CS6464: Concepts in Statistical Learning Theory

SOFTWARE ASSIGNMENT 3

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

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

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 all three methods together in a single plot for each feature separately. A sample plot for one coefficient is shown in Figure 1.
- Evaluation of the models with Residual Sum of Squares (RSS) 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.

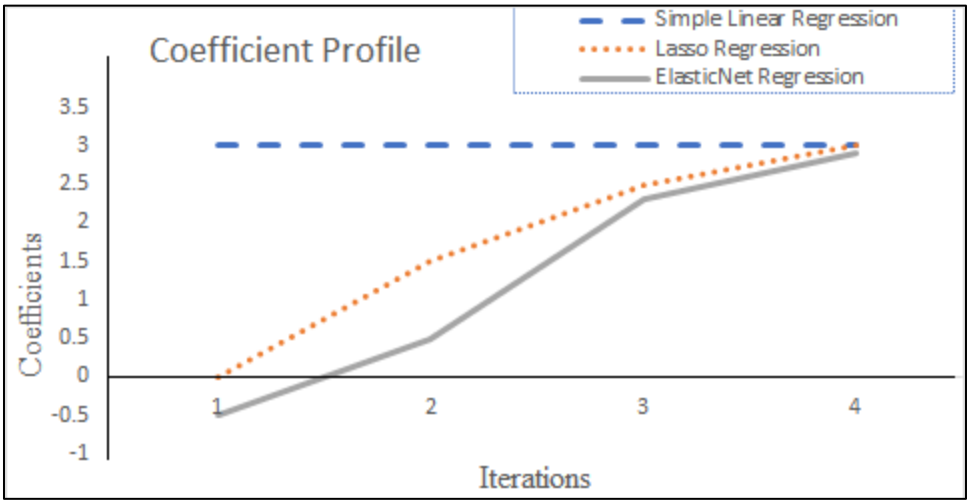

Figure 1: Coefficient profile plotting the coefficient weights vs iterations for a single feature

Grp no	Method 1	Method 2	Method 3
1	Backward Stepwise Regression	Lasso regression	Simple linear regression
2	Ridge Regression	ElasticNet Regression	Kernel Regression
3	ElasticNet Regression	Backward Stepwise Regression	Simple linear Regression
4	Forward Stepwise Regression	Lasso Regression	Polynomial Regression
5	Ridge Regression	Backward Stepwise Regression	Kernel Regression
6	Lasso Regression	Ridge Regression	Kernel Regression
7	Forward Stepwise Regression	ElasticNet Regression	Polynomial Regression

Table 1: Group wise allotment of regression methods

GROUP INFORMATION

Group no	Member 1	Member 2
1	CS17M041	CS17M013
2	CS17D012	ME14B034
3	CS17S010	CS17S009
4	CS18E002	CS18E001
5	CS17M024	CS17M049
6	CS17M015	CS17M044
7	CS17M051	EE18E004

Table 2: Group Assignment

NOTE:

- No Demo is required.
- Submit report online in PDF; details about report will be shared soon.