# Homework 9

Snigdha Peddi

## Exercises (ISLR)

**Use set.seed(20219) in each exercise to make results reproducible.**

**1. Question 6.8.3 pg 260, Suppose we estimate the regression coefficients in a linear regression model by minimizing below equation for a particular value of s. For parts (a) through (e), indicate which of i. through v. is correct. Justify your answer.**

$$\sum_{i=1}^{n} ( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})$$

Subject to

$$\sum_{j=1}^{p} |\beta_j| \leq s$$

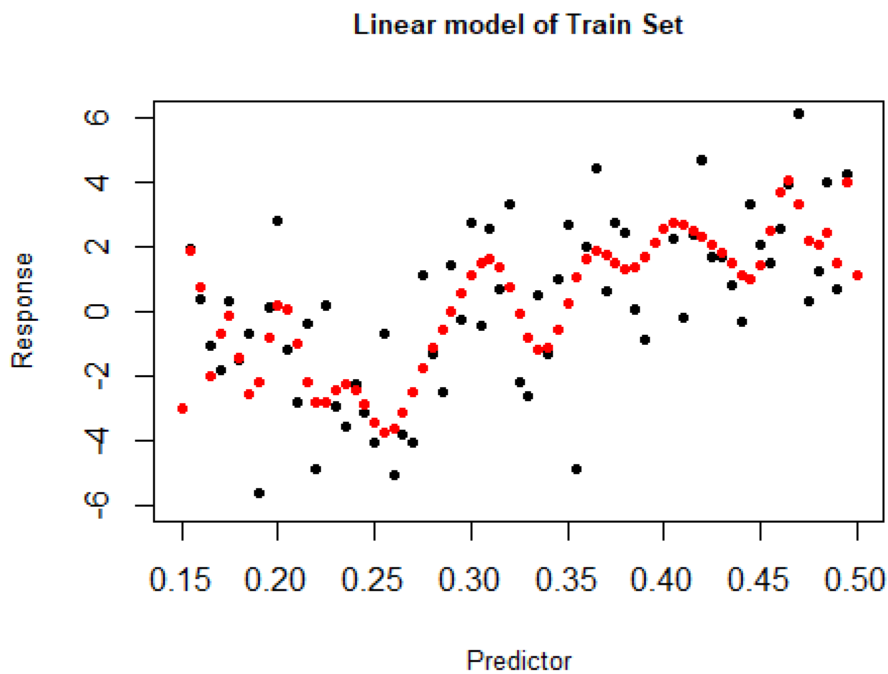
**(a)** As we increase s from 0, the training RSS will:

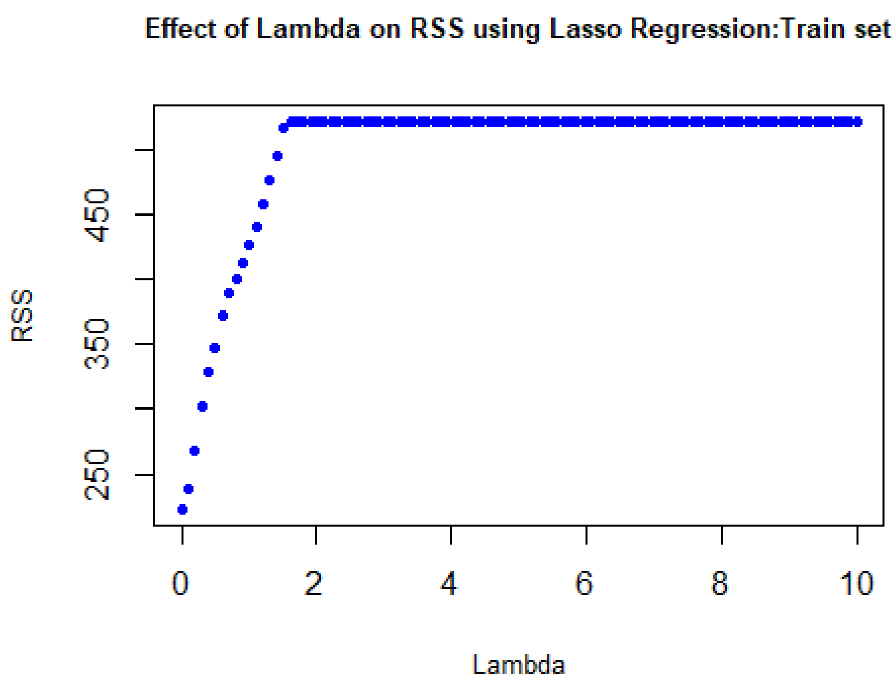
i. Increase initially, and then eventually start decreasing in an inverted U shape.
ii. Decrease initially, and then eventually start increasing in a U shape.
iii. Steadily increase.
iv. Steadily decrease.
v. Remain constant.

**Answer:**

The answer is option III, Steadily decreases.

