Subjective Questions – Advanced Regression Assignment

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:

Based on our final analysis we got to know the Optimal value of Alpha is:

- Lasso Regression = 0.001
- Ridge Regression = 8

When we double value of alpha R2 Score decreases, but there is no significance change in the value.

The most important predictor features are:

For Lasso:

1) Neighborhood_Edwards
2) Fireplaces_1
3) OverallQual_9
4) KitchenAbvGr_2
5) Neighborhood_ClearCr
6) Neighborhood_StoneBr
7) OverallQual_8
8) OverallCond_8
9) Exterior1st_CBlock
10) MSSubClass_75

For Ridge:

1) Neighborhood_Edwards
2) Fireplaces_1
3) OverallQual_9

4) OverallCond_9
5) OverallCond_8
6) Neighborhood_ClearCr
7) Neighborhood_StoneBr
8) KitchenAbvGr_2
9) Exterior1st_CBlock
10) RoofStyle_Gable

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:

We have made the model using both Ridge and Lasso and the optimal lambda values are:

- Ridge = 8
- Lasso = 0.001

Hence, for the final presentation I would choose Lasso as it helps in features selection and also penalizes for adding more features.

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:

After building Lasso Regression model with value of alpha 0.001 now in new model, following below are the new Top 5 variables:

1) Neighborhood_StoneBr

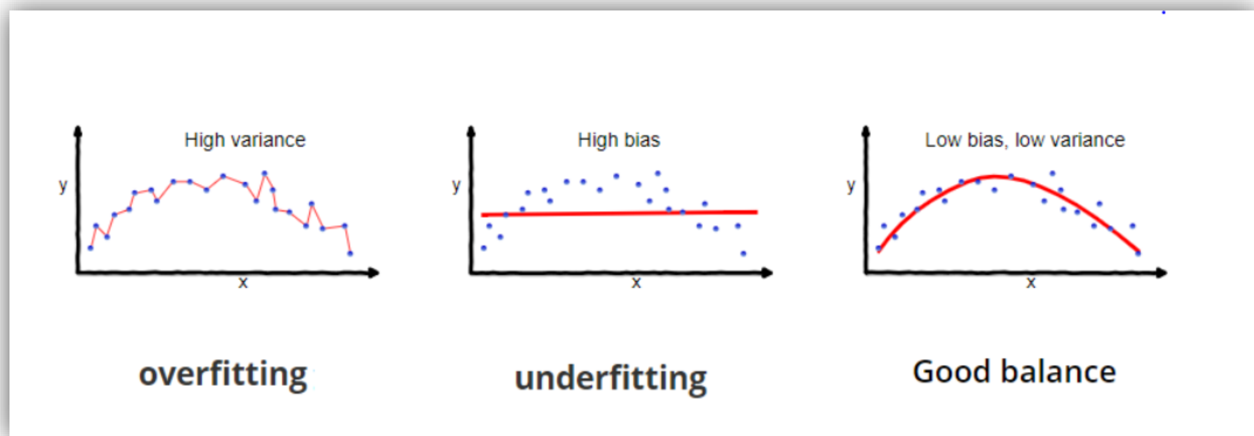
2) OverallQual_8
3) OverallCond_8
4) OverallQual_9
5) Exterior1st_CBlock

