

Subjective Questions

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 optimum value of alpha for ridge regression is '9' and '0.0001' for lasso regression. Predictors are same but the coefficient of these predictor has changed after doubled the alpha values and five predictor variables, after change implemented, as follows:

- 1stFlrSF
- 2ndFlrSF
- OverallQual
- OverallCond
- BsmtFinSF1

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 is better option than ridge regression. Lasso regression also help us in feature eliminate any of the features that not important and make model 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:

five most important predictor variables

- | | |
|---------------|----------|
| • BsmtFinSF1 | 0.087255 |
| • LotArea | 0.073028 |
| • BsmtUnfSF | 0.072591 |
| • GarageArea | 0.065092 |
| • KitchenQual | 0.059210 |

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:

When test accuracy is greater than training score then model considered as robust and generalized. It is also important to make sure the data in the test set is appropriate and should be representative of the wider population. The model should be accurate for datasets other than the ones which were used during training. Model should be evaluated multiple times to estimate the performance. Too much importance should not be given to the outliers so that the accuracy predicted by the model is high. To ensure that this is not the case, the outliers analysis needs to be done and only those which are relevant to the dataset need to be retained. Those outliers which it does not make sense to keep must be removed from the dataset. If the model is not robust, it cannot be trusted for predictive analysis.