

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 lasso regression, the alpha value is 1. In Lasso and Ridge regression as alpha value increases, the slope of the regression line reduces and becomes horizontal.

The most important predictor variables after we double the alpha values are as below.

GrLivArea

OverallQual_8

OverallQual_9

Functional_Typ

Neighborhood_Crawfor

Exterior1st_BrkFace

TotalBsmtSF

CentralAir_Y

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 :-

- The model we will choose to apply will depend on the use case.
- If we have too many variables and one of our primary goal is feature selection, then we will use Lasso.
- If we have too many variables and one of our primary goal is feature selection, then we will use Lasso.

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 :-

Top 5 features in Lasso model which we have to drop and build the model again are as below

OverallQual_9, GrLivArea, OverallQual_8, neighborhood_Crawfor and Exterior1st_BrkFace

After dropping top 5 lasso predictors, we get the following new top 5 predictors as below.

2ndFlrSF, Functional_Typ, 1stFlrSF, MSSubClass_70, Neighborhood_Somerst

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 :-

- A model is robust when any variation in the data does not affect its performance much.
- A generalizable model is able to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.
- To make sure a model is robust and generalizable, we have to take care it doesn't overfit. This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but fail to pick up the patterns in unseen test data.
- In other words, the model should not be too complex in order to be robust and generalizable.
- If we look at it from the perspective of Accuracy, a too complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease.

- In general, we have to find strike some balance between model accuracy and complexity. This can be achieved by Regularization techniques like Ridge Regression and Lasso.