As s increases from 0 and becomes large then coefficient estimates can be large (large slope) resulting in lower Residual sum of squares (lasso regression line can be closer to the linear regression line). When s is sufficiently large then RSS+lasso penalty becomes equal to RSS. Hence with increase in s the RSS decreases. The Lambda and s are related such that increase in lambda corresponds to decrease in s. From the plot illustrating the relation between lambda and RSS shows that with increase in lambda the RSS increases and reaches a point where the RSS remains constant.This indicates that increase in s will decrease the RSS and at a value of s it becomes equal to the least square solution.




**(b)** Repeat (a) for test RSS.

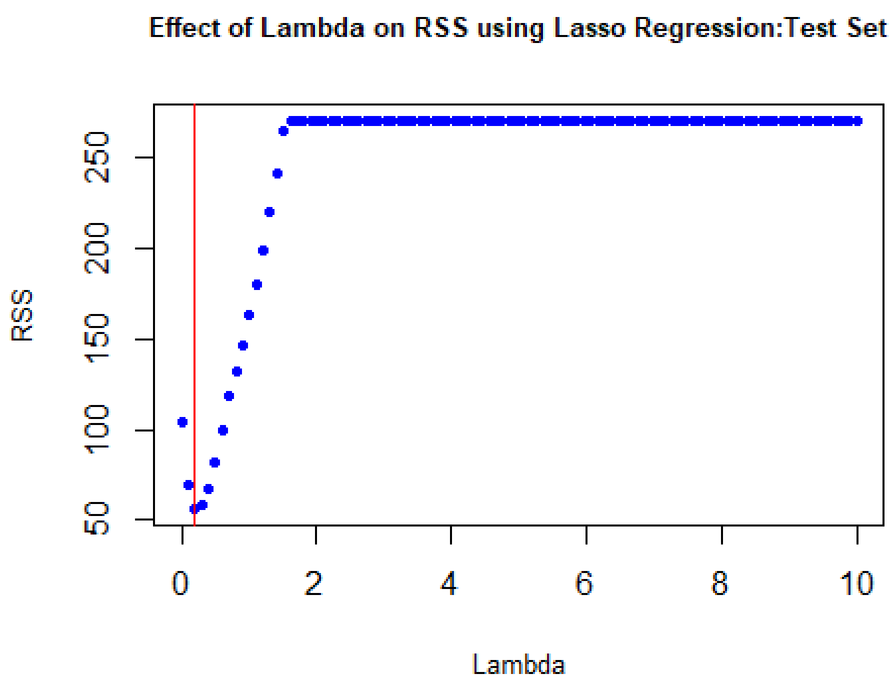
i. Increase initially, and then eventually start decreasing in an inverted U shape.
ii. Decrease initially, and then eventually start increasing in a U shape.

iii. Steadily increase.
iv. Steadily decrease.
v. Remain constant.

**Answer:**

The answer is option II, Decrease initially, and then eventually start increasing in a U shape.

With increase in S the flexibility of model increases, RSS decreases initially and then increases until a point where the model overfits.



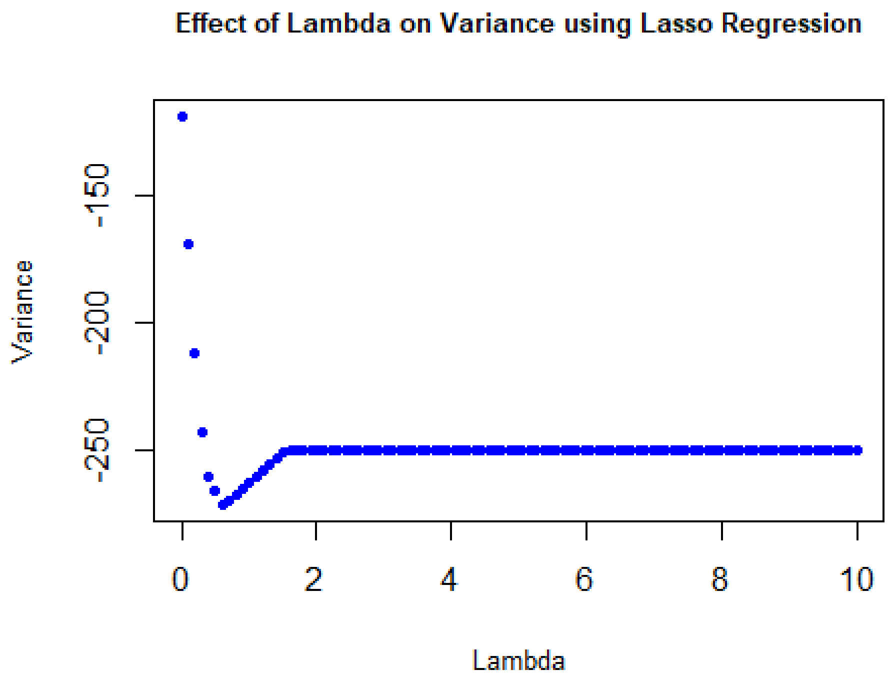

**(c)** Repeat (a) for variance.

i. Increase initially, and then eventually start decreasing in an inverted U shape.
ii. Decrease initially, and then eventually start increasing in a U shape.
iii. Steadily increase.
iv. Steadily decrease.
v. Remain constant.

