DS 301: HW 7

Problem 1: Concept Review

a. For the lasso regression model, what is the bias of $\hat{\beta}_{\text{lasso}}$ when $\lambda = 0$?

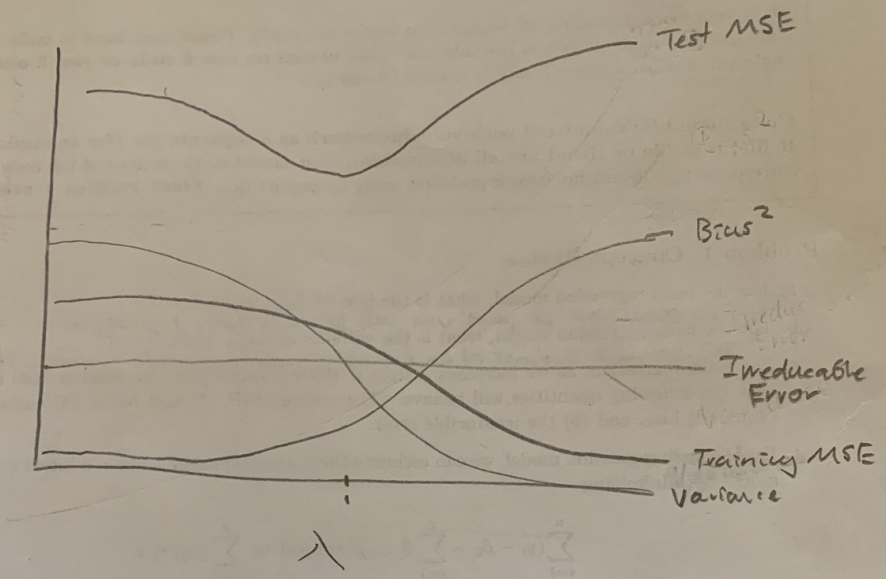
When ⍀ = 0. The variance is high but there is no bias.

b. For the lasso regression model, what is the variance of $\hat{\beta}_{\text{lasso}}$ when $\lambda = \infty$?

When ⍀ = ∞, the bias will increase based on the log of the ⍀. However, variance will eventually be reduce to zero.

c. For ridge regression, as we increases λ from 0, draw a single plot (by hand is fine) showing how the following quantities will behave: (1) training MSE, (2) test MSE, (3) variance, (4) (squared) bias, and (5) the irreducible error.

