My Name (myNetID)

IE598 MLF F18

Module 4 Homework (Regression)

Part 1: Exploratory Data Analysis

Describe the data sufficiently using the methods and visualizations that we used previously in Module 3 and again this week. Include any output, graphs, tables, heatmaps, box plots, etc. Label your figures and axes. DO NOT INCLUDE CODE!

Split data into training and test sets. Use random\_state = 42. Use 80% of the data for the training set. Use the same split for all models.

Part 2: Linear regression

Fit a linear model using SKlearn to all of the features of the dataset. Describe the model (coefficients and y intercept), plot the residual errors, calculate performance metrics: MSE and R2.

Part 3.1: Ridge regression

Fit a Ridge model using SKlearn to all of the features of the dataset. Test several settings for alpha. Describe the model (coefficients and y intercept), plot the residual errors, calculate performance metrics: MSE and R2. Which alpha gives the best performing model?

Part 3.2: LASSO regression

Fit a LASSO model using SKlearn to all of the features of the dataset. Test several settings for alpha. Describe the model (coefficients and y intercept), plot the residual errors, calculate performance metrics: MSE and R2. Which alpha gives the best performing model?

Part 3.3: Elastic Net regression

Fit an ElasticNet model using SKlearn to all of the features of the dataset. Test several settings for l1\_ratio. Describe the model (coefficients and y intercept), plot the residual errors, calculate performance metrics: MSE and R2. Which l1\_ratio gives the best performing model?

Part 4: Conclusions

Write a short paragraph summarizing your findings.

Part 5: Appendix

Link to github repo