

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 value of alpha for ridge regression is 5.4, while for Lasso regression optimal value is 163.

If we double the value of alpha, then the ridge coefficients will change, and the Rsquare score of the model reduces to 0.809 from 0.810.

Similarly, when the value of alpha is doubled for Lasso, the coefficient values change and also some new predictor variables get added to the model (coefficient becomes non-zero). The Rsquare score of the model reduces to 0.810 from 0.812.

After the change is implemented:

- For Lasso regression, the most important predictor variable is **GrLivArea**.
- For Ridge regression, the most important predictor variable is **OverallQual**.

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

It is better to use the Lasso regression as it has feature-selection capability so it has selected the most relevant features in the model which makes the model simpler as compared to Ridge regression.

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 5 most important features (with positive correlation) for predicting the price as per LASSO regression are:

1. Above Ground Living Area
2. Overall qualification
3. Roof Material - Wood Shingles
4. Neighborhood - North Ridge
5. Full Bathrooms above Grade

The 5 most important features with negative correlation to the house price as per LASSO regression are:

1. KitchenAbvGr
2. Lot Shape - Irregular
3. KitchenQual_TA
4. KitchenQual_Gd
5. KitchenQual_TA

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

In order to make the model robust and generalisable, we have to ensure that there is no overfitting of the data. However, when we try to reduce the overfitting, it will also reduce the accuracy of the model to some extent, as the error with respect to the training data increases when we reduce the fitting to the training data.