Regressione robusta e quantile

Indice

Motivazione gestionale

Il modello

Caso di studio: lo storico della panetteria

La lettura duale: i punti di appoggio

Regressione quantile e scorta di sicurezza

Selezione delle caratteristiche con un budget

Esercizi

Obiettivi (1/2)

Alla fine di questo capitolo sarai in grado di:

- formulare la regressione con scarti assoluti come LP;
- leggerne il duale, riconoscendo i punti di appoggio della stima e la proprietà del quantile;

Obiettivi (2/2)

Alla fine di questo capitolo sarai in grado di:

- stimare la quantità da produrre condizionata alle condizioni del giorno;
- selezionare le variabili esplicative rilevanti con un budget sui coefficienti, leggendone il prezzo ombra.

Prerequisiti

- Da questo modulo, Capitolo 7 «Il Newsvendor e le sue varianti»: la regola del quantile e il frattile critico
- Da questo modulo, Capitolo 2 «Richiami: programmazione lineare e analisi di sensitività»: il duale di un LP e i prezzi ombra

Motivazione gestionale

Introduzione: Motivazione gestionale

- Tutti i modelli visti finora partono da parametri noti: i costi della produzione, i prezzi orari della ricarica, gli scenari del newsvendor
- Quei numeri, però, qualcuno li deve stimare
- La stima è essa stessa un problema di ottimizzazione, che si risolve con lo stesso solver del resto del corso

Prevedere una quantità

- Qui usiamo l'ottimizzazione per *prevedere una quantità*
- 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
  - pesando in modo asimmetrico gli scarti in eccesso e in difetto si punta a un *quantile* della domanda invece che al suo centro
- È esattamente il numero che serve al newsvendor

Riepilogo

- Anche i parametri dei modelli vanno stimati, e la stima è un problema di ottimizzazione
- Con gli scarti assoluti la regressione è un LP
- Pesi asimmetrici sugli scarti danno un quantile: il numero che serve al newsvendor

Il modello

Introduzione: Il modello

- Formuliamo la regressione quantile come un LP
- Vediamo come linearizzare il valore assoluto con due variabili non negative
- Un esempio a mano mostra che, senza caratteristiche, il modello restituisce il quantile empirico

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) (1/2)

Simbolo	Tipo	Significato
n	$\in \mathbb{Z}_{\geq 1}$	numero di osservazioni storiche, indicizzate da $i \in \{1, 2, \ldots, n\}$
p	$\in \mathbb{Z}_{\geq 0}$	numero di caratteristiche osservate, indicizzate da $j \in \{1, 2, \ldots, p\}$
x_{ij}	$\in \mathbb{Q}$	valore della caratteristica j nell'osservazione i

Dati (input del modello) (2/2)

Simbolo	Tipo	Significato
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 (1/2)

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

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

Variabili decisionali (2/2)

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

$$\begin{cases} u_i = \text{scostamento in eccesso dell'osservazione } i \\ v_i = \text{scostamento in difetto dell'osservazione } i \end{cases}$$

$$\forall i \in \{1, 2, \ldots, n\}$$

Modello LP di regressione quantile

Modello LP di regressione quantile

$$\min \quad \sum_{i=1}^{n} \left[\tau\, u_i + (1-\tau)\, v_i \right] \tag{10.1a}$$

$$\text{soggetto a} \quad \sum_{j=1}^{p} w_j\, x_{ij} + b + u_i - v_i = y_i, \quad \forall i \in \{1, 2, \ldots, n\}, \tag{10.1b}$$

$$w_j \gtreqless 0, \quad \forall j \in \{1, 2, \ldots, p\}, \tag{10.1c}$$

$$b \gtreqless 0, \tag{10.1d}$$

$$u_i,\ v_i \geq 0, \quad \forall i \in \{1, 2, \ldots, n\}. \tag{10.1e}$$

Descrizione del modello LP

- La funzione obiettivo lineare 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 di scostamento definiscono lo scostamento di ogni osservazione come differenza fra le sue due parti, $u_i - v_i = y_i -$ previsione (n vincoli lineari)
- Gli ultimi vincoli definiscono le variabili: coefficienti e intercetta liberi, scostamenti non negativi

