Homework 1

Due on 02/25/2026

In this exercise, we predict the sale price of a house based on various characteristics. The training data are in "housing_train.csv", and the test data are in "housing_test.csv". The response is in the column "Sale_price", and other variables can be used as predictors. The variable definitions can be found in "dictionary.txt".

(a) Fit a lasso model on the training data. Report the selected tuning parameter and the test error. When the 1SE rule is applied, how many predictors are included in the model?

(b) Fit an elastic net model on the training data. Report the selected tuning parameters and the test error. Is it possible to apply the 1SE rule to select the tuning parameters for elastic net? If the 1SE rule is applicable, implement it to select the tuning parameters. If not, explain why.

(c) Fit a partial least squares model on the training data and report the test error. How many components are included in your model?

(d) Choose the best model for predicting the response and explain your choice.

(e) If R package "caret" was used for the lasso in (a), retrain this model using R package "glmnet", and vice versa. Compare the selected tuning parameters between the two software approaches. Should there be discrepancies in the chosen parameters, discuss potential reasons for these differences.