

Problem Statement - Part II

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?

Answer1:

Optimal value of alpha for ridge: **7**

Optimal value of alpha for ridge: **100**

After make the double alpha for ridge and lasso i.e. **14 and 200**

For Ridge: *Coeff values are increasing as alpha will increase. r^2 score of train data is also drop from 0.796 to 0.7906*

For Lasso: *As alpha value increased more features removed from model. But r^2 score is also dropped by 1% in both test and train data*

Top Features: *Neighborhood_NoRidge', 'Neighborhood_NridgHt', 'Neighborhood_Crawfor', 'OverallQual', 'HouseStyle_2Story'*

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 2:

We will choose **Lasso** as its giving **feature selection** option also. It has removed unwanted features from model without affecting the model accuracy. Which makes are model generalized and simple and accurate.

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 3:

Top 5 features are Neighborhood_NoRidge', 'Neighborhood_NridgHt', 'Neighborhood_Crawfor', 'OverallQual', 'HouseStyle_2Story'. Now top most features are: **Next top 5 features** after dropping 5 main predictors 1stFlrSF ,LotConfig_CulDSac2, Neighborhood_StoneBr3, Exterior2nd_ImStucc622, HouseStyle_2Story

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer 4:

To make model robust and generalizable 3 features are required:

1. **Model accuracy** should be $> 70-75\%$: In our case its coming 79%(Train) and 80%(Test) which is correct.
2. **P-value** of all the features is < 0.05
3. **VIF** of all the features are < 5

Thus we are sure that model is robust and generalizable.