Ridge Regression

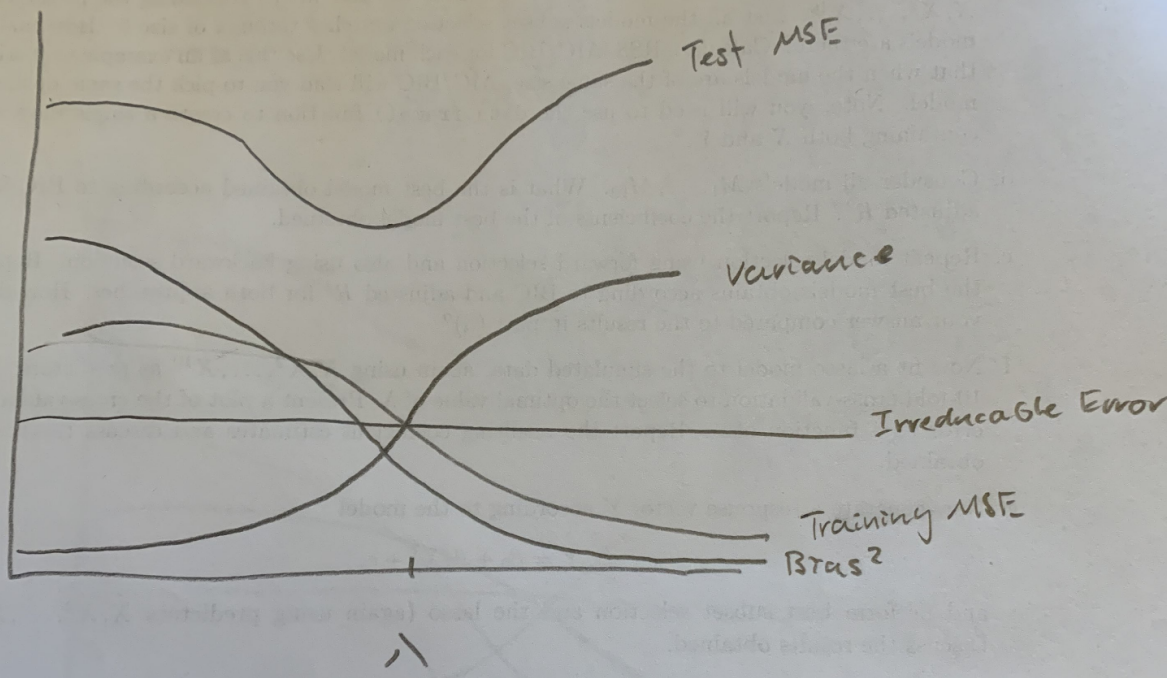


1. Training MSE
2. Test MSE
3. Variance
4. Squared Bias
5. irreducable error

d. For the lasso regression model, we can estimate the regression coefficients in a linear regression model by minimizing

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 \text{ subject to } \sum_{j=1}^{p}|\beta_j| \leq s.$$

This is **exactly equivalent** to the formulation presented in class, except now instead of a penalty λ we have a constraint controlled by s. As we increase s from 0, draw a single plot showing how the following quantities will behave: (1) training MSE, (2) test MSE, (3) variance, (4) (squared) bias, and (5) the irreducible error.


Test MSE
Variance
Irreducable Error
Training MSE
Bias²

Problem 2: Simulation Studies

a).

b).

c).

