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:
For Lasso regression :-Curve plot between negative mean absolute error and alpha we see that model try to penalise more and reduces most of the coefficients to zero. As per our curve plot we see that negative mean absolute error is quite low at alpha = 0.3 and stabilises thereafter, but I have chose **optimal value of alpha to be 0.01** to balance the trade-off between Bias-Variance and to get the coefficients of smallest features.

For Ridge regression :-From the curve plot we see that as the value of alpha increases from 0 the error terms decreases ,also the train error shows increasing trend when alpha increase and the test error is minimum when the value of alpha is 2,So I have taken the **optimal value of alpha to be 2** for ridge regression.

If we double the value of alpha for lasso regression it will penalise our model more and more coefficients will be reduced to zero and r2 square value will also decrease.

Similarly if we double alpha value for ridge regression it will try to make the model more generalised and simpler.It will try to fit every data of the dataset and we will get more error for train and test.

The most important predictor variables after change is implemented are as below:
For Lasso regression:

1.MSZoning_RL
2.MSZoning_FV
3.GrLivArea
4.MSZoning_RH
5.MSZoning_RM
6.SaleCondition_Normal
7.Neighborhood_Crawfor
8.SaleCondition_Partial
9.Exterior1st_BrkFace
10.Neighborhood_StoneBr

For Ridge regression:

1.OverallQual
2.GrLivArea
3.LotArea
4.LotFrontage
5.OverallCond

6.TotalBsmtSF
7.BsmtFinSF1
8.GarageArea
9.Fireplaces
10.LotShape

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:
While using any regression technique it is important to regularize coefficients and improve the prediction accuracy with low variance in order to make the model more robust and generalizable.

If we focus on the features provided by ridge and lasso regression technique:

Ridge regression: uses a tuning parameter called lambda as the penalty is square of magnitude of coefficients which is identified by cross validation.Residual sum of squares should be small by using the penalty.The penalty is lambda times sum of squares of coefficients hence coefficients that have higher value get penalized first.As the value of lambda is increased the variance in the model is decreased bias remains constant.

Lasso regression: uses a tuning parameter called lambda as the penalty is absolute value of magnitude of coefficients which is identified by cross validation.As the lambda value increases Lasso shrinks the coefficients towards zero and it reduces the variables exactly to zero .It also does variable selection.when lambda value is small it performs simple linear regression and as lambda value increases shrinkage takes place and variable with zero value are ignored by the model.

Ridge regression might includes all the predictors variables in the final model and this may not affect the accuracy of the predictions but can make model interpretation challenging when the number of predictors is very large.

Therefore I will choose Lasso regression technique over Ridge 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:
The Five most important predictor variables in current Lasso model are :

1. GrLivArea
2. OverallQual
3. OverallCond
4. GarageArea
5. BsmtFullBath

The Five most important predictor variables after excluding the above variables are :

1.MSSubClass
2.MSZoning
3.KitchenQual
4.TotalBsmtSF
5.Exterior1st_BrkFace

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?

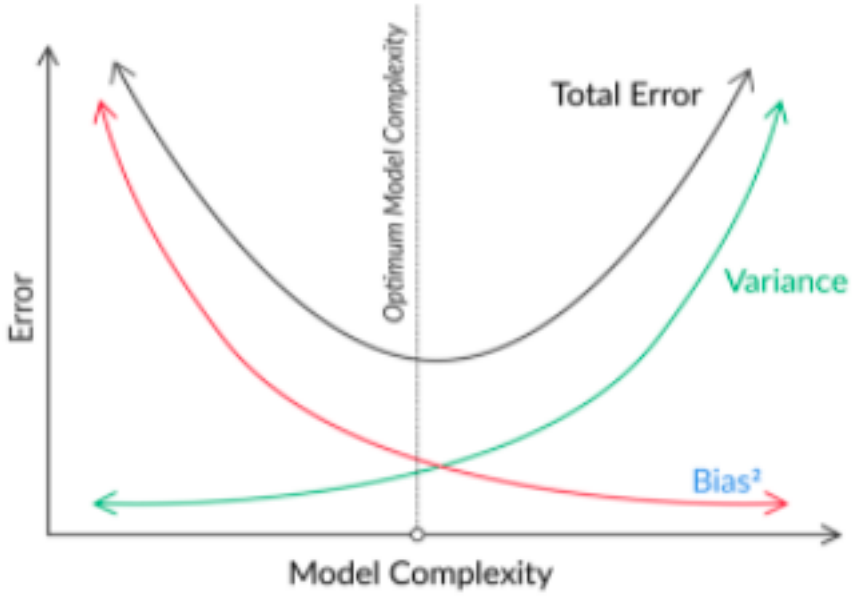
Answer:
As we know a model should be as simpler as possible though the accuracy will decrease due to this but it will be more robust and generalisable.A model is robust means the testing error of the model is consistent with training error and model performs well with enough stability even after adding some noise to the dataset .Therefore robustness or generalzability of a model is a measure of its successful application to datasets other than the used for training and testing.

By implementing Regularization techniques ,we can control the trade-off between model complexity and bias which is connected to robustness of a model. Regularization helps in penalising the coefficients for making the model too complex thereby allowing only the optimal amount of complexity to the model. It helps in controlling the robustness of the model by making the model optimal simpler.So in order to make model more robust and generalizable one has to make sure that there is balance between keeping the model simple and not making it too naive to be any use. Also making the model simpler leads to Bias-variance Trade-off:

- A simpler model that abstract out some pattern followed by the datapoints given is unlikely to change wildly even if more points are added or removed.

- A complex model will need to change for every little change in the dataset and therefore it very unstable and sensitive to any change in training data.

Bias help us to quantify how accurate is the model likely to be on a test data.A complex model can do the an accurate job prediction provided there has to be enough training data.Model that are too naive for eg:one model that give same results for all



test inputs and makes no discrimination whatsoever has a very high bias as its expected error across all test inputs are very high. Variance is the degree of change in the model itself wrt changes in the training data.
Thus accuracy of model can be maintained by keeping balance between Bias and Variance as its minimizes the total error. Also accuracy and robustness may be at the odds to each other as too much accurate model can be liable to over fitting hence it can be too much accurate on train data but fails when it is tested with actual data.