Subjective Questions - Advanced Regression Assignment

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 Regression is 0.1 and for lasso it was around 0.001 . For lasso regression if we double the value of alpha from 0.001 to 0.002 then we observe a fall in accuracy for both training and test sets .Whereas for Ridge regression the accuracy figures remained approximately the same as before . For lasso regression the new important predictors are ['Condition2_PosN', 'GrLivArea', 'Neighborhood_Crawfor', 'OverallQual', 'BldgType_Twnhs' . For Ridge regression the predictors remain the same as 'Condition2_PosN', 'Exterior1st_BrkComm', 'MSZoning_FV', 'MSZoning_RL', 'MSZoning_RH' .

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 : We have used grid search cross validation to find the optimal value of lambda . Since we are dealing with a regression problem , we have provided an R2 score as the metric for comparison between different models with different hyper parameter values. The best estimator is the one where we get the lowest error and highest value of r2 on test set . So in our case the best value of alpha for ridge regression was 0.1 and for lasso regression it was 0.001 .

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 5 most important predictors , we created a new model with different set of features . Now the 5 most important predictors are 'BldgType_Twnhs', 'GrLivArea', 'Neighborhood_MeadowV', 'OverallQual', 'RoofStyle_Gambrel'.

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 : To make a model robust and generalisable , there are a couple of things which we need to keep in mind . For example in case of Regression problem , instead of using R2 score we can use adjusted R2 score as it tells how well your features explain the target variable and also it tells you whether adding a certain feature actually improves the model . We need to

ensure that our model is stable and has low bias and low variance . To achieve such a state we can employ strategies like K-fold cross validation . On top of all this , we can always add some sort of regularisation that penalizes the model by introducing some bias that prevents the model from overfitting the data . Although it might increase the training error but it provides a model with generalisation capability .