PP3 REPORT

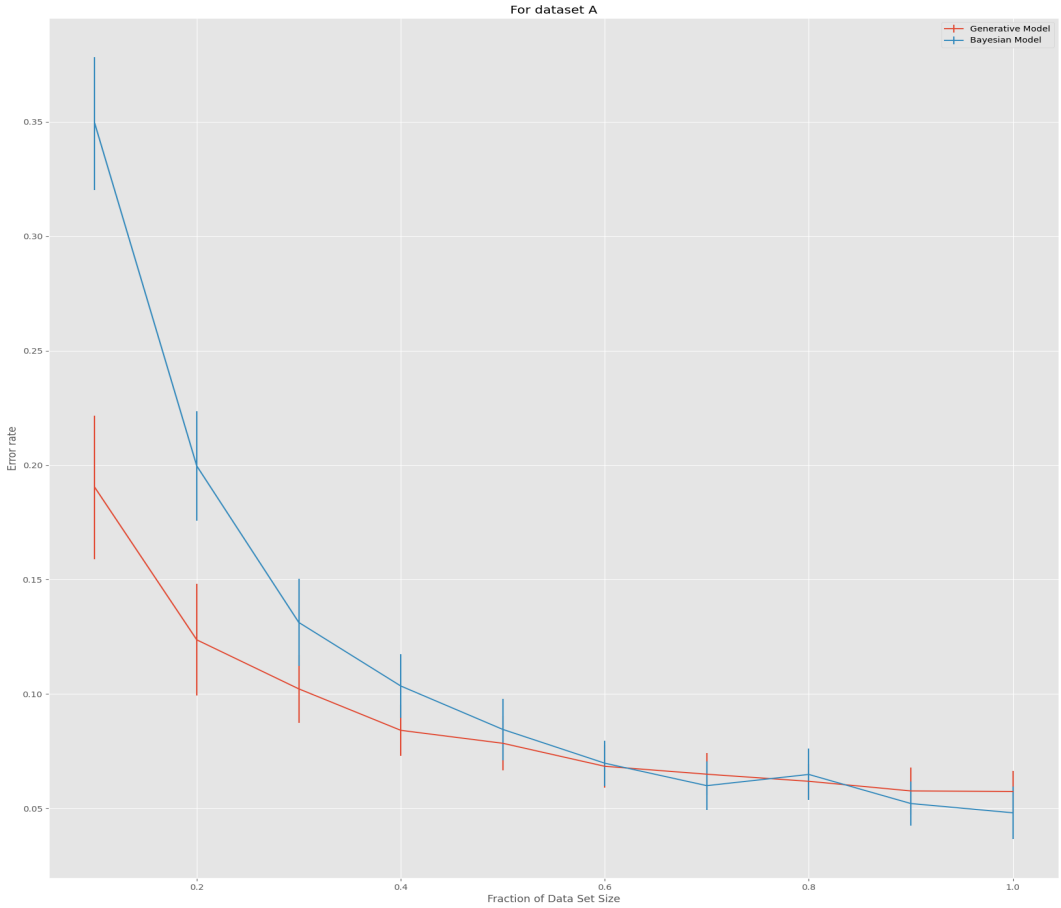
- This project has two tasks, first, it compares a generative model and a discriminative model (using Newton Update) on three datasets: A, B, and USPS. Also, we check our weights on the iris dataset. And in the second task, we compare the updated quality of Newton-Raphson and Gradient Descent

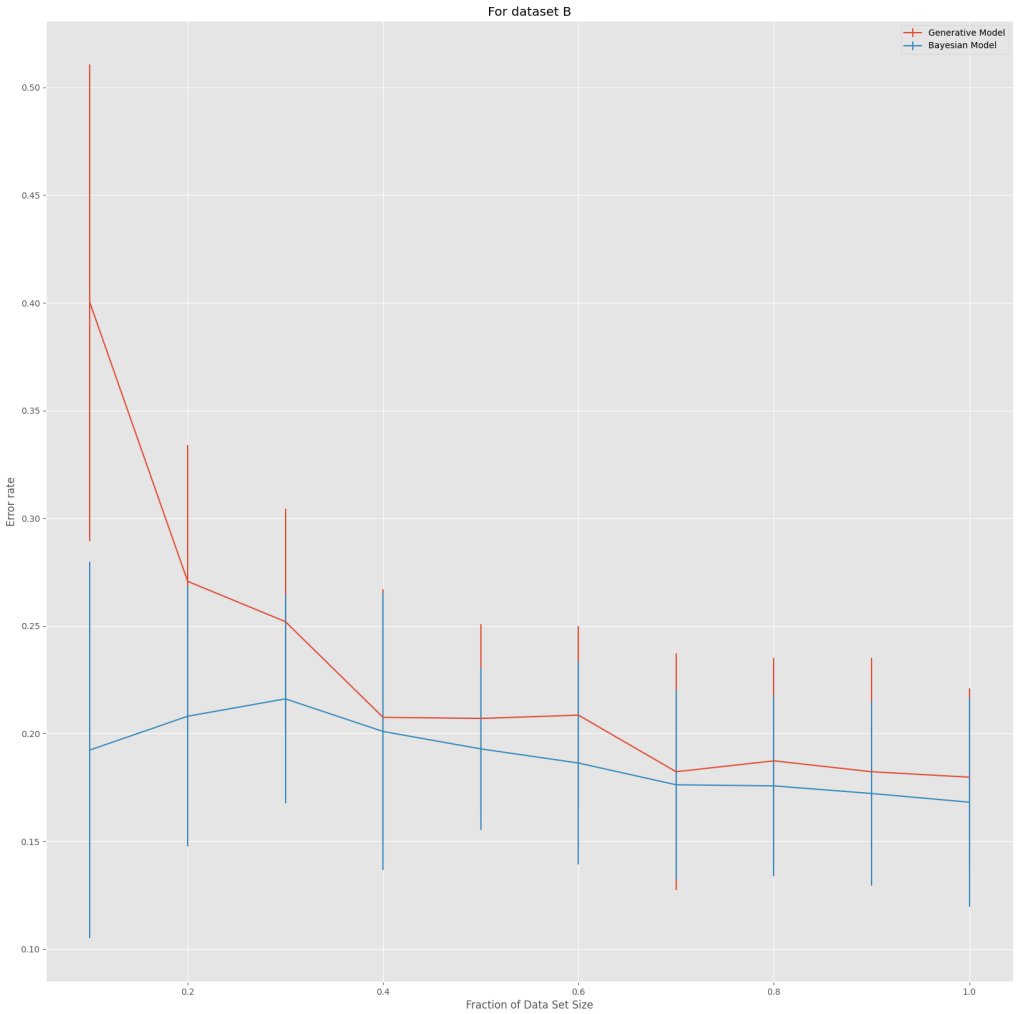
TASK-1:
Here we had to compare a generative model and a discriminative model

