Subjective Question & Answer

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- The optimal value of alpha for ridge is 6 and for lasso is 0.0003. And when we double these values, I saw the model performance remained same in both cases but the value of Mean Squared Error for ridge we observed that is slightly reduced from 0.01359 to 0.01358 also for lasso The Mean Squared Error is slightly reduced from 0.01345 to 0.01326.

AFTER CHANGES LASSO: The most important predictors are 'GrLivArea', 'MSZoning_RL', 'OverallQual', 'TotalBsmtSF', 'MSZoning_RM', 'OverallCond', 'MSZoning_FV', 'Foundation_PConc', 'GarageCars', 'BsmtFinSF1'

AFTER CHANGES RIDGE: The most important predictors are 'GrLivArea', 'MSZoning_RL', 'OverallQual', 'TotalBsmtSF', 'MSZoning_RM', 'MSZoning_FV', 'OverallCond', 'Foundation_PConc', 'GarageCars', 'BsmtFinSF1'

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- We will go with the lasso because as we know lasso helps in unwanted feature reduction and as we can clearly see the coefficient value of one of the lasso's features became 0 so Lasso has a better edge over Ridge. And also lasso have the low Mean Squared as compared to ridge.

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- TotalBsmtSF, FullBath, HalfBath, OverallCond, OverallCond this are the five most important predictor variables after creating new model by excluding the five most important predictor variables.

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- We will ensure that the model will not overfitting and as simple as possible

Simpler models are usually more 'generic' and are more widely applicable

Where as the complex model will overfit the data. Like small variations in the dataset create complex model to confused and it get overfitted.

But we are having underfitting problem with too simple model, Simple models have low variance, high bias and complex models have low bias, high variance so we have to minimise some variance by compromising some little amount of bias to get the best model. Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train. Simpler models are more robust. Complex models tend to change wildly with changes in the training data set. Regularization can be used to make the model simpler. Regularization helps to strike the delicate balance between keeping the model simple and not making it too naive to be of any use. For regression, regularization involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model. Also, Making a model simple lead to Bias-Variance Trade-off: