Advanced Regression – Assignment II

Question 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer 1

The optimal values for Lasso is 187.38

```
Out[107]: GridSearchCV(cv=5, estimator=Lasso(),
                       param_grid={'alpha': array([1.00000000e-05, 1.26185688e-05, 1.59228279e-05, 2.00923300e-05,
                 2.53536449e-05, 3.19926714e-05, 4.03701726e-05, 5.09413801e-05,
                 6.42807312e-05, 8.11130831e-05, 1.02353102e-04, 1.29154967e-04,
                 1.62975083e-04, 2.05651231e-04, 2.59502421e-04, 3.27454916e-04,
                 4.13201240e-04, 5.21400829e-04, 6.57933225e-04, 8....
                 1.20450354e+03, 1.51991108e+03, 1.91791026e+03, 2.42012826e+03,
                 3.05385551e+03, 3.85352859e+03, 4.86260158e+03, 6.13590727e+03,
                 7.74263683e+03, 9.77009957e+03, 1.23284674e+04, 1.55567614e+04,
                 1.96304065e+04, 2.47707636e+04, 3.12571585e+04, 3.94420606e+04,
                 4.97702356e+04, 6.28029144e+04, 7.92482898e+04, 1.00000000e+05])},
                       return_train_score=True, scoring='neg_mean_absolute_error',
                       verbose=1)

In [108]: model_cv.best_estimator_

Out[108]: Lasso(alpha=187.3817422860383)
```

The optimal value using Ridge is 117.68

```
Out[111]: GridSearchCV(cv=5, estimator=Ridge(alpha=100.0),
                       param_grid={'alpha': array([1.00000000e-05, 1.26185688e-05, 1.59228279e-05,
                 2.53536449e-05, 3.19926714e-05, 4.03701726e-05, 5.09413801e-05,
                 6.42807312e-05, 8.11130831e-05, 1.02353102e-04, 1.29154967e-04,
                 1.62975083e-04, 2.05651231e-04, 2.59502421e-04, 3.27454916e-04,
                 4.13201240e-04, 5.21400829e-04, 6.5793...
                 1.20450354e+03, 1.51991108e+03, 1.91791026e+03, 2.42012826e+03,
                 3.05385551e+03, 3.85352859e+03, 4.86260158e+03, 6.13590727e+03,
                 7.74263683e+03, 9.77009957e+03, 1.23284674e+04, 1.55567614e+04,
                 1.96304065e+04, 2.47707636e+04, 3.12571585e+04, 3.94420606e+04,
                 4.97702356e+04, 6.28029144e+04, 7.92482898e+04, 1.00000000e+05])},
                       return_train_score=True, scoring='neg_mean_absolute_error',
                       verbose=1)

In [112]: model_cv.best_estimator_

Out[112]: Ridge(alpha=117.68119524349991)
```

An increase in Alpha will lead to the decreased variance and increased bias.

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer 2

We will consider Ridge Regression as the R2 value for Train and Test data is better in case of Ridge.

Question 3

After building the model, you realized that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Answer 3

210 rows × 2 columns

```
In [125]: betas.nlargest(5, ['Ridge'])
```

Out[125]:

	Ridge	Lasso
36	14672.770328	15986.304880
10	11970.244427	3591.016480
37	11763.259396	12915.748977
109	10088.832321	10104.357367
28	9767.479859	11479.802954

```
In [132]: betas.nlargest(5, ['Lasso'])
```

Out[132]:

	Ridge	Lasso
36	14672.770328	15986.304880
8	6973.848585	14396.839714
7	7826.031620	14202.119220
37	11763.259396	12915.748977
28	9767.479859	11479.802954

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer 4

To make sure that model is robust and generalizable, we need to check that there is balance between variance and bias. This helps ensure that we do not underfit or overfit the model.

Regularization techniques like Ridge and Lasso are used to create a good balance between the two.