

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

The optimal lambda/alpha value in current case of Ridge and Lasso are:

- ☐ Ridge - 3
- ☐ Lasso - 0.0001

If we change the alpha in both ridge and lasso to double, the coefficients will change.

In case of Ridge, as we increase the value of alpha the Negative Mean Absolute Error slightly start decreasing at first and then it starts decreasing more. When we try to double the alpha in ridge regression i.e. equal to 6 then the model will apply more penalty to the curve and try to make the model more simpler and generalize.

In case of Lasso, as we increase the value of alpha the Negative Mean Absolute Error slightly start increasing slightly at first and then it starts decreasing.

The most important variable after the changes has been implemented for ridge regression and Lasso regression are as follow:

RIDGE	LASSO
MSZoning_RL	MSZoning_RL
GrLivArea	GrLivArea
OverallQual	MSZoning_RM
MSZoning_RM	MSZoning_FV
MSZoning_FV	OverallQual
TotalBsmtSF	TotalBsmtSF
OverallCond	OverallCond
Foundation_PConc	Foundation_PConc
Garage_cars	Garage_cars
Neighborhood_Crawfor	MSZoning_RH

Question 2

You have determined the optimal value of λ for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer

Lasso should perform better in situations where only a few among all the predictors that are used to build our model have a significant influence on the response variable. So, feature selection, which removes the unrelated variables, should help. But Ridge should do better when all the variables have almost the same influence on the response variable.

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 are:

- ☐ MSZoning_RL
- ☐ GrLivArea
- ☐ OverallQual
- ☐ MSZoning_RM
- ☐ MSZoning_FV

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 robust model will continue to make accurate predictions even when faced with challenging situations. In other words, robustness ensures that a model can generalize well to new unseen data.
- ☐ The model's ability 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**. The model should not be complex, it should be simpler in order to make sure that the model is robust and generalizable.
- ☐ When it comes to consider **Accuracy**, a complex model have high accuracy, in a way to make sure that the model is robust and generalizable we will need to keep the variance low which means addition of bias i.e Accuracy decreases which can be achieved using regularisation(Ridge and Lasso).