**Answer:**

The answer is option III, Steadily increase.

With increase in s from 0 model flexibility is increasing (corresponding to decrease in lambda). The bias is decreasing and the variance increases.When s is sufficiently large then the coefficient estimate becomes equal to least square solution then the variance remains constant.



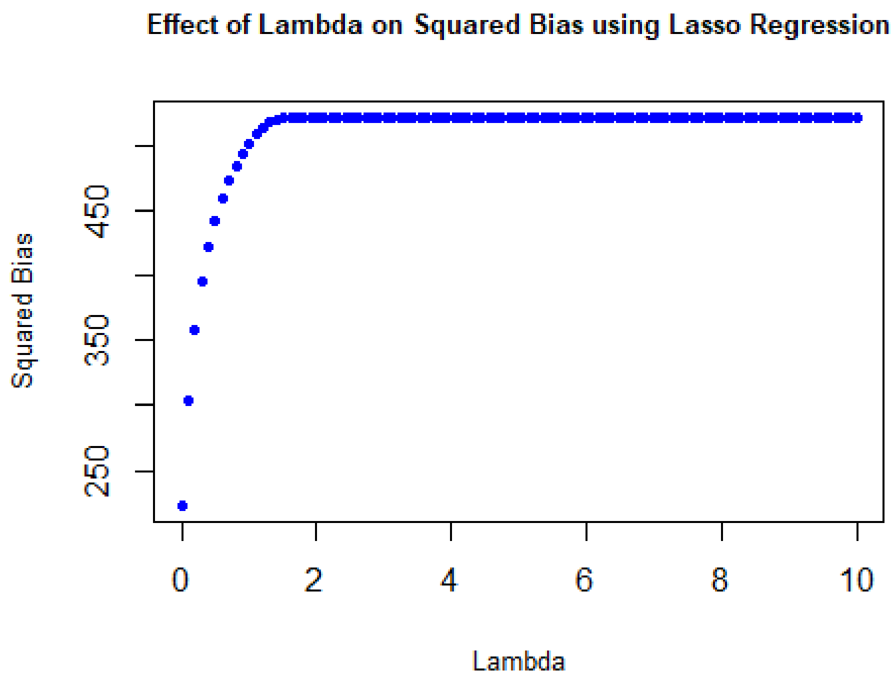

**(d)** Repeat (a) for (squared) bias.

i. Increase initially, and then eventually start decreasing in an inverted U shape.
ii. Decrease initially, and then eventually start increasing in a U shape. iii.Steadily increase.
iii. Steadily decrease.
iv. Remain constant.

**Answer:**

The answer is option IV, Steadily decrease.

As the s increases from 0 the lasso regression line is close to linear regression line and hence has a low bias.The plot below shows that with increase in lambda bias increases which corresponds to decrease in bias with increasing s. And bias remains constant when s is sufficiently large where RSS is equal to least squares solution.




**(e)** Repeat (a) for the irreducible error.

i. Increase initially, and then eventually start decreasing in an inverted U shape.
ii. Decrease initially, and then eventually start increasing in a U shape. iii.Steadily increase.
iii. Steadily decrease.
iv. Remain constant.

**Answer:**

The answer is option V, Remain constant.

Irreducible error is a noiSe in the system which may be due to the unexplained independent variables or response variation. Changing the flexibility of the system does not change the error and changes in S (lambda) do not effect the error and remains constant.

**2. Question 6.8.4 pg 260, Suppose we estimate the regression coefficients in a linear regression model by minimizing below equation for a particular value of λ. For parts (a) through (e), indicate which of i. through v. is correct. Justify your answer.**

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}) + \lambda\sum_{j=1}^{p}\beta_j^2$$

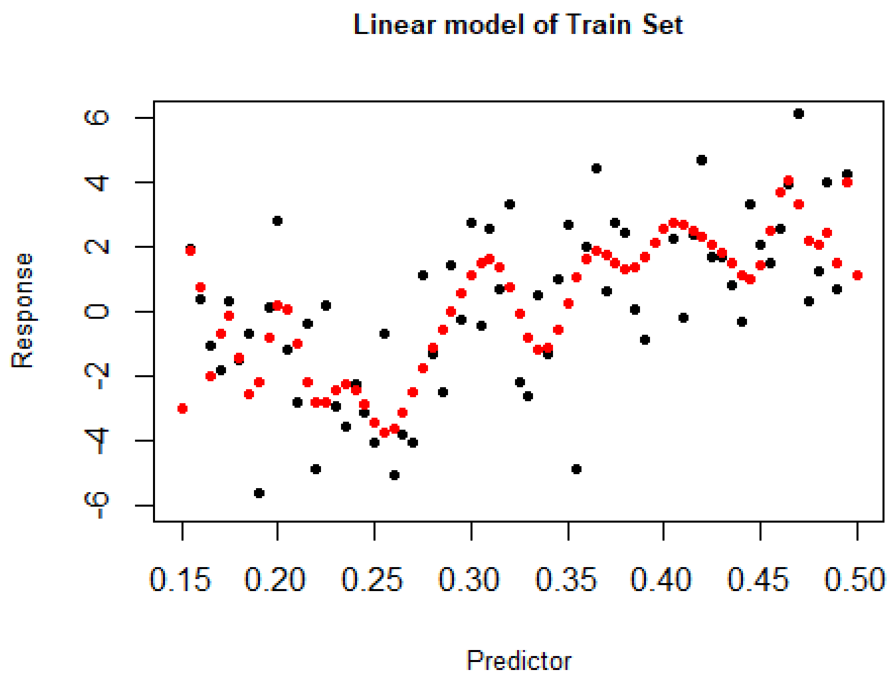
**(a)** As we increase λ from 0, the training RSS will:

