Simple Linear Regression

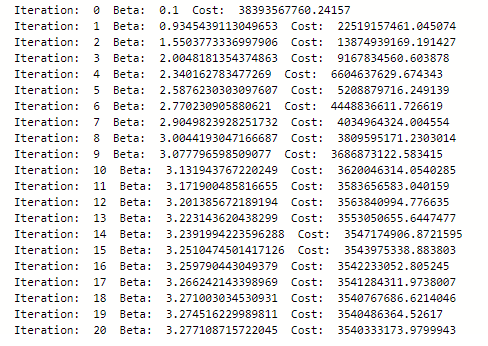
The following exercise is a simple linear regression using batch gradient descent written from scratch without any usage of any libraries in pyspark. The data set used is attached to this document. The data set comprises of the details from New York City taxi trip reports from the year 2013. I will be finding the weights of the linear regression problem (beta) with reducing our loss function (cost).

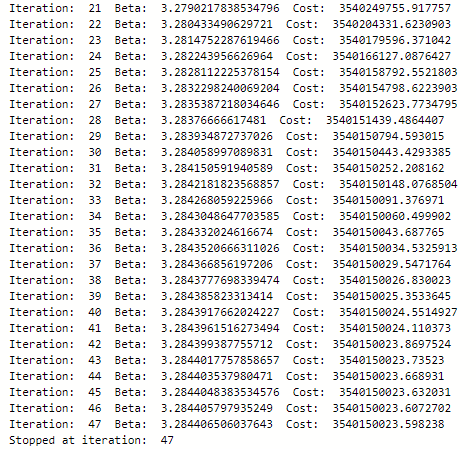
The independent variable chosen is the trip distance while dependent variable was the total amount.

The following results are calculated on the cloud (GCP).

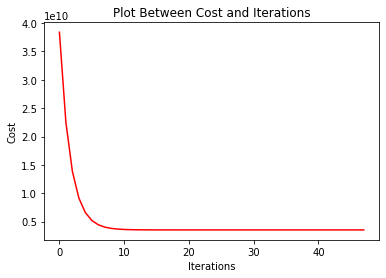
Slope: 3.2884 Time Taken: 1hr 24 seconds

Output:





Plot of the cost in each iteration



The Jupyter Notebook attached is done on the smaller version of the data set. The final weights can be used to predict on independent variables.