**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 below are the alpha values for ridge and lasso regression:

**Ridge Regression:** We have plotted Alpha Value vs. Negative Mean Abs. Error to find the optimum alpha value, which turns seems to be 2 but we have chosen an Aplha value of 13.22 as the best value as the error terms at Alpha=13.22 is the least among others.

**Lasso Regression:** The optimum value for Alpha is 10.

When we double the value alpha value for both Ridge and Lasso regression, then the coefficient value will tend to get smaller and will help with model generalization and will help prevent overfitting but very large value of alpha might lead to underfitting of the model. For lasso regression when alpha is doubled more co-efficient of predictors will tend to become zero/insignificant, it will help prevent overfitting but very large value of alpha might lead to underfitting of the model trained.

The top predictors post doubling alpha values will now be :

**Ridge Regression:**

RoofMatl\_\*

2ndFlrSF

TotalBsmtSF

MSZoning\_RL

GrLivArea

OverallQual

**Lasso Regression:**

RoofMatl\_\*

2ndFlrSF

1stFlrSF

TotalBsmtSF

**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?

* To create a more generalized model, we need to keep a check on the model complexity and performance on unseen/test data. The model should be able to perform with similar accuracy for both test and train data. In ridge regression the model is having better accuracy but as we increase the value of lambda the variance in model decreases but the model retains all the variables in the final model unlike using RFE or Lasso regression which might not help us creating more generic and performant model.
* In Lasso regression as we tune and increase alpha/lambda value the value of co-efficients shrinks and tends to become zero, this helps us feature elimination/selection and making more robust and generic model.

**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?

* The top 5 variables after removing the prior most top 5 variables are :

GrLivArea

MSZoning\_RL

OverallQual

HouseStyle\_1Story

Neighborhood\_NoRidge

**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?

* A model should be as simple as possible in order to make it more robust and generalizable. We can select the optimum complexity of a model by studying Bias-Variance trade off.

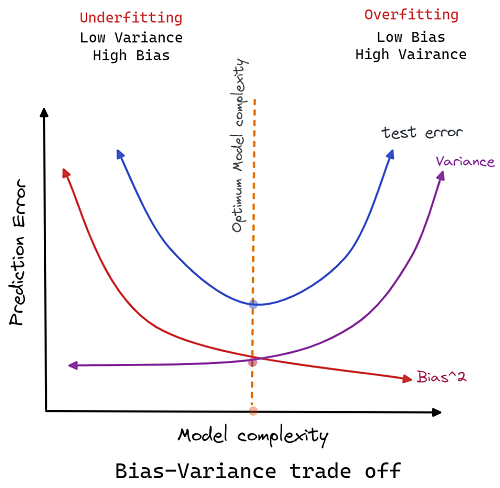


Image Reference: <https://blog.devgenius.io/bias-variance-trade-off-c4139fc798b7>

Making a model generic and robust will have the below effects on the model accuracy:

**Accuracy:** The accuracy will be lower than other models on test and train datasets but the model will be able to perform on unseen datasets and will be the best fit overall.

**Bias:** Bias is the error in the model and when model is generic the bias will be more.

**Variance:** It refers to the degree of changes in model with some test data w.r.t to changes in the test data. A generic model will be able to perform better than an overfitted or underfitted model which is the case for a balanced model/generic and robust model. We would need to perform the regularization to prevent model becoming too complex or overfitting.