Assignment-II

Q1. 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. Optimum value of alpha:

1. Ridge: 18.33
2. Lasso: 0.0051

Changes after the doubling the alpha:

1. Coefficients of Ridge and Lasso have moved very slightly towards 0.
2. There is also very slight increase in train error which is degrading training accuracy signifying that model is compromising bias to handle variance. There is also very slight improvement in test accuracy.

Most Important Predictor: OverallQual for Ridge
GrLivArea for Lasso

Q2. 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. Optimal Lambda values:

1. Ridge: 18.33
2. Lasso: 0.0051

Grid Search K Cross Validation is used with shuffle to determine lambda for both Ridge and Lasso as part for hyper-parameter tuning. This technique provides best lambda from the given set of lambdas based on accuracy of validation set.

The Chosen Model is **Ridge**. Features selected from RFE model were supplied to Ridge.

Reason: The train and test scores of Ridge are very close to each other signifying that the model is pretty stable on the other hand there is considerable difference in accuracies for Lasso.

	Ridge	**Lasso**	**RFE**
R2 Test	0.799692	0.863763	0.798942
R2 Train	0.809491	0.926348	0.808706

Q3. 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 new 5 most important predictors after removing previous top 5 features for selected Ridge model are:

1. SaleType_New
2. GarageQual_TA
3. BsmtCond_TA
4. BsmtCond_Gd
5. Neighborhood_MeadowV

Q4. 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. Generally simple models are more robust and generalisable. To make model robust and generalisable:

1. Identification of class of problem being solved and figuring out the right algorithm to use is primary step.
2. Outlier detection and removal is necessary to make the model more generalisable.
3. Identification of important features and their right transformations.
4. Regularization to reduce variance.
5. Using validation techniques and Grid Search CV for hyper-parameter tuning.
6. Finding right balance of bias and variance.

Implication for Accuracy: Both training and test accuracy should be high and close to each other. The ideal state should be low bias and low variance. Bias is error made by model in prediction and variance is how stable is model to change in input features.

For a stable and robust model test and train accuracies should be close to each other. Higher train and low test accuracy is overfitting and Low train accuracy is certainly under-fitted model.