Advance Regression Subjective Questions & Answers

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 to 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 lambda for Ridge Regression = 9.0
Optimal value of lambda for Lasso = 0.001

The changes in the model if we choose to double in the value of alpha for both ridge and lasso:

- **Changes in Ridge Regression metrics:**

    1. R2 score of train set decreased from 0.94 to 0.93
    2. R2 score of test set remained same at 0.93

- **Changes in Lasso metrics:**

    1. R2 score of train set decreased from 0.92 to 0.91
    2. R2 score of test set decreased from 0.92 to 0.91

In case of ridge that will lower the coefficients and in case of Lasso there would be more less important features coefficients turning 0.

The most important predictor variables after the change is:

GrLivArea 1.09
OverallQual_Very Good 1.08
OverallQual_Excellent 1.07
Neighborhood_Crawfor 1.07
Functional_Typ 1.07
Exterior1st_BrkFace 1.06
OverallCond_Excellent 1.05
TotalBsmtSF 1.05
CentralAir_Y 1.05
SaleCondition_Alloca 1.04

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 model we will choose to apply will depend on the use case.
- If we have too many variables and one of our primary goal is feature selection, then we will use **Lasso**.
- If we don't want to get too large coefficients and reduction of coefficient magnitude is one of our prime goals, then we will use **Ridge Regression**.
- In Our Case, As we Got good score for both the models so we can go with Lasso regression as it results in model parameters such that lesser important features coefficients become zero.

  Ridge: Train :93.9 Test :92.7.4 and Lasso : Train :92.0 Test :92.0

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 predictor variables now?

Answer: The following are the top most important predictors after dropping five most important predictor variables. ..

```
2ndFlrSF                          0.10
Functional_Typ                    0.08
1stFlrSF                          0.08
MSSubClass_2-STORY 1945 & OLDER   0.07
Neighborhood_Somerst              0.07
```

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:

- When a model's performance is mostly unaffected by variations in the data, it is considered robust.
- When fresh data is taken from the same distribution as the model's original data set, it can be appropriately adjusted to build a generalizable model.
- We must ensure that a model does not overfit for it to be reliable and broadly applicable. This is because to the extremely high variance of an overfitting model, which significantly impacts the model's prediction even with modest changes in the data. A model of this kind will detect every pattern in training data, but it won't recognize patterns in test data that hasn't been seen.
- Put otherwise, for a model to be reliable and applicable, it shouldn't be overly intricate.
- When considering accuracy, a model that is overly complex will have an extremely high accuracy. Therefore, we will need to reduce variance, which will introduce bias, to strengthen and broaden the applicability of our model. Increased bias implies a decline in accuracy.
- We need to find a way to balance the complexity and accuracy of the models. Regularization strategies like Lasso and Ridge Regression can help achieve this.