Student Name: Mohit Patil **Email-ID:** mohitz4418@gmail.com

Assignment Based Subjective Questions.

(Advanced Regression – Surprise Housing)

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 changes is implemented?

Answer –

The optimal value of alpha for Ridge Regression Model is – 0.2

And,

The optimal value of alpha for Lasso Regression Model is – 0.001

If we choose to double the value of alpha for both ridge and lasso then –

1. The values of feature coefficient are slightly decreased.
2. Due to that some features changes their position in top 10 features based on coefficient. And some features changed.
3. The value of alpha is not optimal for model so the coefficient of feature decreases.
4. In lasso regression model the R-Squared value of train and test data is decreased due to doubling the alpha value.
5. The bias of model would be high
6. In lasso some of the coefficient value become zero, but in case of ridge, the coefficient become close to zero but not zero.

After we double the value of alpha in top features in Ridge Regression Model are: -

1. OverallQual
2. MSZoning_RM
3. MSZoning_RL
4. MSZoning_RH
5. LotConfig_FR3
6. SaleCondition_Alloca
7. Functional_Maj2
8. SaleType_Oth
9. GarageCars
10. Exterior2nd_Stucco

After we double the value of alpha in top features in Lasso Regression Model are: -

1. OverallQual
2. GarageCars
3. OverallCond
4. BsmtFinSF1
5. House_Age
6. SaleType_Oth
7. Condition1_PosA
8. Neighborhood_OldTown
9. SaleCondition_Partial
10. Exterior2nd_Stucco

After we doubled the value of alpha in both models most important predictor variables are: -

1. OverallQual
2. MSZoning
3. LotConfig
4. SaleCondition
5. Functional
6. SaleType
7. GarageCars
8. BSmtFinSF1
9. House_Age
10. Condition1

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?

Answer –

1. In both ridge and lasso model some important predictive features are same.
2. But from both model the R-Squared score for Train data is higher in Ridge regression and R-Squared score for Test data is higher in Lasso regression model.
3. Lasso regression model is performing well on unseen data. So, we will be using the Lasso Regression Model
4. And also, the Lasso is also helpful for feature selection.
5. The optimal value of lasso model is smaller than ridge regression model. So, we will choose the lasso to apply.

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 variables now?

Answer -

After we create the model after excluding the top five features for lasso regression model. We got the below features as most important predictor variables in the lasso model.

1. BsmtFinSF1
2. House_Age
3. Neighborhood_OldTown
4. SaleCondition_Alloca
5. Condition1_Norm

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?

Answer –

1. A model needs to be made robust and generalizable so that they are not impacted by outliers in the training data.

2. The model should also be generalizable so that the test accuracy is not lesser than the training score.

3. The model should be accurate for datasets other than the ones which were used during training.

4. A model is considered to be robust if the model is stable, i.e. It does not change drastically upon changing the training set.

5. Too much weightage should not be given to the outliers so that the accuracy predicted by the model is high. To ensure that this is not the case, the outliers analysis needs to be done and only those which are relevant to the dataset need to be retained.

6. Those outliers which it does not make sense to keep must be removed from the dataset. This would help increase the accuracy of the predictions made by the model.

7. If the model is not robust, it cannot be trusted for predictive analysis.

8. The model is considered generalizable if it does not overfits the training data, and works well with new data.

9. Its implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data i.e., the accuracy does not change much for training and test data.

10 Here are some changes we can make in model to make it robust and more generalizable:

a) Use a model that's resistant to outliers

b) Use a more robust error metric.

Here are some changes we can make to our data:

a) Transform your data and

b) Remove the outliers.