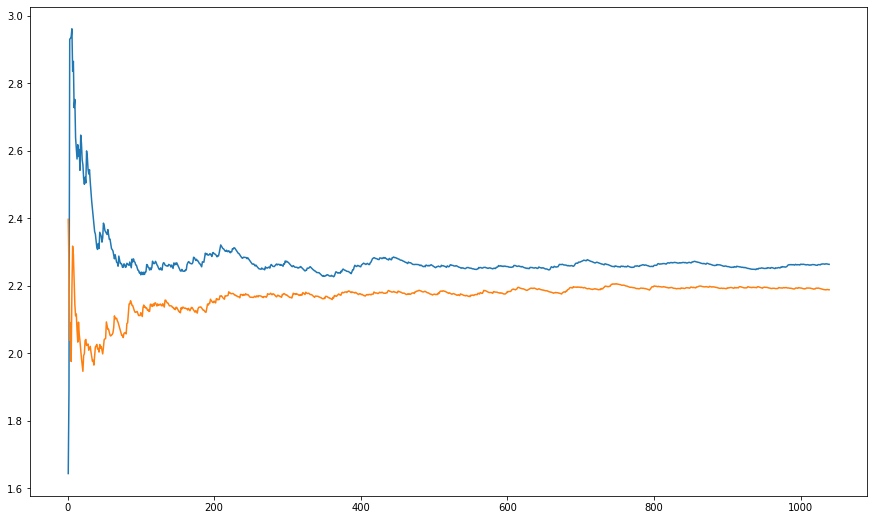
Documentation for Logistic Regression on Wine Quality Data.

1. Data is imputed to fill out the missing elements.
2. Data is normalised, and then separated into Test, Train and Validation sets.
3. Learning curve for the separated train and validation data is plotted. This results in understanding that current data is suffering from high bias problem.

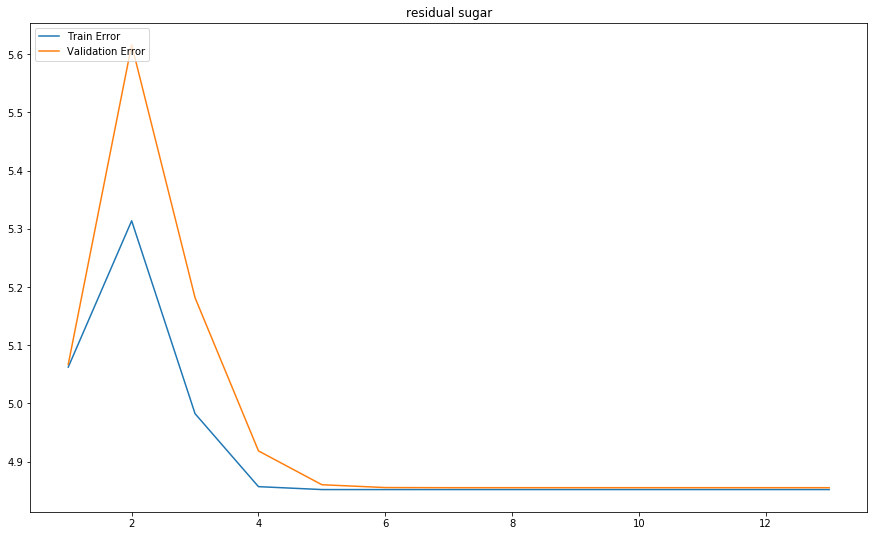


BLUE – Training Error

Orange – Validation Error

As we can see, both converge, but there is net error present, therefore it is over bias problem.

1. Now we plot the Training error and validation error, for finding out the optimal degree for each feature. The error is plotted with degree on X axis.



The above graph is for ‘residual sugar’. As e can see from graph, it tends to perform best with a power of 5, with min val and train error.

Similarly, best degree is found for every feature, and then various experiments are done accordingly.

**ALL Graphs**

1. Now we find optimal values of lambda for various data sets we are taking.