My Name Biao Feng(myNetID: biaof2)

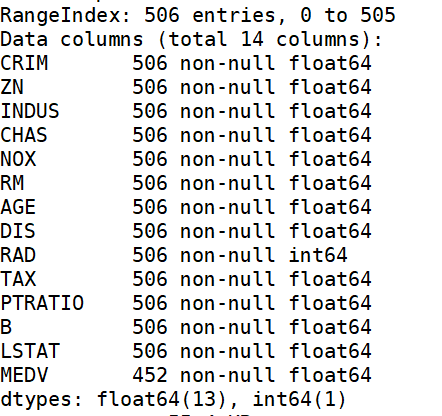
IE598 MLF F18

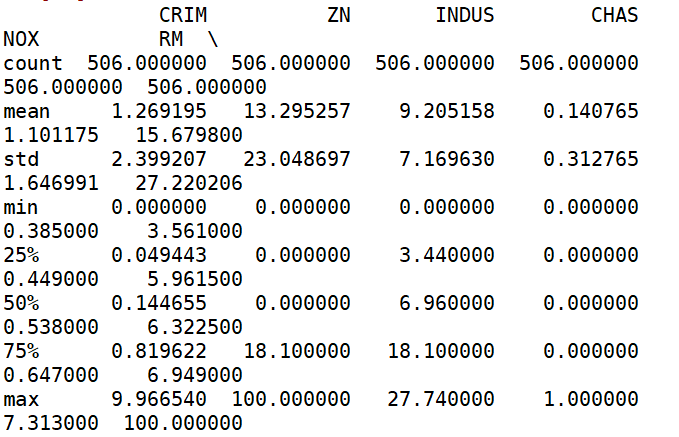
Module 4 Homework (Regression)

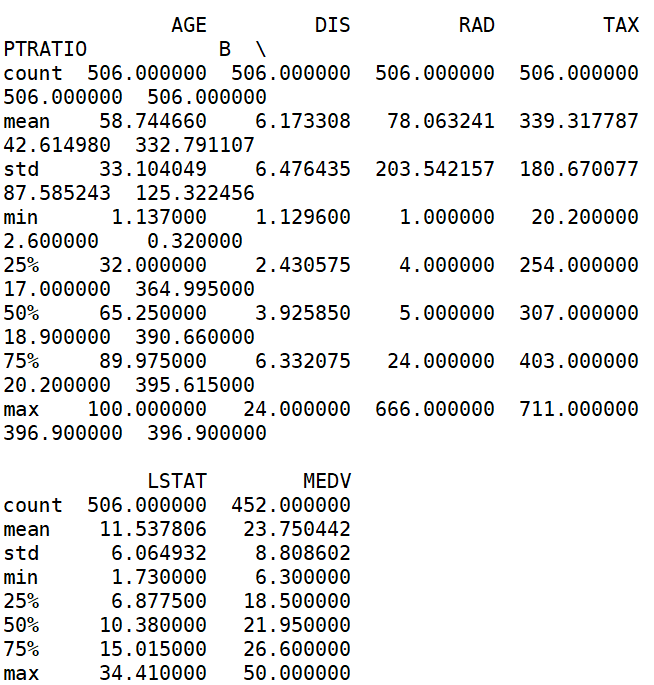
Part 1: Exploratory Data Analysis

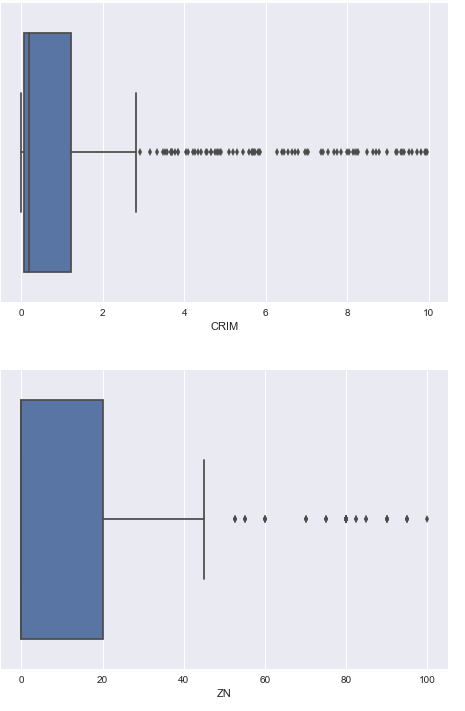
Describe the data sufficiently using the methods and visualizations that we used previously in Module 3 and again this week. Include any output, graphs, tables, heatmaps, box plots, etc. Label your figures and axes. DO NOT INCLUDE CODE!

