Problem Statement - Part II

SHASHWATH S

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?

The alpha value is used to fine tune the model. This value can become anything starting from 0. But, the optimum value of alpha or the best value of alpha should be the value which gives lesser magnitude of coefficients. Also, the optimal value of alpha should results in lesser R-Squared and RSS Values.

If the alpha is double, then there will be a slight difference between RSS Score between training and test sets.

The most important predictor variable after changing alpha value is : **WdShngl** (Wood Shingles)

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?

The best alpha we got for ridge regression is 20 & The best value we got for alpha in lasso regression is 0.001. After fitting a ridge & lasso regression models using these alpha values we got the difference between R2 Score of Training & Test Set is very minimal in Ridge Regression. Hence we choose to apply Ridge Regression for Given Data Set.

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?

Both lasso & Ridge regression will give different important predictor variable.

The 5 Important predictor variables given by lasso regression which we can exclude are.

1. Partial (Condition of sale)
2. TA (Garage Condition)
3. WdShake (Roof Meterial)
4. GasW (Type of Heating)
5. BrkFace (Exterior covering on house)

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?

A Model is robust and generalisable when the model understands training data well and works well with unseen data.

Accuracy of model also implies that the variance of results observed should be minimal when working with unseen data.