Lab 3
Zheyi Yi

# Part 1.

This part record performance as a function of increasing training set sieze for A, B USPS sample.

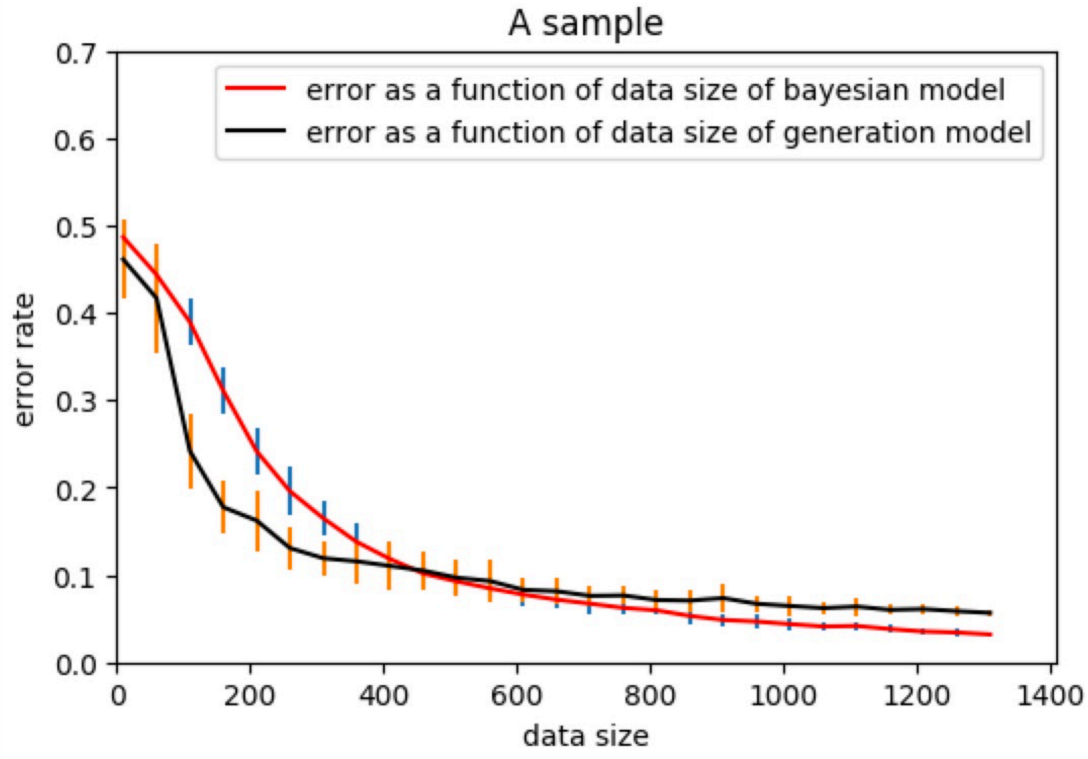



Figure 1
Figure 1 showed that the error rate generation model

Figure 1 showed the generative model performs better than Bayesian model in the beginning. Because the data does not conform to the Gaussian generative model but the data is linearly separable. For linear separable, it was expected that there are less outliers samples existing. Generative model performs better when less malformed data exists. Another is that Bayesian model assume a prior from Gaussian distribution but A sample does not conform to Gaussian model.

