

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:

Ans: In the case of ridge regression:- When plotting the curve between negative mean absolute error and alpha we can see that as the alpha value increase from 0 the error term decrease and the train error shows increasing trend with the value of alpha increases .when alpha is 2 the test error is minimum so decided to continue with value of alpha as 2 for our ridge regression.

For lasso regression I decided to keep very small value of alpha that is 0.01, with increase in the value of alpha the model try to penalize more and try to make most of the coefficient value zero. Initially it came as 0.4 in negative mean absolute error and alpha. When we double the value of alpha for our ridge regression the model will apply more penalty on the curve and try to make the model more generalized i.e. making model more simpler.

When we increase the value of alpha to double for lasso we try to penalize more our model and more coefficient of the variable will reduced to zero, when we increase the value of alpha then r^2 square also decreases.

The most important variable after the changes is implemented for ridge regression are :-

1. MSZoning_FV
2. MSZoning_RL
3. Neighborhood_Crawfor
4. MSZoning_RH
5. MSZoning_RM
6. SaleCondition_Partial
7. Neighborhood_StoneBr
8. GrLivArea
9. SaleCondition_Normal
10. Exterior1st_BrkFace

The most important variable after the changes has been implemented for lasso regression are:-

1. GrLivArea
2. OverallQual

3. OverallCond
4. TotalBsmtSF
5. BsmtFinSF1
6. GarageArea
7. Fireplaces
8. LotArea
9. LotArea
10. LotFrontage

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

Lasso regression would prove better option it would help in feature elimination and the 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 most important predictor variables which will be excluded are as below:-

1. GrLivArea
2. OverallQual
3. OverallCond
4. TotalBsmtSF
5. GarageAre

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:

The model must 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 is the model, more the bias but has less variance and is more generalizable.

Its implication in terms of accuracy is that a robust and generalisable model will actually perform equally well on both the training and test data i.e. the accuracy does not change much for training and test data.