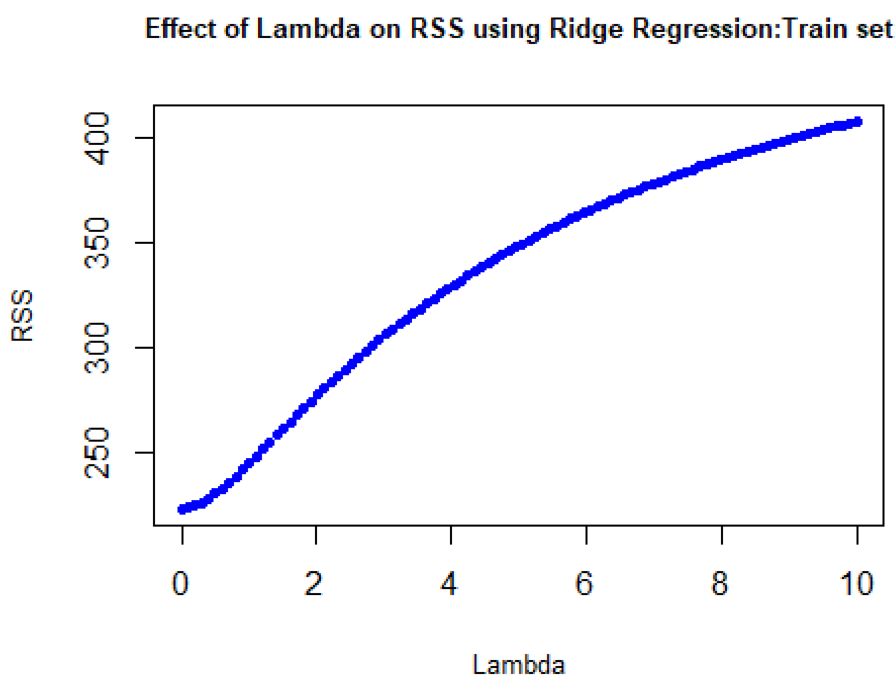
i. Increase initially, and then eventually start decreasing in an inverted U shape.
ii. Decrease initially, and then eventually start increasing in a U shape.
iii. Steadily increase.
iv. Steadily decrease.
v. Remain constant.

**Answer:**

The answer is option III, Steadily increases.

As the Lambda increases the residual sum of squares and ridge penalty decreases. The RSS increases as the bias is introduced into the system and reducing the slope of the linear model.




**(b)** Repeat (a) for test RSS.

i. Increase initially, and then eventually start decreasing in an inverted U shape.
ii. Decrease initially, and then eventually start increasing in a U shape.

iii. Steadily increase.
iv. Steadily decrease.
v. Remain constant.

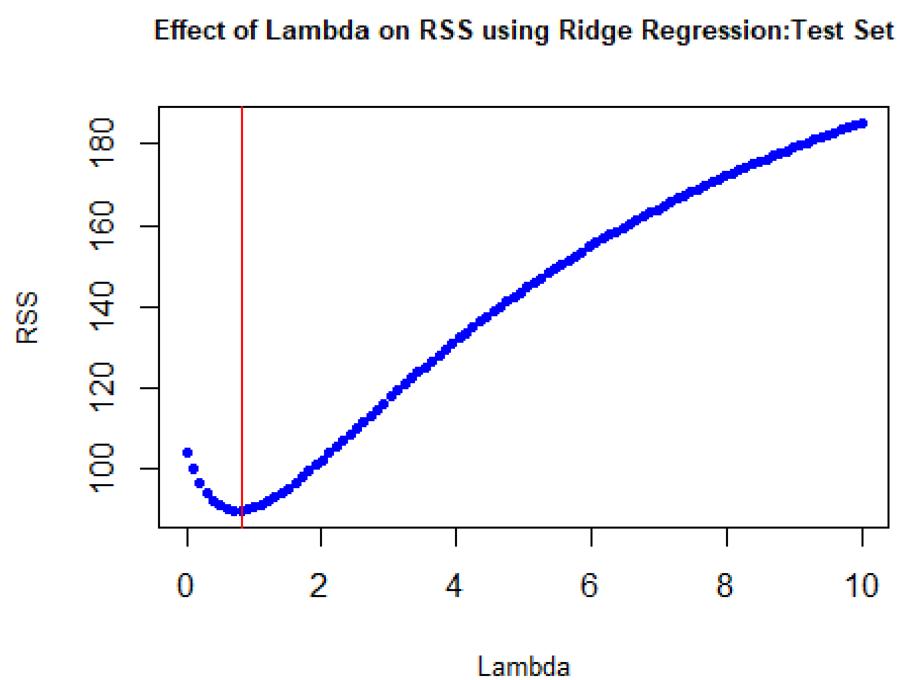
**Answer:**

The answer is option II, Decrease initially, and then eventually start increasing in a U shape.

