

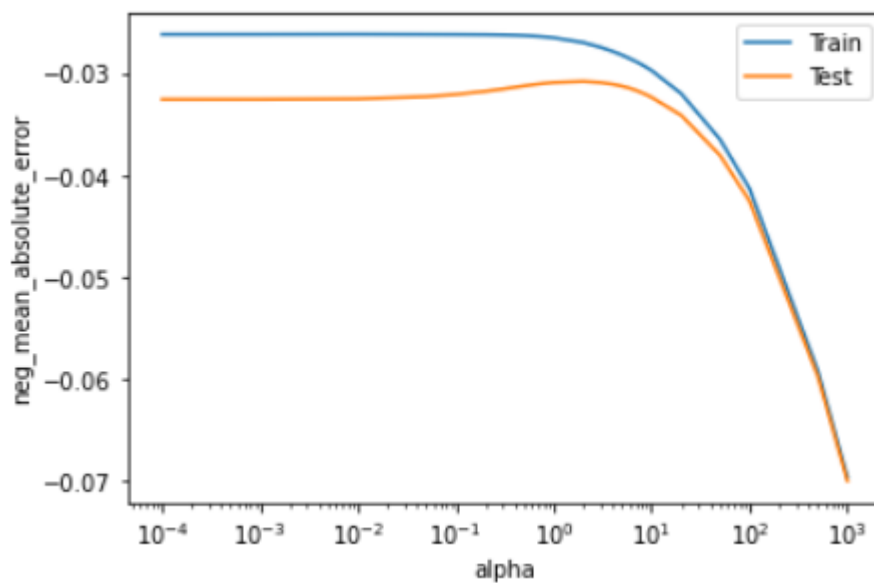
Subjective Questions

Question 1.

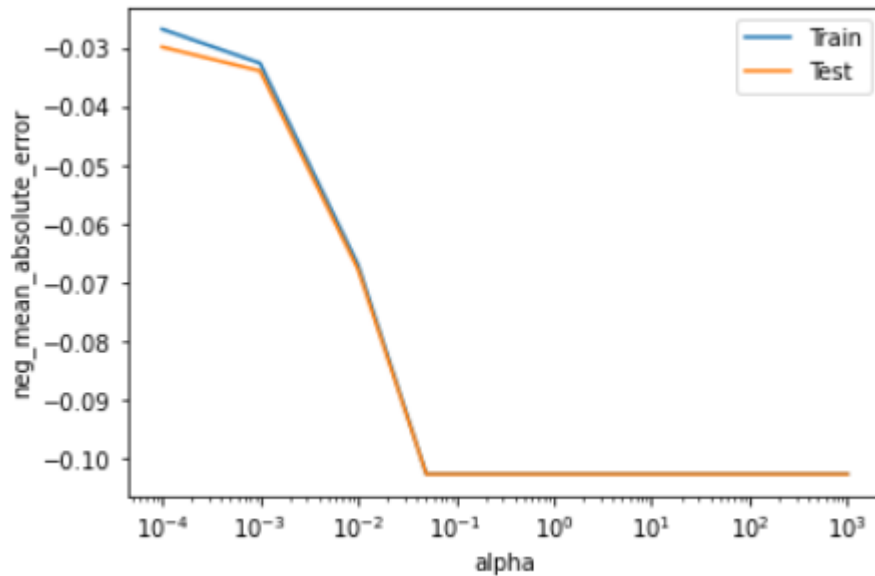
- a. What is the optimal value of alpha for ridge and lasso regression?
- b. What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?
- c. What will be the most important predictor variables after the change is implemented?

Answer a.

The optimal value of alpha for Ridge regression is 2.0



The optimal value of alpha for Lasso regression is 0.0001



Answer b.

The changes happened in model after doubling the value of alpha in Lasso and Ridge Regression are:

1. Lasso Regression:

Value of R2 Score - Decrease in R2 Score

At alpha = 0.0001

for train data at $\alpha(0.0001) = 0.9081992163518053$

for test data at $\alpha(0.0001) = 0.8908987843911962$

At alpha = 0.0002

for train data at $\alpha(0.0002) = 0.9041555455441971$

for test data at $\alpha(0.0002) = 0.8886248277399393$

The best features will change:

Top best features, At alpha = 0.0001

1. GrLivArea
2. OverallQual
3. GarageCars
4. OverallCond
5. LotArea

Top best features, At alpha = 0.0002

1. GrLivArea
2. OverallQual
3. GarageCars
4. OverallCond
5. FullBath

2. Ridge Regression:

Value of R2 Score - Decrease in R2 Score

At alpha = 2.0

for train data at $\alpha(2.0) = 0.9092692793844319$

for test data at $\alpha(2.0) = 0.8863677574710318$

At alpha = 4.0

for train data at $\alpha(4.0) = 0.905758623496098$

for test data at $\alpha(4.0) = 0.8821668437653378$

The best features will change:

Top best features, At alpha = 2.0

1. OverallQual
2. GrLivArea
3. 1stFlrSF
4. GarageCars
5. OverallCond

Top best features, At alpha = 4.0

1. OverallQual
2. GrLivArea
3. 1stFlrSF
4. GarageCars
5. 2ndFlrSF

Answer c.

Most important predictor variables after the change is implemented are:

Ridge Regression

Top best features, At alpha = 4.0

1. OverallQual
2. GrLivArea
3. 1stFlrSF
4. GarageCars
5. 2ndFlrSF

Lasso Regression

Top best features, At alpha = 0.0002

1. GrLivArea
2. OverallQual
3. GarageCars
4. OverallCond
5. FullBath

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.

We conclude above that:

1. Lasso Regression: At alpha = 0.002 --- Test R2 score will be much less than Train R2 score. which shows more complexity in model. At alpha = 0.001, Test R2 score will not be much less than Train R2 score. which shows less model complexity / Overfitting.
2. Ridge Regression: At alpha = 4.0 and alpha = 2.0 --- Model does not show any much significant change in R2 score.

After comparing the models - Lasso and Ridge Regression at different alpha values. We conclude that Lasso Regression at alpha = 0.0001 shows less model complexity / Overfitting. which means a balance between Variance and Bias. Lasso Regression is a better option as it does the feature elimination and model will be more robust.

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 top five feature after removing previous top five features in Lasso Regression are as follows:

1. 1stFlrSF
2. 2ndFlrSF
3. BsmtQual
4. Neighborhood_StoneBr

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.

1. We can make our model robust and generalisable by making sure that it does not Overfit. Because, when a model overfits - it has very high Variance. This results in worst predictive power of model.
2. Overfitting model remembers all the patterns in the train data. But, it will fail to identify patterns in unseen data.
3. For the accuracy of model - Overfitting model has high variance and we have to balance the accuracy and complexity of model. For Overfitting model, we will try to reduce variance and add up the bias. It will decrease accuracy and we can achieve a balance between Variance and Bias.

This can be achieved by using Regularization Techniques like Lasso Regression and Ridge Regression.