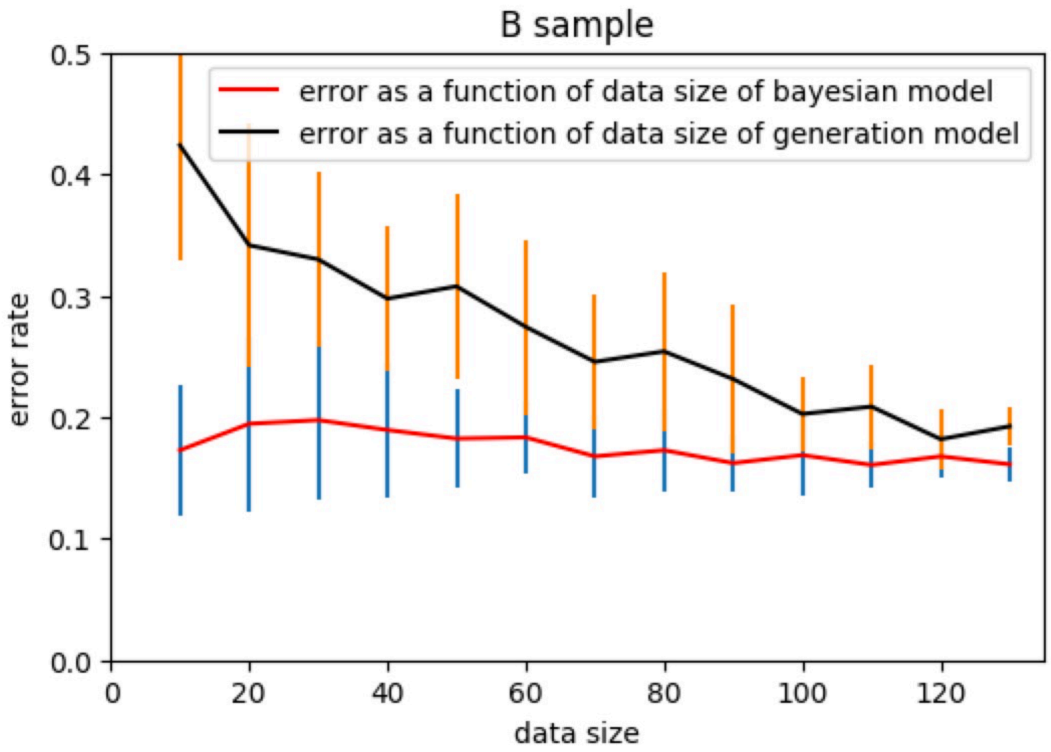



Figure2 shows that error rate of Bayesian model is lower than the generation model. Although B is generated from Gaussian model, it has different covariance structure. So it is quadratic separation instead of linear one. It might be the prior of the Bayesian model give it better performance.

