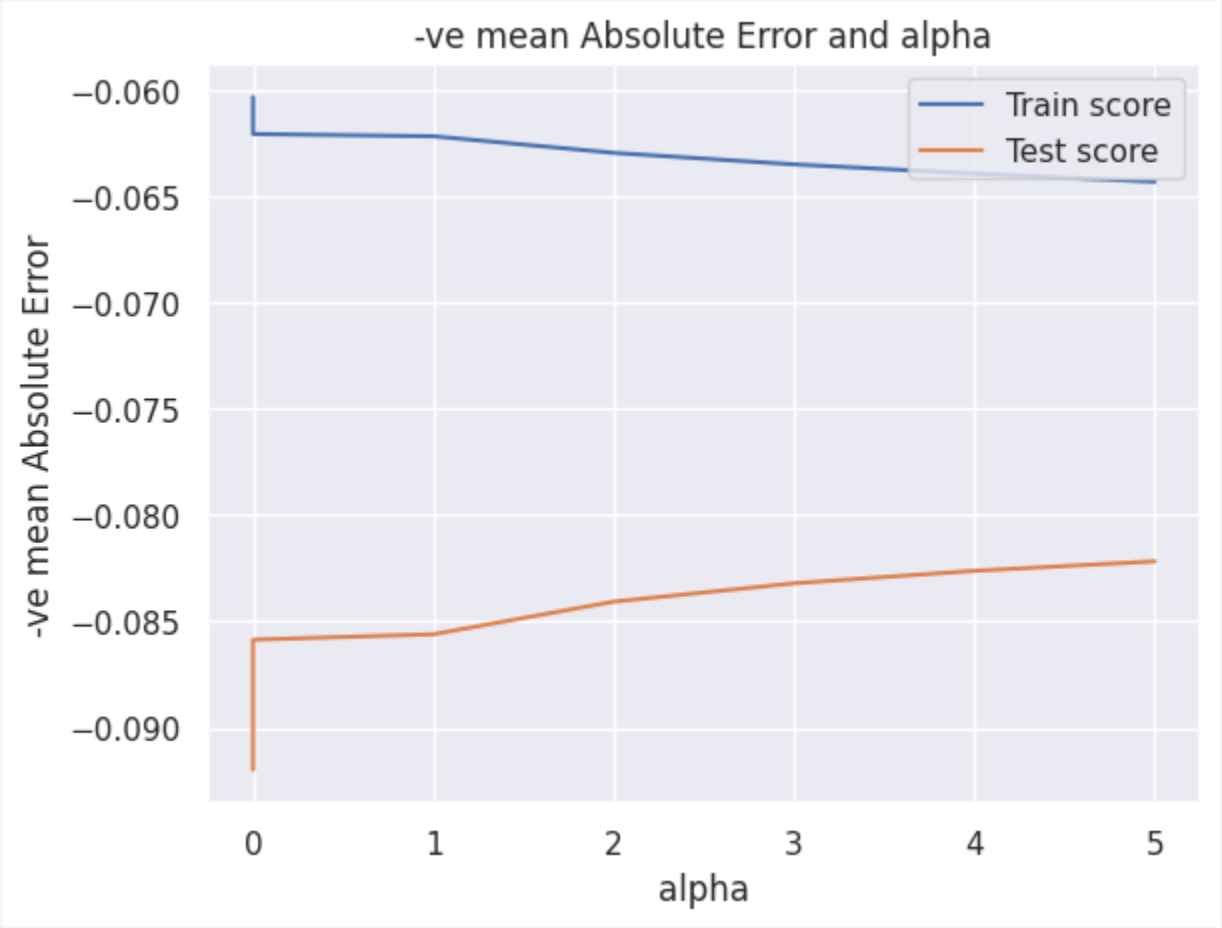
Question-1:

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:

Ridge Regression:

The optimal alpha value for ridge regression is 2



For ridge when we double the alpha value i.e. the predictivity percentage changes to decrease from