As the Lambda increases the residual sum of squares and ridge penalty decreases and thereafter increases as the model becomes overfit beyond a certain lambda value.There is bias variance trade off. At a Lambda of 0.81 the RSS is at lowest and thereafter the Bias increases.

*Lamda with lower RSS*

| Lambda | RSS |
|---|---|
| 0.808 | 89.864 |



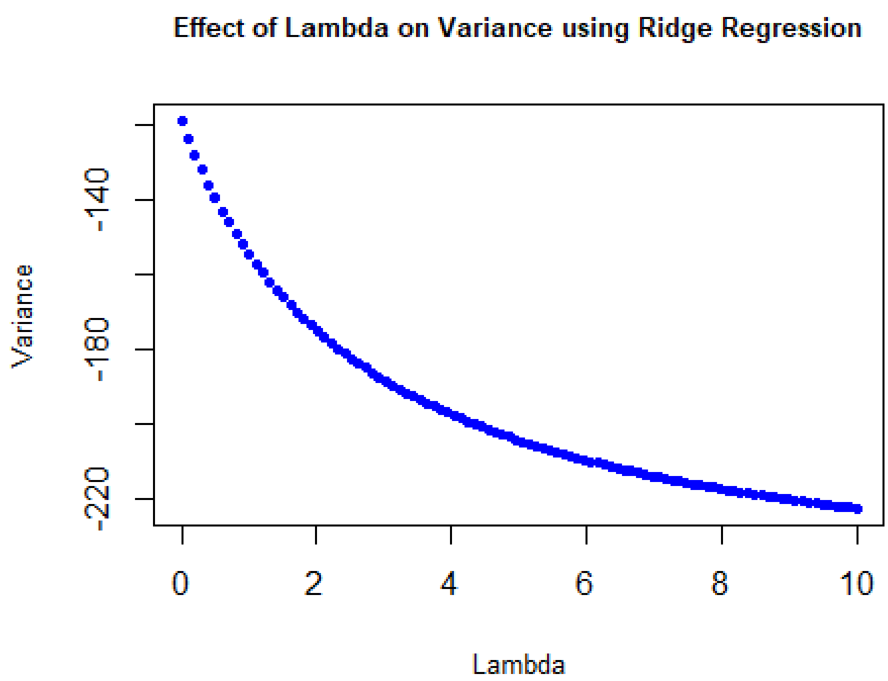

**(c)** Repeat (a) for variance.

i. Increase initially, and then eventually start decreasing in an inverted U shape.
ii. Decrease initially, and then eventually start increasing in a U shape.
iii. Steadily increase.
iv. Steadily decrease.
v. Remain constant.

**Answer:**

The answer is option IV, Steadily decrease.

As the Lambda increases the residual sum of squares and ridge penalty decreases shrinking the value of coefficient reducing the variance.



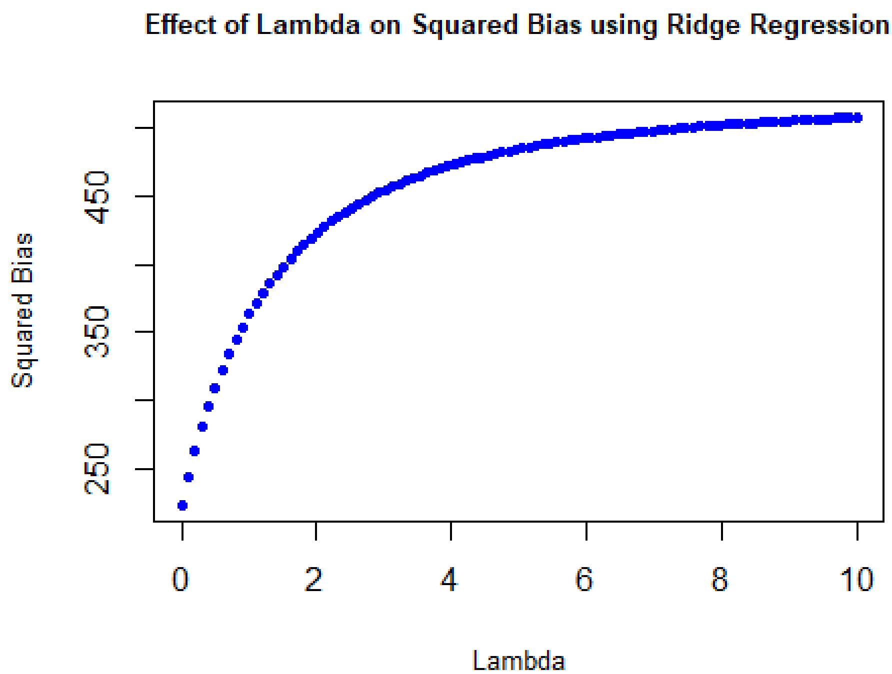

**(d)** Repeat (a) for (squared) bias.

i. Increase initially, and then eventually start decreasing in an inverted U shape.
ii. Decrease initially, and then eventually start increasing in a U shape. iii.Steadily increase.
iii. Steadily decrease.
iv. Remain constant.

