VIT – AP University Amaravati

Lab 8

Course Title: Fundamentals of Datascience
Instructor's name: Dr. Aastha
Course Code: MAT4201

Class: Msc. Datascience
Semester:Fall 2023-24

For the given dataset, perform the following analysis in python.

- Consider price as the dependent variable and the remaining variables as independent.
- Construct the correlation heat map for all the variables. What do you observe?
- Divide the data such that 75% data is train and remaining is test.
- Fit multiple linear regression model, ridge regression and lasso regression with ($\alpha = 1$), the tuning parameter on the train data.
- Extract the estimated coefficient values for each model with respect to the independent variables as output. What are the changes you observe in the coefficient values?
- Use the fitted models on the test data to predict price and compute the MSE for each model. Which model gives the best prediction?
- Create three adjacent barplots with respect to each estimated coefficient value corresponding to each independent variable. Use green, red and blue respectively for linear, ridge and lasso.
- Is lasso eliminating the correlated variables?