

Homework 2, due September 9th, 11:59pm

September 2, 2020

1. In this problem we use the `abalone` dataset available on Canvas. The dataset is about predicting the age of the abalone from its physical measurements. Use the first 7 variables as predictors and the 8-th as the response.

Report all results as the average of 10 random splits with 80% of data for training and 20% for testing.

- a) OLS regression, analytic, by solving the normal equations, with $\lambda = 0.0001$. Report the average training and test R^2 . (2 points)
- b) Regression tree of maximum depth 1, 2, ..., up to 7, for a total of 7 regression trees. Plot the average training and test R^2 vs the tree depth. (2 points)
- c) Random forest regression with 10, 30 and 100 trees. Report the average training and test R^2 in each case. (3 points)