Train: 0.934741360608845; Test: 0.9044754555045873

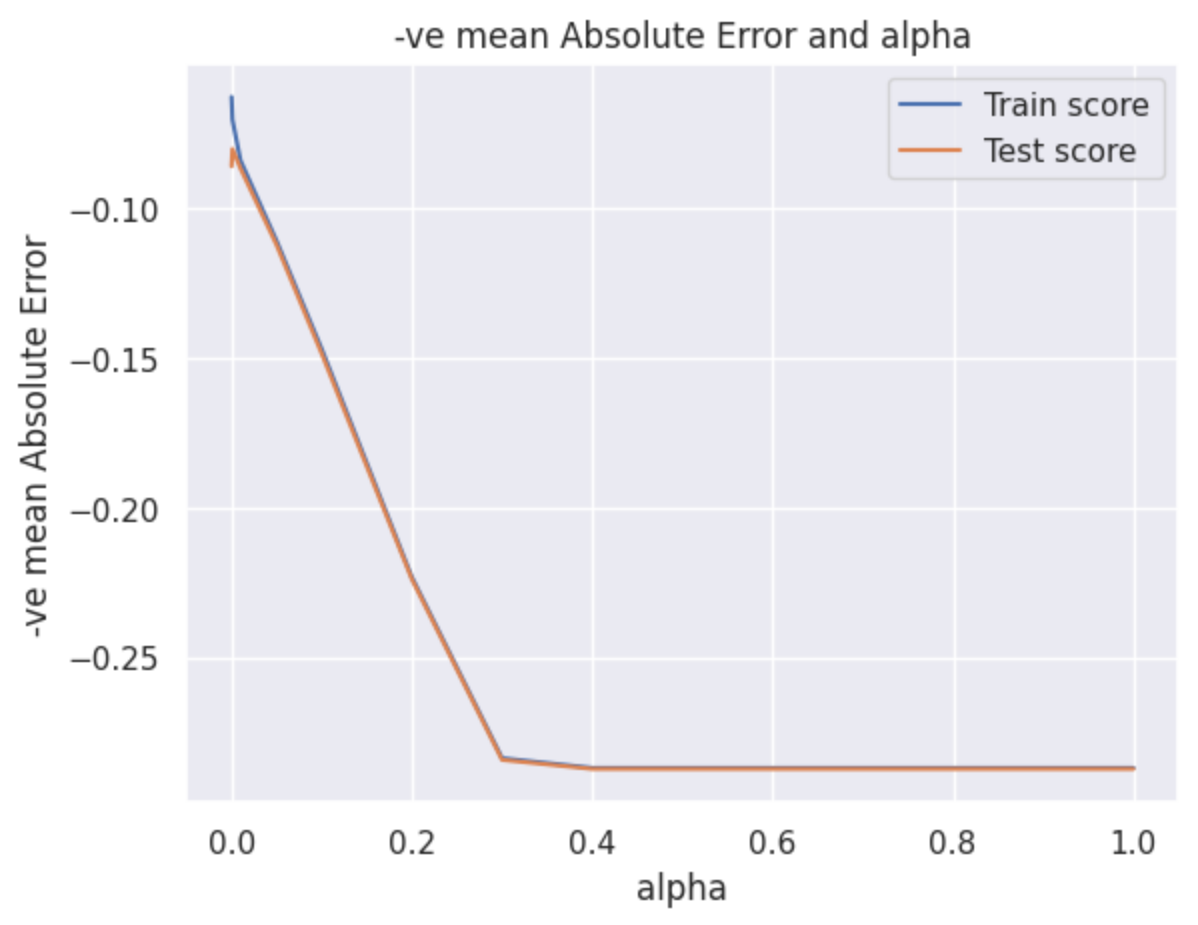
To

Train: 0.9327938972990905; Test: 0.9055924229977949

We see coefficients of the feature variables decreasing and nearing zero

The most important variables from ridge regression model are
Neighborhood_StoneBr
Neighborhood_Crawfor
GrLivArea
SaleCondition_Normal
Condition1_PosN
SaleCondition_Partial
Exterior1st_BrkFace
OverallQual
SaleType_ConLD
Foundation_PConc
2ndFlrSF
MSZoning_RH
OverallCond
TotalBsmtSF

Lasso Regression:
The optimal alpha value for lasso regression is 0.01



For lasso when we double the alpha value i.e. the predictivity percentage changes to decrease from
Train: 0.8857372526638595; Test: 0.8902256227665828
to
Train: 0.8725316654729705; Test: 0.8802515703856408

The RMSE : 0.11509135493253878 decreases slightly to RMSE : 0.11441649534930953

We see coefficients of the feature variables decreasing and nearing zero

The most important variables from lasso regression model are
GrLivArea
OverallQual

TotalBsmtSF
GarageArea
OverallCond
BsmtFinSF1
LotArea
Fireplaces
BsmtFullBath
WoodDeckSF
KitchenAbvGr
PropAge

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:
In my view a model should be simple and should produce consistent predictions. Here the number of features to select are high and lasso makes selection of features and drives the coefficients to zero unlike ridge. Lasso produces a simple model with fewer features and it is more interpretable unlike Ridge.
As out of so many feature only few are relevant for prediction, we should choose lasso over ridge.

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 are
GrLivArea
OverallQual

TotalBsmtSF
GarageArea
OverallCond

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 ensure model is robust amd generalisable below are some of the consideration to be taken into account at different phases of model building
1. Gather Diverse and real world scenario training and test data
2. Don’t over complex model
3. Cross validation
4. Base Varience trade-off, which includes regularization, hyperparameter tuning