Assignment Part-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?

Ans. Optimal Value of alpha for ridge regression is 0.7 and for Lasso is 0.0001. If we choose double the value of alpha for ridge regression we observed coefficients value decreases and for lasso some more coefficients turns to 0 which are less important. Top 5 most important predictors are as below.

For Ridge

OverallQual	0.481
OverallCond	0.297
TotalBsmtSF	0.480
2ndFlrSF	0.224
TotRmsAbvGrd	0.339
GarageArea	0.243
MSZoning_FV	0.217

For Lasso

OverallQual	0.561
OverallCond	0.329
TotalBsmtSF	0.610
TotRmsAbvGrd	0.343
GarageArea	0.261

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?

Ans. Both ridge and Lasso model train and test accuracy score is quite similar so we choose Lasso over Ridge because with Lasso regression we can do feature selection as well.

Question 3

After building the model, you realised 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?

Ans. After dropping five most important predicators and again retrain the model we observed below are the five most important predictors.

For Ridge

GarageArea	0.321
2ndFlrSF	0.277
FullBath	0.273
BedroomAbvGr	0.269
LotArea	0.236

For Lasso

GarageArea	0.360
LotArea	0.323
2ndFlrSF	0.303
BedroomAbvGr	0.280
FullBath	0.275

Question 4

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

Ans. A Model is robust when any variation in data doesn't affect its performance. And a model is generalisable when it performs similar with new unseen data. Model should be as simple as possible. Means Bias and Variance should be as low as possible. To make sure a model is robust and generalizable, we have to take care it doesn't overfit. This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but fail to pick up the patterns in unseen test data. we look at it from the perspective of Accuracy, a too complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease. We have to find some balance between Bias and Variance and this can be achieved by Regularization Techniques like Ridge and Lasso.