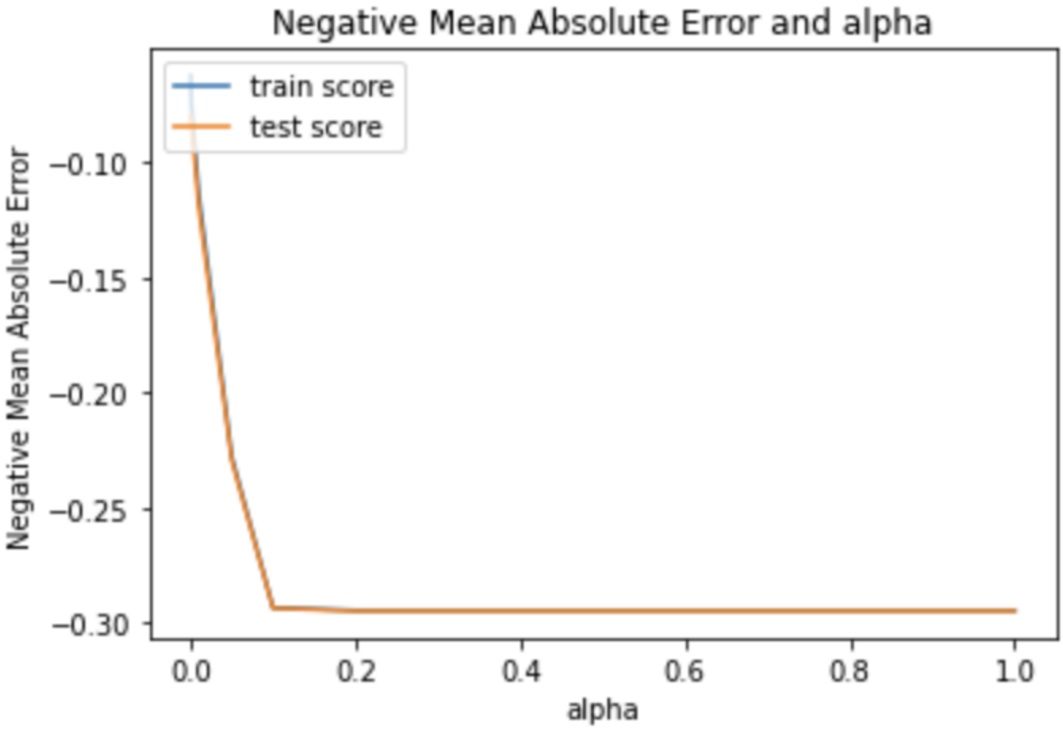
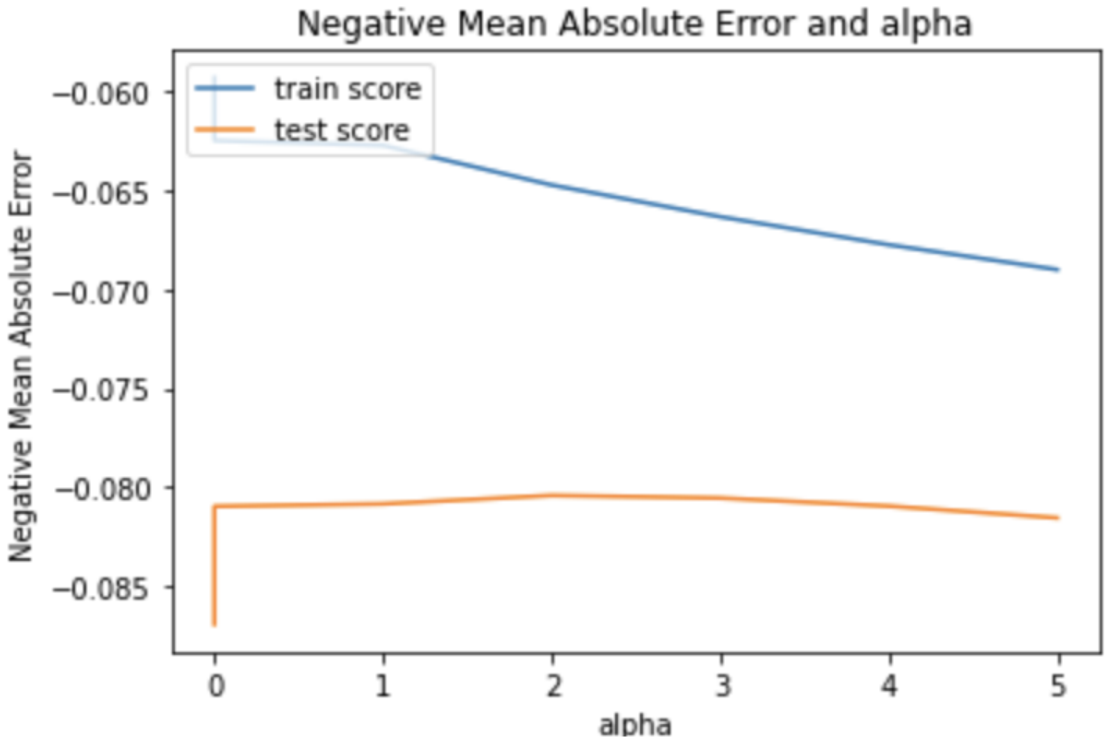
1. What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

   Ans: In Lasso regression:

   I have decided to keep very small value that is 0.01, when we increase the value of alpha the model try to penalize more and try to make most of the coefficient value zero.



   Ridge regression:- When we plot the curve between negative mean absolute error and alpha, we see that as the value of alpha increase from 0 the error term decrease and the test error is showing increasing trend when value of alpha increases .when the value of alpha is 2 the test error is minimum so we decided to go with value of alpha equal to 2 for our ridge regression.



Most predictor variable after the mentioned the changes of making the alpha as 4 (double the initial) for **ridge** regression:

OverallQual	0.309
GrLivArea	0.285
1stFlrSF	0.267
OverallCond	0.234
Neighborhood_IDOTRR	-0.091
Age	-0.104
Neighborhood_MeadowV	-0.104

Most predictor variable for lasso regression after doubling the alpha as 0.2:

GrLivArea	0.447
YearRemodAdd	0.217
Foundation_PConc	0.134
GarageType_Attchd	0.113
FireplaceQu	-0.020
LotShape_Reg	-0.029
MasVnrType_None	-0.076

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?
Ans:
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.

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?
Ans:
Those 5 most important variables are :-

1. GrLivArea (Above grade (ground) living area square feet)
2. OverallQual (Rates the overall material and finish of the house)
3. YearRemodAdd (Remodel date)
4. BsmtFinSF1 (Type 1 finished square feet)
5. Age.

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?

Ans:

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