Assignment Part -II

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?

The optimal value is chosen such that, when **alpha** is 0 , **Lasso regression** produces the same coefficients as a linear **regression**. When **alpha** is very ,very large, all coefficients are zero.

As the value of alpha increases, the model complexity reduces. Though higher values of alpha reduce overfitting, significantly high values can cause underfitting as well (eg. alpha = 5). Thus alpha should be chosen wisely. A widely accept technique is cross-validation, i.e. the value of alpha is iterated over a range of values and the one giving higher cross-validation score is chosen.

When we double the alpha value-and take a look back again at the cost function for ridge regression.

$$\min\left(\left||Y - X(\theta)|\right|_2^2 + \lambda||\theta||_2^2\right)$$

Earlier we noticed, we come across an extra term, which is known as the penalty term. λ given here, is actually denoted by alpha parameter in the ridge function. So by changing the values of alpha, we are basically controlling the penalty term. Higher the values of alpha, bigger is the penalty and therefore the magnitude of coefficients are reduced.

We can see that as we increased the value of alpha, coefficients were approaching towards zero, but if you see in case of lasso, even at smaller alpha's, our coefficients are reducing to absolute zeroes. Therefore, lasso selects the only some feature while reduces the coefficients of others to zero. This property is known as feature selection and which is absent in case of ridge.

Mathematics behind lasso regression is quiet similar to that of ridge only difference being instead of adding squares of theta, we will add absolute value of Θ.

$$\min\left(\left||Y - X\theta|\right|_2^2 + \lambda||\theta||_1\right)$$

Here too, λ is the hypermeter, whose value is equal to the alpha in the Lasso function.

The most important predictors after the change has been implemented are-

We had got these variables in lasso :

GrLivArea: Above grade (ground) living area square feet

OverallQual: Rates the overall material and finish of the house

OverallCond : Rates the overall condition of the house

TotalBsmtSF: Total square feet of basement area

BsmtFinSF1 : Type 1 finished square feet

When compared to these in ridge because of change in co-efficient value.

GrLivArea : Above grade (ground) living area square feet

OverallQual : Rates the overall material and finish of the house

MSZoning_RL : Zoning classification of the sale - Residential Medium Density

GarageType_Attchd : Garage Type-Attached to home

2ndFlrSF : Second floor square feet

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?

The final variables from both lasso and ridge regression models are same from assignment.

Rsquared score from ridge – Train is 91.95% and test is 89.09%

Rsquared score from lasso- Train is 90.05% and test is 87.36%. These scores are also almost same.

I would choose to use lasso regression model .

Where lasso gives an additional advantage of feature selection. That is least important and multicollinearity variables are eliminated by co-efficient becoming zero.

In the present model, after rfe with 50 variables, lasso automatically eliminated 10 of them by making co-efficient to zero. So rest of them being the same, I would choose to apply lasso regression , given computational cost is not an issue.

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?

We can plot a curve to analyze these coefficients , From our assignment we observed the new 5 most important variables are –

1.KitchenQual 2.GarageArea 3.Foundation_PConc 4. Fireplaces 5. Garage Area

-By using Standardized coefficients - Statistical software calculates standardized regression coefficients by first standardizing the observed values of each independent variable and then fitting the model using the standardized independent variables.

-Change in R-squared for the last variable added to the model.

Lasso Regression usually shrinks the less important feature's coefficient to zero thus, removing some features altogether. This thereby helps in performing feature selection.

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?

If we consider a simple model over the complex model we can make sure that a model is robust and generalizable. There is no specific definition of complexity of a model. However there are few ways oflooking the complexity of a model:

1. The number of parameters required to specify the model completely. Lesser the parameters, simplerthe model is.
2. The degree of function required to specify the model completely. Lesser the degree of thefunction, simpler the model is.
3. The depth or size of a decision tree. Lesser the depth, simpler the model is.
4. We have lesser complexity when lesser the size taken by the best – possible representation of the model.

As we know that, simple models have low variance, high bias and complex models have low bias, high variance. "Variance" is how sensitive is the model to input data and "Bias" is the deviation from the expected, ideal behaviour i.e. how much error the model is likely to make in the test dataset.

Although, in practice, we often cannot have a low bias and low variance. As the model complexity goes up, the bias reduces while the variance increases, hence the trade- off is required and the phenomenon is referred to as the bias-variance trade-off.

The implications on the accuracy of the model: Robust and generalizable models are simple models are not very accurate compared to their counterpart – complex model. And that is because simple models have low variance, high bias. Bias qualifies for the accuracy of the model - how accurate is the model likely to be on future (test) data. This is because the simple model doesn’t memorize entire training dataset, so when the future (test) data appears – it is bound to make some error.