

Advance Regression Assignment

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?

Ans:

- The Optimal Value of Alpha for Ridge is 5 and lasso is 0.001
- If we double the value of ridge and lasso, ridge will be 10 and lasso will be 0.002. As alpha values are very small, there would not be significant difference.
- SaleCondition For Ridge and GrLivArea for Lasso.

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?

Ans:

I will be choosing Lasso over Ridge. Although the accuracy for both Lasso and Ridge are same, Lasso is good option to choose as it helps remove the variables and it penalises more.

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?

Ans:

- OverallQual
- CentralAir
- MSZoning_RL
- MSZoning_FV
- SaleCondition_Normal

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?

Ans:

As per Occam's Razor, given two models that show similar 'performance' in the finite training or test data, we should pick the one that makes fewer on the test data due to following reasons: -

- Simpler models are usually more 'generic' and are more widely applicable and widely used
- Simpler models require less training samples for good training than the more complex model and hence are easier to train.
- Simpler models are more robust and good. Complex models tend to change wildly with changes in the training data set it may lead to overfitting. Simple models have low variance, high bias and complex models have low bias and high variance. Simpler models make more errors in the training set.
- A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data.
- A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed. Bias quantifies how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Thus accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error.