# of models: 9 sumsets of each size up to 8

```
x10     FALSE     FALSE
9 subsets of each size up to 8
Selection Algorithm: exhaustive
          x1  x2  x3  x4  x5  x6  x7  x8  x9  x10
1  ( 1 ) " " " " "*" " " " " " " " " " " " " "
1  ( 2 ) " " " " " " " " " "*" " " " " " " " "
1  ( 3 ) " " " " " " " " " " " " "*" " " " " "
1  ( 4 ) " " " " " " " " " " " " " " " "*" " "
1  ( 5 ) "*" " " " " " " " " " " " " " " " " "
1  ( 6 ) " " " " " " " " " " " " " " " " " "*"
1  ( 7 ) " " " " " " " " " " " " " "*" " " " "
1  ( 8 ) " " " " " " " " " " "*" " " " " " " "
1  ( 9 ) " " "*" " " " " " " " " " " " " " " "
2  ( 1 ) " " " " "*" " " " " " " " " "*" " " "
2  ( 2 ) " " " " "*" " " " " " "*" " " " " " "
2  ( 3 ) " " " " "*" " " " " " " " " " " " "*"
2  ( 4 ) " " "*" "*" " " " " " " " " " " " " "
2  ( 5 ) "*" " " "*" " " " " " " " " " " " " "
2  ( 6 ) " " " " "*" "*" " " " " " " " " " " "
2  ( 7 ) " " " " "*" " " "*" " " " " " " " " "
2  ( 8 ) " " " " "*" " " " " " " " "*" " " " "
2  ( 9 ) " " " " "*" " " " " " " " " " "*" " "
3  ( 1 ) " " "*" "*" "*" " " " " " " " " " " "
3  ( 2 ) " " "*" "*" " " " " " "*" " " " " " "
3  ( 3 ) " " "*" "*" " " " " " " " " "*" " " "
3  ( 4 ) " " "*" "*" " " " " " " " " " " "*" "
3  ( 5 ) "*" " " "*" " " "*" " " " " " " " " "
3  ( 6 ) " " " " "*" "*" " " "*" " " " " " " "
3  ( 7 ) " " " " "*" " " "*" " " " " " " " "*"
3  ( 8 ) " " " " "*" " " "*" " " " " "*" " " "
3  ( 9 ) " " " " "*" "*" " " " " " " " "*" " "
4  ( 1 ) " " "*" "*" "*" " " "*" " " " " " " "
4  ( 2 ) " " "*" "*" "*" " " " " " " " "*" " "
4  ( 3 ) " " "*" "*" "*" " " " " " " " " " "*"
4  ( 4 ) " " "*" " " "*" "*" "*" " " " " " " "
4  ( 5 ) " " "*" "*" " " " " " "*" " " "*" " "
4  ( 6 ) " " "*" "*" " " " " " "*" " " " " "*"
4  ( 7 ) " " "*" "*" "*" " " " " " " " " "*" "
4  ( 8 ) " " "*" "*" "*" " " " " " "*" " " " "
4  ( 9 ) " " "*" "*" "*" "*" " " " " " " " " "
5  ( 1 ) " " "*" "*" "*" " " "*" " " "*" " " "
5  ( 2 ) " " "*" "*" "*" " " "*" " " " " " "*"
5  ( 3 ) " " "*" "*" "*" " " " " " " "*" " "*"
5  ( 4 ) "*" "*" "*" "*" " " "*" " " " " " " "
5  ( 5 ) " " "*" "*" "*" " " "*" " " " " "*" "
5  ( 6 ) " " "*" "*" "*" " " "*" "*" " " " " "
5  ( 7 ) " " "*" "*" "*" "*" "*" " " " " " " "
5  ( 8 ) "*" "*" "*" "*" " " " " " " "*" " " "
5  ( 9 ) " " "*" "*" " " " " " "*" " " "*" " "*"
6  ( 1 ) " " "*" "*" "*" " " "*" " " "*" " "*"
6  ( 2 ) " " "*" " " "*" "*" "*" " " "*" " "*"
6  ( 3 ) " " "*" "*" "*" " " "*" " " "*" "*" " "
6  ( 4 ) "*" "*" "*" "*" " " "*" " " "*" " " " "
6  ( 5 ) " " "*" "*" "*" " " "*" "*" "*" " " " "
6  ( 6 ) " " "*" "*" "*" "*" "*" " " "*" " " " "
6  ( 7 ) " " "*" " " "*" "*" "*" "*" "*" " " " "
6  ( 8 ) " " "*" " " "*" "*" "*" " " "*" "*" " "
6  ( 9 ) " " "*" "*" "*" " " "*" " " " " "*" "*"
7  ( 1 ) "*" "*" "*" "*" " " "*" " " "*" " " "*"
7  ( 2 ) " " "*" "*" "*" " " "*" " " "*" "*" "*"
7  ( 3 ) " " "*" "*" "*" " " "*" "*" "*" " " "*"
7  ( 4 ) " " "*" "*" "*" "*" "*" " " "*" " " "*"
7  ( 5 ) "*" "*" "*" "*" "*" "*" " " "*" " " " "
7  ( 6 ) "*" "*" "*" "*" " " "*" "*" "*" " " " "
7  ( 7 ) "*" "*" "*" "*" "*" "*" " " " " " " "*"
```

```
print("aic")
which.min(AIC)
print("bic")
which.min(BIC)
```

```
[1] "aic"
1
1
[1] "bic"
1
1
```

```
Selection Algorithm: exhaustive
         x1  x2  x3  x4  x5  x6  x7  x8  x9  x10
1  ( 1 ) " " " " "*" " " " " " " " " " " " " "
1  ( 2 ) " " " " " " " " "*" " " " " " " " " "
1  ( 3 ) " " " " " " " " " " " " " "*" " " " "
1  ( 4 ) " " " " " " " " " " " " " " " "*" " "
1  ( 5 ) "*" " " " " " " " " " " " " " " " " "
```

When the model predictor size is 1, both AIC and BIC picked model 1. Model 1 best selection is x3 (x^3) as we build y off of x^3 as that is the highest curve.