**Answer:**

The answer is option III, Steadily increase.

As the Lambda increases the residual sum of squares and ridge penalty decreases,decreasing the flexibility by introducing bias(shrinking the coefficient values).




**(e)** Repeat (a) for the irreducible error.

i. Increase initially, and then eventually start decreasing in an inverted U shape.
ii. Decrease initially, and then eventually start increasing in a U shape. iii.Steadily increase.
iii. Steadily decrease.
iv. Remain constant.

**Answer:**

The answer is option V, Remain constant.

Irreducible error is a noiSe in the system which may be due to the unexplained independent variables or response variation. Changing the flexibility of the system does not change the error and changes in lambda do not effect the error and remains constant.

**3. We will now try to predict permeability in the *rock* data set.**

```
##
## Dimensions of Rock dataset: 48 4
```

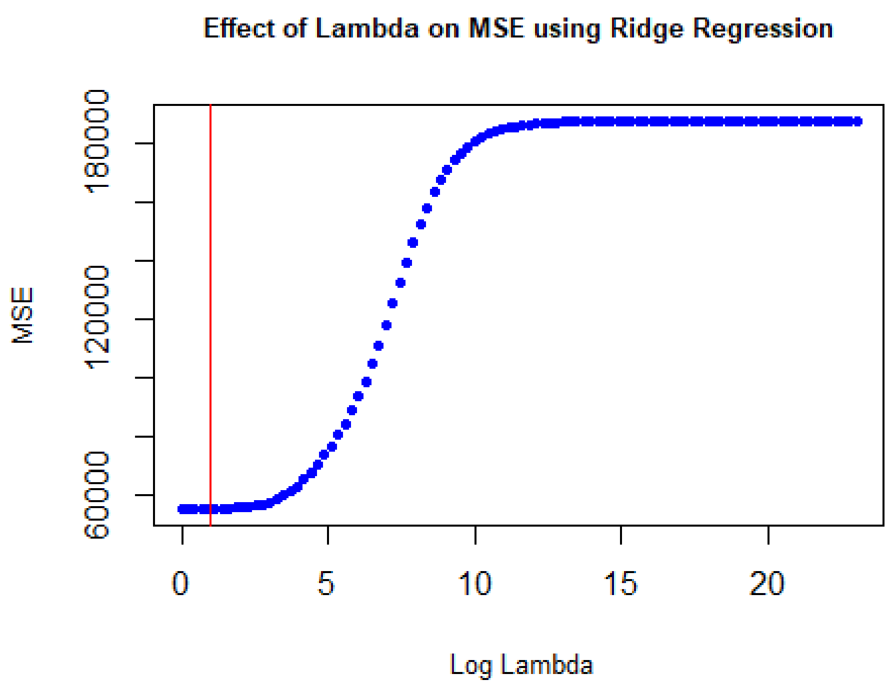
**a)** Use lasso and ridge regression methods to fit the model:

$$permeability \sim area + perimeter + shape$$

Compare the two methods. Present and discuss results for the two approaches.

**Answer:**

*glmnet()* function is used to fit the Ridge and Lasso regression models.The Mean square error was calculated for all the lambda values and the lambda at which the MSE is lower was identified for both the models.Both the models gave a best MSE at lambda of 1. However, the ridge method has a better MSE of 55489.9 compared to 55492.1 of lasso regression. This indicates that all the variables used for the models are useful.In both cases as the lambda increases the MSE increases. At around a lambda of 15 for ridge fit and 6 for lasso fit the MSE became consistent with increasing lambda indicating that the coefficient estimates are not sensitive to independent variables.

