Problem Statement – Part II

Assignment Part – II

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?

Ans: In case of ridge regression:- If we plot the curve between the negative mean absolute error and alpha we find that as the value of alpha increases from 0 the error term decreases and rail error shows an increasing trend with the value of alpha high .When alpha has a value of 2 test error is minimum so we decided to go with a value of alpha equal to 2 for our ridge regression.

For the lasso regression I decided to set a small value of 0.01, while increasing the alpha value to penalize the model more and keep most coefficient values to zero initially at a negative mean absolute error and alpha came out as 0.4.

When we double the alpha value for our ridge regression we take an alpha value equal to 10. The model will apply more penalties on the curve and try to generalize the model more which makes the model weaker and less focused there will fit any data of the data set. from the graph we can see that if alpha is 10 we get more errors for both test and train.

Similarly, if we increase the value of alpha for the lasso, we are trying to penalize our model more and reduce more coefficients of the variable to zero, even less when we increase the value of our r2 squared.

The most important variable after the changes has been implemented for lasso regression are as follows:-

1. GrLivArea
2. OverallQual
3. OverallCond
4. TotalBsmtSF
5. BsmtFinSF1
6. GarageArea
7. Fireplaces
8. LotArea
9. LotArea
10. LotFrontage

The most important variable after the changes has been implemented for ridge regression are as follows:-

1. MSZoning_FV
2. MSZoning_RL

3. Neighborhood_Crawfor
4. MSZoning_RH
5. MSZoning_RM
6. SaleCondition_Partial
7. Neighborhood_StoneBr
8. GrLivArea
9. SaleCondition_Normal
10. Exterior1st_BrkFace

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?

Ans: It is important to regularize coefficients and improve the prediction accuracy also with the lower in variance and making the model interpretably.

Ridge regression makes use of a tuning parameter called lambda as the penalty is rectangular of importance of coefficients that is diagnosed through move validation. Residual sum or squares must be small using the penalty. The penalty is lambda instances sum of squares of the coefficients, subsequently the coefficients that have extra values gets penalized. As we increase the fee of lambda the variance in model is dropped and bias remains consistent. Ridge regression includes all variables in the very last version not like Lasso Regression.

Lasso regression makes use of a tuning parameter called lambda as the penalty is absolute value of significance of coefficients which is recognized by means of pass validation. As the lambda fee will increase Lasso shrinks the coefficient towards 0 and it makes the variables precisely same to zero. Lasso also does variable choice. When lambda value is small it plays easy linear regression and as lambda cost increases, shrinkage takes location and variables with zero value are neglected through the model.

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?

Ans: Those 5 predictor variables that will have to be excluded are :-

1. TotalBsmtSF
2. GarageArea
3. GrLivArea
4. OverallQual
5. OverallCond

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?

Ans: The model ought to be as simple as possible, although its accuracy may be reduced but it will be robust and generalizable. It also can be understood with the aid of Bias-Variance trade-off. The less difficult the model, the extra biased but less variance and generalizable. The implication for accuracy is that robust and generalizable models will perform equally properly for training and testing facts ie. Accuracy does not exchange lots for education and testing data.