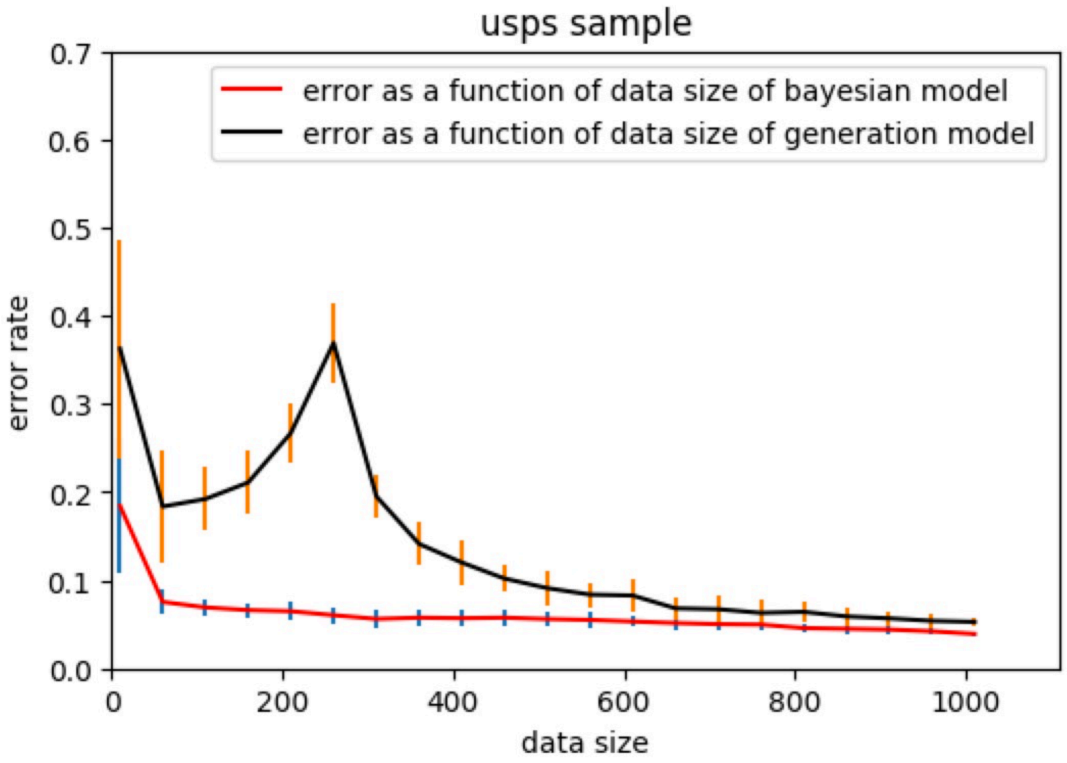



Figure 3 showed that there is fluctuation on generation model in the beginning. It might be some data is very near that it is hard to differ for generative model on small samples. The Bayesian model performs better than generative model.

Part 2.

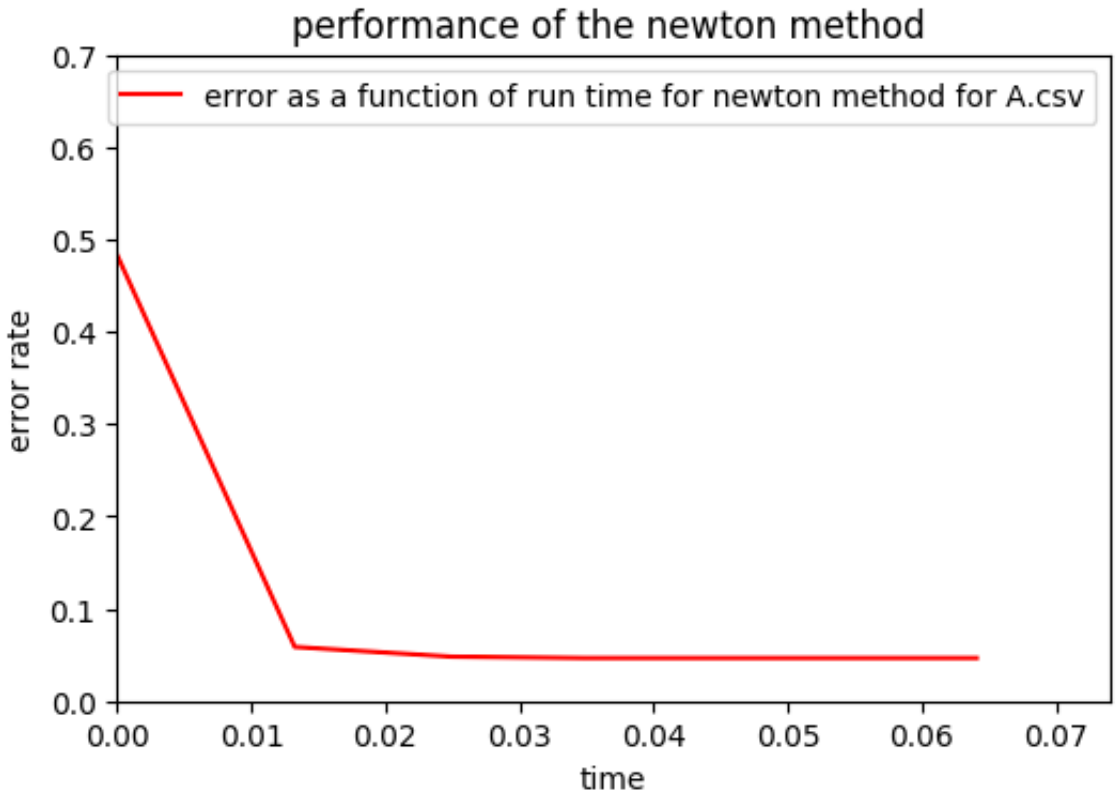

performance of the newton method
error as a function of run time for newton method for A.csv
error rate
time
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07

