

ASML: terzo laboratorio

Regressione polinomiale e locale

Leonardo Egidì

Ottobre 2022

Università di Trieste

Laboratorio 3

L'obiettivo è costruire modelli di regressione polinomiale, shrinkage e locale per esaminare la correlazione tra il livello di antigene prostatico specifico e una serie di misure cliniche negli uomini che stavano per ricevere una prostatectomia radicale. Ecco una descrizione delle variabili che troverete nel file:

VARIABILE	DESCRIZIONE
lpsa	livello di antigene prostatico specifico
lcavol	$\log(\text{volume cancerogeno})$
lweight	$\log(\text{peso prostatico})$
age	età
lbph	$\log(\text{quantità di iperplasia prostatica benigna})$
svi	invasione delle vescicole seminali
lcp	$\log(\text{penetrazione capsulare})$
gleason	punteggio di Gleason
pgg45	percentuale Gleason con punteggio 4 o 5

Laboratorio 3 (cont.)

- Caricare i dati `prostate.csv`. Valutare la correlazione tra le variabili per saggiare possibili effetti non lineari.
- Stimare una regressione polinomiale con `lpsa` come variabile risposta. Si ottiene qualche miglioramento rispetto a un modello lineare? Suggerimento: usare la funzione `anova`.
- Produrre una stima del modello polinomiale con solo $y = \text{lpsa}$ in funzione di $x = \text{lcavol}$, e disegnare la curva stimata.
- Usare uno dei modelli selezionati e produrre un grafico *forest plot* per i coefficienti, usando la funzione `coefplot` del pacchetto `arm`.
- Stimare una regressione a gradini usando la funzione `cut` e interpretare. Potete specificare i cutpoints tramite l'opzione `breaks`.
- Stimare una regressione di tipo lasso, usare la funzione `coefplot` per visualizzare i coefficienti stimati e interpretare (non serve calcolare gli s.e.). Confrontare i risultati con le altre stime.
- Stimare una regressione locale con solo $y = \text{lpsa}$ in funzione di $x = \text{lcavol}$ usando la funzione `loess`. Suggerimento: usare diversi valori per l'argomento `span`.