

ASML: sesto laboratorio

Metodi di insieme

Leonardo Egidì

Novembre 2024

Università di Trieste

L'obiettivo è familiarizzare con i metodi di insieme.

In classe:

- Considerare prima i dati sulle assicurazioni `TL.csv` e poi quelli sui tumori prostatici `prostate.csv`. Previa suddivisione delle osservazioni in training e test set, condurre una ricca e documentata analisi a fini predittivi usando metodi bagging, boosting e random forests con i pacchetti `gbm` e `randomForest` - in alternativa/aggiunta, usare `caret`. Qual è il metodo migliore per prevedere le rispettive variabili risposta, e quali sono i parametri di tuning adottati? Produrre anche grafici di importanza delle variabili.

Per la consegna:

- Applicare una ricca analisi simile a quella fatta in classe e commentare i risultati usando i dati `Hitters` del pacchetto `ISLR2`, usando come al solito la variabile `log(Salary)` come variabile risposta. Qual è il metodo più potente per prevedere il logaritmo del salario? Quali sono le variabili più importanti? Condurre un'analisi comparativa per più metodi di apprendimento (bagging, RF, boosting, ed eventualmente anche quelli già utilizzati in altri laboratori). Per quanto riguarda il boosting, far variare i tre iperparametri visti a lezione e commentare. Aiutarsi anche con comparazioni grafiche.