Perché due variabili per un solo scostamento (1/2)

- Lo scostamento $y_i - \left(\sum_j w_j x_{ij} + b\right)$ 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

Perché due variabili per un solo scostamento (2/2)

- All'ottimo almeno una delle due è nulla
- Se fossero entrambe positive, ridurle della stessa quantità lascerebbe soddisfatto il vincolo di scostamento 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

Esempio: Senza caratteristiche, torna il quantile empirico (1/4)

- Con $p = 0$ resta la sola intercetta e il modello sceglie un unico numero b
- Sette giorni di vendite osservate:

$$88,\ 96,\ 104,\ 112,\ 120,\ 132,\ 148 \text{ pezzi}$$

- La panetteria del newsvendor ha $c_u = 9$ e $c_o = 4$, quindi

$$\tau = \alpha^* = 9/13 = 0{,}6923$$

Esempio: Senza caratteristiche, torna il quantile empirico (2/4)

Passo 1 — che cosa costa muovere b

- Spostando b di una unità verso l'alto si risparmiano τ su ogni osservazione che sta *sopra* b
- E si pagano $1 - \tau$ in più su ogni osservazione che sta *sotto*
- Conviene salire finché le osservazioni sopra sono tante da compensare

Esempio: Senza caratteristiche, torna il quantile empirico (3/4)

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

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

Esempio: Senza caratteristiche, torna il quantile empirico (4/4)

Passo 3 — i duali

$$\tilde{\boldsymbol{\pi}} = \left(-\frac{4}{13}, -\frac{4}{13}, -\frac{4}{13}, -\frac{4}{13}, -\frac{2}{13}, \frac{9}{13}, \frac{9}{13}\right)$$

- $\tau = 9/13$ sulle osservazioni sopra la stima, $-(1-\tau) = -4/13$ su quelle sotto, un valore interno sull'osservazione interpolata
- La somma è nulla e il valore duale $\sum_i \tilde{\pi}_i\, y_i = 680/13$ coincide con quello primale

Senza caratteristiche il modello restituisce il **quantile empirico**: la regola del quantile del newsvendor, scritta come LP. Con le caratteristiche diventa un quantile *condizionato* alle condizioni del giorno.

Riepilogo

- La regressione quantile è un LP: scostamento $= u_i - v_i$, con $u_i, v_i \geq 0$
- L'obiettivo paga τ gli eccessi e $1 - \tau$ i difetti
- All'ottimo almeno una tra u_i e v_i è nulla
- Senza caratteristiche si ottiene il quantile empirico di livello τ

Caso di studio: lo storico della panetteria

Introduzione: Lo storico della panetteria

- Applichiamo il modello a sessanta giorni di vendite della panetteria
- Stimiamo la retta due volte: su tutti i giorni e sui soli giorni normali
- Vedremo che l'LP resiste ai giorni anomali

