sum_i r_i^2$ con r_i libera e $\sum_j w_j x_{ij} + b + r_i = y_i$).

```
LP (mediana)          : domanda = 209,77 - 3,63 * temperatura
QP (minimi quadrati): domanda = 202,34 - 3,61 * temperatura
Scarto assoluto medio sui 55 giorni normali: LP 13,08 pezzi, QP 14,81 pezzi
QP senza i 5 giorni di sciopero: domanda = 213,58 - 3,72 * temperatura
Distanza dall'intercetta "pulita" 213,58: LP 3,81 pezzi, QP 11,24 pezzi
```

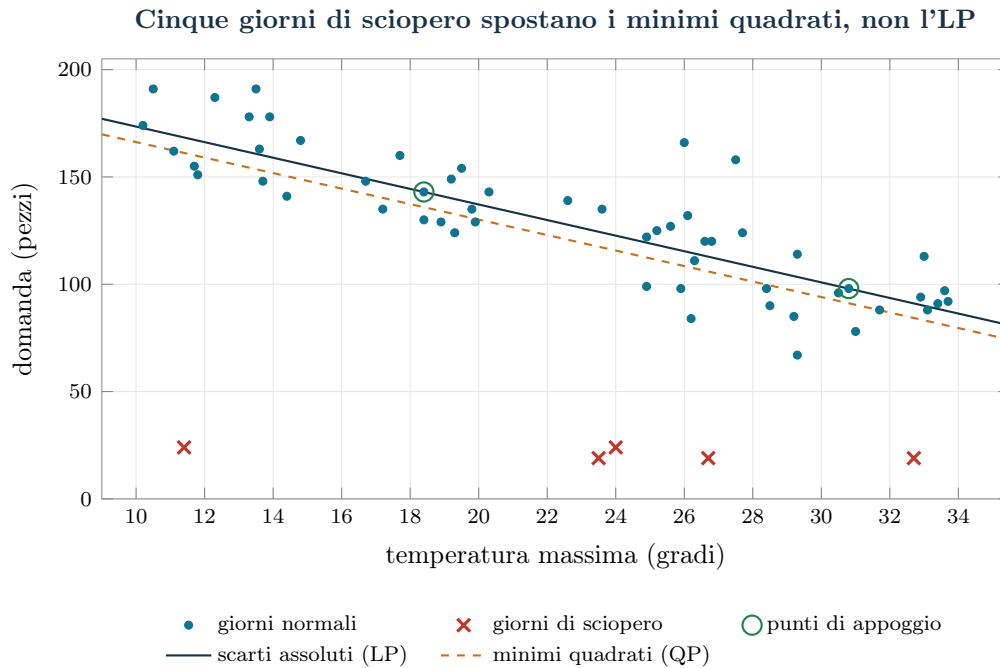


Figura 16.1: I 60 giorni della panetteria: la retta dell'LP (blu) ignora i cinque giorni di sciopero, quella dei minimi quadrati (arancione tratteggiata) ne viene tirata verso il basso di oltre 7 pezzi sull'intercetta. I due punti cerchiati in verde sono i punti di appoggio, gli unici che la retta dell'LP interpola.

Insight manageriale

Un solo dato anomalo su dieci sposta i minimi quadrati perché il loro obiettivo paga lo scarto *al quadrato*: uno scarto di 100 pesa come diecimila scarti di uno. L'LP paga lo scarto *una volta sola* e la retta resta dove sono i dati normali: arriva da sola alla stessa risposta che il QP dà soltanto dopo che qualcuno ha ripulito a mano lo storico. Cambiare obiettivo, qui, non è una raffinatezza tecnica: è la differenza tra un modello che va sorvegliato e uno che si può lasciare girare ogni settimana.

16.4 La lettura duale: i punti di appoggio

Associando un duale π_i a ciascun vincolo di scostamento (16.1b), la condizione di ottimalità sulle variabili u_i e v_i (non negative) limita ogni duale a una finestra che dipende solo da τ :

$$-(1 - \tau) \leq \tilde{\pi}_i \leq \tau, \quad \forall i \in \{1, 2, \dots, n\}, \quad (16.2)$$

mentre l'ottimalità sulle variabili libere w_j e b dà le due condizioni

$$\sum_{i=1}^n \tilde{\pi}_i x_{ij} = 0 \quad \forall j \in \{1, 2, \dots, p\}, \quad \sum_{i=1}^n \tilde{\pi}_i = 0. \quad (16.3)$$

Gli scarti complementari dicono dove si trova ciascun duale:

Posizione dell'osservazione	Variabile positiva	Duale
sopra la stima (y_i maggiore)	$\tilde{u}_i > 0$	$\tilde{\pi}_i = \tau$
sotto la stima (y_i minore)	$\tilde{v}_i > 0$	$\tilde{\pi}_i = -(1 - \tau)$
sulla stima (interpolata)	$\tilde{u}_i = \tilde{v}_i = 0$	$-(1 - \tau) < \tilde{\pi}_i < \tau$

Le osservazioni interpolate sono i **punti di appoggio**: sono le uniche il cui duale non è incollato a un estremo, e sono loro a determinare la stima. In generale sono $p+1$, tante quante le variabili libere del modello: spostare una qualunque delle altre osservazioni, purché resti dal suo lato della stima, non cambia nulla.

```
Osservazioni sopra la retta: 29, sotto: 29, interpolate: 2 <- p + 1 = 2
pi = +0,5 (punto sopra la retta): 29 punti
pi = -0,5 (punto sotto la retta): 29 punti
somma dei duali = 0,000000
somma dei duali per la temperatura = 0,000000
valore duale 616,9960 = valore primale 616,9960 (dualita' forte)
```

Il duale dell'intercetta è la proprietà del quantile

La condizione $\sum_i \tilde{\pi}_i = 0$ è il duale dell'intercetta, ed è la proprietà del quantile scritta in forma algebrica: sostituendo i valori della tabella, $\tau n_+ - (1 - \tau) n_- + \sum_{i \text{ interpolate}} \tilde{\pi}_i = 0$, dove n_+ e n_- contano le osservazioni sopra e sotto la stima. A meno dei $p+1$ punti di appoggio la stima lascia quindi sotto di sé una frazione τ delle osservazioni: con $\tau = 1/2$, metà sopra e metà sotto (29 e 29 nell'esempio). La condizione $\sum_i \tilde{\pi}_i x_{ij} = 0$ dice la stessa cosa caratteristica per caratteristica: il segno degli scostamenti non deve avere alcun legame residuo con i dati.

Insight manageriale

I punti di appoggio sono, per la regressione, quello che i support vector del capitolo 15 sono per la classificazione: pochissime osservazioni sostengono la soluzione e le altre potrebbero sparire senza cambiarla. La differenza è che qui il problema è un LP e non un QP, ma la lettura è identica — ed è la stessa complementarità del capitolo 2.

Lo scarto stesso è uno strumento diagnostico: stimando il modello con tutte e tre le caratteristiche sui 60 giorni, i giorni normali hanno scarto al più 23,2 pezzi e i cinque giorni di sciopero almeno 67,5. Una soglia a 40 pezzi li isola esattamente, senza sapere nulla degli scioperi.

```
Modello a tre caratteristiche sui 60 giorni:
domanda = 276,46 - 3,97*temperatura - 23,12*prezzo + 28,51*weekend
Scarto assoluto: 23,2 pezzi al massimo nei giorni normali, 67,5 al minimo negli
anomali
Giorni con scarto oltre 40 pezzi: [12, 23, 30, 42, 47] -> sono i giorni di
sciopero
```

16.5 Regressione quantile e scorta di sicurezza

Sui 55 giorni normali stimiamo il modello a tre caratteristiche per diversi valori di τ . L'ultima colonna verifica la proprietà del quantile: la frazione di osservazioni che cade sotto la stima.

τ	intercetta	temperatura	prezzo	weekend	quota sotto la stima
0,10	269,29	-4,04	-24,71	+37,05	0,073
0,50	272,01	-3,85	-22,24	+28,74	0,455
9/13 = 0,6923	281,19	-3,96	-23,21	+31,47	0,673
0,90	287,49	-3,57	-25,69	+27,98	0,873

