# Assignment: Advance Regression

Question 1:

Answer:

If a model has training accuracy 97% and test accuracy 48%, it is called **over-fitting model.** This type of model has **high variance Low bias** trad-off.

Reason:

Model complexity could be reason of this behavior of model. Model **complexity** might be cause of train model very large number of predictors.

Solution:

We need to make **model very simple**. We have to use method which can penalized large number of independent variables.

We can use here **generalization and regularization** to make model simple.

**Ridge and Lasso** are good way to make simple model.

Question 2:

Answer:

**Difference L1 and L2 regularisation in regression:**

Main difference is : L1 Regularization is known as Lasso regression which use absolute magnitude value to penalize complex model. L1 feature selection internally.

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

L2 Regularization is known as Ridge regression which use square value of magnitude in cost function to make simple model.

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

Q3:

Ans:

L1: y = 39.76x + 32.648628

I would like go with L1 model.

L1 model less complex to compare than model L2. In regularization, our motive is reduce constant value.

Q4:

Ans:

I will  compare model score of train and test dataset. If both scores are approx. same then model is robust and generalizable.

Max number of variables are cause of complex model.

Test trained model using unseen dataset then check model variance and bias trade-off.

High variance and low bias is way to found generalization of the model.

Q5 :

Answer: