

ASML: quinto laboratorio

Alberi di regressione e MARS

Leonardo Egidì

Novembre 2022

Università di Trieste

Laboratorio 5

L'obiettivo è familiarizzare con gli alberi di regressione, capendone il partizionamento ricorsivo binario, la scelta dei nodi e le tecniche di potatura. Alla fine, si chiede di stimare un modello di tipo MARS.

- Importare i dati sulle assicurazioni nel file `TL.csv` (già usati per Laboratorio 2) e creare una partizione in training e test set a piacere (usare sempre: $LFACE = \log(FACE)$ e $LINCOME = \log(INCOME)$).
- Stimare un primo albero di regressione usando la funzione `tree` del medesimo pacchetto. Quali sono le variabili influenti? Quanti sono i nodi terminali? Produrre un grafico dell'albero. Interpretare i risultati e le stime. Confrontare con i risultati `lm`, `ridge` e `lasso`.
- Ora potare l'albero e calcolare la larghezza ottimale del sotto-albero tramite CV. Produrre un plot per l'albero potato.
- Ora prevedere e calcolare MSE. Produrre un grafico simile a quello di Figura 8.5 del libro ISL (anche contenuta nella slide 21 della lezione sugli alberi). Commentare i risultati.
- Stimare un modello MARS per `LFACE`, interpretare i risultati e compararli con quelli precedenti (valutare l'inclusione di interazioni mediante argomento `degree`). Quanti sono i termini totali del modello? Produrre anche un plot sull'importanza delle variabili (analogo a quello di slide 32 della lezione)