# Learning to Rank using Linear Regression

Haoquan Zhao

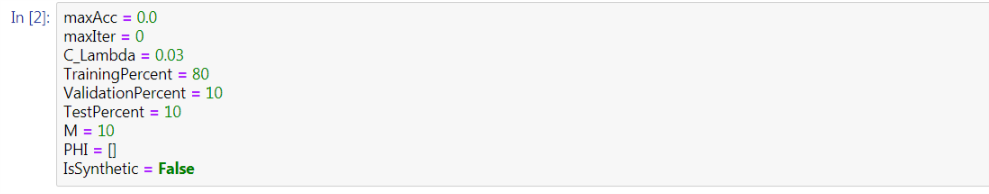
50189594

haoquanz@buffalo.edu

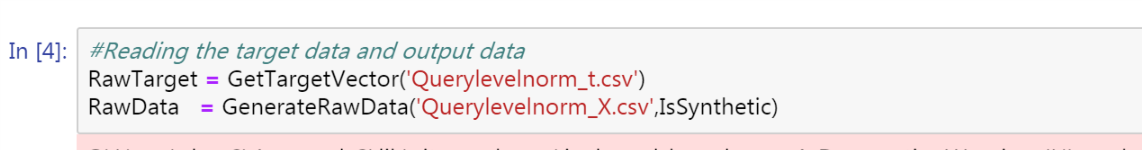
# 2.Introduction.

In this part, we apply linear regression on a synthetic dataset from microsoft LETOR 4.0, train our model on part of it and evaluate the performance on another part, then by tuning Hyper-Parameters, we get to know with which value of k can we get the minimum error. And apply this into test part of dataset. And finally, by using Root Mean Square error to evaluate the solution on a test part of the dataset, we could find out how dose the solution work.

