Report
Name: Prabha Pandey
Roll Number: 2018201053

Assignment 7

Regularization:

The process of reducing over-fitting, by increasing bias is regularization. In case of linear regression, if the model is over-fitted on training data, then a small increase in bias decreases the variance drastically.
A tuning parameter λ controls the strength of the penalty. λ is basically the amount of shrinkage.

Regularization can be done using two approaches:
1) Lasso Regression
2) Ridge Regression

1) Lasso Regression:

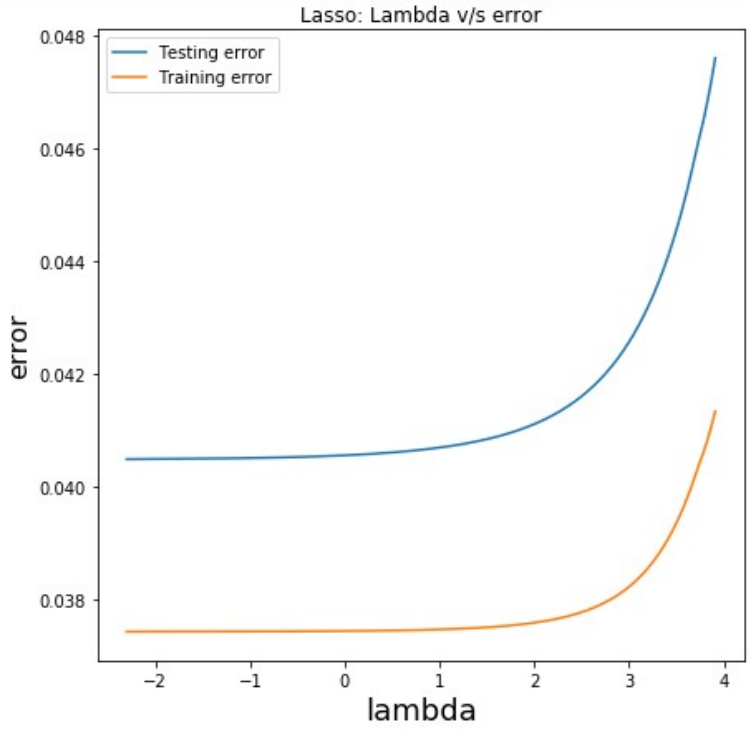
Formula used to update the theta value in case of lasso regression is as follows:

$$\theta_j := \theta_j - \alpha.1/m\Sigma(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} - \alpha * \lambda * sign(w_j)/2m$$

Some Observations:

- When λ = 0, no parameters are eliminated. The estimate is equal to the one found with linear regression.
- As λ increases, more and more coefficients are set to zero and eliminated.

Graph Obtained:



2) Ridge Regression:

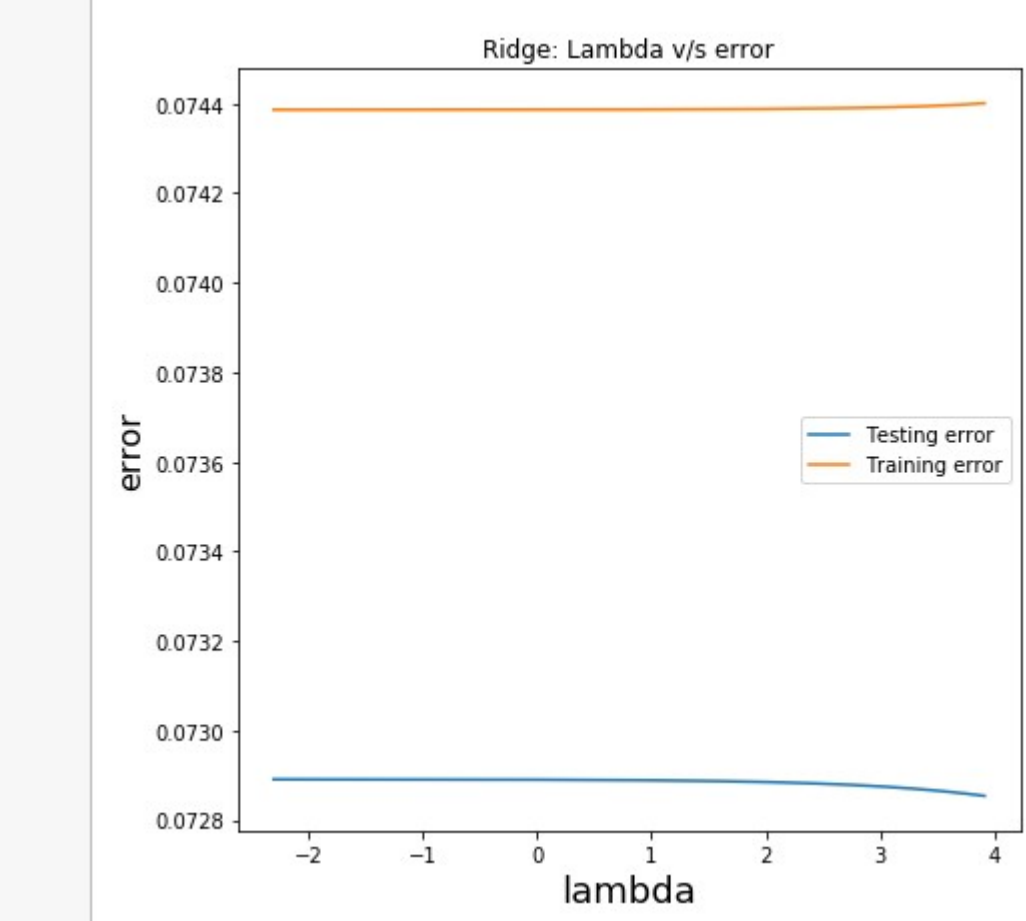
Formula used to update the theta value in case of lasso regression is as follows:

$$\theta j := \theta j - \alpha.1/m\Sigma(h\theta(x(i)) - y(i))x(i)j - \alpha * \lambda * wj/2m$$

Some Observations:

- When λ = 0, no parameters are eliminated. The estimate is equal to the one found with linear regression.
- As λ increases, more and more coefficients approaches zero.

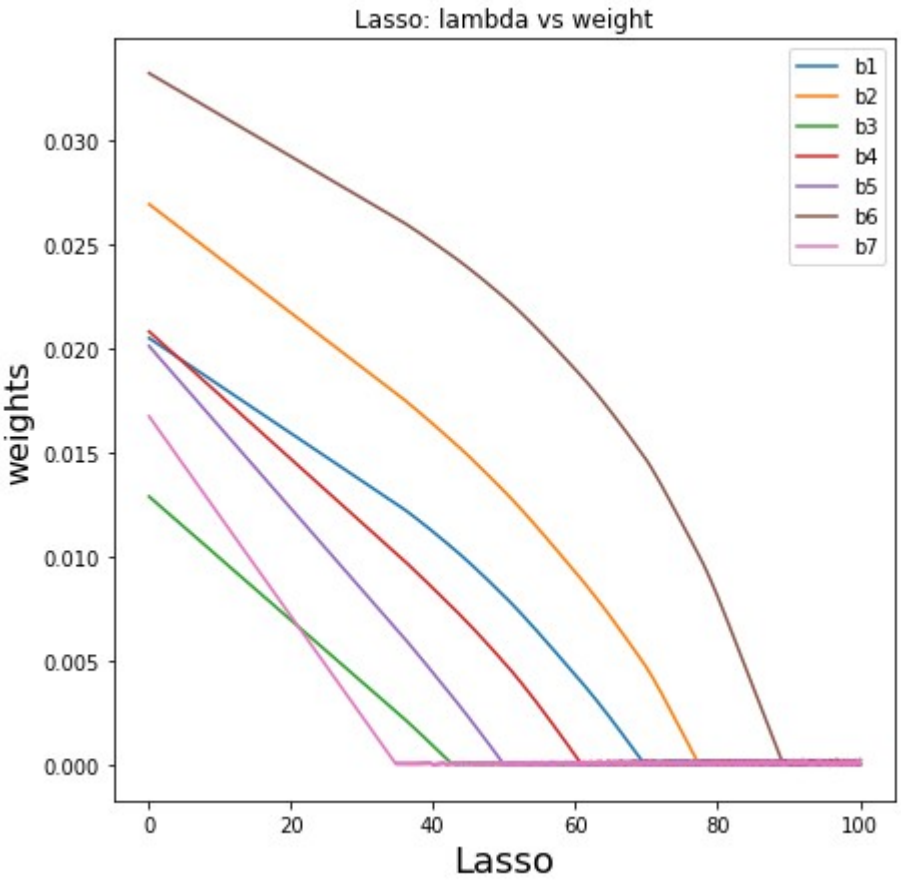
Graph Obtained:

