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?

Answer 1

The optimal values of alpha for ridge and lasso regression are 0.8 and 0.0001 respectively.

If we choose double the value of alpha for both ridge and lasso the R2 Score of both Train and Test decreases where as the MSE slightly increases.

As alpha moves towards higher values, the shrinkage penalty increases, pushing the coefficients further towards 0, which may lead to model under fitting. Choosing an appropriate alpha becomes crucial: If it is too small, then we would not be able to solve the problem of over fitting, and with too large a alpha, we may actually end up under fitting.

We basically want models that do not over fit the data, but they should be able to identify underlying patterns in it. Hence, an appropriate choice of lambda becomes crucial. This can be achieved through hyper parameter tuning.

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?

Answer 2

The optimal values of alpha for ridge and lasso regression are 0.8 and 0.0001 respectively. I have chosen the Ridge model because,

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.933599	0.890851	0.889290
1	R2 Score (Test)	0.827328	0.857822	0.835283
2	RSS (Train)	1.231765	2.024742	2.053716
3	RSS (Test)	1.589386	1.308696	1.516164
4	MSE (Train)	0.035027	0.044907	0.045228
5	MSE (Test)	0.060726	0.055104	0.059311

From the above Metric the Ridge Regression is showing better than Lasso Regression as R2 Score(Test) is higher, RSS and MSE is comparatively lower than Lasso.

The R2 Score of Ridge is 0.89 in training set where as nearly 0.86 in test set and the difference is not more than 4.

What we need is lowest total error, i.e., low bias and low variance, such that the model identifies all the patterns and is also able to perform well with unseen data.

So we conclude that our model is robust and generalisable.

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?

Answer 3

1stFlrSF, MSZoning_RH, Exterior1st_CBlock, Exterior2nd_CBlock, HouseStyle_2.5Unf

are the five important predictor variables which are 0 coefficient in lasso model.

- Lasso pushes the model coefficients towards 0 in order to handle high variance, just like Ridge regression. But, in addition to this, Lasso also pushes some coefficients to be exactly 0 and thus performs variable selection.
- This variable selection results in models that are easier to interpret.

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?

Answer 4

As in Occams Razor simpler models are robust and generalisable and requires few

assumptions and less data for learning.

Simpler models make more errors in the training set. In case of over fitting or if a model is complex it memorises the data rather than intelligently learning the underlying trends in it.

This is because it is possible to memorise data, and this is a problem because the real test

happens on unseen, real-world data.

Where R2 score is more higher in training and in test set it is much lower.

On the other hand simple models have low variance, high bias and complex models have

low bias, high variance.

Therefore to make the model more robust and generalizable, make the model simple but

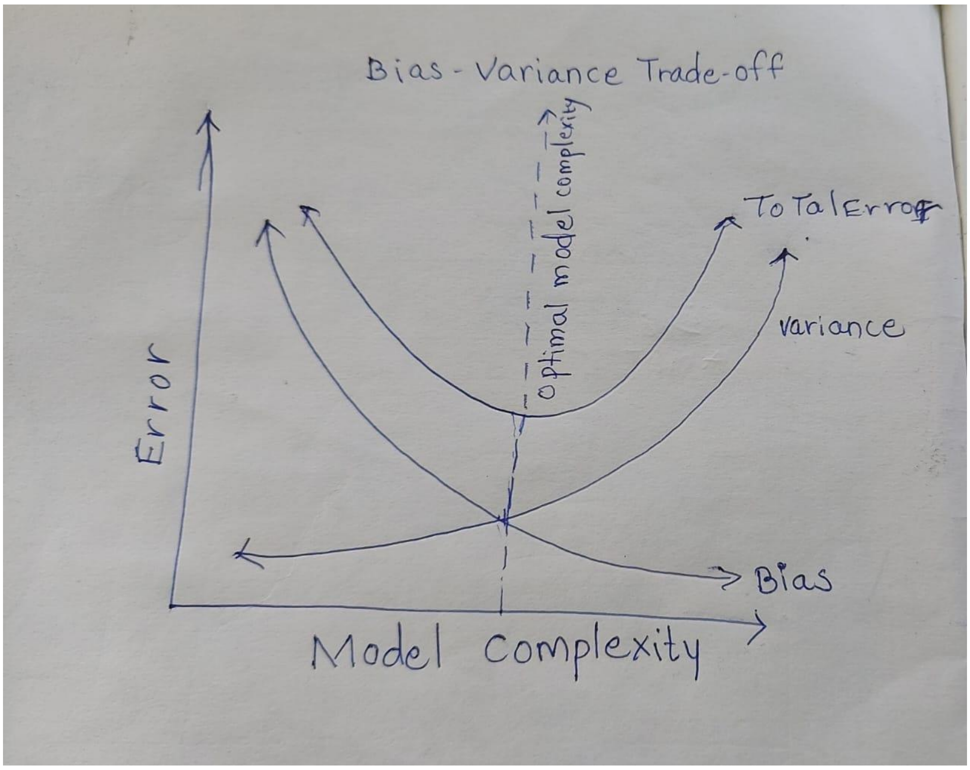
not simpler which will not be of any use.

Regularization can be used to make the model simpler.

Regularization helps to strike the delicate balance between keeping the model simple and not

making it too naive to be of any use. For regression, regularization involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model. Thus accuracy of the model can be maintained by keeping the balance between Bias and

Variance as it minimizes the total error as shown in the below graph



In complex model R2 score is more higher in training and in test set it is much lower. But

in our case R2 score in training and test set difference is not more than 4.

RSS and MSE are also lower.

It required to have lowest total error, i.e., low bias and low variance, such that the model

identifies all the patterns and is also able to perform well with unseen data.

So we conclude that our model is robust and generalisable.