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:

The optimal values of lambda are:

- Ridge: 10
- Lasso: 100

The r2 score for the optimal values of lambda comes out to be as:

	R2 Score Train	R2 Score Test
Ridge (alpha=10)	0.9102708640356322	0.8291066996854972
Lasso (alpha=100)	0.9127400949466069	0.8226115465655931

When the values of lambda are doubled:

	R2 Score Train	R2 Score Test
Ridge (alpha=20)	0.8992839099644752	0.8289147747773998
Lasso (alpha=200)	0.9001663484100466	0.8255244923413052

When the value of lambda is doubled, there is a slight difference in the r2 scores for both ridge and lasso.

The important variables after the value of alpha is doubled are:

- OverallQual
- RoofMatl
- Neighborhood
- SaleCondition
- GarageCars
- Functional

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:

The optimal values of lambda for ridge is 10 and for lasso is 100. When building models based on these values the there isn't a significant difference in the performance of both models. The r2_score for the test data shows no significant difference.

However as when you look at the model parameters, Ridge does not make the coefficients zero, while on the other hand Lasso does make the coefficients of quite a few variables zero, thus helping in feature selection. Hence it would be better to use Lasso regression with lambda set to 100.

Question-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?

Answer:

After the top 5 features are dropped the next most important predictors become MSSubClass, OverallQual, KirchenQual, BsmtQual and LotShape.

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:

A model can be considered robust and generalizable if it is shows no drastic change in performance when the training set is changed, i.e. the model should not overfit on the training data and should be able to handle new/unseen data properly.

When it comes to accuracy, a model which is robust and generalizable should perform equally well on both the training and test data.