d).

```
Selection Algorithm: exhaustive
         x1  x2  x3  x4  x5  x6  x7  x8  x9  x10
1  ( 1 ) " " " " "*" " " " " " " " " " " " " "
1  ( 2 ) " " " " " " " " "*" " " " " " " " " "
1  ( 3 ) " " " " " " " " " " " " " "*" " " " "
1  ( 4 ) " " " " " " " " " " " " " " " "*" " "
1  ( 5 ) "*" " " " " " " " " " " " " " " " " "
```

```
[8] "obj"
       1        1        1        1        1        1        1        1
1775.716 1776.837 1777.978 1778.881 1779.944 1785.360 1785.682 1785.978
       1        2        2        2        2        2        2        2
```

Based on the BIC and Adjusted R^2, we can see that Mode1 is the best selection. Model 1 best selection is x3 (x^3) as we build y off of x^3 as that is the highest curve.

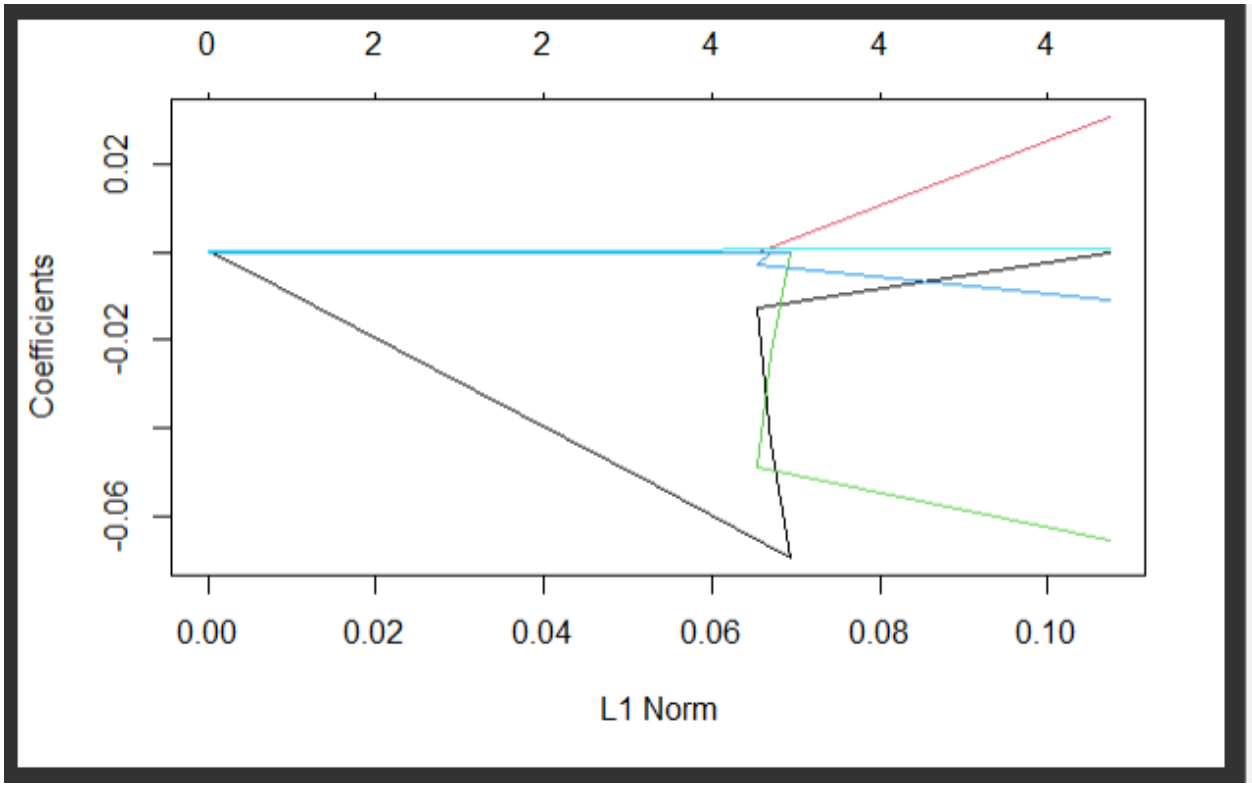
e).

```
[1] "forward selection"
[1] "bic"
1
1
[1] "R^2"
[1] 72
```

```
[1] "backward selection"
[1] "bic"
1
1
[1] "R^2"
[1] 72
```

Based on the result for both forward and backward selection, the result is identical to what we got in d.

