Assignment-based Subjective Questions :

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?

ANS: The optimal value of ridge and lasso regression are as follows:

Ridge: 20.0 , Lasso: 0.001.

If we double the alpha value, the accuracy changes and becomes low. As alpha increases, magnitude of coefficients are pushed closer to zero.

2) You have determined the optimal value of lambda for ridge and lasso regression during the assignment. No, which one will you choose to apply and why?

ANS: It is critical to regularise coefficients in order to improve prediction accuracy while also decreasing variance and making the model interpretable.

Lasso regression employs a tuning parameter known as lambda as the penalty, which is the absolute magnitude of the coefficient as determined by cross validation. Lasso shrink the coefficient towards zero as the lambda value increases, making the variable exactly equal to 0. Lasso can also perform variable selection.

When lambda is small, the model performs simple linear regression, as lambda increases, shrinkage occurs, and variables with 0 values are ignored by the model.

Ridge regression employs a tuning parameter known as lambda as the penalty in the square of magnitude of the coefficients as determined by cross validation. Using the penalty, the residual sum or square should be small. The penalty is lambda times the sum of squares of the coefficients, so the coefficients with higher values are penalised. As we increase the value of lambda, the variance in the model decreases while the bias remain constant. In contrast to Lasso Regression, Ridge Regression includes all variables in the final model.

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?

ANS: If this occurs, we must select the next five most important predictors.

They are as follows:

1) GarageArea
2) Neighborhood_StoneBr
3) 1stFlrSF
4) RoofMatl_Tar&Grv
5) BsmtFinType2

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?

ANS: The model should be as simple as possible, as this will reduce accuracy while increasing robustness and generalizability. The bias-variance trade off can also be used to understand it. The simpler the model, the greater the bias, but lower the variance and generalizability. Its accuracy implication is that a robust and generalizable model will perform equally well on both training and test data.

Variance is an error that occur in a model when it attempts to learn from the data. The high variance mean model will perform well on training data because it has been very well trained on this data, but it performs extremely poor on test data, because the model has never seen this data before.

To avoid Overfitting and Underfitting of data, it is critical to maintain a balance of bias and variance.

A bias in a model occurs when the model is too weak to learn from the data. A high bias indicates That the model is unable to learn the details from the data. On training and testing data, the model performs poorly.