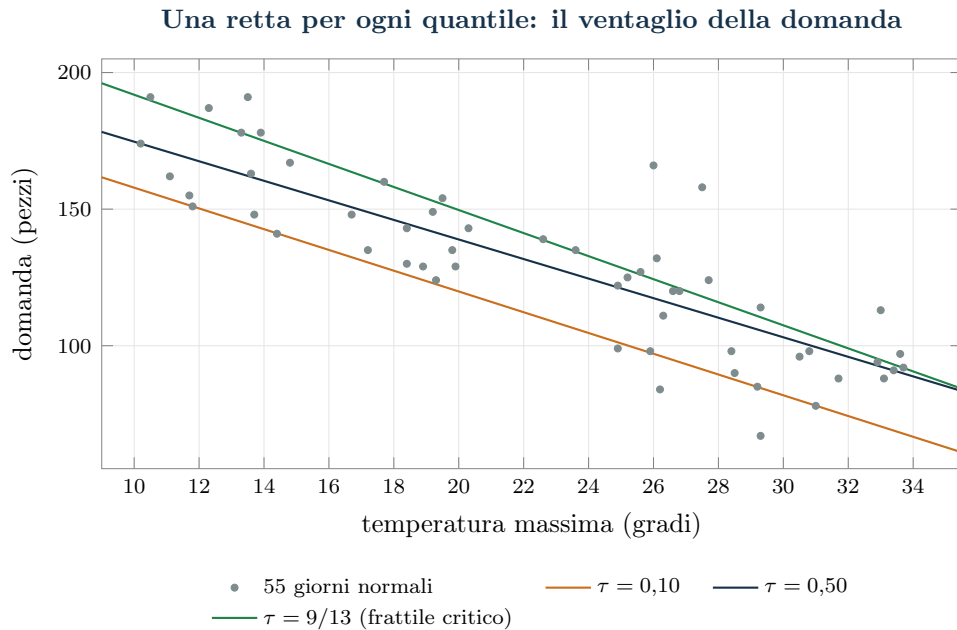


Figura 16.2: Le rette stimate con la sola temperatura per tre valori di τ : ciascuna lascia sotto di sé la frazione corrispondente di osservazioni. La retta verde, con τ pari al frattile critico della panetteria, è la quantità da produrre giorno per giorno.

Per un giorno caldo di weekend a prezzo pieno (28 gradi, prezzo 3,00) il modello dà due numeri diversi a seconda di τ :

domanda mediana prevista	126,3 pezzi	
quantile 0,6923 previsto	132,1 pezzi	<- quantita' da produrre
scorta di sicurezza	5,8 pezzi	
quantile 0,6923 senza caratteristiche	148,0 pezzi	

Insight manageriale

La regola del quantile del capitolo 12 resta valida, ma il numero a cui si applica cambia: il quantile calcolato su tutto lo storico dice di produrre 148 pezzi, quello *condizionato* alle condizioni del giorno ne chiede 132. Sedici pezzi al giorno di invenduto in meno, senza toccare i costi né la regola decisionale: solo usando l'informazione che si ha già. La differenza fra la stima mediana e quella al frattile critico, 5,8 pezzi, è la scorta di sicurezza — e il modello la calcola direttamente, senza passare da ipotesi sulla distribuzione.

16.6 Selezione delle caratteristiche con un budget

Aggiungendo caratteristiche il modello migliora sempre sui dati storici, anche quando sono irrilevanti: con abbastanza libertà finisce per inseguire il rumore. Un modo per governare il fenomeno è mettere un *budget* sui coefficienti. Servono p nuove variabili non negative,

$$z_j = \text{budget di coefficiente assegnato alla caratteristica } j, \quad \forall j \in \{1, 2, \dots, p\},$$

e il dato $t \in \mathbb{Q}_{\geq 0}$, budget complessivo. Al modello (16.1) si aggiungono i vincoli

$$-z_j \leq w_j \leq z_j, \quad \forall j \in \{1, 2, \dots, p\}, \quad (16.4a)$$

$$\sum_{j=1}^p z_j \leq t, \quad (16.4b)$$

$$z_j \geq 0, \quad \forall j \in \{1, 2, \dots, p\}. \quad (16.4c)$$

I vincoli (16.4a) costringono z_j a essere almeno il valore assoluto di w_j (all'ottimo vale esattamente quello, perché z_j consuma budget); il vincolo (16.4b) distribuisce il budget fra le caratteristiche. Il modello resta un LP.

Il budget somma coefficienti di caratteristiche misurate in unità diverse: gradi, euro, indicatori 0/1. Senza standardizzare, il vincolo (16.4b) confronterebbe grandezze incommensurabili e la selezione dipenderebbe dalle unità di misura. Qui ogni caratteristica è portata a media nulla e deviazione standard unitaria, quindi ogni coefficiente si legge come *pezzi per deviazione standard* della caratteristica.