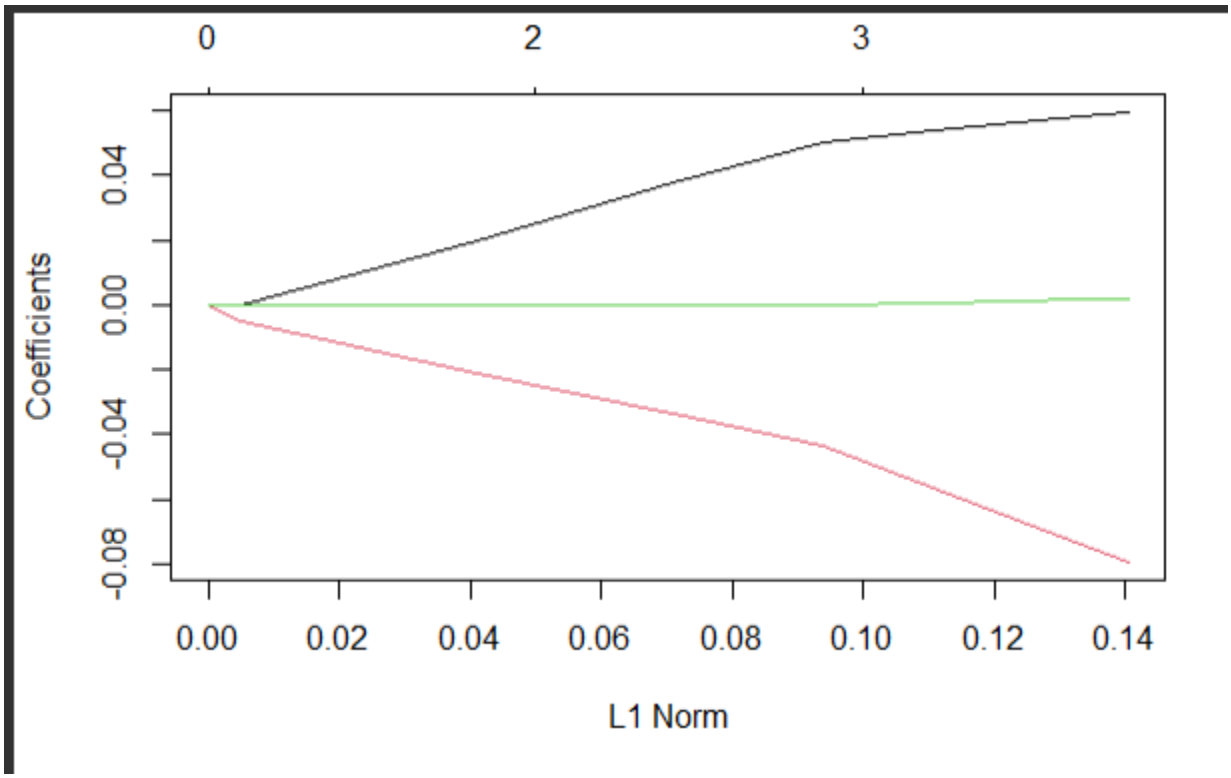
f).



Based on our model, we can see that our coefficients start at 0 (when lambda is 0) and transform into a range of -0.07 to 0.01 (as lambda increases). We can see that as lambda increases, the coefficient values become more scattered and unpredictable.

g).

```
[1] "------------"
[1] "aic"
1
1
[1] "bic"
1
1
[1] "R^2"
[1] 72
```



When the model predictor size is 1, both AIC and BIC picked model 1. Model 1 best selection is x7 (x^7) as we build y off of x^7 (that is the highest curve).

Based on our model, we can see that our coefficients start at 0 (when lambda is 0) and transform into a range of -0.09 to 0.05 (as lambda increases). We can see that as lambda increases, the coefficient values become more scattered and unpredictable.