Figure 1

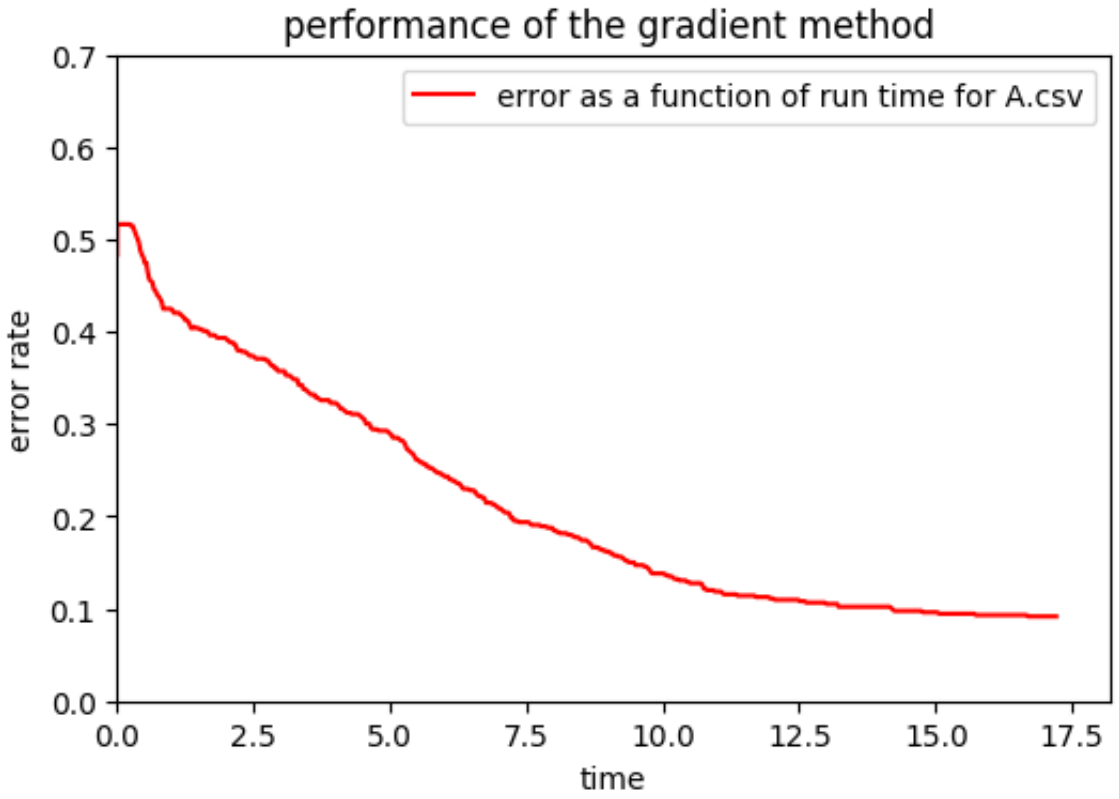

performance of the gradient method
error as a function of run time for A.csv
error rate
time
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

