Assignment Part-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?

- Optimal value of lambda for Ridge Regression = 10

  Optimal valueof lambda for Lasso Regression = 0.0001

- If you choose double the value of alpha
  1. In case of Lasso Regression R2 value is decreasing and coefficients for more less important features(around 20 features) turning into zero .
  2. In case of Ridge Regression R2 value is decrasing and coefficients are moving towards zero.

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?

- R2 for Ridge Regression:
  1. Training R-Square = 0.86
  2. Testing R-Square = 0.86
- R2 for Lasso Regression:
  1. Training R-Square = 0.87
  2. Testing R-Square = 0.86

From above score we can see that R-square for Lasso Regression is higher that R-square for Ridge regression, So we can go with Lasso Regression as prediction for Lasso are 1% greater that ridge 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?

- Five most important predictor variables in lasso regression are SalesCondition_Partial, SalesCondion_Normal, SalesCondition_Family, SalesCondition_Alloca and SaleCondition_AdjLand

```
('GarageFinish_RFn', 0.03),
('GarageFinish_Unf', 0.03),
('SaleType_CWD', 0.03),
('SaleType_Con', 0.04),
('SaleType_ConLD', 0.04),
('SaleType_ConLI', 0.04),
('SaleType_ConLw', 0.04),
('SaleType_New', 0.04),
('SaleType_Oth', 0.05),
('SaleType_WD', 0.06),
('SaleCondition_AdjLand', 0.06),
('SaleCondition_Alloca', 0.07),
('SaleCondition_Family', 0.07),
('SaleCondition_Normal', 0.13),
('SaleCondition_Partial', 0.28)]
```

- After excluding most important variables we found that below predictors are most important that is Sale

  ('SaleType_ConLD', 0.06),

  ('SaleType_ConLI', 0.06),

  ('SaleType_ConLw', 0.07),

  ('SaleType_New', 0.07),

  ('SaleType_Oth', 0.13),

  ('SaleType_WD', 0.28)]

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?

- Any model will be robust and generalisable when it is not overfitted it means there is no much difference in r-square value of training and testing dataset.
- We can avoid overfitting by using Ridge and Lasso Regularization
- Also in reverse it should not be underfitted that means r-square for training and testing data is too low.
- We can avoid underfitting by increasing number of features and complexity of model
- If your model is complex and having high number of features, then model will have high accuracy. So, to make our model robust and generalisable, we will have to decrease variance which will lead to some bias. And addition of bias will decrease accuracy of model.
- In general, we have to balance between model accuracy and complexity. This can be achieved by Regularization techniques like Ridge and Lasso Regression.