

Introduzione al Machine Learnig

Prof. Fabio Divino
Tutorial 1: Linear regression

Considerare il dataset **boston_data.RData** contenente informazioni raccolte dal Census Service degli Stati Uniti in merito agli alloggi nella zona di Boston. In particolare, relativamente a 506 records sono riportate le seguenti 14 variabili:

1. **crim**: per capita crime rate by town;
2. **zn**: proportion of residential land zoned for lots over 25,000 sq.ft;
3. **indus**: proportion of non-retail business acres per town;
4. **chas**: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise);
5. **nox**: nitrogen oxides concentration (parts per 10 million);
6. **rm**: average number of rooms per dwelling;
7. **age**: proportion of owner-occupied units built prior to 1940;
8. **dis**: weighted mean of distances to five Boston employment centres;
9. **rad**: index of accessibility to radial highways;
10. **tax**: full-value property-tax rate per \$10,000;
11. **ptratio**: pupil-teacher ratio by town;
12. **black**: $1000(B_k - 0.63)^2$, where B_k is the proportion of blacks by town;
13. **lstat**: lower status of the population (percent);
14. **medv**: median value of owner-occupied homes in \$1000s.

Quesiti

A) Considerando al massimo 5 fra le prime 13 variabili, determinare il modello di regressione lineare ottimo in termini predittivi utilizzando i seguenti datasets di training e test, rispettivamente:

1. **boston_training** (400 record);
2. **boston_test** (106 records).

Suggerimento: utilizzare una procedura *backward*, con eliminazione delle variabili meno significative utilizzando il criterio del *p-value*.

B) Fare un'analisi semantica dei risultati.

Risultati

Dopo aver svolto il tutorial, definire il modello ottimo nell'array **model** e salvare il data environment come file

nome_cognome.Rdata (rigorosamente tutto in minuscolo!)

Inviare il file a

fabio.divino@unimol.it

indicando nel testo del messaggio le 5 variabili individuate nel modello ottimo e la relativa analisi semantica dei risultati.