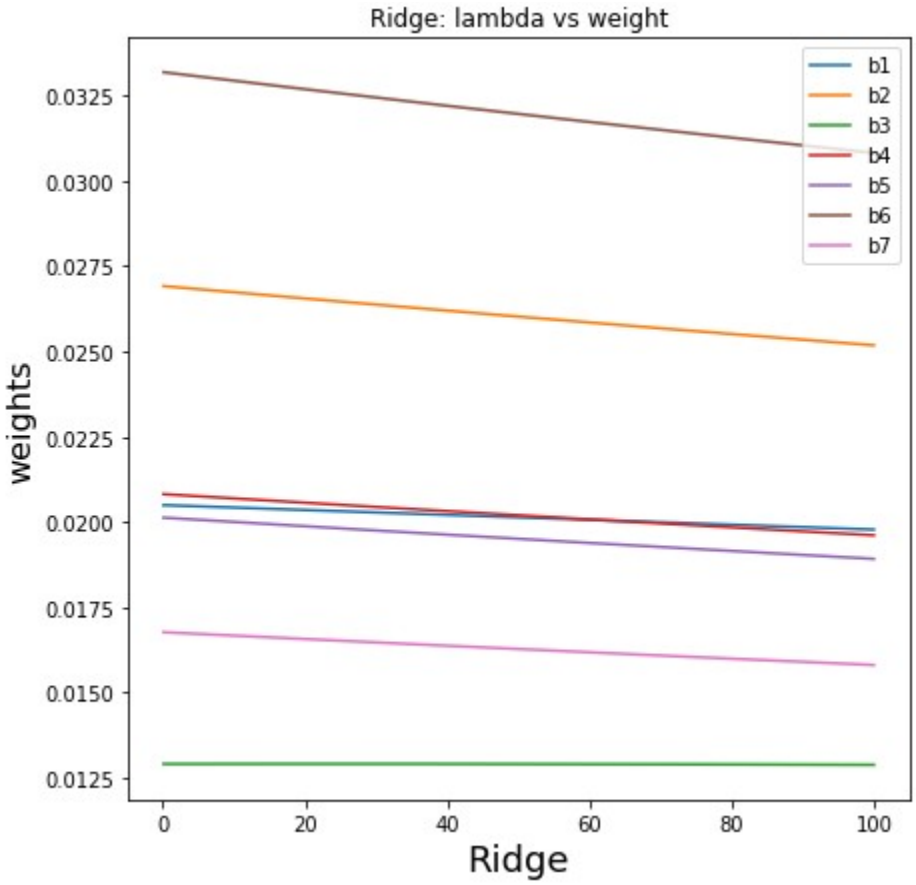


3) How the hyper-parameter λ plays a role in deciding between bias and variance.

From the above 2 graphs, it is evident that, as the value of λ increases, bias is introduced, complexity of model is reduced and variance (hence over-fitting) reduces, test error reduces,.
And as λ decreases, bias decreases, complexity increases, over fitting comes into picture, i.e. variance increases.
But after a certain point large value of λ makes the weights negligible, hence both bias and variance increase.

4) How the two different regularization techniques affect regression weights in terms of their values and what are the differences between the two.





Observation:

From the above 2 graphs, it can be noted that, as the lambda value increases the weights decrease.
This decrease is drastic in case of Lasso regression compared to ridge regression because, lasso regression is L1 regularization, so it makes weights close to zero faster in comparison to ridge which is L2 regression.

5) K-fold cross validation