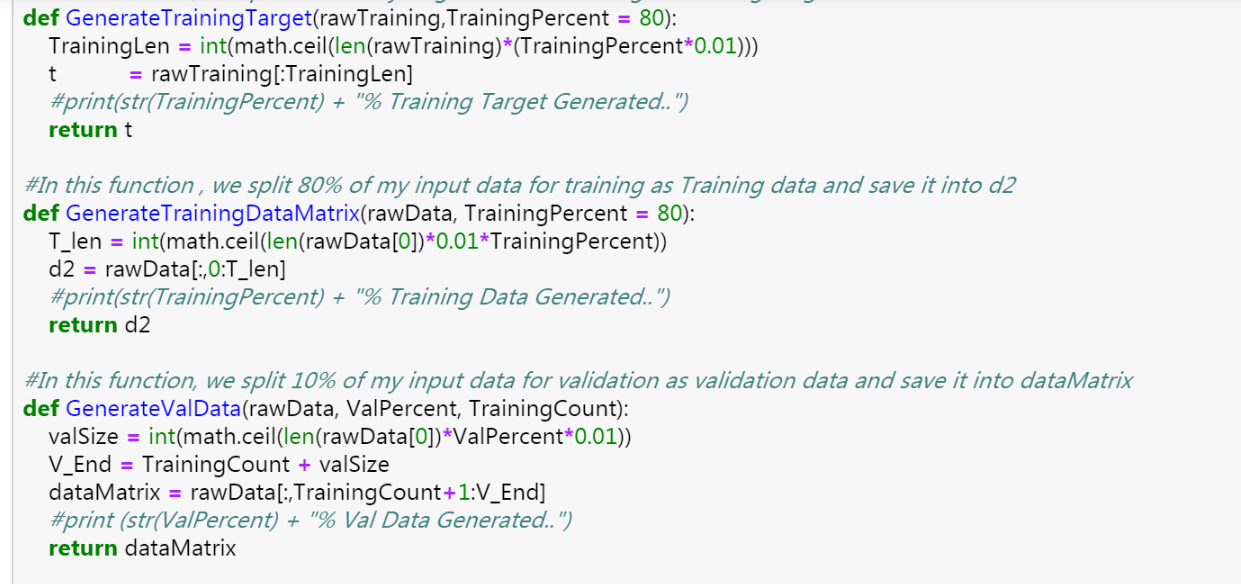
# 3. LeToR Dataset



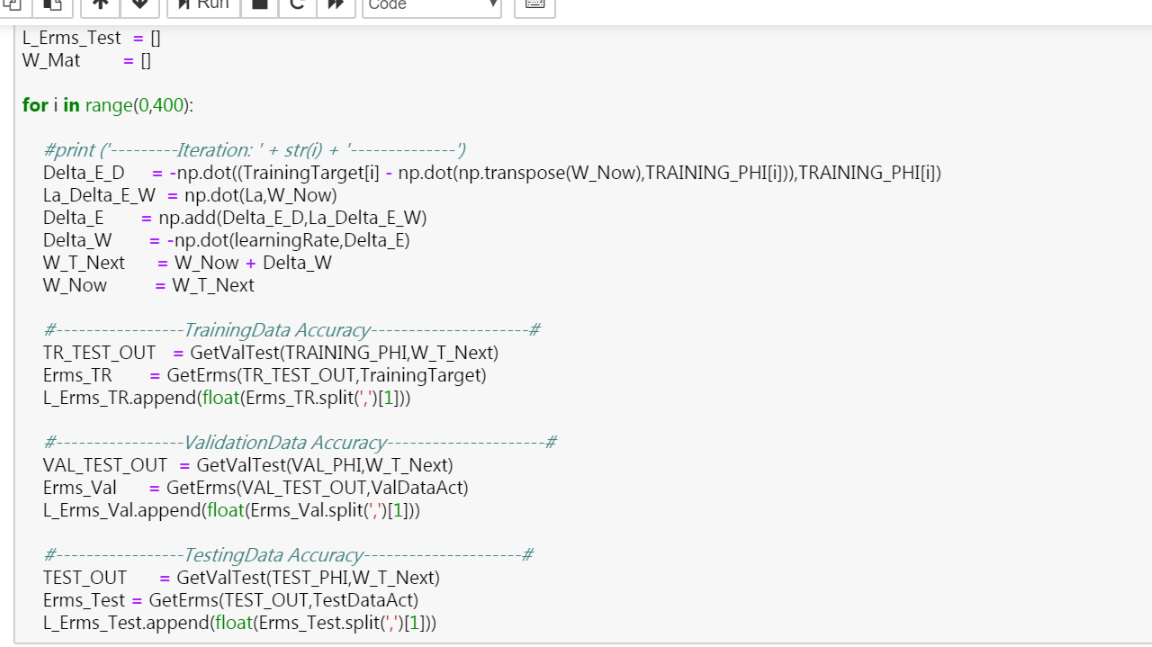
In this part, we initialized the parameter that we need to use to do LeToR.



In this part, we reading the dataset we need to use to do LeToR.

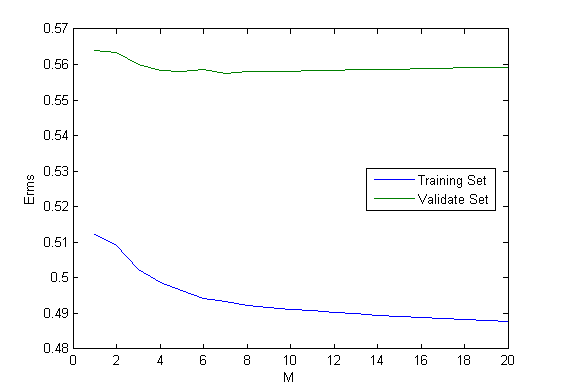


In this part, we do Partition to decide the whole dataset into three parts. The first one is used to train our model, so that we could generate the optimal parameter set as expected. The second part is called valid part, which is used as a confirmation that the model we build by training the train part dataset is reliable. The last part is reserved as test dataset, with which we could evaluate the efficiency of our model. 80% of the dataset is used as train data, 10% of the dataset is used as valid data, and the rest 10% is used for test. In the dataset, there are 69623 query-document pairs each consisting of 46features.



Then we define how dose SGD works in the dataset, firstly, we take a random initial value and by using, we get to know how those values changes in the whole dataset. In which, we also get, which is called weight update, it goes against with direction of gradient of the error.

# Comparison of different M



# Com