I dati

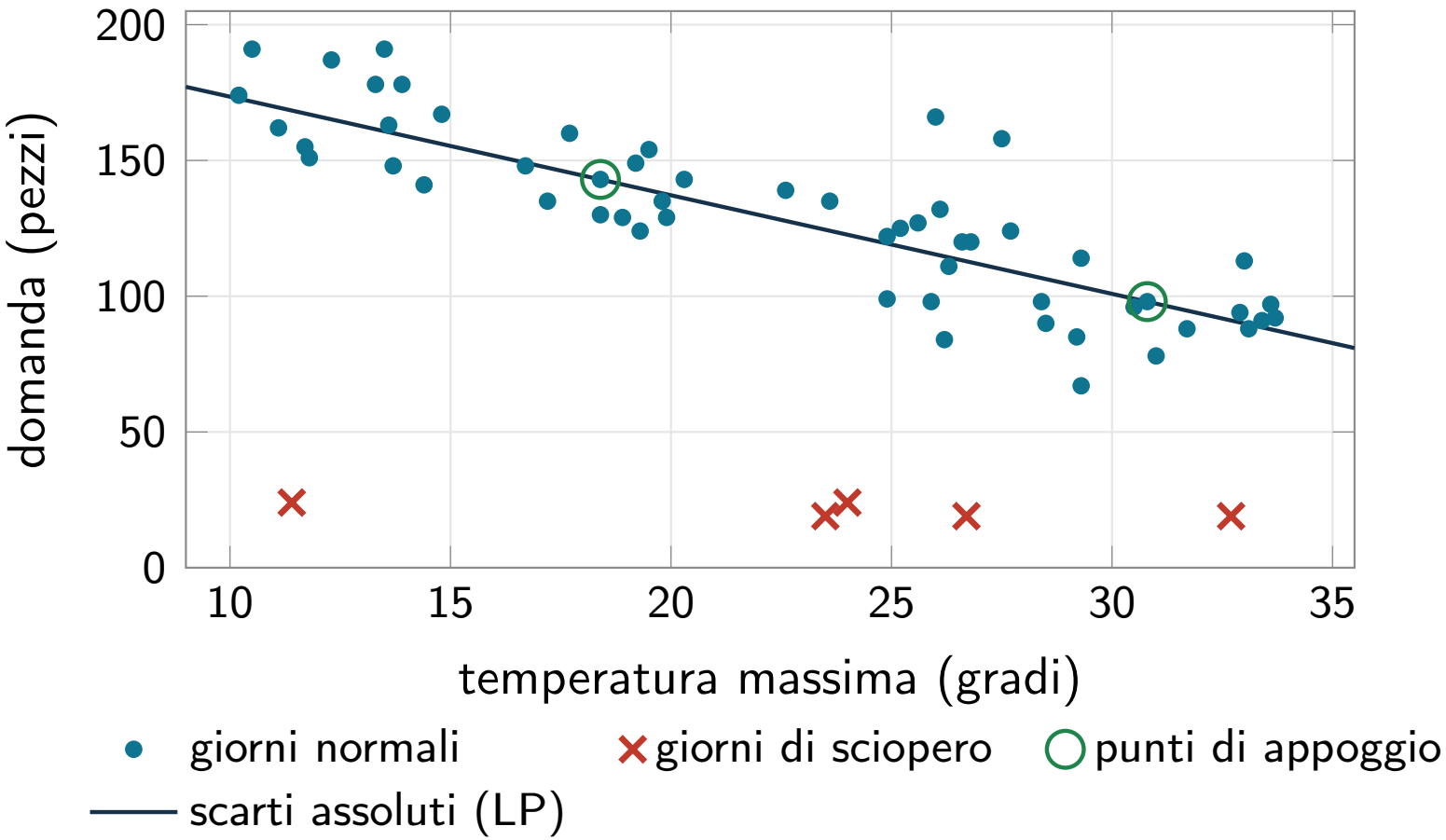
- Sessanta giorni di vendite della panetteria: per ogni giorno la domanda in pezzi, la temperatura massima, il prezzo praticato e l'indicazione di weekend
- Cinque di quei giorni cadono durante due scioperi dei trasporti: il centro è deserto e le vendite crollano a una ventina di pezzi
- Stimiamo prima la relazione con la sola temperatura ($p = 1$), che si vede in un grafico

Scarti assoluti con e senza i giorni di sciopero

modello ($\tau = 1/2$)	retta stimata
LP sui 60 giorni	domanda $= 209{,}77 - 3{,}63 \cdot$ temperatura
LP senza i 5 giorni di sciopero	domanda $= 210{,}49 - 3{,}58 \cdot$ temperatura

- Scarto assoluto medio sui 55 giorni normali: 13,08 pezzi
- Distanza fra le due intercette: appena 0,72 pezzi

Cinque giorni di sciopero non spostano la retta dell'LP



Perché l'LP resiste ai dati anomali

- L'LP paga ogni scarto *una volta sola*, in proporzione alla sua grandezza
- La retta resta dove sono i dati normali: il modello arriva da solo alla risposta che si avrebbe ripulendo a mano lo storico
- I due punti cerchiati in verde sono i punti di appoggio, gli unici che la retta interpola
- Minimizzare gli scarti *al quadrato* porta ai minimi quadrati, un modello quadratico: nel modulo *Modellazione PNL*, «Regressione ai minimi quadrati»

Riepilogo

- Scarti assoluti (LP): la retta resta dove sono i dati normali
- Con e senza i giorni di sciopero le intercette differiscono di 0,72 pezzi
- Un modello che non va sorvegliato: si può lasciare girare ogni settimana

