CS6464: Concepts in Statistical Learning Theory SOFTWARE ASSIGNMENT 2

PROBLEM STATEMENT

The assignment aims at predicting house prices given training and test house data of 20-dimensional features and comparing the performance of various regression methods.

TASKS:

Two regression models (one row for each group) as specified in Table 1 have to be trained using the training data (available in the file named "**kc_house_train_data.csv**") and the house prices should be predicted for the test data (available in the file named "**kc_house_test_data.csv**"). Perform 10-fold cross-validation. Compare the prediction quality between the three methods allotted.

INPUT DATA

- 20-dimensional housing data for training, 17385 samples
- 20-dimensional housing data for testing, 4230 samples

OUTPUT

- Compute the regression weights and interpret them based on the methods allotted.
- Plot the coefficient profiles of top 5 interesting features based on the largest change of the coefficients over iterations (as in Fig. 3.10 (a) in Hastie's book). Plot the coefficient profiles of each method separately. (Note: By iterations, we mean the iterations of the optimization function adopted (as in LASSO, ElasticNet, etc), or the steps (as in Stepwise regression).
- Evaluation of the models with Residual Sum of Squares (RSS) (or MSE) metric using the computed regression weights, predictors and outcome.

HINTS FOR EXCELLENCE

Additional observations and visualizations of the data and the attributes of the trained models will be given extra credit.

Grp no	Method 1	Method 2	Method 3
1	Simple linear regression	Lasso regression	Kernel Regression
2	Lasso regression	Forward Stepwise Regression	Polynomial Regression
3	Ridge Regression	Backward Stepwise Regression	Kernel Regression
4	Ridge Regression	ElasticNet Regression	Backward Stepwise Regression
5	Simple linear regression	Forward Stepwise Regression	Kernel Regression
6	Ridge Regression	Backward Stepwise Regression	Polynomial Regression
7	Simple linear regression	Lasso Regression	Polynomial Regression
8	Ridge regression	ElasticNet Regression	Kernel Regression
9	Simple linear regression	Ridge Regression	Polynomial Regression
10	Lasso regression	Forward Stepwise Regression	Backward Stepwise Regression
11	Ridge Regression	Forward Stepwise Regression	Kernel Regression
12	Lasso regression	ElasticNet Regression	Kernel Regression
13	ElasticNet Regression	Forward Stepwise Regression	Polynomial Regression
14	Simple linear regression	ElasticNet Regression	Backward Stepwise Regression
15	Lasso regression	Polynomial Regression	Forward Stepwise Regression
16	Simple linear regression	Polynomial Regression	Kernel Regression
17	Ridge Regression	ElasticNet Regression	Polynomial Regression
18	Lasso regression	Forward Stepwise Regression	Kernel Regression
19	Simple linear regression	Backward Stepwise Regression	Lasso Regression

Table 1: Group-wise allotment of regression methods

GROUP INFORMATION

Group No.	Member 1	Member 2
1	CS19M017	CS19M062
2	CS19M042	CS19M023
3	CS19M051	CS19M052
4	ED18B001	ED18B032
5	CS18S040	CS19D002
6	CS19M038	CS19M036
7	CS19M067	CS19M066
8	CS19M029	CS19M010
9	CS19M060	CS19M064
10	CS19M045	CS19M044
11	CS19M049	CS19M039
12	CS19M016	CS19M055
13	CS19M031	CS19M030
14	CS19M033	CS18M016
15	CS19D002	CS18S040
16	CS19M048	CS19M047
17	CS19M028	CS19M024
18	CS19M011	
19	CS19M050	

Table 2: Group Assignment