

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 is 0.1 and R2-score associated with it is 0.88

The Optimal value of Alpha for Larso is 0.001 and R2-score associated with it is 0.85.

After considering double the value of Alpha

I.e Alpha for Ridge is 0.2 results in R2-score of 0.78

Alpha for lasso is 0.002 results in R2-score of 0.87

After doubling Alphas ,The most important predictor variables are as follows:

| Lasso Co-Efficient after Doubling | |
|-----------------------------------|----------|
| OverallQual | 0.102857 |
| GrLivArea | 0.101971 |
| Neighborhood_Crawfor | 0.068405 |
| Neighborhood_NridgHt | 0.057354 |
| Neighborhood_Somerst | 0.055390 |

| Ridge Alpha Co-Efficient after Doubling | |
|---|----------|
| MSZoning_RL | 0.368570 |
| MSZoning_FV | 0.363079 |
| MSZoning_RH | 0.361342 |
| MSZoning_RM | 0.326664 |
| RoofMatl_Membran | 0.319111 |

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:

After comparing R2-score and MSE of both models i will choose Ridge as it is giving better R2-score compared with lasso.

MSE of both models are almost same. Lasso will be great when it comes to feature reduction

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:

After excluding the five most important predictor variables and building a model

The new R2-score is 0.69

The new Mean squared error is 0.00298

New Top 5 predictor variables are:

- LotFrontage
- Total_porch_sf
- HouseStyle_2.5Unf
- HouseStyle_2.5Fin
- Neighborhood_Veenker

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 should be generalized so that the test accuracy is not lesser than the training score. The model should be accurate for datasets other than the ones which were used during training. 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.