Alle tre caratteristiche vere se ne aggiungono quattro senza alcun legame con la domanda (follower sui social, pioggia in una regione vicina, indice di borsa, giorno del mese). Al crescere del budget si osserva quali entrano.

budget	temperatura	prezzo	weekend	follower	pioggia	borsa	giorno_del_mese
10	-10,00	0,00	0,00	0,00	0,00	0,00	0,00
20	-18,16	-1,51	0,00	0,00	-0,33	0,00	0,00
30	-22,43	-4,45	2,35	0,00	-0,77	0,00	0,00
45	-26,98	-8,22	9,09	0,00	-0,63	0,00	-0,09

budget 5:	scarto medio	22,37	pezzi, prezzo ombra	-19,952	pezzi per unita'
budget 10:	scarto medio	18,83	pezzi, prezzo ombra	-18,665	pezzi per unita'
budget 20:	scarto medio	13,84	pezzi, prezzo ombra	-10,658	pezzi per unita'
budget 30:	scarto medio	10,63	pezzi, prezzo ombra	-7,271	pezzi per unita'
budget 45:	scarto medio	7,19	pezzi, prezzo ombra	-4,197	pezzi per unita'

Il prezzo della complessità: errore contro budget

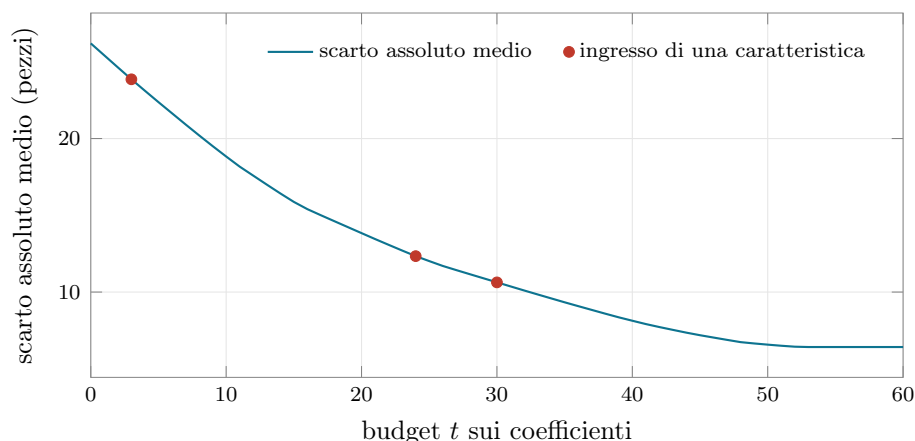


Figura 16.3: Lo scarto assoluto medio in funzione del budget sui coefficienti: una curva convessa e decrescente, come il valore ottimo di un LP al variare del termine noto. I punti rossi segnano il budget a cui entrano, nell'ordine, temperatura ($t = 3$), prezzo ($t = 24$) e weekend ($t = 30$).

Insight manageriale

Il prezzo ombra del vincolo (16.4b) è il **prezzo della complessità**: quanti pezzi di errore si risparmiano per ogni unità di budget in più. Vale $-19,95$ pezzi per unità quando il budget è scarso e $-4,20$ quando è abbondante: le prime unità comprano la temperatura, che spiega quasi tutto; le ultime comprano rumore. Le tre caratteristiche vere entrano nell'ordine della loro importanza — temperatura, prezzo, weekend — mentre nessuna delle quattro irrilevanti supera i 2 pezzi per deviazione standard finché il budget non diventa palesemente eccessivo. Il budget da scegliere non è quello che minimizza l'errore storico, ma quello oltre il quale il prezzo ombra diventa trascurabile.

16.7 Esercizi**Esercizio 16.1 — quantile a mano**

Con i sette giorni dell'esempio (88, 96, 104, 112, 120, 132, 148) e $\tau = 1/2$, calcolare a mano la mediana, il valore ottimo dell'obiettivo e i sette duali; verificare con lo script che la somma dei duali sia nulla.

Esercizio 16.2 — punti di appoggio

Stimare il modello con la sola temperatura e con $\tau = 1/2$, poi spostare di $+30$ pezzi un'osservazione che sta sopra la retta e rieseguire: la stima cambia? E se si sposta di -30 pezzi uno dei due punti di appoggio? Spiegare con gli scarti complementari.

Esercizio 16.3 — costo dell'errore asimmetrico

La panetteria valuta di svendere l'inventario a fine giornata, portando c_o da 4 a 2 €. Ricalcolare α^* , rieseguire la regressione quantile e dire di quanti pezzi cambia la produzione del giorno tipo.

Esercizio 16.4 — budget e convessità

Tracciare il valore ottimo dell'LP con budget per $t \in \{0, 1, \dots, 60\}$ e verificare che sia convesso e lineare a tratti; individuare i punti di rottura e mostrare che coincidono con i budget a cui il prezzo ombra cambia valore.

Esercizio 16.5 — validazione

Dividere i 55 giorni normali in stima (40) e verifica (15). Per $t \in \{5, 10, \dots, 55\}$ tracciare lo scarto assoluto medio sui due insiemi: quello di stima è monotono in t ? E quello di verifica? Quale budget si sceglie?

