ASSIGNMENT: ADVANCED REGRESSION

Pratik Hatwalne

Q1. 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 optimal value of alpha for Ridge and Lasso regression varies based on the specific dataset. Typically, alpha is selected using cross-validation techniques like GridSearchCV.

Doubling the value of alpha for Ridge and Lasso will increase the regularization strength, leading to a simpler model with more coefficients pushed toward zero.

After doubling alpha, the most important predictor variables will be those that have survived the increased regularization. These variables will have more significant coefficients and contribute more to the model's predictions.

Q2. 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?

After performing hyper parameter tuning and building the models based on Ridge and Lasso regression, we come to the conclusion that Lasso regression performs better altogether than the Ridge regression model due to increased efficiency, lesser number of predictor variables and the significantly less amount of mean absolute error.

Q3. 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?

After removing the 5 most important predictor variables (the variables which have the highest coefficients in the detailed statistics) , on building a new model we come to know that the next 5 most important predictors are: Foundation_Slab, SaleCondition_Alloca, 2ndFlrSF, TotalBsmtSF, Street_Pave.

Q4. 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?

- The model is expected to be as simple as possible and simpler models are considered as more generic models, though their accuracy will be decreased but it would increase the model’s robustness

- this can be understood from the bias variance tradeoff. The simpler the model the more the bias but less variance becoming generalizable. Whereas more complex models will have high variance but resulting in a very low bias.
- Sometimes underfitting and overfitting are the problems associated with the model. Hence it is important to have balance in bias and variance to avoid such problems. This balance is obtained with a process known as regularization
- Regularization helps in managing the model complexity by essentially shrinking the coefficients towards 0, this avoids the model becoming too complex, thus reducing the risk of overfitting
- Regularization should be used to keep the model optimum simpler. It penalizes the model if it becomes more complex
- Regularization method helps to achieve the bias variance trade off. It compromizes by increasing bias to a optimum position where the total error in general is minimum.
- This point is also know as the optiumum model complexity where the model is sufficient simpler to be generalizable but complex enough to be robust