Figure 2

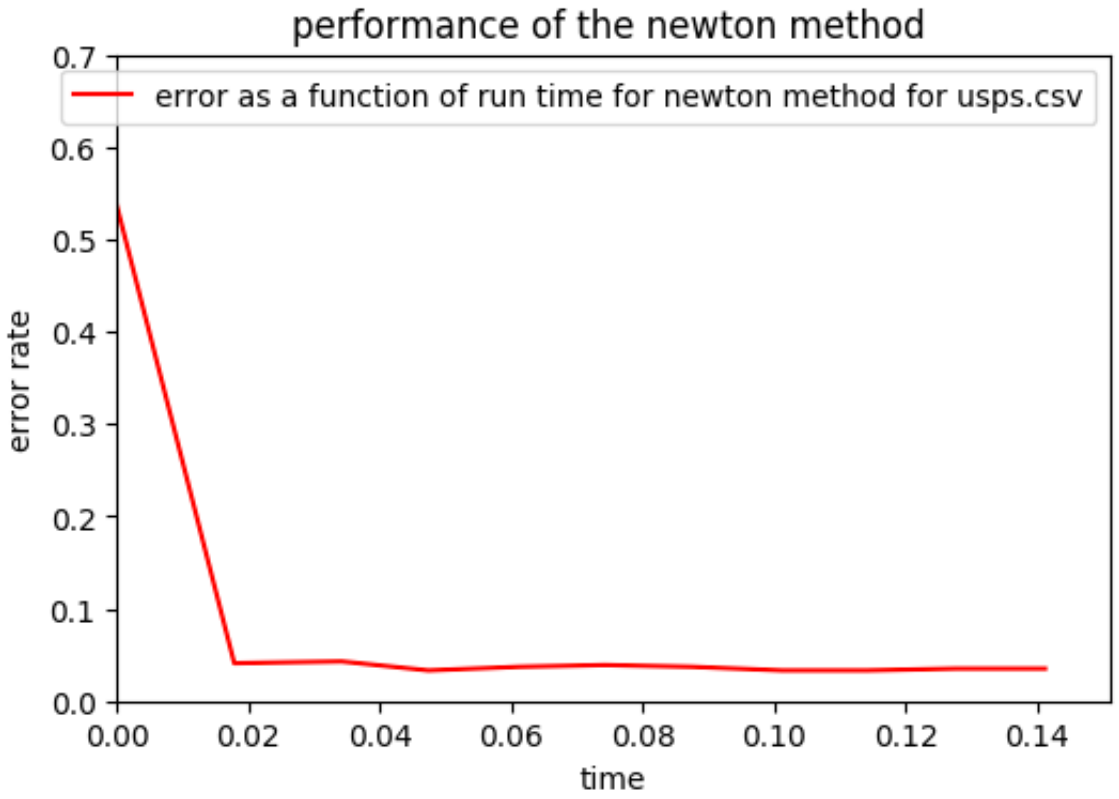

performance of the newton method
error as a function of run time for newton method for usps.csv
error rate
time
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14

