ASSIGNMENT PART-II - 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?

Answer:

- Optimal value of alpha through Ridge Regression is 0.5
- Optimal value of alpha through Lasso Regression is 0.0002

On doubling the value of alpha below changes were observed-

	RIDGE		LASSO	
	alpha = 0.5	alpha = 1.0	alpha = 0.0002	alpha = 0.0004
Mean Squared Error	0.0187	0.0189	0.0190	0.0191
R- squared Train	0.9180	0.9170	0.9163	0.9131
R- squared Test	0.8573	0.8561	0.8539	0.8547

Observation:

- The general trend seen after doubling alpha for both Ridge and Lasso is that the mean squared error terms have increased.
- And the R-squared on the training dataset is decreased.
- Also, there is a small reduction in the beta coefficients

Reason: This may be due to the fact that regression algorithms trade off a little bias to see a significant reduction in variance. So, when we increase the alpha(lambda) value, we can see an increase in bias and the model turns to become simpler and hence, we see a little reduction in the model performance. On increasing alpha, Ridge will shrink the beta coefficients more and Lasso will reduce a greater number of predictors by pushing their coefficients to absolute zero.

Most Important Predictors after the change:

Ridge	Lasso
GrLivArea	GrLivArea
OverallQual	OverallQual
LotArea	OverallCond
OverallCond	LotArea
MSZoning_FV	GarageCars
MSZoning_RL	1stFlrSF
GarageCars	BsmtQual
1stFlrSF	SaleType_New
GarageQual	GarageQual
BsmtQual	BsmtFinSF1

<u>Question 2:</u>

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?

<u>Answer</u>:

- Performance wise, both Ridge and Lasso models are at par with each other. However, I will prefer Lasso as it tends to push beta coefficients to absolute zero and helps in feature reduction. However, Ridge can only shrink the beta coefficients towards zero but never achieves it and so we end up having lot many predictor variables even if they are not relevant.
- Moreover, interpretation of model is better in Lasso as it helps in feature selection.
- In this particular dataset, there are a lot of predictor variables and so it makes sense to do feature selection which is provided by Lasso.

<u>Question 3:</u>

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?

<u>Answer</u>:

Variables that I excluded are-

- 'GrLivArea'
- 'OverallQual'
- 'LotArea'
- 'OverallCond'
- 'GarageCars'

The five most important predictor variables after dropping above columns and rebuilding Lasso model are-

- **'1stFlrSF' :** First Floor square feet
- **'FullBath' :** Full bathrooms above grade
- **'GarageQual' :** Garage quality
- **'LotFrontage' :** Linear feet of street connected to property
- **'BsmtQual' :** Evaluates the height of the basement

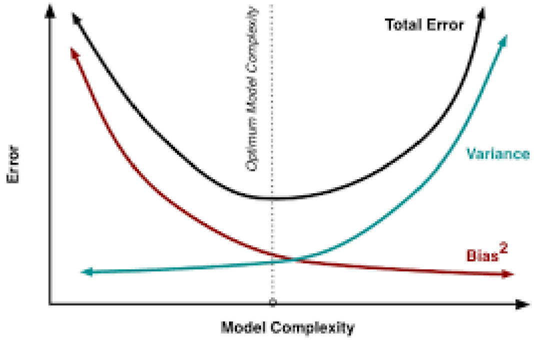
Question 4:

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?

Answer:

To make a model robust and generalisable we must first solve the problem of bias-variance trade off in linear regression using regularisation. There are certain regularization techniques like Ridge, Lasso and Elastic Net which can be used to achieve this. These basically reduce the variance in the model by allowing some bias.

In Linear Regression, we obtain our Beta-coefficients by using the OLS(Ordinary Least Square) method where we try to minimize sum of the squares of the residuals. Now, when we fit our model on the training dataset, the model tries to learn almost the entire dataset and tends to overfit. However, when we apply regularisation techniques, it adds a penalty term which tries to keep the model simple.



If we run OLS, without regularisation then our model will be complex i.e., low bias and high variance and will be more towards the right side of above graph.

In case of Ridge, the penalty term is highlighted in the below image:

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

In case of Lasso, the penalty term is highlighted in the below image:

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

For both Ridge and Lasso, the choice of lambda becomes utmost necessary because if we keep lambda too high then model will become too naïve and would underfit. On the other hand, if we keep lambda small then model will become complex and overfit.

To get the best lambda, we use Grid Search Cross Validation and look at the mean train and test scores of different models created for different lambda. The best performing model has the highest accuracy while maintaining minimum difference between train and test scores.