Assignment -II

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 value for Lasso Regression is 10.0 and for Ridge regression is 20.0 as per the model.
   - The basic idea of generalization is to manage model complexity by essentially shrinking the model coefficients towards zero. when we double the value of alpha, it will apply more penalty on the curve and try to make the model more and more generalized
   - In the case of Ridge regression, an increase in the shrinkage penalty, pushes the model coefficients toward zero but they will never be zero
   - In the case of Lasso Regression, a high shrinkage penalty can cause the model coefficients to zero
   - But an increase in the value of alpha can lead to underfitting of the model
   - The most important variables after the change for Ridge Regression
     i. MSZoning_FV
     ii. MSZoning_RL
     iii. Neighborhood_Crawfor
     iv. MSZoning_RH
     v. MSZoning_RM
     vi. SaleCondition_Partial
     vii. Neighborhood_Stone Br
     viii. GrLivArea
     ix. SaleCondition_Normal
     x. Exterior1st_BrkFace
   - The most important variable after the changes has been implemented for lasso regression are as follows:
     i. GrLivArea
     ii. Overall Qualare
     iii. OverallCond
     iv. TotalBsmtSF
     v. BsmtFinSF1
     vi. GarageArea
     vii. Fireplaces
     viii. LotArea
     ix. LotArea
     x. LotFrontage

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?

- In Ridge Regression, the cost function is given by the formula:

Cost function, ~~for~~ OLS $= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

Penalty

Cost function for Ridge $\quad \text{Cost} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$

- If the shrinkage penalty =0, then there will be no shrinkage of model coefficients
- Higher the shrinkage penalty, the model coefficients will be pushed to zero
- But care should be taken not to underfit the model
- With Ridge regression, for each value of lambda , we get different set of model coefficients. Hence choosing an appropriate value of lambda becomes crucial.
- In Lasso Regression, the cost function is given by the formula:

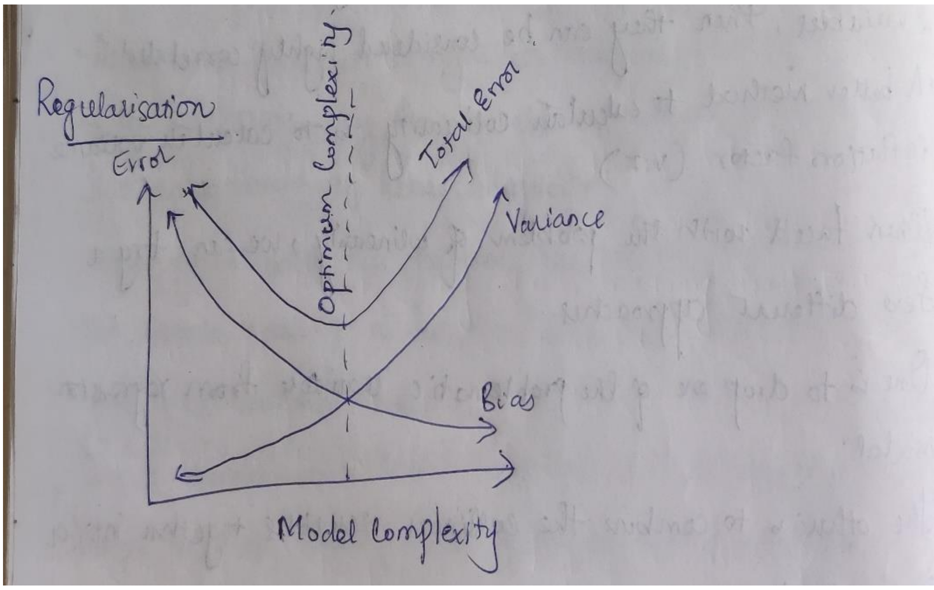
$$\text{Cost} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

- With the increase in the value of lambda, variance reduces with slight compromise in bias
- Lasso also pushes the model coefficients towards 0 in order to handle high variance, just like Ridge Regression.
- But in addition to this, Lasso also pushes some coefficients to be exactly zero and thus performs variable selection.
- This variable selection results in models that are easier to interpret.
- Hence, I would choose Lasso regression

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?
    - GrLivArea
    - OverallQual
    - OverallCond
    - TotalBsmtSF
    - GarageArea

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?
    - The model should be generalisable and robust. It should be able to perform well on the unseen data
    - Else it will lead to overfitting where model performs very well on trained data but fails miserably on test data
    - One way to avoid overfitting is regularization.

- This can be understood with the help of bias-variance trade off graph



- As we see in the graph, ass the model complexity increases, bias decreases but variance increases
- Ideally, we want to reduce both variance and bias, so that the total error on the model is reduced.
- Regularisation helps us in managing model complexity by essentially shrinking the model coefficient estimates towards zero
- This discourages the model from becoming too complex and thus avoiding the risk of overfitting
- When we use regularisation, we add penalty term to model's cost function
    i. Cost Function= RSS+Penalty
- Adding this penalty term in the cost function helps suppress or shrink the magnitude of the model coefficients towards zero
- This discourages the creation of more complex models
- When we add this penalty and try to get the model parameters that optimise the updated cost function(RSS+Penalty), the coefficients that we may get given the training data may not be the best (may be more biased)
- Although with this minor compromise in terms of bias, the variance of this model may see a marked reduction.
- Essentially, with regularisation, we compromise by allowing a little bias for a significant gain in variance