Assignment: Advanced Regression 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 value for alpha chosen for **Ridge Regression is 10**, and for **Lasso Regression is 0.01**.
Higher the alpha, most of the feature coefficients would become zero.
However if we double the value of alpha in this case, the mean squared error would increase, and the most important predictor variables would be **changed**.

Important 5 Predictor Variables
LASSO Regression
Before Doubling:
['OverallQual', 'OverallCond', 'MSZoning_RL', 'MSZoning_RM', 'CentralAir_Y']
After Doubling:
['OverallQual', 'OverallCond', 'GrLivArea', 'CentralAir_Y', 'GarageType_Attchd']

RIDGE Regression
Before Doubling:
['OverallQual', 'MSZoning_RL', 'MSZoning_RM', 'RoofMatl_CompShg', 'RoofMatl_Tar&Grv']
After Doubling:
['OverallQual', 'OverallCond', 'GrLivArea', 'MSZoning_RL', 'RoofMatl_CompShg']

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 value for lambda chosen for **Ridge Regression is 10**, and for **Lasso Regression is 0.01**.
I would like to choose Lasso Regression rather than Ridge Regression as there is **feature selection** which was **done by the Lasso Regression.**

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 excluding these top 5 most important predictor variables:
['OverallQual', 'OverallCond', 'MSZoning_RL', 'MSZoning_RM', 'CentralAir_Y']

We get the below result as the 5 most important predictor variables:
Result: ['BsmtCond', '1stFlrSF', 'BedroomAbvGr', 'GarageType_Attchd', 'SaleCondition_Normal']

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 the model more robust and generalisable we would have to deal with bias and variance of the model and make sure the model does not underfit, that is the model should not have high bias and low variance and also have to make sure that the model does not overfit that is low bias and high variance.
The model should be right in the middle of both the cases that is having low bias and low variance.
Then we can call the model a very robust and generalised model.
Accuracy may be comparatively high in case we go for the generalized model on the test data.
Although, if we try to overfit the data then training accuracy would be very good, but then would affect the test accuracy.