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 of alpha for ridge regression in the assignment was 1 and lasso regression was 0.0001. If the value of alpha is doubled, we are compromising more bias for variance reduction. Which would mean that the model would become more generalizable and might have a possibility of under fitting. So, if we increase the alpha value to double, we might get a higher training error. If the model underfits, which is unlikely for an alpha of 2 or 0.0002, we might also see a higher test error. We also expect that the parameters shrink more towards 0. The higher the value of alpha, the more the coefficients are shrinked towards 0

When we did this in the assignment notebook. We see a drop in both training and test R2 scores for both lasso and ridge regression which is expected as we have deviated from the optimal alpha value. But the drop is very less. We also observed that most parameter coefficients shrink towards 0, which is expected on higher values of alpha.

The most important predictor variables on the model for double values of optimal alpha are

a. Ridge regression → OverallQual, GrLivArea, GarageCars, OverallCond, MSSubClass_60
b. Lasso regression → GrLivArea, OverallQual, GarageCars, OverallCond, Neighborhood, NeighborhoodNridgHt

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?**

We found that both ridge and lasso regression contribute to simplifying and generalizing a model reasonably for optimal values of alpha. In the assignment notebook we found out that the test r2 score was much better for lasso regression model than the ridge regression model. Our lasso regression model performed better on unseen data.

Not only this, lasso regression performs feature selection and eliminates features that are not relevant to the model by equating its coefficients to 0. Ridge regression can only push it close to 0 and not 0, that too only on higher values of alpha. Hence this effects model interpretability.

I would choose lasso regression for 2 reasons

a. Better performance on unseen data
b. Better interpretation of the model. Since, lasso performs feature selection, and eliminates irrelevant features from the model. It gives better understanding of which of the predictor variables are important to the model and hence this can help in making better business decisions with clarity.

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?

Five most important variables → GrLivArea, OverallQual, OverallCond, GarageCars, MSSubClass_160

After removing the above 5 variables and retraining

TotalBsmtSF, Neighborhood_NoRidge, LotArea, FullBath, Neighborhood_StoneBR

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?

A machine learning model can be made robust and generalizable by choosing the right variance and bias for the model. This concept is called bias variance trade off. Bias is the training error made on a model. We do want the error on the training data set to be low, but not to a point where the model has just memorized the data.

If a model memorizes data, it is highly unlikely to do well on unseen data or test data. A model must be judges on its ability to perform on unseen data and not training data. Such a model is said to have low bias, but high variance. High variance because it is sensitive to changes in the data and has not understood or identified any patterns on the data which is going to lead to a bad test score.

To strike the right balance between both bias and variance, we use regularization. Regularization adds a penalty term to the cost function, which causes the model to trade some bias for variance reduction. This causes the parameter coefficients to be shrunk towards 0, hence making the model less susceptible to variations in the values of a particular predictor. There are 2 types of regularizations possible

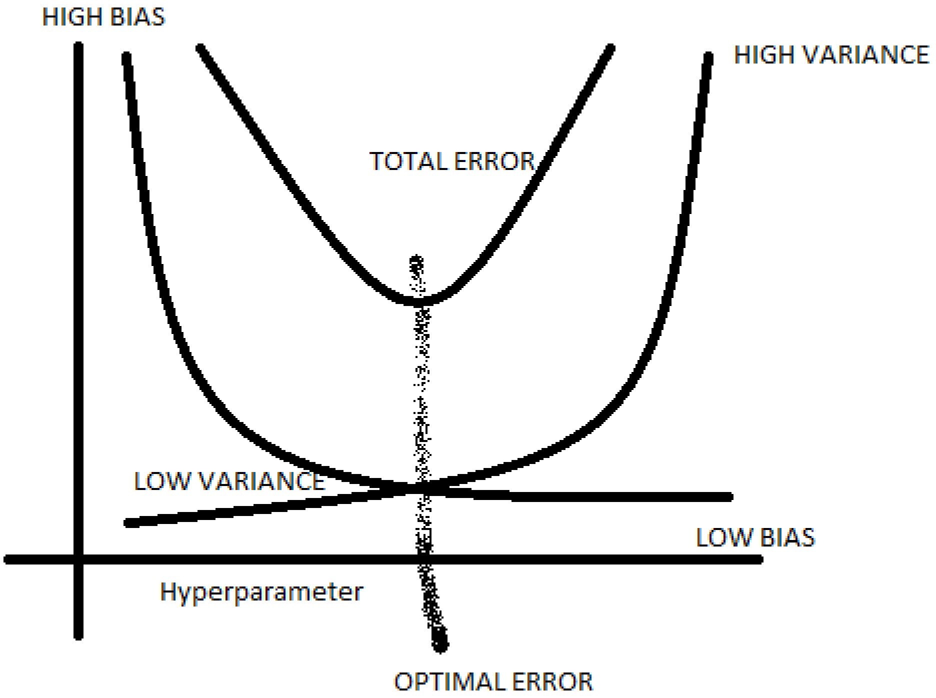
a. Ridge: - Adds the square of the coefficient times alpha to the cost function. This type of regularization causes coefficients to be shrunk close to 0 on large values of alpha, but not exactly 0.
b. Lasso: - Adds the modulus of the coefficient times alpha to the coefficients. This causes a lot of feature coefficients to be 0 on optimal alpha, hence performing feature selection and better model interpretability.

The implications on accuracy

Non regularized and complex models can tend to overfit. The accuracy score would be good on the training data set, but would be very less on the test data set.

Regularized models with optimal value of alpha perform well on both training and test data set.

Regularized models with extremely large values of alpha can cause the model to underfit, which causes it to have low accuracy on both test and training data sets.


HIGH BIAS
HIGH VARIANCE
TOTAL ERROR
LOW VARIANCE
LOW BIAS
Hyperparameter
OPTIMAL ERROR