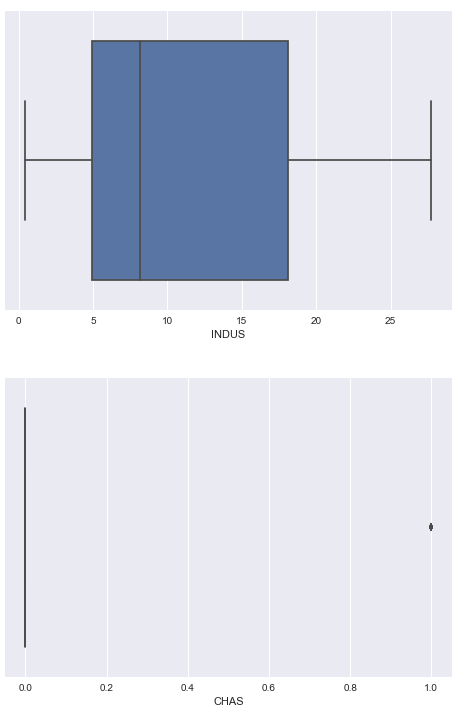
Split data into training and test sets. Use random\_state = 42. Use 80% of the data for the training set. Use the same split for all models.

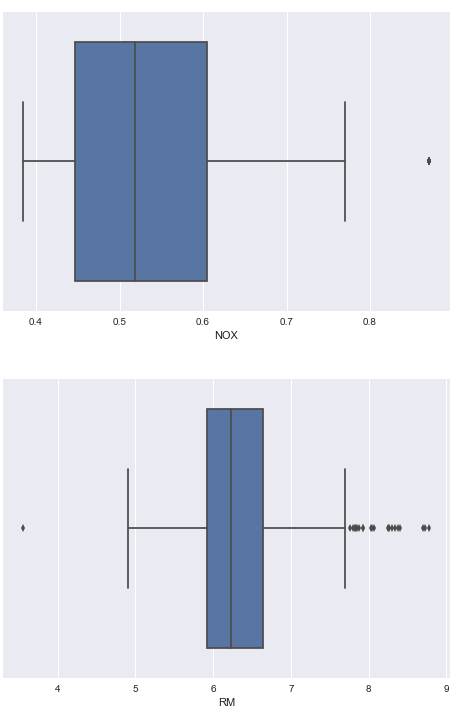


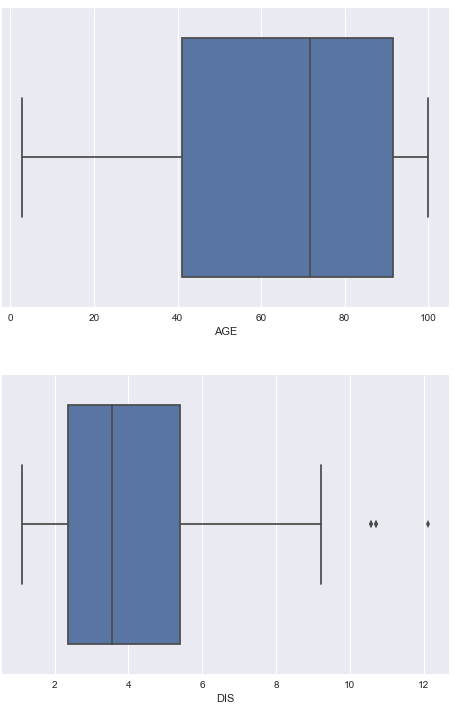


