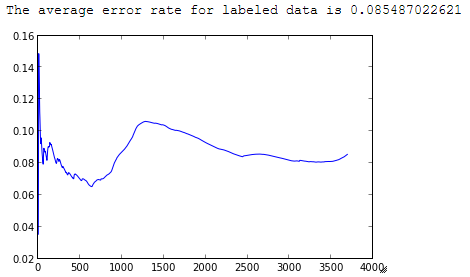
## learning\_rate = 0.3/t

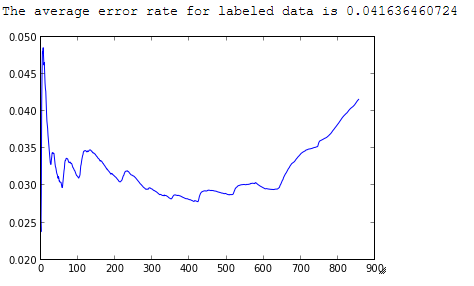
On the whole data set after Dec 20th



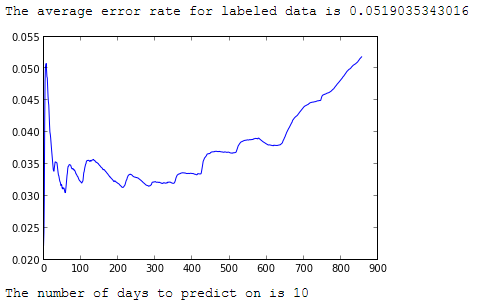
The first peak is due to initial training, and the second is around year new eve, when new data come.

Since it is going to take a while for entire data set, we used the results from the first 10 days as evalucaton of different learning rate, and then choose the one with lowest error rate. Then apply it on the whole data. For first ten days, we have 900\*1000 + tweets in total.

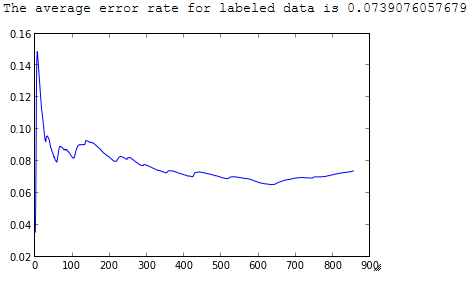
## learning\_rate = 0.1



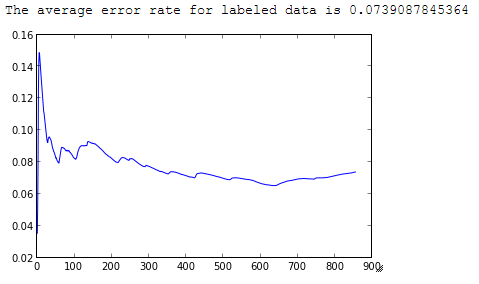
## learning\_rate = 0.5



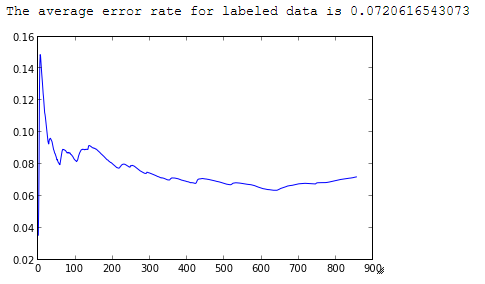
## learning\_rate = 1.0



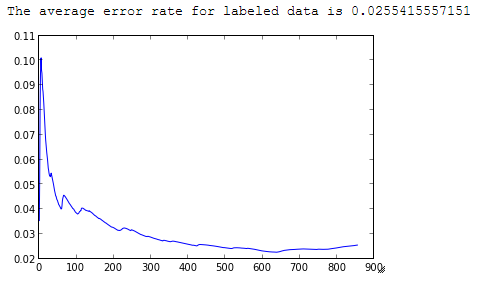
## learning\_rate = 0.3/t



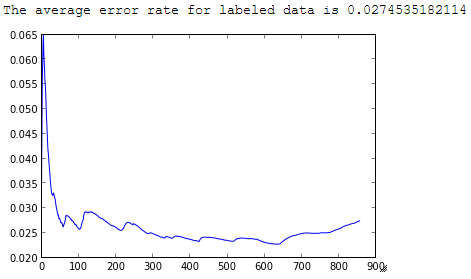
## learning\_rate = 1.0/t



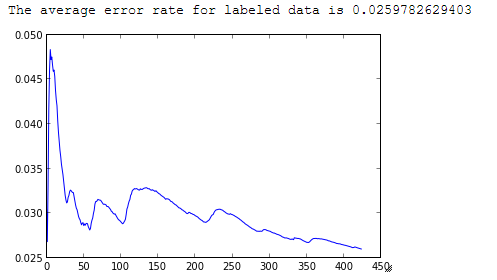
## learning\_rate = 2.0/t



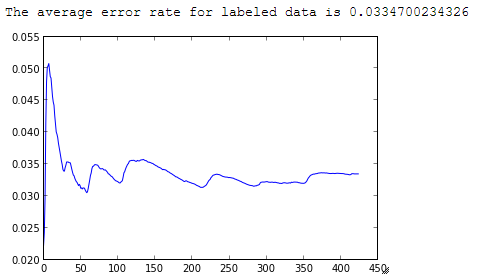
## learning\_rate = 0.3/sqrt(t)



## learning\_rate = 1/sqrt(t)



## learning\_rate = 5/sqrt(t)



It is stable around more than 0.03