La lettura duale: i punti di appoggio

Introduzione: La lettura duale

- Associamo un duale a ogni vincolo di scostamento
- Gli scarti complementari dicono quali osservazioni determinano la stima: i **punti di appoggio**
- Il duale dell'intercetta è la proprietà del quantile

La finestra dei duali

Associando un duale π_i a ciascun vincolo di scostamento:

- l'ottimalità sulle variabili non negative u_i e v_i limita ogni duale a una finestra che dipende solo da τ:

$$-(1-\tau) \;\leq\; \tilde{\pi}_i \;\leq\; \tau, \qquad \forall i \in \{1, 2, \ldots, n\}$$

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

$$\sum_{i=1}^{n} \tilde{\pi}_i \, x_{ij} = 0 \quad \forall j \in \{1, 2, \ldots, p\}, \qquad \sum_{i=1}^{n} \tilde{\pi}_i = 0$$

Scarti complementari

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)	$\tilde{u}_i = \tilde{v}_i = 0$	$-(1-\tau) < \tilde{\pi}_i < \tau$

I punti di appoggio

- 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

I duali sullo storico della panetteria

Retta con la sola temperatura, $\tau = 1/2$:

osservazioni sopra la retta ($\pi = +0{,}5$)	29
osservazioni sotto la retta ($\pi = -0{,}5$)	29
osservazioni interpolate ($= p + 1$)	2
somma dei duali	0,000000
somma dei duali per la temperatura	0,000000

Valore duale 616,9960 = valore primale 616,9960 (dualità forte)

Il duale dell'intercetta è la proprietà del quantile (1/2)

- La condizione $\sum_i \tilde{\pi}_i = 0$ è il duale dell'intercetta
- 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

Il duale dell'intercetta è la proprietà del quantile (2/2)

- 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

Pochi punti sostengono la stima

- Pochissime osservazioni, i punti di appoggio, sostengono la soluzione
- Le altre potrebbero sparire senza cambiarla, purché restino dal loro lato della stima
- È la stessa complementarietà vista nei richiami

Lo scarto come strumento diagnostico

- Stimiamo il modello con tutte e tre le caratteristiche sui 60 giorni:

  $$\text{domanda} = 276{,}46 - 3{,}97\,\text{temp.} - 23{,}12\,\text{prezzo} + 28{,}51\,\text{weekend}$$

- Scarto assoluto: al più 23,2 pezzi nei giorni normali, almeno 67,5 nei cinque giorni di sciopero
- Una soglia a 40 pezzi isola esattamente i giorni $12, 23, 30, 42, 47$: sono i giorni di sciopero
- Il modello li trova senza sapere nulla degli scioperi

Riepilogo

- Ogni duale vive nella finestra $[-(1-\tau), \tau]$
- Le osservazioni interpolate ($p+1$ in generale) sono i punti di appoggio e determinano la stima
- $\sum_i \tilde{\pi}_i = 0$ è la proprietà del quantile
- Lo scarto del modello robusto individua i giorni anomali

Regressione quantile e scorta di sicurezza

Introduzione: Regressione quantile

- Cambiando τ si stima un quantile diverso della domanda
- Con τ pari al frattile critico del newsvendor si ottiene direttamente la quantità da produrre
- La differenza con la mediana è la scorta di sicurezza

