Assignment-based Subjective Questions

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?

- The optimal lambda value in case of Ridge and Lasso is as below:
  - Ridge - 1.0
  - Lasso - 0.0001
- And when the value of Alpha is doubled for Ridge regression, The value of alpha becomes 2.0. And the Mean squared Error remains same with a slight variation. Here, the model built with the alpha doubled will apply more penalty on the curve and make the model more generalized. Thus, the model becomes more simpler and fails to fit every data given in the dataset.
- Similarly, when the alpha value is doubled for Lasso regression, The value of alpha becomes 0.0002. And the Mean squared Error remains same with a slight variation. It also penalizes the model even more and a greater number of coefficients of the variable is shrunk towards Zero.
- Also, by increasing the value of Alpha, the r2 score will be decreased.
- The most important predictor variables after the value of alpha is doubled in the ridge regression are as follows,
  - MSZoning_FV
  - MSZoning_RL
  - 1stFlrSF
  - 2ndFlrSF
  - GarageType_BuiltIn
  - OverallQual

- GarageType_Attchd
- MSZoning_RM
- Foundation_PConc
- Neighborhood_NridgHt

- The most important predictor variables after the value of alpha is doubled in the lasso regression are as follows,

  - MSZoning_FV
  - MSZoning_RL
  - 1stFlrSF
  - MSZoning_RM
  - 2ndFlrSF
  - OverallQual
  - GarageType_BuiltIn
  - GarageType_Attchd
  - Neighborhood_Crawfor
  - Neighborhood_NridgHt

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?

The model built with Lasso regression is preferred than the ridge regression. Since, the lasso regression performs feature selection as when the lambda value increases, Lasso shrinks the coefficient of the variable towards zero and makes the variable exactly equal to zero and it is neglected by the model. Also, Lasso has the low Mean square value as compared to Ridge regression.

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?

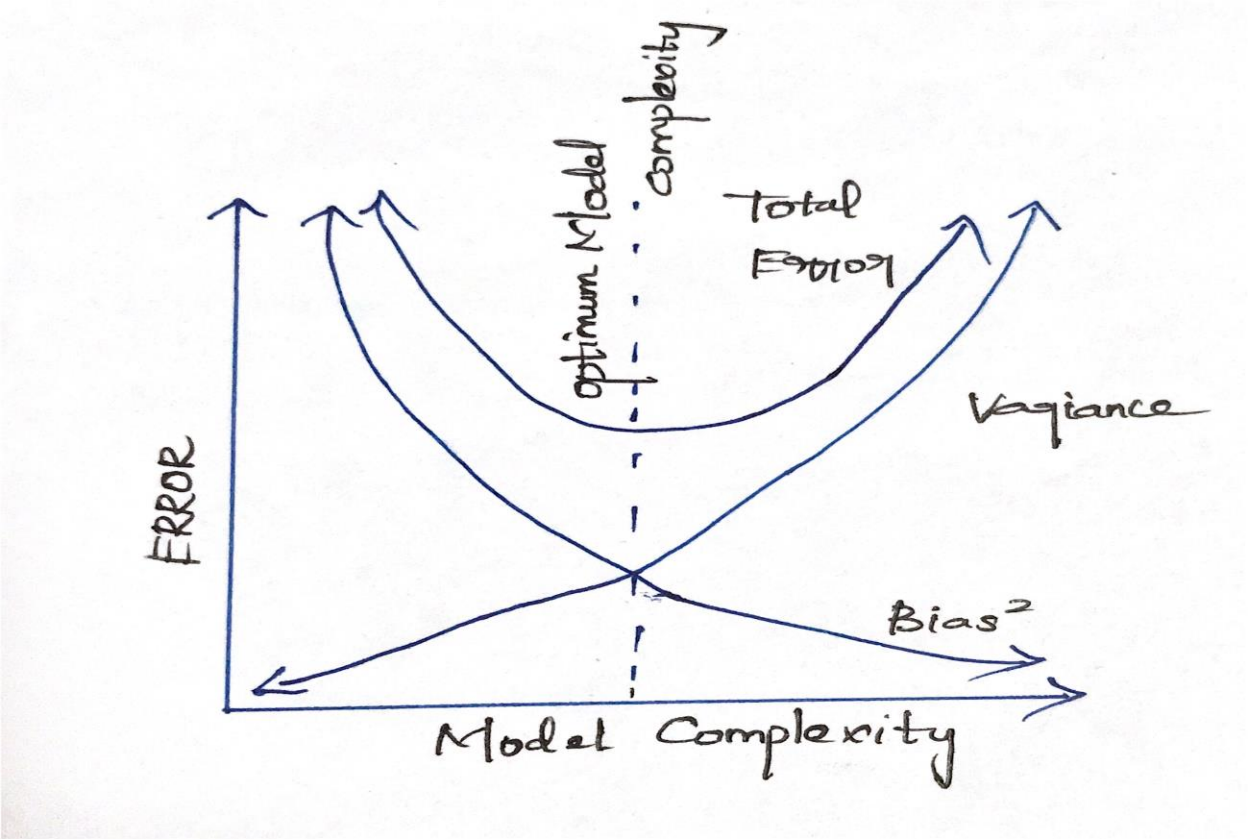
The important 5 predictor variables after the exclusion in the lasso model are,

- OverallQual
- Exterior1st_BrkFace
- Neighborhood_Crawfor
- Neighborhood_NridgHt
- GarageArea

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?

- The model is expected to be as simple as possible and simpler models are considered as more 'generic', though its accuracy will be decreased but it will be more robust.
- This can be understood from the Bias-Variance trade-off. The simpler the model the more the bias but less variance becoming generalizable. Whereas the complex model will have high variance and low bias.
- Sometimes underfitting and overfitting are the problems associated with the model. Hence, it is important to have balance in Bias and Variance to avoid such problems. This is possible with **"Regularization".**
- Regularization helps in managing the model complexity by essentially shrinking the coefficients towards zero. This avoids the model becoming too complex, thus reducing the risk of overfitting.
- Making a model simple lead to Bias-Variance trade off.



Variance has to be minimized by compromising some little amount of bias to get the best model.