

Homework 1

Due on 02/22/2022

In this exercise, we predict the sale price of a house using its other characteristics. The training data are in the file “housing_train.csv”, and the test data are in “housing_test.csv”. The response is in the column “Sale_price”. Among the 25 feature variables, some are numeric features, such as living area square feet or first floor square feet, and some are categorical features, such as the overall material and finish of the house or kitchen quality. A detailed description of the variables is in “dictionary.txt”.

- (a) Fit a linear model using least squares on the training data. Is there any potential disadvantage of this model?
- (b) Fit a lasso model on the training data and report the test error. When the 1SE rule is applied, how many predictors are included in the model?
- (c) Fit an elastic net model on the training data. Report the selected tuning parameters and the test error.
- (d) Fit a partial least squares model on the training data and report the test error. How many components are included in your model?
- (e) Which model will you choose for predicting the response? Why?