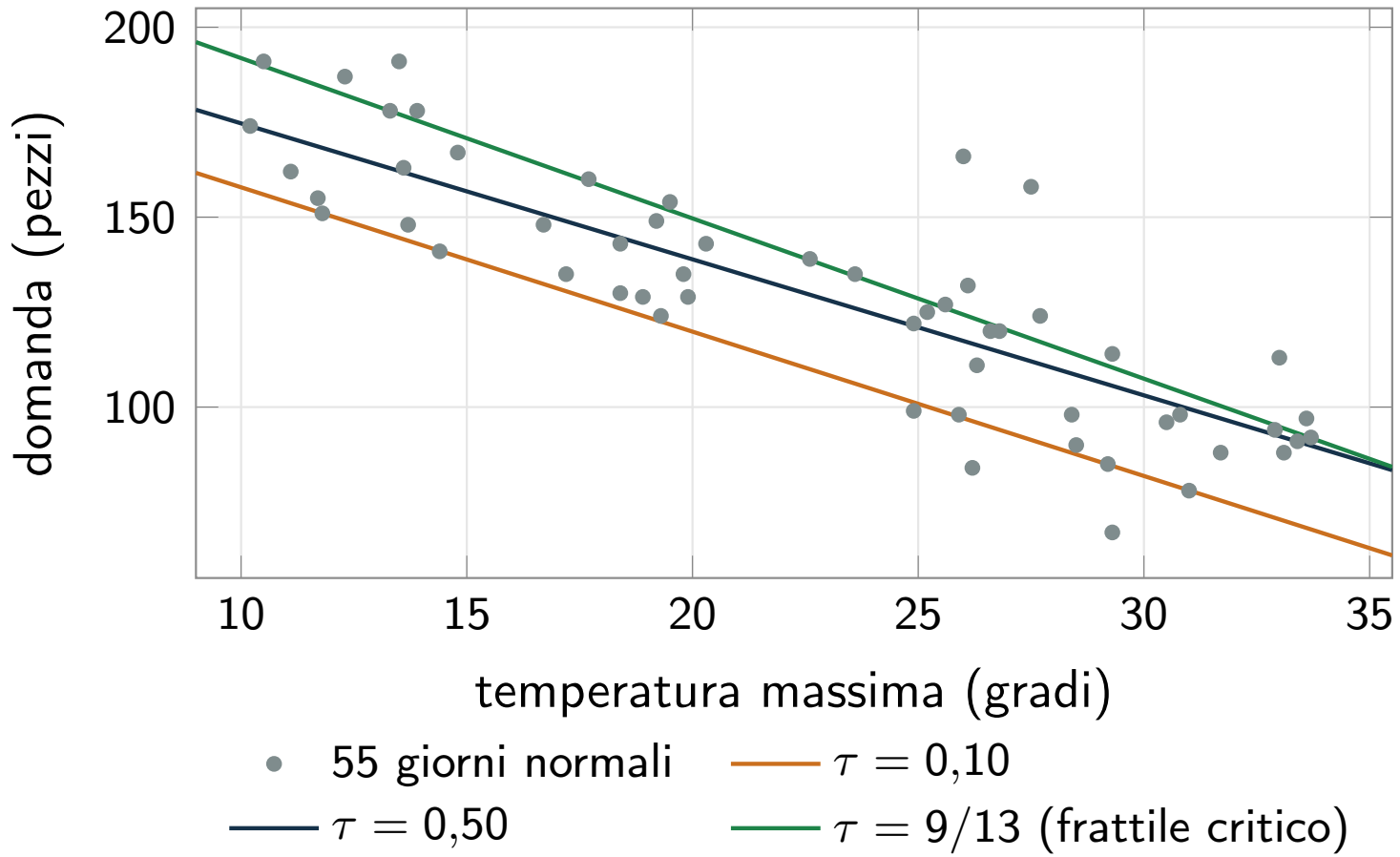
Il modello a tre caratteristiche per diversi τ

Sui 55 giorni normali; l'ultima colonna verifica la proprietà del quantile.

τ	intercetta	temp.	prezzo	weekend	quota sotto
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	281,19	−3,96	−23,21	+31,47	0,673
0,90	287,49	−3,57	−25,69	+27,98	0,873

- $9/13 = 0{,}6923$ è il frattile critico della panetteria
- "quota sotto": frazione di osservazioni che cade sotto la stima

Una retta per ogni quantile: il ventaglio della domanda



Come leggere il grafico

- 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

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 (quantità da produrre)	132,1 pezzi
scorta di sicurezza	5,8 pezzi
quantile 0,6923 senza caratteristiche	148,0 pezzi

Il quantile condizionato (1/2)

- La regola del quantile del newsvendor 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

Il quantile condizionato (2/2)

- La differenza fra la stima mediana e quella al frattile critico, 5,8 pezzi, è la **scorta di sicurezza**
- Il modello la calcola direttamente, senza passare da ipotesi sulla distribuzione