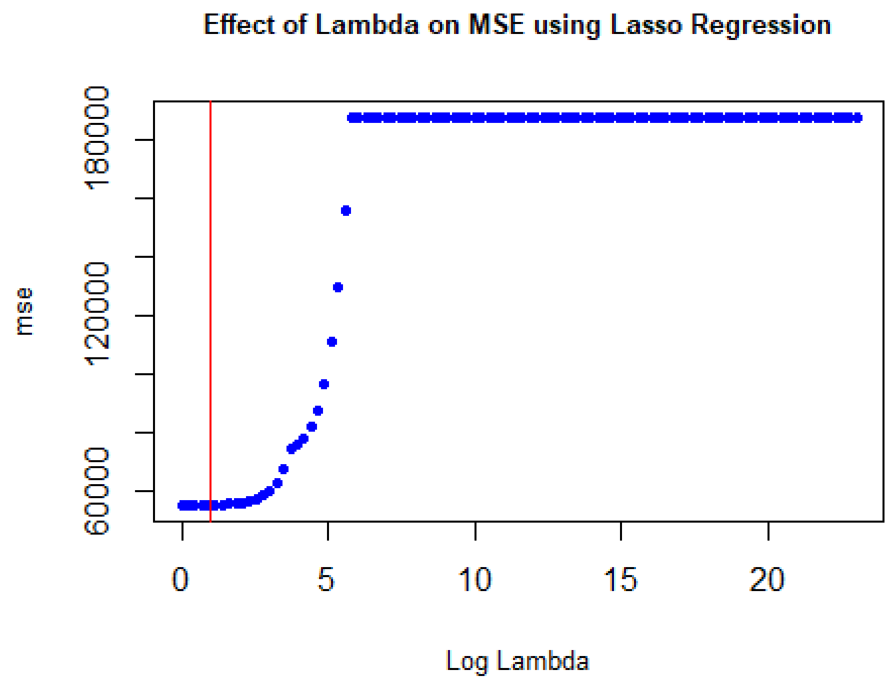



```
## Ridge Regression Model for Rock Data:

## glmnet(x = rok.x, y = rok.y, alpha = 0, lambda = grid[i], standardize =
TRUE)

## The MSE of the Ridge Regression model at the best Lambda: 55489.9

##
##
## Labmda with lower MSE(Ridge Regression Fit): 1
```




```
## Lasso Regression Model for Rock Data:

## glmnet(x = rok.x, y = rok.y, alpha = 1, lambda = grid[i], standardize =
TRUE)

## The MSE of the Lasso Regression model at the best Lambda: 55492.08

##
##
## Labmda with lower MSE(Lasso Regression Fit): 1
```

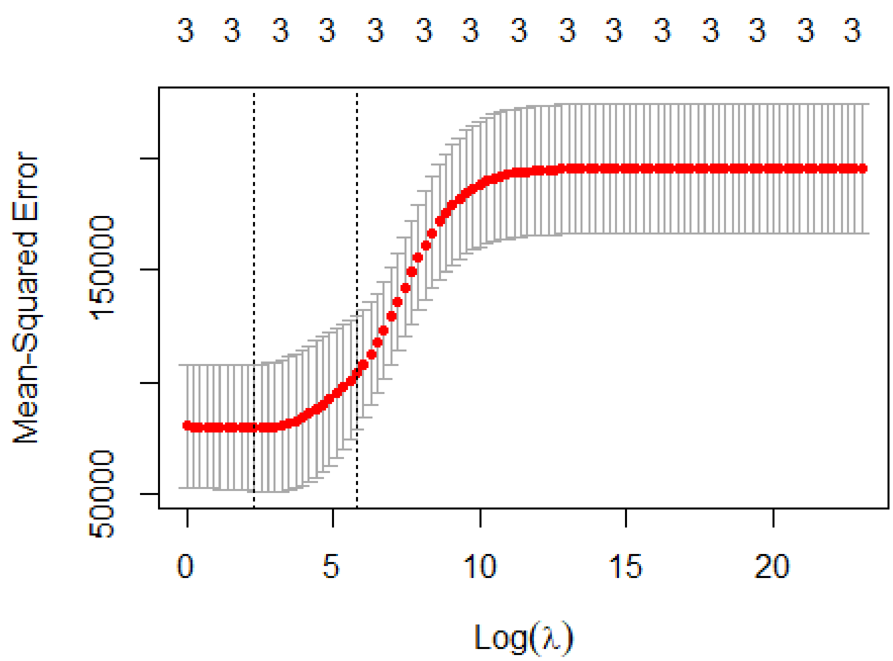
**b)**Evaluate model performance of both lasso and ridge regression using validation set error, cross validation, or some other reasonable alternative, as opposed to using training error. State explicitly which method you choose to evaluate the error and justify your choice.

**Answer:**

I have used Leave One Out Cross Validation to evaluate my models as the number of observations in the Rock data are less (48). I have used LOOCV to reduce the bias and randomness as LOOCV yields same results when performed multiple times. I have used cv.glmnet() function to perform the LOOCV for both the models. I have standardized the x variables prior to fitting the models as the values of the independent variables are at different scale. A grid of lambda is pre-selected for the models. The Ridge regression fit has a lower MSE of 79326 (achieved at a lambda of 10.2) compared to the MSE of Lasso regression fit which is 80419 (achieved at a lambda of 1).Plots of both the models show the same trend. As the lambda increases the MSE increases and plateaus out where the effect of the independent variables is not significant on the response variable. The model summaries also show that for Lasso fit, at the lambda that gives the most regularized model where the MSE is within one standard deviation the number of variables with zero coefficients is 2. After evaluating the models with LOOCV I beleive that Ridge regression fit is better than Lasso regression Fit.

```
## Ridge regression Fit for Rock data and LOOCV:

##
## Call:  cv.glmnet(x = rok.x, y = rok.y, lambda = 10^seq(0, 10, length =
100),      type.measure = "mse", nfolds = length(rok.y), alpha = 0,
standardize = TRUE, family = "gaussian")
##
## Measure: Mean-Squared Error
##
##      Lambda Index Measure    SE Nonzero
## min    10.2    90   79326 28361       3
## 1se   335.2    75  104243 25346       3
```

