**Advanced Linear Regression Subjective Questions**

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 - The optimal value for ridge and lasso regression are as follows

Ridge = 3.0

Lasso = 0.0001

If we double the alpha value for ridge and lasso the model starts underfitting, which is understandable because we know that on higher value of alpha the model starts underfitting.

We know if we increase the lambda/alpha we start generalising the model which good for simpler use case and in case of lasso as we increase the alpha value more and more variable coefficient reaches to zero by decreasing the overall r2 score

The most important predictor variables after the change are for Ridge

LotArea

MasVnrArea

BsmtFinSF1

BsmtUnfSF

2ndFlrSF

GrLivArea

GarageArea

WoodDeckSF

OpenPorchSF

The most important predictor variables after the change are for Lasso

BsmtFinSF1

GrLivArea

GarageArea

1. 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 –

During the development process according to the results received I will be using the Ridge regression because in case of ridge the test and train score differences is not more than 5% and what standard says that the difference between the train and test r2 score percent should not be greater than 5%.

also, we know that as the lambda value increases Lasso shrinks the coefficient towards zero and it make the variables exactly equal to 0. Lasso also does variable selection.

When lambda value is small it performs simple linear regression and as lambda value increases, shrinkage takes place and variables with 0 value are neglected by the model.

whereas in case of ridge as we increase the value of lambda the variance in model is dropped and bias remains constant.

The results I got is given below

-->So, for Ridge it comes

The train r2 = 0.93

The test r2 = 0.88

-->So, for Ridge it comes

The train r2 = 0.94

The test r2 = 0.84

We can clearly see the model fitting in case of Ridge is very impressive which would end up giving us good results.

1. 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 - The five most important variable that might get excluded is

BsmtUnfSF

GrLivArea

GarageArea

OverallCond

2ndFlrSF

1. 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 –

The model should be as simple as possible, though its accuracy will decrease but it will be more robust and generalisable. The simpler the model the more the bias but less variance and more generalizable. Its implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data i.e., the accuracy does not change much for training and test data.

It can be also understood using the Bias-Variance trade-off