Riepilogo

- Ogni τ dà una retta che lascia sotto di sé circa una frazione τ delle osservazioni
- Con τ = frattile critico si ottiene la quantità da produrre, condizionata alle condizioni del giorno
- Scorta di sicurezza = quantile al frattile critico − mediana, senza ipotesi sulla distribuzione

Selezione delle caratteristiche con un budget

Introduzione: Selezione delle caratteristiche

- 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

Nuove variabili e nuovo dato

Servono p nuove variabili non negative:

$$z_j = \text{budget di coefficiente assegnato alla caratteristica } j$$

$$\forall j \in \{1, 2, \ldots, p\}$$

e il dato $t \in \mathbb{Q}_{\geq 0}$, budget complessivo.

I vincoli di budget

Al modello LP di regressione quantile si aggiungono i vincoli

$$-z_j \leq w_j \leq z_j, \qquad \forall j \in \{1, 2, \ldots, p\}$$

$$\sum_{j=1}^{p} z_j \leq t$$

$$z_j \geq 0, \qquad \forall j \in \{1, 2, \ldots, p\}$$

- I primi costringono z_j a essere almeno $|w_j|$ (all'ottimo vale esattamente quello, perché z_j consuma budget)
- Il vincolo $\sum_j z_j \leq t$ distribuisce il budget fra le caratteristiche
- Il modello resta un LP

Standardizzare le caratteristiche

Osservazione 1: Unità di misura diverse

Il budget somma coefficienti di caratteristiche misurate in unità diverse: gradi, euro, indicatori $0/1$*. Senza standardizzare, il vincolo* $\sum_j z_j \leq t$ *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.

Quattro caratteristiche irrilevanti

- 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

Coefficienti al crescere del budget (1/2)

Caratteristiche vere:

budget	temperatura	prezzo	weekend
10	−10,00	0,00	0,00
20	−18,16	−1,51	0,00
30	−22,43	−4,45	2,35
45	−26,98	−8,22	9,09

Coefficienti al crescere del budget (2/2)

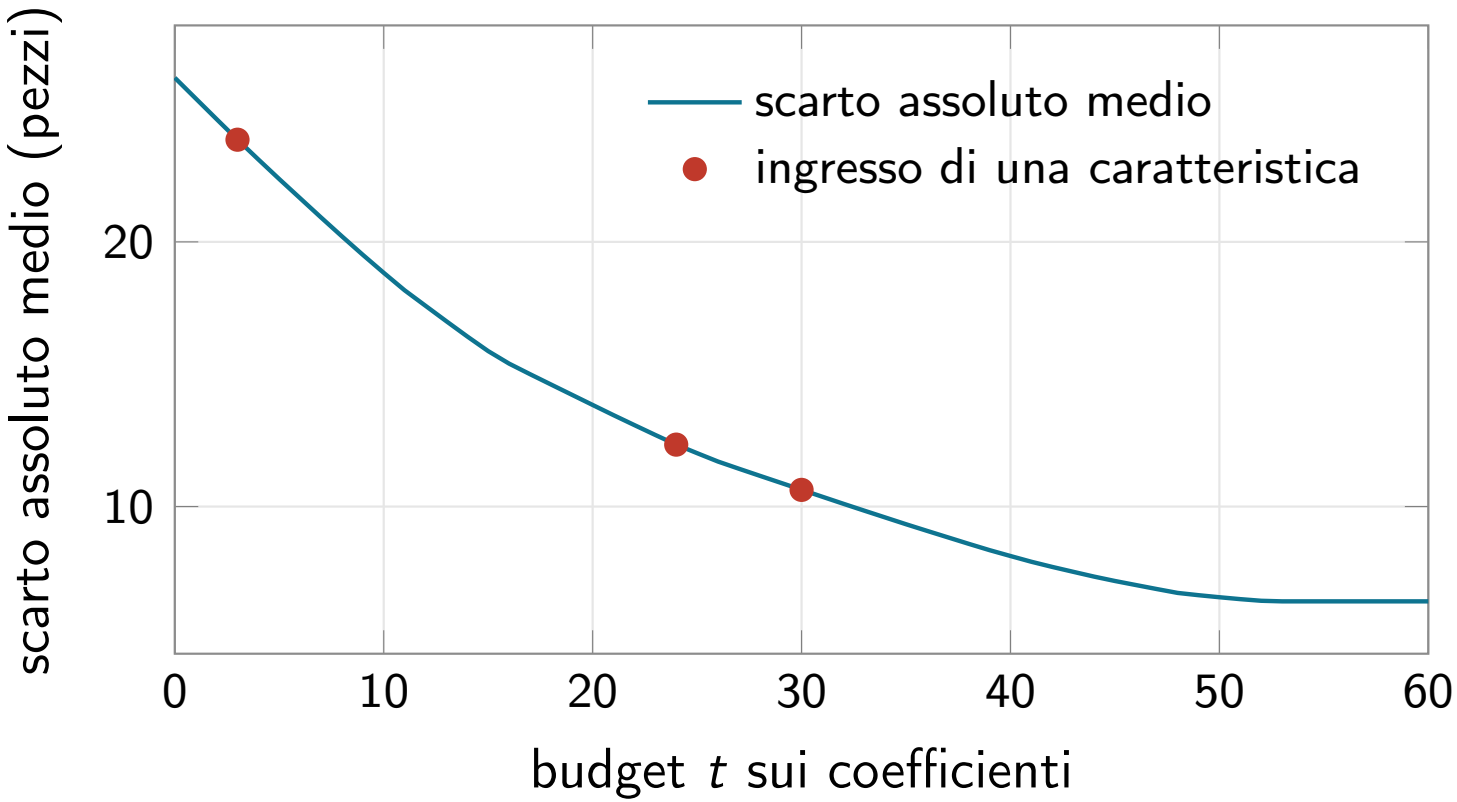
Caratteristiche irrilevanti:

budget	follower	pioggia	borsa	giorno del mese
10	0,00	0,00	0,00	0,00
20	0,00	−0,33	0,00	0,00
30	0,00	−0,77	0,00	0,00
45	0,00	−0,63	0,00	−0,09

Scarto medio e prezzo ombra del budget

budget	scarto medio (pezzi)	prezzo ombra (pezzi per unità)
5	22,37	−19,952
10	18,83	−18,665
20	13,84	−10,658
30	10,63	−7,271
45	7,19	−4,197

Il prezzo della complessità: errore contro budget



Come leggere il grafico

- 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 ($t = 30$)

Il prezzo della complessità (1/2)

- Il prezzo ombra del vincolo $\sum_j z_j \leq t$ 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

Il prezzo della complessità (2/2)

- Le tre caratteristiche vere entrano nell'ordine della loro importanza: temperatura, prezzo, weekend
- 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

Riepilogo

- Un budget t sulla somma dei $|w_j|$ seleziona le caratteristiche, restando in un LP
- Le caratteristiche vanno standardizzate prima
- Il prezzo ombra del budget è il prezzo della complessità
- Si sceglie il budget oltre il quale il prezzo ombra diventa trascurabile

Esercizi

Introduzione: Esercizi

- Gli esercizi mettono alla prova i concetti del capitolo
- Si risolvono a mano o con lo script del laboratorio e Gurobi
- Toccano il quantile a mano, i punti di appoggio, i costi asimmetrici, la convessità del budget e la validazione

Esercizio: 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: punti di appoggio

- Stimare il modello con la sola temperatura e con $\tau = 1/2$
- 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: costo dell'errore asimmetrico

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

Esercizio: budget e convessità

- Tracciare il valore ottimo dell'LP con budget per $t \in \{0, 1, \ldots, 60\}$
- 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: validazione

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

Riepilogo finale (1/2)

Il modello:

- Regressione quantile come LP: $u_i, v_i \geq 0$ pagati τ e $1 - \tau$
- Senza caratteristiche: il quantile empirico

Robustezza:

- Gli scarti assoluti ignorano i giorni anomali
- Lo scarto del modello robusto individua le anomalie

Riepilogo finale (2/2)

Lettura duale:

- Duali nella finestra $[-(1-\tau), \tau]$; $p+1$ punti di appoggio
- $\sum_i \tilde{\pi}_i = 0$ è la proprietà del quantile

Decisioni:

- τ = frattile critico: quantità da produrre e scorta di sicurezza
- Budget sui coefficienti: il prezzo ombra è il prezzo della complessità