Figure 3

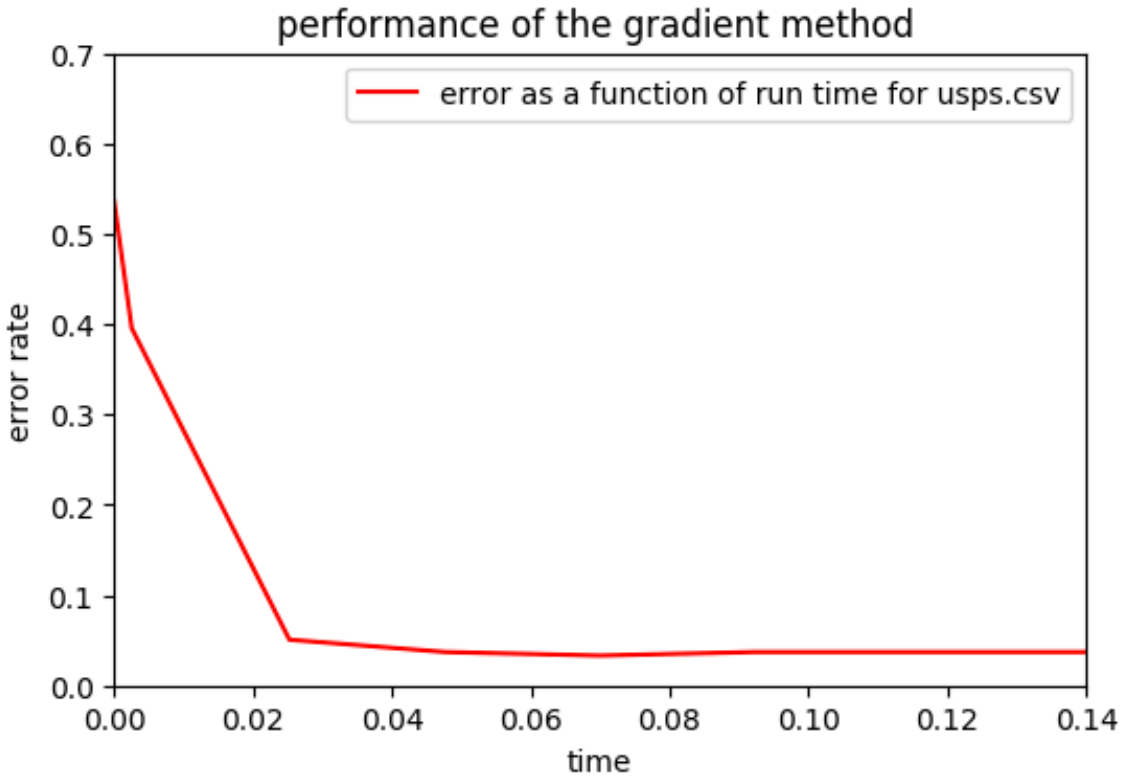

performance of the gradient method
error as a function of run time for usps.csv
error rate
time
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14