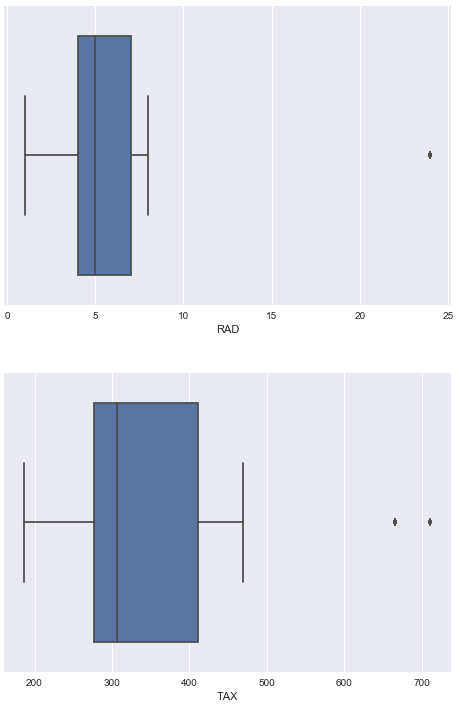


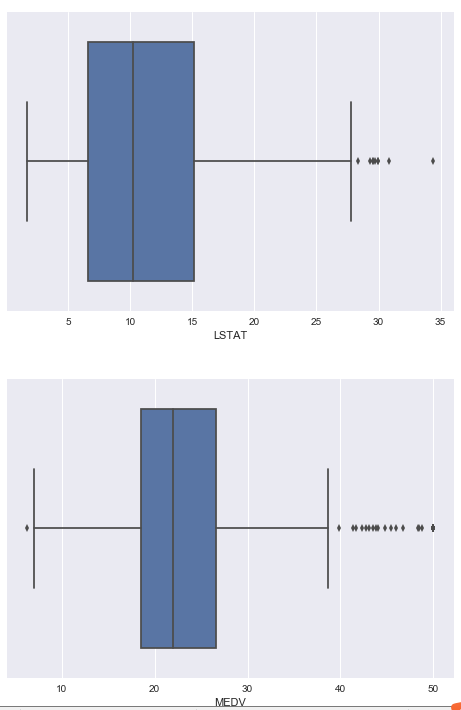


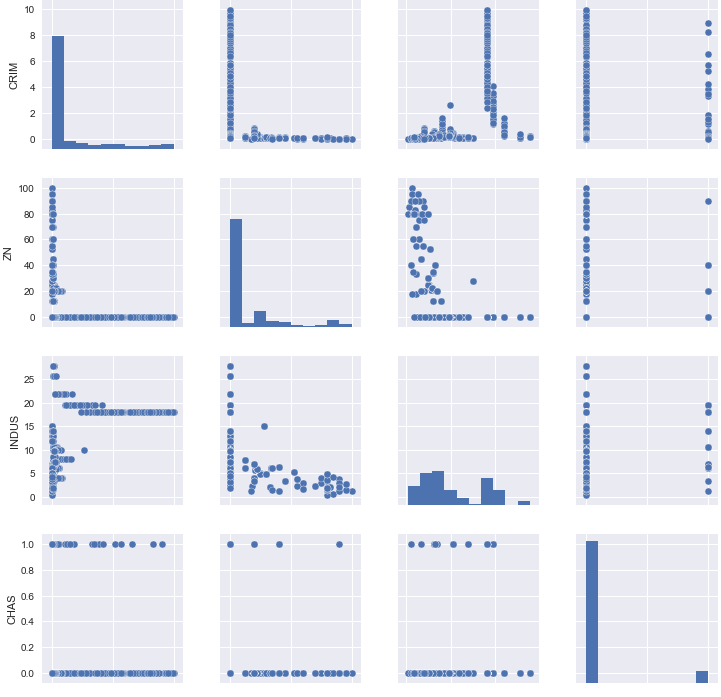


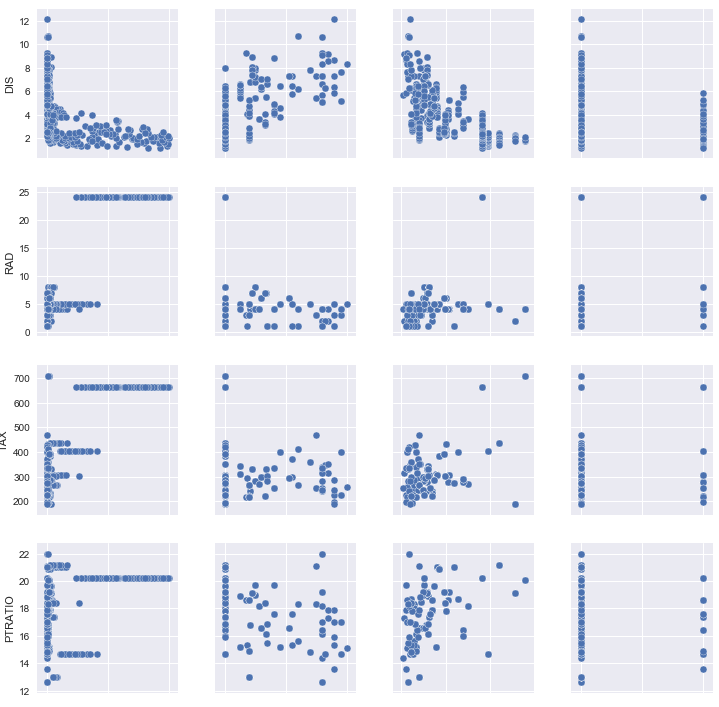




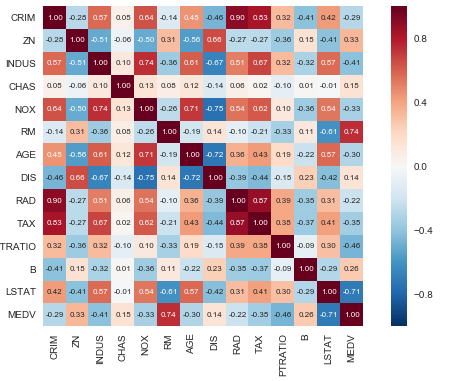


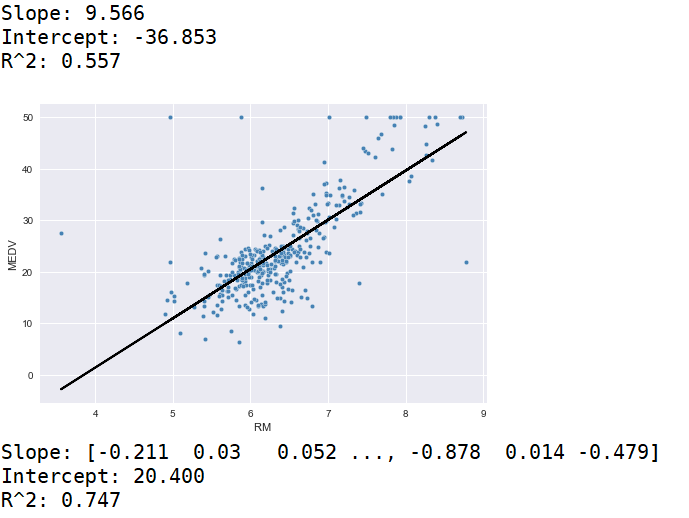








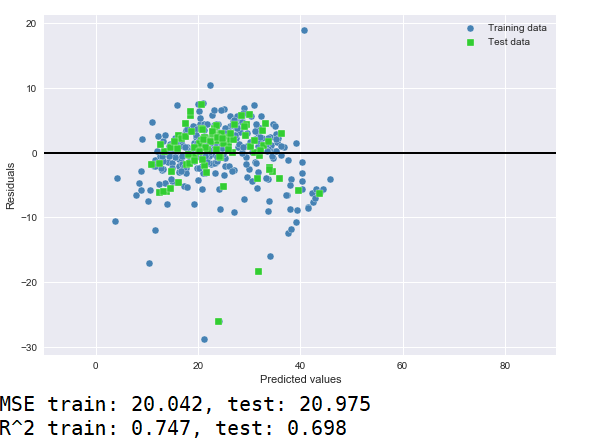




From the graph, we can see that there are total 506 data. The target is MEDV columns, and there are other 13 feature columns. In every column, we can find that the cout, mean, std, min, 25%, 50%, 75% and max. From heatmap, we can see the correlation between two features.