Parte V

Organizzazione del laboratorio

Organizzazione del laboratorio

17.1 Percorso essenziale in quattro laboratori

	Contenuto	Obiettivi di apprendimento
Lab 1	Produzione e scorte (cap. 4)	formulare un LP multiperiodale; leggere duali, slack e range; verificare un prezzo ombra per perturbazione
Lab 2	Markowitz (cap. 6)	costruire un QP convesso; tracciare una frontiera; discutere la fragilità delle stime
Lab 3	Pricing <i>oppure</i> budget (capp. 7–8)	modellare funzioni non lineari; studiare concavità; verificare le KKT numericamente
Lab 4	Progetto a scelta	supply chain, ricarica EV, localizzazione, code, Newsvendor, CVaR o SVM; presentazione manageriale dei risultati

I capitoli 12–15 possono sostituire i laboratori 3–4 in un’edizione orientata all’incertezza e al machine learning, oppure diventare approfondimenti per progetti individuali.

17.2 Struttura della consegna

Ogni gruppo consegna un notebook o un breve report (max 8 pagine) con questa struttura obbligatoria, che ricalca la scaletta dei capitoli:

1. **Problema e ipotesi** — il contesto in una pagina, le semplificazioni dichiarate esplicitamente;
2. **Modello** — insiemi, parametri, variabili, vincoli, obiettivo, con la spiegazione di ogni vincolo (non solo la formula);
3. **Dati** — origine, unità di misura, eventuale generazione;
4. **Risultati** — valore ottimo, decisioni principali, vincoli attivi;
5. **Sensitività** — protocollo del capitolo 2, sezione 2.3: almeno prezzi ombra verificati, una one-at-a-time e una frontiera di trade-off;
6. **Raccomandazione manageriale** — massimo dieci righe, senza formule: cosa fare, quanto vale, quanto è robusto.

17.3 Criteri di valutazione

Dimensione	Peso	Che cosa si valuta
Correttezza della formulazione	30%	vincoli giusti e giustificati, unità coerenti, convessità discussa
Implementazione e verifica	25%	codice riproducibile; stato del solver controllato; almeno una verifica indipendente (esempio a mano, perturbazione, caso limite)
Analisi di sensitività	25%	protocollo completo; duali interpretati con i range di validità; trade-off tracciati
Interpretazione e comunicazione	20%	raccomandazione chiara; figure leggibili; onestà sui limiti del modello

17.4 Domande tipiche di discussione

Le domande con cui difendere la consegna — e che conviene porsi *prima*:

- Quale risorsa conviene aumentare per prima, e quanto si può pagare per essa?
- Qual è il costo di una promessa di servizio più ambiziosa?
- La soluzione resta credibile se i dati cambiano del 5%? Quale parametro è il più fragile?
- Quale compromesso consigliereste a un decisore, e perché proprio quel punto della frontiera?
- Che cosa NON dice il modello? (ipotesi, orizzonte, variabili escluse)

17.5 Errori ricorrenti da evitare

1. Leggere `.X` o `.Pi` senza controllare `m.Status`.
2. Dimenticare `lb=-GRB.INFINITY` sulle variabili libere (b della SVM, η del CVaR).
3. Interpretare un prezzo ombra fuori dal suo intervallo di validità (`SARHS``Low/Up`).
4. Confondere il segno dei duali nei problemi di minimo (vincolo $\leq$: `Pi` $\leq$ 0).
5. Aggiornare il RHS di un vincolo che contiene costanti a sinistra (capitolo 10).
6. Ottimizzare un solo obiettivo quando il problema ne ha due (minimax puro senza costo).
7. Scegliere gli iperparametri (C , γ , λ) guardando il test set o gli scenari usati per decidere.
8. Riportare sei cifre decimali da stime che ballano alla seconda.

17.6 Materiali del corso

dispensa/	questa dispensa (<code>main.tex</code> ; le figure leggono i CSV in <code>figure/dat/</code>)
python/	uno script per capitolo; <code>esegui_tutti.py</code> rigenera dati e figure
dati/	tutti i dati dei casi di studio in CSV
GUIDA_GUROBI.md	guida autonoma al solver
soluzioni/	soluzioni degli esercizi (per i docenti)
slides/	slide dei quattro laboratori (italiano e inglese)

Riproducibilità: `python3 -m pip install gurobipy matplotlib pandas scipy` (scipy serve solo per le funzioni statistiche, non per l'ottimizzazione), poi `python3 python/esegui_tutti.py` e `latexmk -pdf dispensa/main.tex`.