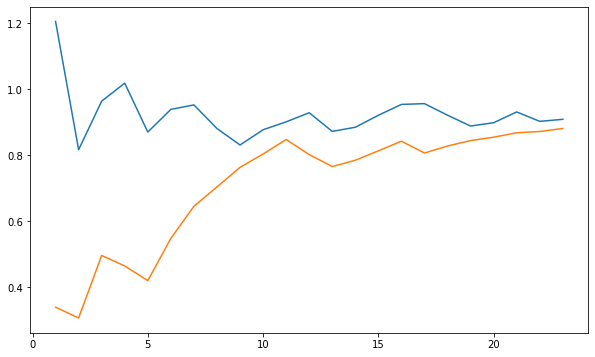
**Data\_analysis** is made solely to understand data better in a more formulated and better way.

Following procedure has been formulated, as per the course of Andrew Ng from coursera.

1. Separate Training, Validation and Test data sets.

Data is separated in data\_seperation.

1. Plot the learning curves, to understand the type of issue.



Learning Curve – Orange is training error, and Blue is val error.

As we can see, over increasing no of samples, Both error converge to a common point, and hence there is an error of .9

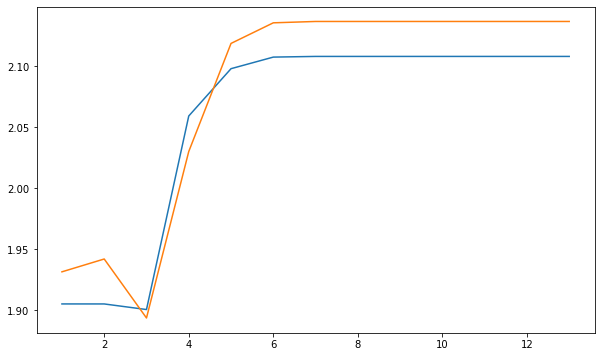
This shows that current data set is suffering from high Bias.

Therefore, either we can add more samples, or we can include polynomial features.

Next we check which polynomial degree would be best for features.

1. Plot the error curves for various degree of features, to obtain the best possible degree for use.

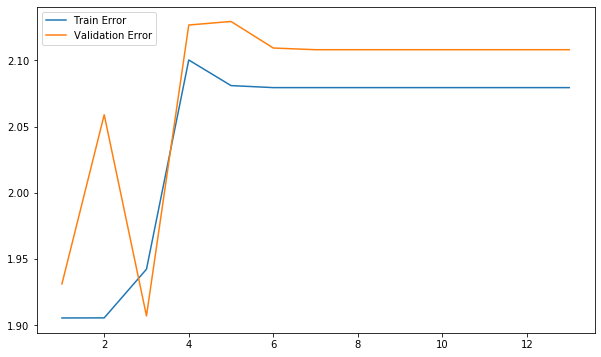
Here we plot errors, with increase in polynomial degree.

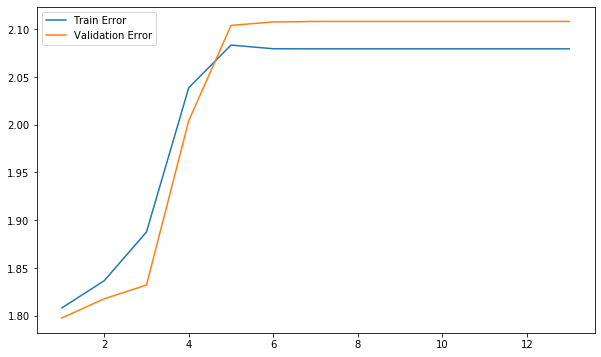


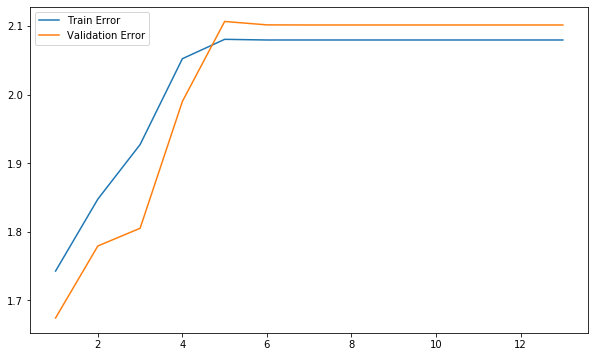
In this curev, blue is the training erroe, and orange is the val error.

Min point of val error, is the optimal degree. We have run this on Sepla Length feature.

Hence we find the optimal degree for this is 3.







These are the degree optimisation plots for -- SepalWidthCm, PetalLengthCm & PetalWidthCm

1. Obtain optimal lambda by plotting error curves for different lambda