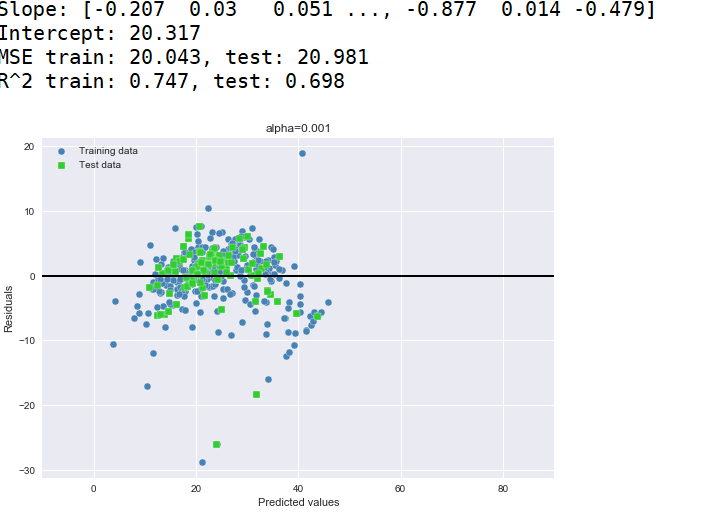
Part 2: Linear regression

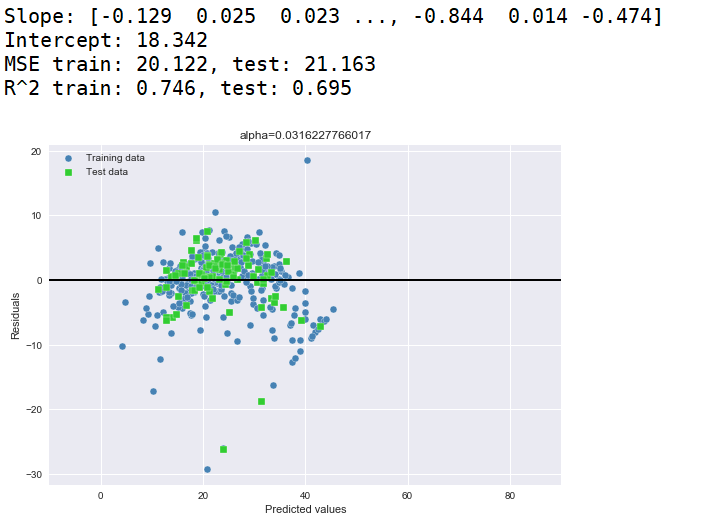
Fit a linear model using SKlearn to all of the features of the dataset. Describe the model (coefficients and y intercept), plot the residual errors, calculate performance metrics: MSE and R2.

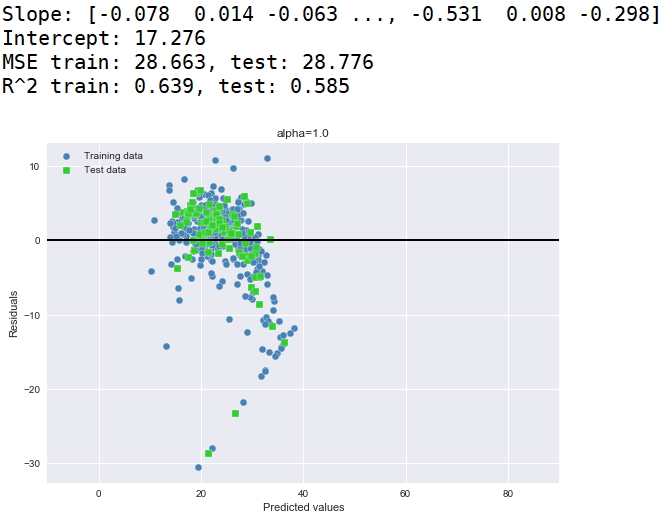


Part 3.1: Ridge regression

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?



