School of Computer Science Engineering and Technology

Course- BTech   Type- Core
Course Code- CSET301   Course Name-AIML
Year- 2022   Semester- Odd
Date- 26-10-2022   Batch- V Sem

Lab Assignment No. 8.1

Exp. No.	Name	CO-1	CO-2	CO-3
8.1	Lass and Ridge (Regression Problem)	✓	✓	--

Objective: To implement Lasso and Ridge technique of regression (using Scikit-learn) to predict price of the house.

Download the dataset from https://www.kaggle.com/code/himaoka/house-simple-svr-support-vector-regression/data?select=train.csv (10)

About Dataset:

The dataset consists of total 81 columns including categorical attribute. Out of 81 columns present in the dataset 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this competition challenges you to predict the final price of each home. The first variable 'Id' can be simply discarded, and the target variable is 'saleprice'.

1. **Data Pre-processing step: (30)**

    a) Take only the train.csv and consider it as you complete data name it as 'house_price.csv'.
    b) Check the number of columns in the dataset house_price.csv.
    c) Check the presence of missing values and handle it if present
    d) Check the presence of Categorical columns. Handle it if present. i.e., Transform categorical features into numerical features. (Hint: Use either one hot encoding, label encoding or any other suitable pre-processing technique).
    e) Scale the numerical columns value using MinMax Scalling.

2. Split the dataset into 80% for training and rest 20% for testing (sklearn.model_selection.train_test_split function) (5)
3. Create ridge regression model using built-in function on the training set with alpha=10 (Hint: Use from sklearn.linear_model import Ridge) (10)
4. Check the performance of the trained model on test set. Find the value of R^2 score. (10)
5. Create lasso regression model using built-in function on the training set with alpha=10 (Hint: Use from sklearn.linear_model import Lasso) (10)

6. Check the performance of the trained model on test set. Find the value of R^2 score. (10)
7. Compare the results of ridge and lasso regressor with previously used MLR.(10)

   Extra: a) Selecting Optimal Alpha Values Using Cross-Validation in Sklearn and go for creating the ridge and lasso model. b) Optimize the performance of the model by removing the noise/outlier from the data

   Suggested Platform: Python: Jupyter Notebook/Azure Notebook/Google Colab.