

ASML: secondo laboratorio

Selezione del modello lineare e metodi di regolarizzazione

Leonardo Egidì

Ottobre 2022

Università di Trieste

Laboratorio 2

L'obiettivo è determinare le caratteristiche famigliari che influenzano l'importo di un'assicurazione. La quantità d'interesse è misurata da FACE, l'importo che l'azienda pagherà in caso di decesso dell'assicurato.

1. Accedere al file `TL.csv` e ispezionare le variabili di interesse.
2. Trasformare le variabili FACE e INCOME in scala logaritmica. Spiegare perché conviene in questo caso.
3. Costruire un modello per $\log(y)$, dove $y = \text{FACE}$ usando metodi di tipo *best subset*, e/o stepwise di tipo backward e forward. Quali sono le variabili che spiegano l'importo assicurato? Suggerimento: usare la funzione `regsubsets` del pacchetto `leaps`.
4. Scegliere il migliore tra i modelli usando errori su set di validazione e di cross-validazione (scegliete voi il numero di folds).
5. Applicare una regressione di tipo ridge e lasso, stimando λ con approccio cross-validation. Quali le conclusioni con i due metodi? Suggerimento: usare la funzione `glmnet` del medesimo pacchetto.