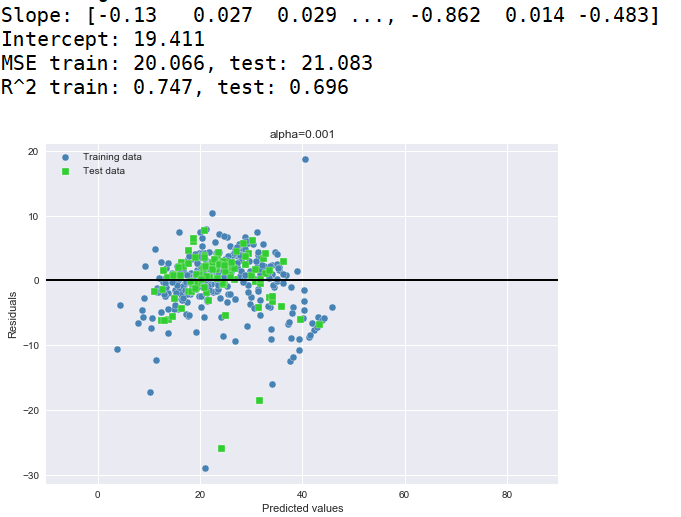


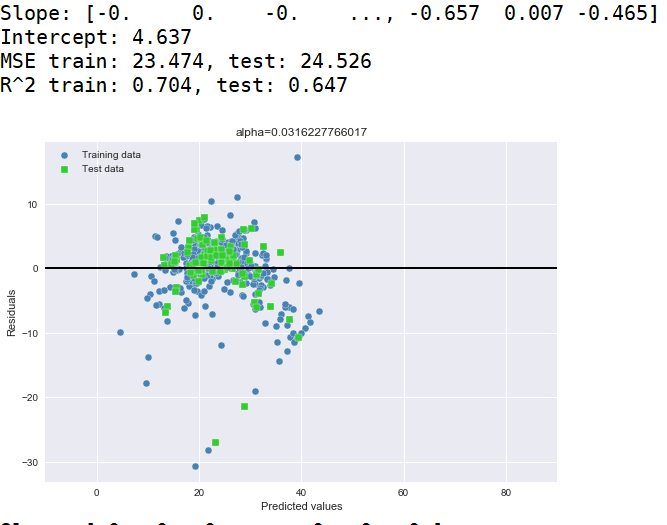


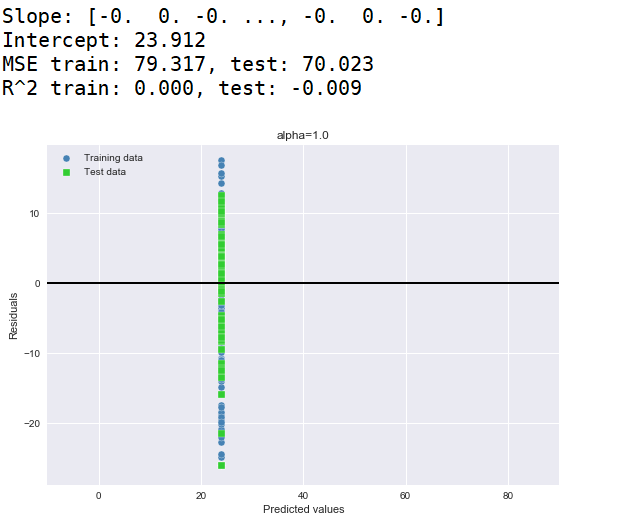
When alpha = 0.031622, it is the best model

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?



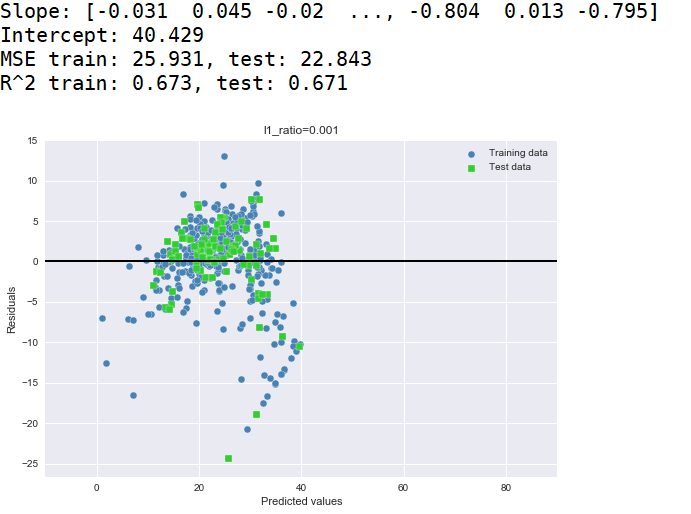


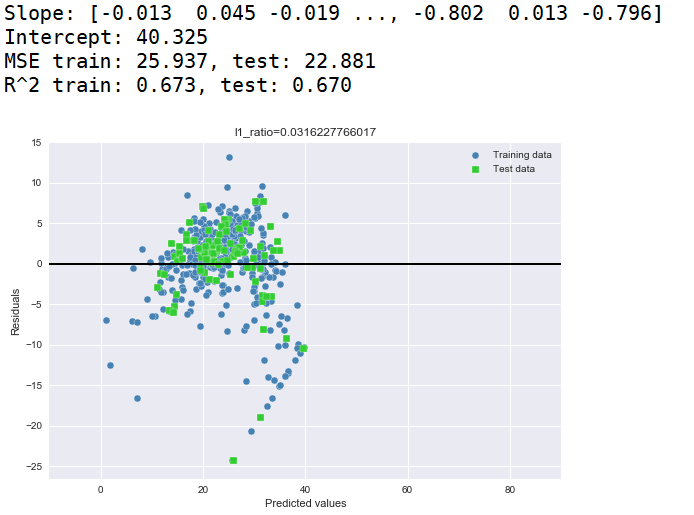


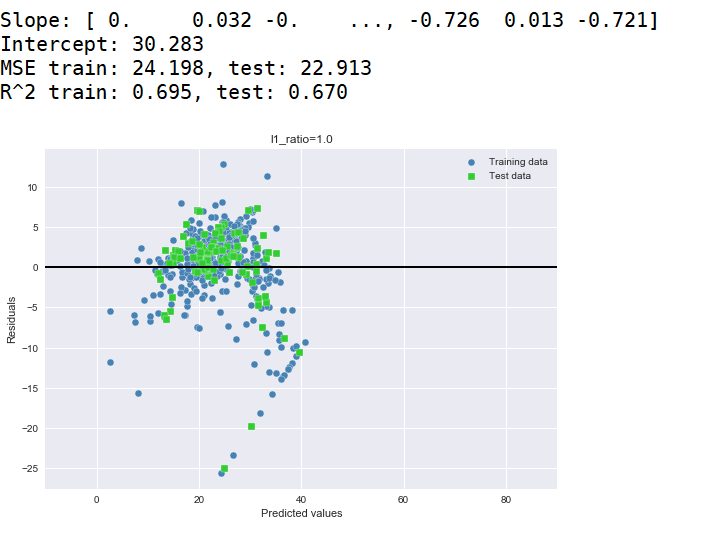
When alpha = 0.001, it is the best model.

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?







When alpha ratio is 0.001, it is the best model.

Part 4: Conclusions

Different regressions will have different R^2 in different alpha. From ridge regression, R^2 is bigger when alpha = 0.031622. From lasso regression, R^2 is bigger when alpha = 0.001. From Elastic Net regression, R^2 is bigger when alpha = 0.001.

Part 5: Appendix

https://github.com/johnfeng123/Biao-Feng