Homework #2

Due on November 10, 2019 at 11:55pm

Professor Hoda Mohammadzade

Amirhossein Yousefi

97206984

Problem 6

Section a:
In this part we use linear regression model is used in order to predict the house price.Firstly we used forward selection which in each step selects the most import feature according to its *z-score* however in backward selection we omit the least significant feature in each step successively.Bellow figure shows feature as they were sorted by these methods.

```
########forward selection :features from weakest to strongest(left to right)
b=['LSTAT','INDUS','ZN','TAX','RAD','CRIM','CHAS','AGE','DIS','NOX','B','PTRATIO','RM']
```

(a) Forward selection

```
######backward selection sorting is bellow(from weakest to storngest(left to right))
a=['AGE','INDUS','CHAS','ZN','TAX','CRIM','RAD','B','NOX','DIS','PTRATIO','RM','LSTAT']
```

(b) Backward selection

Figure 1: Feature significance from weakest to strongest

Section b:
We know than forward selection select sub optimal most significant features.Forward selection result for top three feature is **RM ,PTRATIO, B.** and by backward selection result is **LSTAT,RM,PTRATIO** and other result such as train error and test error and score mentioned in bellow figure.

```
score is  0.7836295385076265
 mean square train error  23.51333444932702
 mean square test error  19.831323672063387
```

(a) Full feature regression

```
score is  0.7228623473287098
 mean square train error  28.35603780904954
 mean square test error  25.40090941218404
```

(b) Top three feature regression by backward selection

```
score is  0.7111927174772167
 mean square train error  36.69465011769556
 mean square test error  26.470483350890515
```

(c) Top three feature regression by forward selection

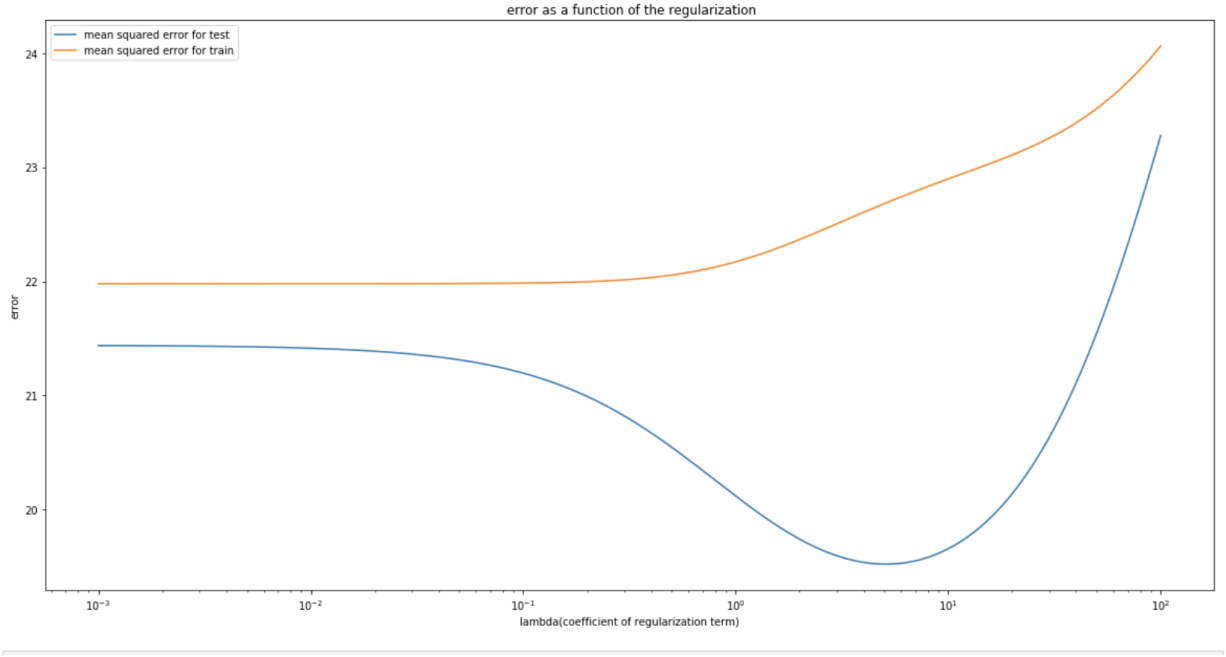
Figure 2: Results

As we can see from result train and test error for full feature regression is lower than the rest and score is also

better but with a little difference in comparison with others means omitted features were insignificant.

Section c:

In order to make model overfitted we used %.95 of data in training process.As λ is a hyper parameter it has to be chosen when test error (validation) reached it's minimum.below figure demonstrate the regularization term impact.



(a) Full feature regression

Adding regularization term causes bias in model but can significantly reduce the variance of model so that generalization error for model is going to be better.Increasing λ causes training error to be increased.The lowest part of blue curve which is test error give us the optimal value of λ.Overally regularization term prevents us from overfitting.