

Question: 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:

- Optimal Value for Alpha for Lasso is 0.0001 and Ridge is 10
- Changes in Ridge Regression metrics:
 - R2 score of train set decreased from 0.89 to 0.87
 - R2 score of test set increase from 0.83 to 0.81
- Changes in Lasso Regression metrics:
 - R2 score of train set decreased from 0.89 to 0.88
 - R2 score of test set decreased from 0.85 to 0.84
- The most important predictor variables after we double the alpha values are:-
 - Neighborhood_NoRidge
 - GrLivArea
 - OverallQual_10
 - OverallQual_9
 - FullBath
 - Neighborhood_Crawfor
 - Neighborhood_NridgHt
 - Fireplaces
 - TotRmsAbvGrd

Question: 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

- If we want feature selection will use **Lasso**
- Else if we want reduction of coefficient magnitude, we will use **Ridge**

Question: 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 now
 - 1stFlrSF
 - 2ndFlrSF
 - GarageArea
 - MasVnrArea
 - Age_11

Question: 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 make sure model is robust and generalized:
 - We have to take care that it doesn't overfit
 - Model should not be complex.
 - Use model which is resistant to outliers
- Implications of the same for the accuracy of the model
 - A more accurate model is a complex one, so we need to introduce some bias which makes the model less complex and less accurate to make it robust and generalisable.
 - So we need to figure out how to keep balance between accuracy and complexity.