Problem 3: Regularized Regression Models

a).

b).

```
20 x 1 sparse Matrix of class "dgCMatrix"
                      s0
(Intercept)  161.99259644
AtBat         -1.94797003
Hits           7.34315842
HmRun          4.15002285
Runs          -2.24162013
RBI           -0.99806148
Walks          6.16995995
Years         -2.95865910
CAtBat        -0.18673104
CHits          0.21580517
CHmRun        -0.04505545
CRuns          1.39868466
CRBI           0.75844959
CWalks        -0.79025591
LeagueN       63.50330429
DivisionW   -116.80326009
PutOuts        0.28138459
Assists        0.37621098
Errors        -3.42003333
NewLeagueN   -25.68140501
  (Intercept)         AtBat          Hits         HmRun          Runs
 161.99259644   -1.94797003    7.34315842    4.15002285   -2.24162013
          RBI         Walks         Years        CAtBat         CHits
  -0.99806148    6.16995995   -2.95865910   -0.18673104    0.21580517
       CHmRun         CRuns          CRBI        CWalks       LeagueN
  -0.04505545    1.39868466    0.75844959   -0.79025591   63.50330429
     DivisionW       PutOuts       Assists        Errors    NewLeagueN
-116.80326009    0.28138459    0.37621098   -3.42003333  -25.68140501
```

c).

```
20 x 1 sparse Matrix of class "dgCMatrix"
                      s0
(Intercept)  5.359257e+02
AtBat        5.443467e-08
Hits         1.974589e-07
HmRun        7.956523e-07
Runs         3.339178e-07
RBI          3.527222e-07
Walks        4.151323e-07
Years        1.697711e-06
CAtBat       4.673743e-09
CHits        1.720071e-08
CHmRun       1.297171e-07
CRuns        3.450846e-08
CRBI         3.561348e-08
CWalks       3.767877e-08
LeagueN     -5.800263e-07
DivisionW   -7.807263e-06
PutOuts      2.180288e-08
Assists      3.561198e-09
Errors      -1.660460e-08
NewLeagueN  -1.152288e-07
  (Intercept)         AtBat          Hits         HmRun          Runs
 5.359257e+02  5.443467e-08  1.974589e-07  7.956523e-07  3.339178e-07
          RBI         Walks         Years        CAtBat         CHits
 3.527222e-07  4.151323e-07  1.697711e-06  4.673743e-09  1.720071e-08
       CHmRun         CRuns          CRBI        CWalks       LeagueN
 1.297171e-07  3.450846e-08  3.561348e-08  3.767877e-08 -5.800263e-07
    DivisionW       PutOuts       Assists        Errors    NewLeagueN
-7.807263e-06  2.180288e-08  3.561198e-09 -1.660460e-08 -1.152288e-07
```

The coefficient values are much smaller with the larger lambda compared to the larger coefficient values with the smaller lambda value.

