Part - 2

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?

The optimum value of alpha
Ridge regression = 7
Lasso regression =120

Ridge regression:

Accuracy before changing alpha for ridge
traindata
R_2score=0.865
test data
R_2score=0.852

Accuracy after changing the alpha for ridge
traindata
R_2score = 0.852
testdata
R_2score = 0.844

Lasso regression:

Accuracy before the changing the alpha
traindata
R_2score = 0.864
testdata
R_2score = 0.854

Accuracy after changing the lambda
traindata
R_2score = 0.853
testdata

R_2score = 0.843
Clearly in both cases accuracy decreases

After changing the model the most important predictors variables are
GrLivArea
OverallQual
Neighborhood_NoRidge
Neighborhood_NridgHt
GarageCars

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?

Ridge regression:

Accuracy before changing alpha for ridge
traindata
R_2score=0.865
test data
R_2score=0.852

$L = \sum(\hat{Y}i - Yi)2 + \lambda \sum \beta^2$

Lasso regression:

Accuracy before the changing the alpha
traindata
R_2score = 0.864
testdata
R_2score = 0.854

$L = \sum(\hat{Y}i - Yi)2 + \lambda \sum |\beta|$

Almost both ridge and lasso have same accuracy.In ridge regression, it can't make coefficients to zero. Penalty term is square of magnitude of co efficients. In lasso regression, it will perform parameter shrinkage to zero and automatic selection features.Penalty term is absolute value of co efficients. Less complex model is more robust. So, in this case lasso preferred over ridge.

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?

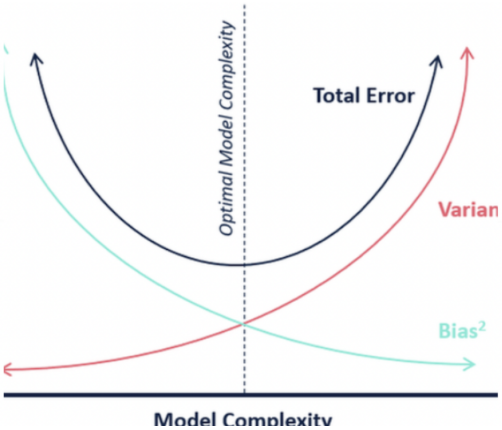
After removing the top five predictors and build the model and next top five predictors are
Next five important predictors after removing the top five

1stFlrSF
2ndFlrSF
GarageArea
BsmtQual
ExterQual

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?

Model should be as simple as possible. Simpler models are more generic and widely accepted.
Even Though simple models makes more training training error, it performs well in the test set.
Simple model is easy to train.
Complex models tend to perform poor when small changes in a data set.
A model can suffer from two things
High Bias
High Variance

Bias:Inability to find relation between data points. The model is weak to learn from data when there is a high bias. If a model has a high bias, it leads underfitting.
Variance:error that occurs due to model sensitivity to small fluctuation in the training data. When there high variance model perform well in train data but poor on test data. If model has high variance it leads to overfitting.
Regularization helps to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting
The accuracy of the model can be maintained by keeping the balance between bias and variance as it minimizes the total error.


Optimal Model Complexity
Total Error
Varian
Bias²
Model Complexity