Question 4

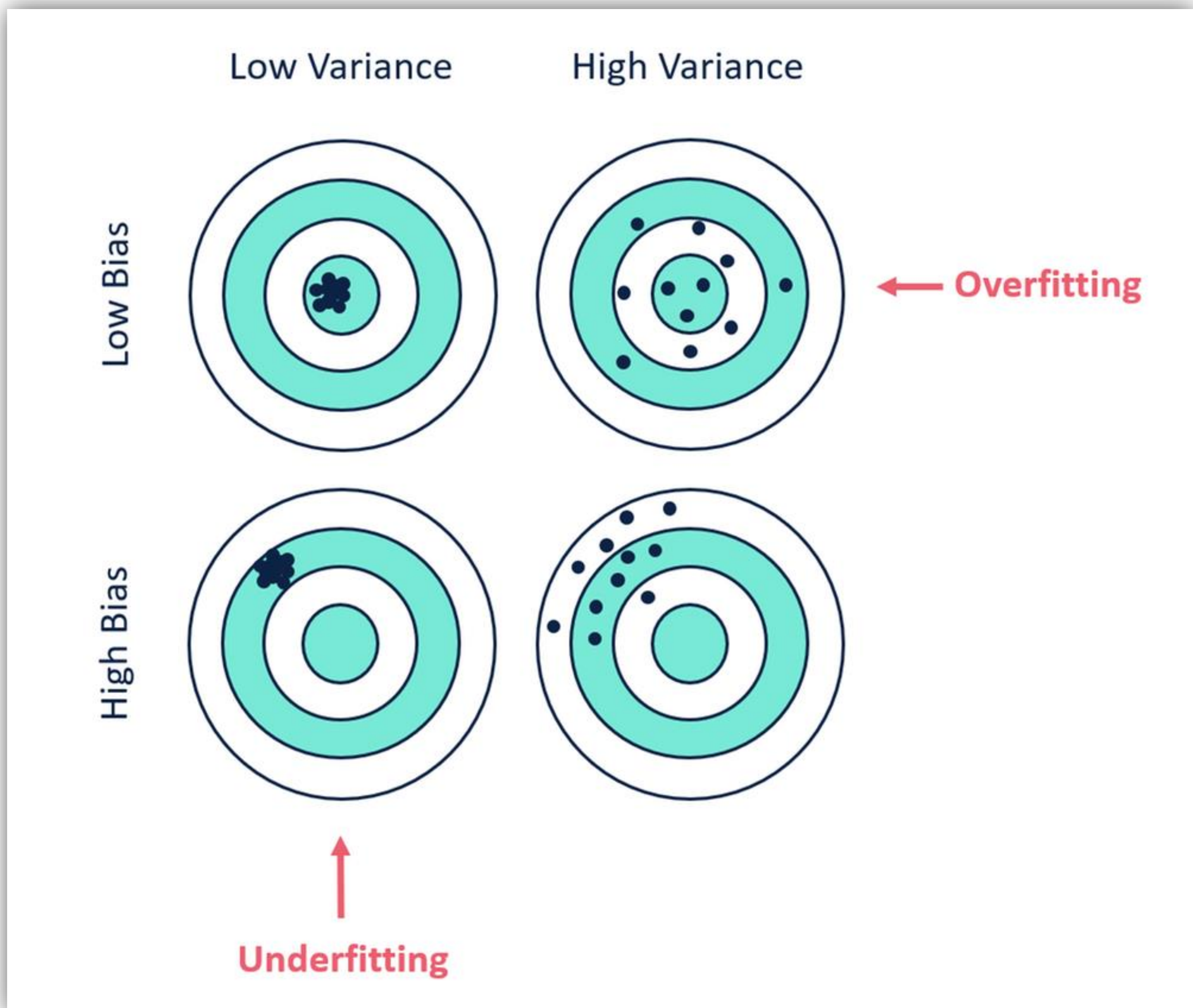
How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer 4:

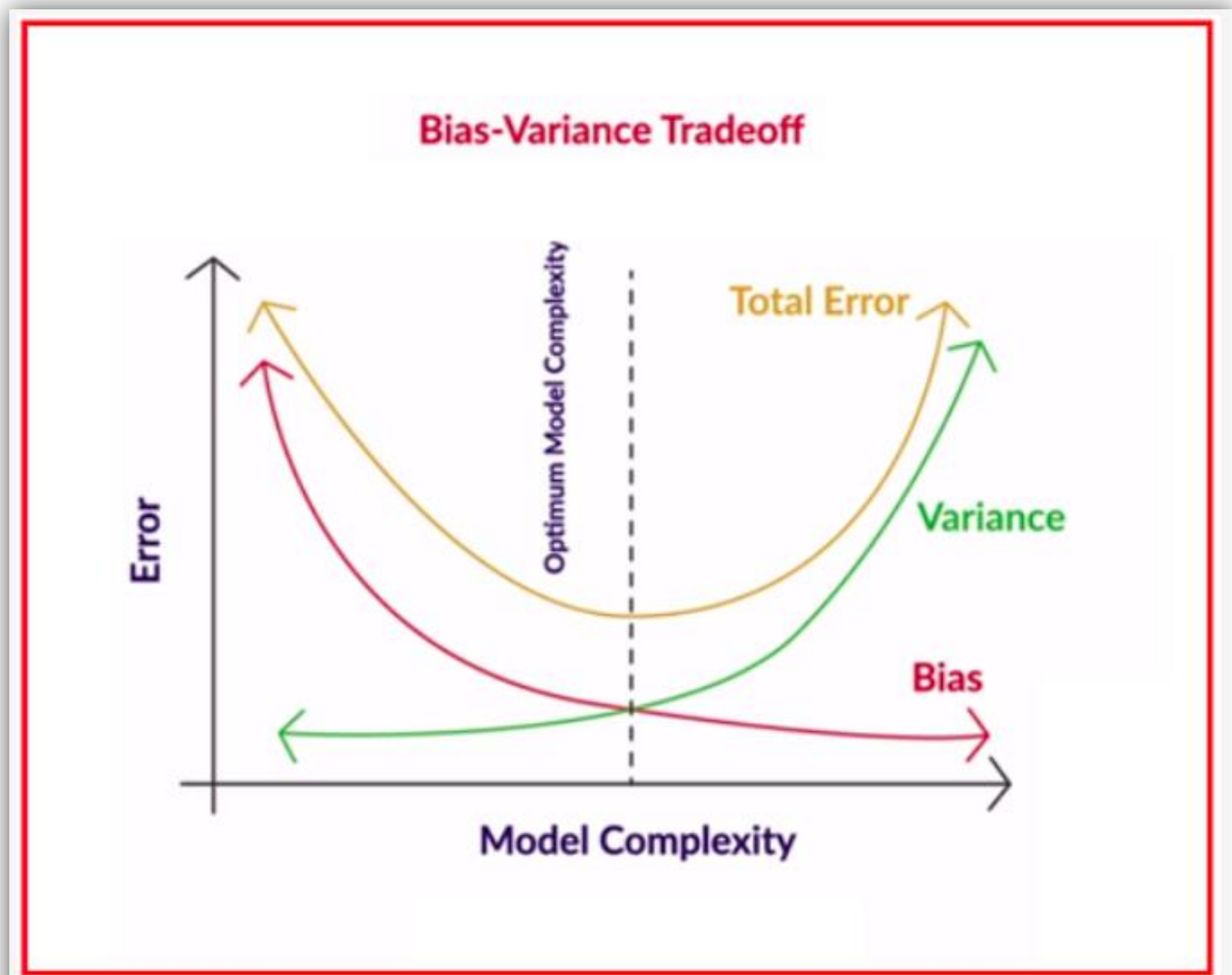
In order to make model robust and generalizable we need to cater Bias and Variance trade-off. The Target is to get Low variance, Low Bias Model but practically it is difficult to achieve it.



- Model with high bias pays very little attention to the training data and oversimplifies the model. It always leads to high error on training and test data.
- Model with high variance pays a lot of attention to training data and does not generalize the data which it has not seen before. As a result, such models perform very well on training data but has high error rates on test data.



- To build a good model, we need to find a good balance between bias and variance such that it minimizes the total error.



Bias Variance Tradeoff

- In Ridge regression, we add a penalty by way of a tuning parameter called lambda which is chosen using cross validation. The idea is to make the fit small by making the residual sum or squares small plus adding a shrinkage penalty. The shrinkage penalty is lambda times the sum of squares of the coefficients so coefficients that get too large are penalized. As lambda gets larger, the bias is unchanged, but the variance drops. The drawback of ridge is that it doesn't select variables. It includes all the variables in the final model.
- In Lasso regression, the penalty is the sum of the absolute values of the coefficients. Lasso shrinks the coefficient estimates towards zero and it has the effect of setting variables exactly equal to zero when lambda is large

enough while ridge does not. Hence, much like the best subset selection method, lasso performs variable selection. The tuning parameter lambda is chosen by cross validation. When lambda is small, the result is essentially the least squares estimates. As lambda increases, shrinkage occurs so that variables that are at zero can be thrown away. So, a major advantage of lasso is that it is a combination of both shrinkage and selection of variables. In cases with very large number of features, lasso allows us to efficiently find the sparse model that involve a small subset of the features.