

Subjective Questions

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 Ridge regression the Mean Absolute Error stabilizes at alpha value 2 and for lasso regression the MAE stabilizes at 0.1 but for better performance alpha 0.01 is taken. Hence the alpha values **as 2 and 0.01** are optimal value **for ridge and lasso respectively**.

After doubled the alpha i.e. 0.02 for Lasso and 4 for ridge the observations are: -

1. For Lasso the Train score , test score and RSME values are not impacted much but the variables **“GrLivArea”, “GarageArea ”** are not top important predictors when alpha is at 0.02
2. For Ridge the train score , test score and RSME values are not impacted that much but all the top 5 predictors variables are different when alpha is at 4

As pre lasso model **“Foundation_PConc”** i.e. foundation with poor condition is most important predictor and as per ridge the **“GrLivArea”** is the most important predictor after we change the alpha.

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 Lasso Regression the train and test score stabilizes after 0.1 but for better performance I used 0.01 as lambda value. Cause the train and test score at 0.01 is better than 0.1.

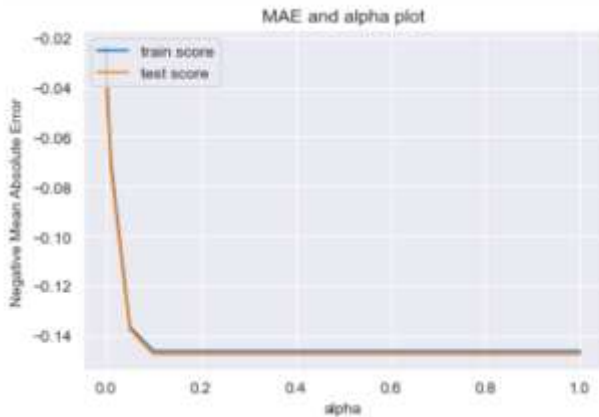


Fig. Screenshot taken from my assignment Jupyter notebook

For Ridge Regression the MAE(mean absolute error) stabilizes when lambda is at 2 , choosing a smaller or larger value for lambda will not fit the model well. Hence I have used lambda for ridge as 2.

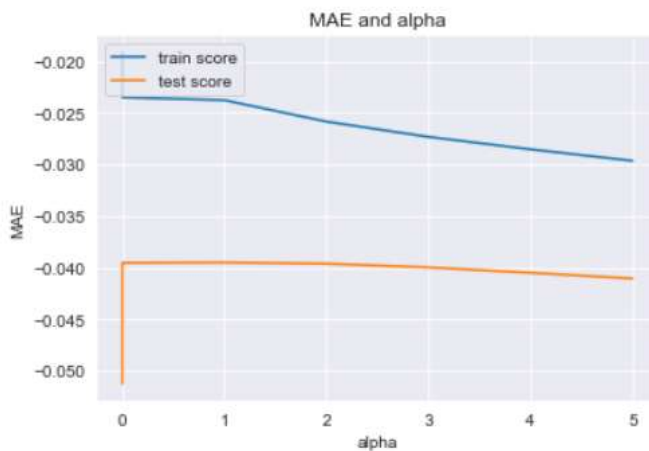


Fig. Screenshot taken from my assignment Jupyter notebook

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 at first are as Follows:-

1. "GrLivArea",
2. "GarageArea",
3. "OverallQual_OQVGood",
4. "Foundation_PConc",
5. "BsmtFinType1_GLQ"

After excluding these variables and building the model again the five most important variables are as bellow:

1. 1stFlrSF
2. SaleDwellingType_SDTsixty
3. GarageCars
4. BsmtFullBath
5. MasVnrType_Stone

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: -

It is important that our model should perform well on both train and test data set that is we can say if model works well on sample data it should work well on unseen data as well. A model with high bias and less variance is considered to more robust and generalizable.

Here with python implementation we R-squared value to decide whether model is good or not. If R-squared is good for train data and it decreases for test data we can say that the model is not robust or it may be over fitting also we can say that there is more variance and low bias.