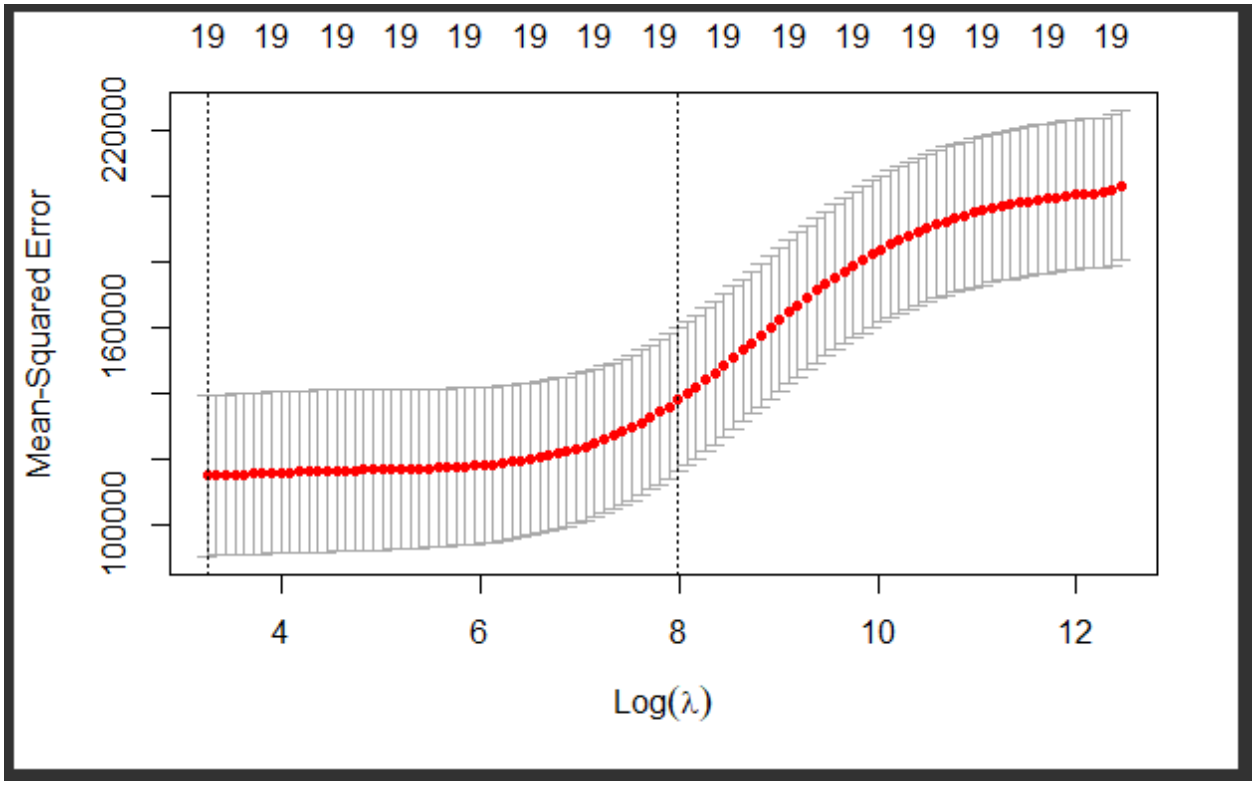
d). As lambda values increase, the coefficient values will become smaller and eventually reach 0 when lambda is infinity.

e).

```
100 19 51.64     26
[1] "lambda = 0.013"
Warning in ridgePredicted - Y :
  longer object length is not a multiple of shorter object length
[1] 493764.7
[1] "lambda = 10^10"
Warning in ridgePredicted - Y :
  longer object length is not a multiple of shorter object length
[1] 476857.6
```

I expect the larger lambda to produce a smaller value compared to the larger value on the smaller lambda. The reason is as lambda gets larger, the coefficient, MSE, and other factors will get smaller.

f).