Figure 4

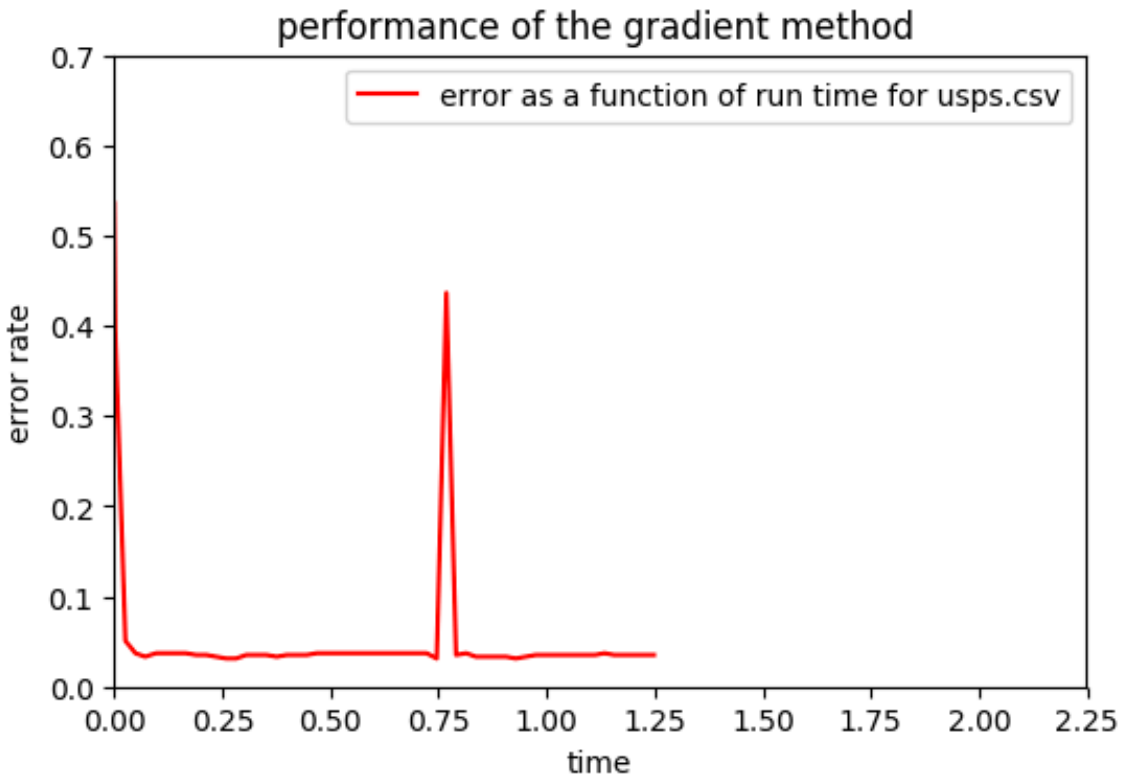
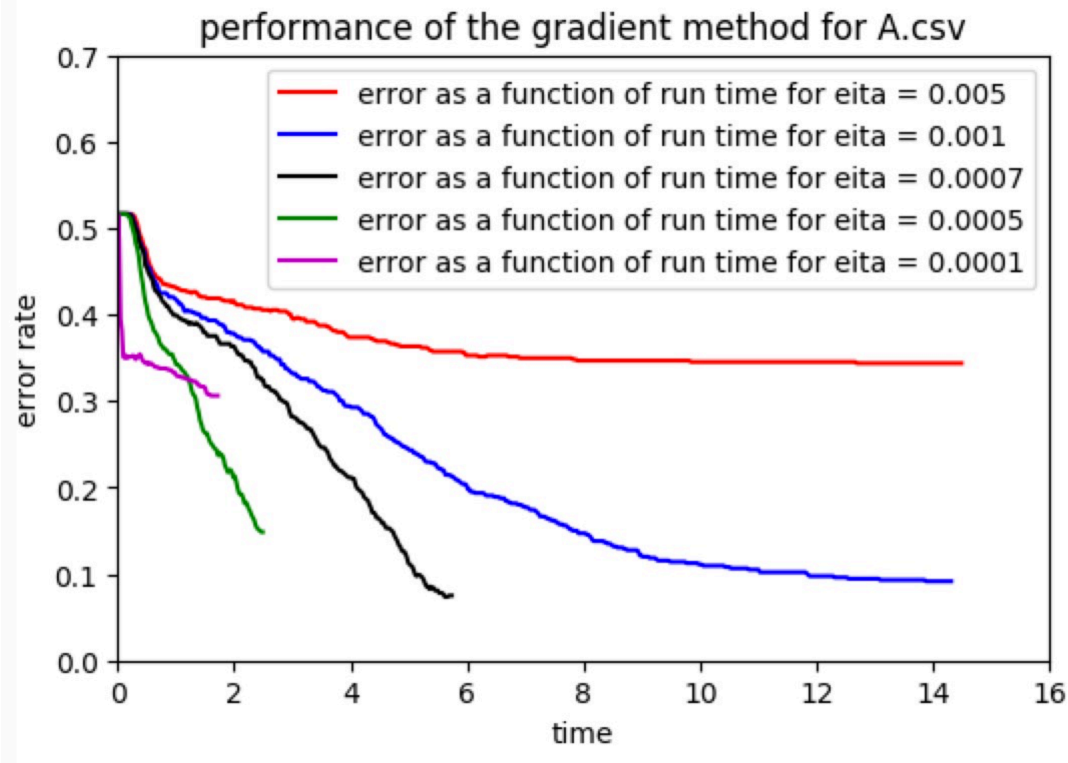



figure 5

From Figure 1 and 2,for A sample, we can see that it spend a few times to finish update w for newton, but for gradient method, it need many many times to finish update. So we can conclude newton update need much less time to finish update than gradient method.

From 3 and 4, for usps sample, we can see although gradient method need much more time to update, but for each time, gradient method is much faster than newton method. Usps sample contain large data so large matrix, newton method needs a large amount time to update one time so we can see their time to finish update doesn't differ much compared figure1 and 2.

Part 3.




I change eita from 0.005 to 0.0001 to see the error rate of A sample. If we see the line of eita = 0.0001. We can see, it stops error rate 0.3. The reason is that the update speed is so low that we don't have enough data to update it to the best w.
If we see the line of eita = 0.005, the error rate come to 0.4 which means the update change beyond the best value every time, so it cannot achieve the best w value.

If we come to the line of eita = 0.001 and 0.007, we can see they have the same lowest error rate which means we update to the best w, and the line of 0.007 is much fast than = eita. So we should find an automatic way to find best eita value every update.