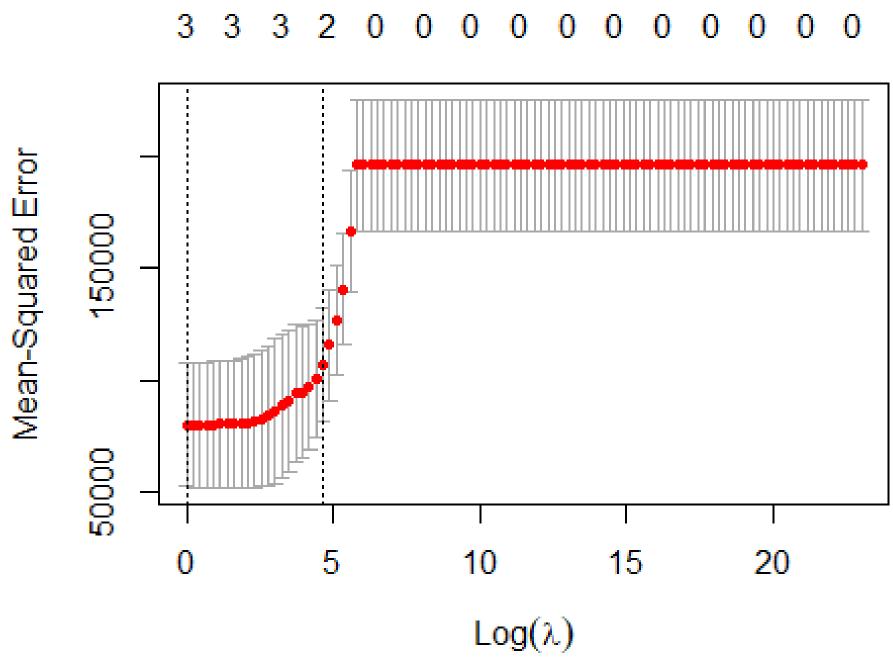



Effect of Lambda on MSE:Ridge Regression Fit

```
## Lasso regression Fit for Rock data and LOOCV:

##
## Call:  cv.glmnet(x = rok.x, y = rok.y, lambda = 10^seq(0, 10, length =
100),      type.measure = "mse", nfolds = length(rok.y), alpha = 1,
standardize = TRUE, family = "gaussian")
##
## Measure: Mean-Squared Error
```

```
##
##      Lambda Index Measure    SE Nonzero
## min    1.0   100   80419 27765       3
## 1se  104.8    80  106680 25228       2
```



Effect of Lambda on MSE:Lasso Regression Fit

**c)** Are there issues with the data that would justify the added complexity of the ridge regression? Justify your answer. Consider the residuals from a linear model.

### Answer:

A linear model is fit for the Rock data set. The training MSE is calculated and was estimated to be 55480. The MSE of the Ridge regression fit is 55490. The ridge regression penalty has introduced more bias to the system and increased the complexity of the model.The lower MSE of the ridge fit is achieved at a lambda of 1.It is introducing very little bias and when lambda is 0 it is essentially equal to the linear fit. The residual values of the linear model are very low compared to ridge regression model. As only 3 variables are used for the model I believe the simple linear model with lower MSE is better then Ridge or Lasso fits.

```
## Summary of Linear Model for Rock Data:

##
## Call:
## glm(formula = perm ~ area + peri + shape, data = rock)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -750.26   -59.57    10.66   100.25   620.91
```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 485.61797  158.40826   3.066 0.003705 **
## area          0.09133    0.02499   3.654 0.000684 ***
## peri         -0.34402    0.05111  -6.731 2.84e-08 ***
## shape       899.06926  506.95098   1.773 0.083070 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 60523.24)
##
##     Null deviance: 9009186  on 47  degrees of freedom
## Residual deviance: 2663023  on 44  degrees of freedom
## AIC: 670.56
##
## Number of Fisher Scoring iterations: 2

## MSE of the Linear model: 55479.64

##
##
## The MSE of the Ridge Regression model(at Lambda=1): 55489.9
```

### References:


- Blog by Trevor Hastie, Junyang Qian, Kenneth Tay on *An Introduction to glmnet*, February 17,2021.
- Video Lecture by Josh Starmer (statquest) on *Ridge,Lasso and Elastic-Net Regression in R*, October 23,2018
- Video Lecture by Experfy on *Learn Lasso and Ridge Regression in R*, December 18,2017.
- Video Lecture by Josh Starmer (statquest) on *Machine Learning Fundamentals: Bias and Variance*, September 17,2018
- Video Lecture by Josh Starmer (statquest) on *Regularization Part 1: Ridge(L2) Regression*, September 24,2018
- Video Lecture by Josh Starmer (statquest) on *Regularization Part 2: Lasso(L1) Regression*, October 1,2018