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 changes are implemented?

   A. : In the case of ridge regression:- When we plot the curve between negative mean absolute error and alpha we see that as the value of alpha increases from 0 the error term decrease and the train error is showing increasing trend when the value of alpha increases .when the value of alpha is 2 the test error is minimum so we decided to go with the value of alpha equal to 2 for our ridge regression. For lasso regression I have decided to keep a very small value that is 0.01, when we increase the value of alpha the model tries to penalize more and try to make most of the coefficient value zero. Initially, it came as 0.4 in negative mean absolute error and alpha.

Out[184]:

	Variable	Coeff
0	constant	11.644
32	MSZoning_FV	0.154
34	MSZoning_RL	0.126
53	Neighborhood_Crawfor	0.117
33	MSZoning_RH	0.106
35	MSZoning_RM	0.096
212	SaleCondition_Partial	0.096
69	Neighborhood_StoneBr	0.094
211	SaleCondition_Normal	0.076
14	GrLivArea	0.073
98	Exterior1st_BrkFace	0.069
73	Condition1_Norm	0.067
138	Foundation_Stone	0.066
4	OverallQual	0.064
208	SaleCondition_AdjLand	0.062
200	SaleType_CWD	0.058
202	SaleType_ConLD	0.057
106	Exterior1st_Stucco	0.054
168	CentralAir_Y	0.052
5	OverallCond	0.051

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?

A. It is important to regularize coefficients and improve the prediction accuracy also with the decrease in variance, and making the model interpretably.

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 or squares should be small by using the penalty. The penalty is lambda times sum of squares of the coefficients, hence the coefficients that have greater values gets penalized. As we increase the value of lambda the variance in model is dropped and bias remains constant. Ridge regression includes all variables in final model unlike Lasso Regression.

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 coefficient towards zero and it make the variables exactly equal to 0. Lasso also does variable selection. When lambda value is small it performs simple linear regression and as lambda value increases, shrinkage takes place and variables with 0 value are neglected by the model.

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?

A. Those 5 most important predictor variables that will be excluded are :-
1. GrLivArea
2. OverallQual
3. OverallCond
4. TotalBsmtSF
5. GarageArea

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?

A. The model should be as simple as possible, though its accuracy will decrease but it will be more robust and generalisable. It can be also understood using the Bias-Variance trade-off. The simpler the model the more the bias but less variance and more generalizable. Its implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data i.e. the accuracy does not change much for training and test data.

Bias: Bias is error in model, when the model is weak to learn from the data. High bias means model is unable to learn details in the data. Model performs poor on training and testing data.

Variance: Variance is error in model, when model tries to over learn from the data. High variance means model performs exceptionally well on training data as it has very well trained on this of data but performs very poor on testing data as it was unseen data for the model. It is important to have balance in Bias and Variance to avoid overfitting and under-fitting of data.