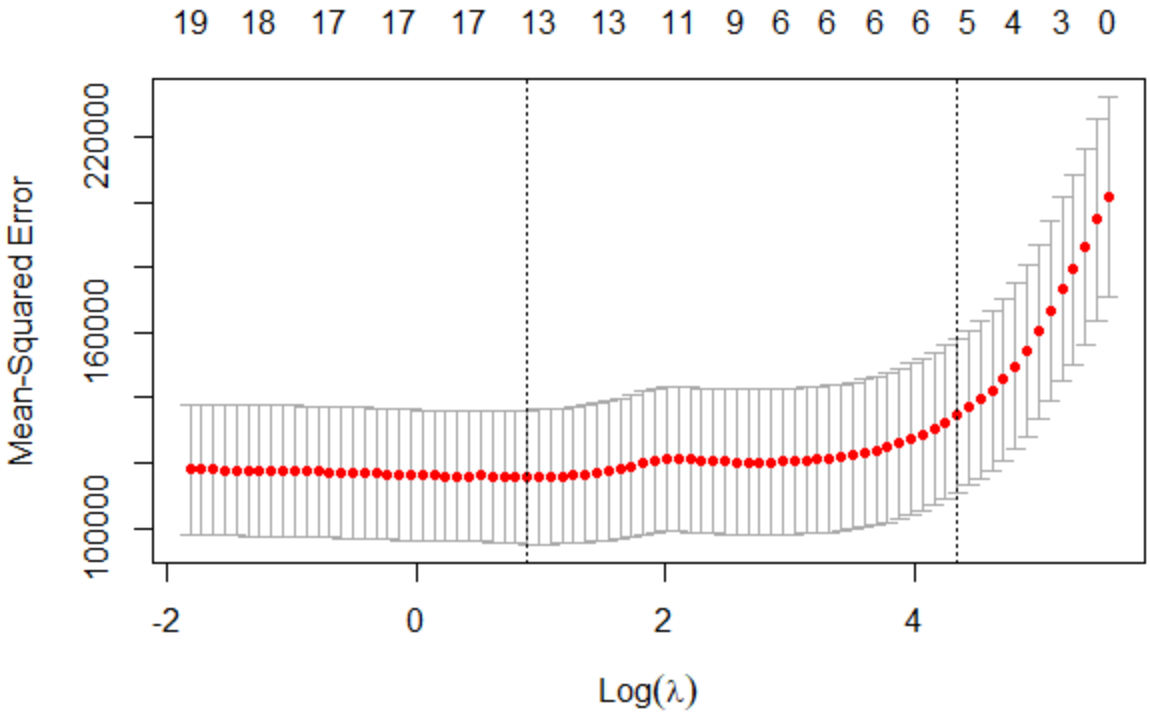

```
      Lambda Index Measure    SE Nonzero
min   25.5    100  115039 24392      19
1se 2935.1     49  138098 21865      19
20 x 1 sparse Matrix of class "dgCMatrix"
```

Lambda min = 25.5

g).

```
[1] 2674.375
```

h).




```
      Lambda Index Measure    SE Nonzero
min    2.67     50  118589 14873      13
1se   69.40     15  132727 15963       6
20 x 1 sparse Matrix of class "dgCMatrix"
                      s1
(Intercept) 127.95694754
```

i).

```
[1] "lambda = 2.44"
Warning in ridgePredicted
  longer object length is
[1] 493764.7
[1] "lambda = 2935.1"
Warning in ridgePredicted
  longer object length is
[1] 481388
[1] "lambda = 2.67"
Warning in ridgePredicted
  longer object length is
[1] 491074.9
[1] "lambda = 69.40"
Warning in ridgePredicted
  longer object length is
[1] 484358.5
```

Ridge min : 1st
Ridge 1se : 2nd

Lasso min : 3rd
Lasso 1se : 4th

Ridge regression is the best as it has the largest min mse.

j).

Ridge has the smallest min intercept of 8.11 compared to lasso's 123.85 min intercept.

k).

Focus on career hits and home run.

Problem 4: Comparing Predictive Models

a).

b).

c). This makes sure that large values are not treated on the same level as a small value.

d).

```
Measure: Mean-Squared Error

    Lambda Index Measure      SE Nonzero
min  376.5   100 1802649  923550      17
1se 2009.1    82 2714057 1505433      17
18 x 1 sparse Matrix of class "dgCMatrix"
                    s1
```

376.5

e).

```
  longer object length i
[1] 917142.5
```

Yes it is a major improvement compare to part be because it has a larger mse than part b.

f).

```
     Lambda Index Measure     SE Nonzero
min  16.71    58 1403665 314317      13
1se 272.38    28 1700017 482789       5
18 x 1 sparse Matrix of class "dgCMatrix"
                        s1
(Intercept) -554.34776536
```

Mean

```
  longer object
[1] 23500032
```

Coefficient

```
                        s1
(Intercept) -554.34776536
PrivateYes     .
Accept         1.15120042
Enroll         0.39172423
Top10perc     26.64599503
Top25perc      .
F.Undergrad    0.02639519
P.Undergrad    .
Outstate       .
Room.Board     .
Books          .
Personal       .
PhD            .
Terminal       .
S.F.Ratio      .
perc.alumni    .
Expend         0.01891522
Grad.Rate      .
```

g).

Ridge model is the best because it has the largest mse value.

Ridge and lasso both have very high mse compared to the linear regression.