ShunTat Lam (stlam2)

IE517 S23

Module 4 Homework (Regression)

Part 1: Exploratory Data Analysis

The data contains 13 explanatory variables and one scalar response(MEDV)

Chart

Description automatically generated

*scatter plot of features.*

Chart, bar chart

Description automatically generated

*heatmap of correlations between features*

(In EDA part only parts of the “x” are selected, since trying to run all of them ruined the laptop and resulted to a forced -shut down.)

Part 2: Linear regression

The first regression trying is using all 13 x, into a regular linear regression.

The models’ coefficients and y intercept are shown below

Text

Description automatically generated

The residual

Chart, scatter chart

Description automatically generated

The models MSE and R^2:



apparently the model is not good. There is undesired pattern in the residual plot, the MSE is high.

(Tried multiple methods,including .score and manually calculating R^2 but always get a negative value.)

Part 3.1: Ridge regression

The next trying is to use ridge. The graph below shows the changing of weight in predictors with the changing of alpha. Looks like alpha does not affect much.

Chart

Description automatically generated

The coefficients and intercept:

Text

Description automatically generated

MSE and R-squares:



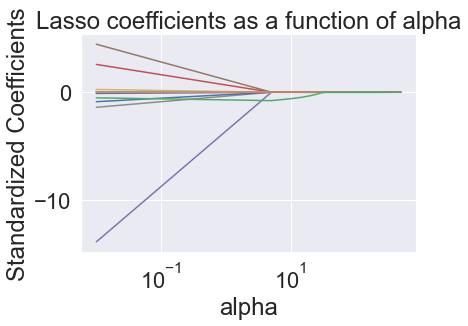
The residual:

Chart, scatter chart

Description automatically generated

Based on the MSE, ridge performs better.

Part 3.2: LASSO regression



The issue of lasso is it seems like the alpha setting results to set all predictors’ weight to 0.

Text, chat or text message

Description automatically generated



So, just use the ridge

Part 4: Conclusions

When dealing with multiple x, ridge performs much better than the regular regression with no penalty. For lasso, it seems like alphas are hard to deal with, sometimes they just result to all weights = 0

Part 5: Appendix

Link to github repo