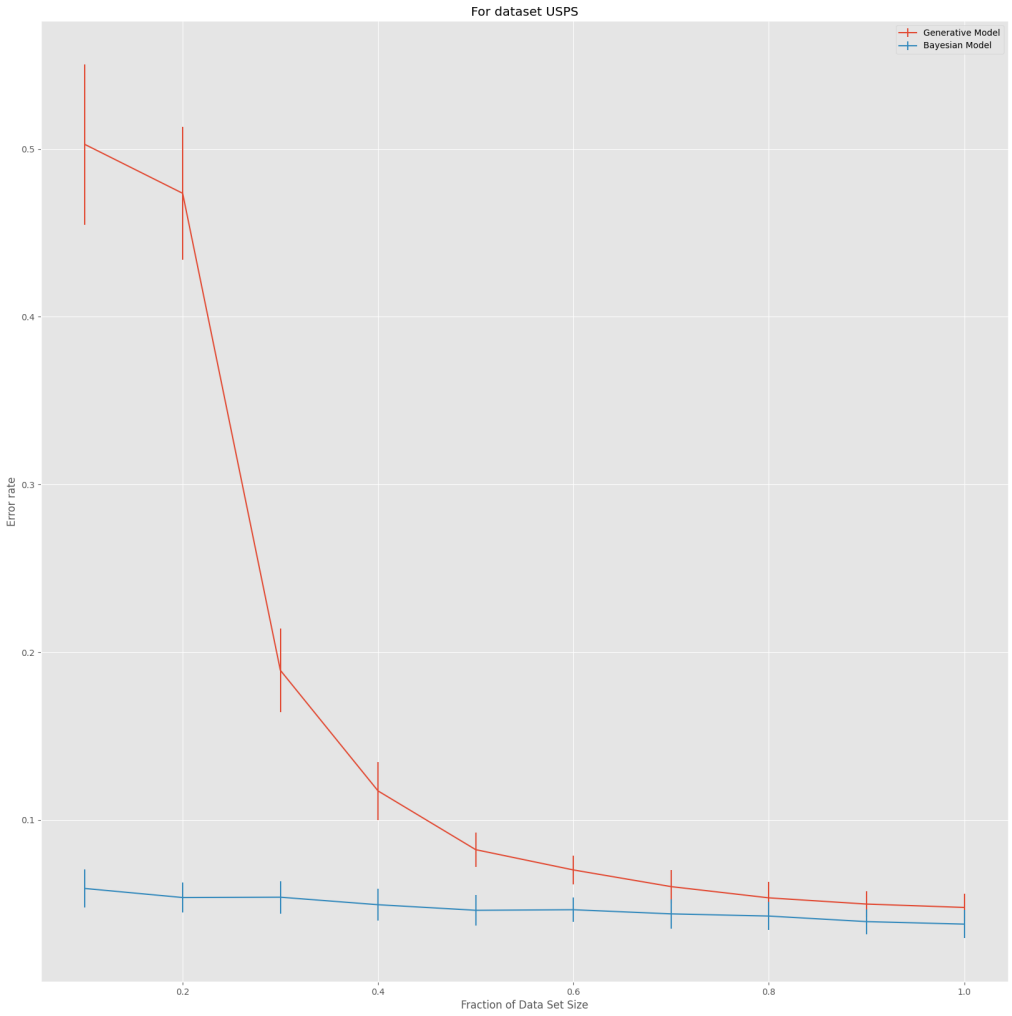
Observations

- For dataset B and USPS, the discriminative model performed well as compared to the generative model except in dataset A. As our assumption of shared covariance, thus a linearly separable line is wrong, so our discriminative model didn’t perform well. But for the rest, the discriminative model performed well
- Also, as the dataset size increases, the performance of both these models starts to converge. Also, we can see that the standard deviation of the errors for all 30 runs also

decreases as we increase the dataset size. It signifies that as we increase the dataset size, we overfit less, thus the performance is better with lower variance on the test set. The mean of the error also decreases as the size increases.

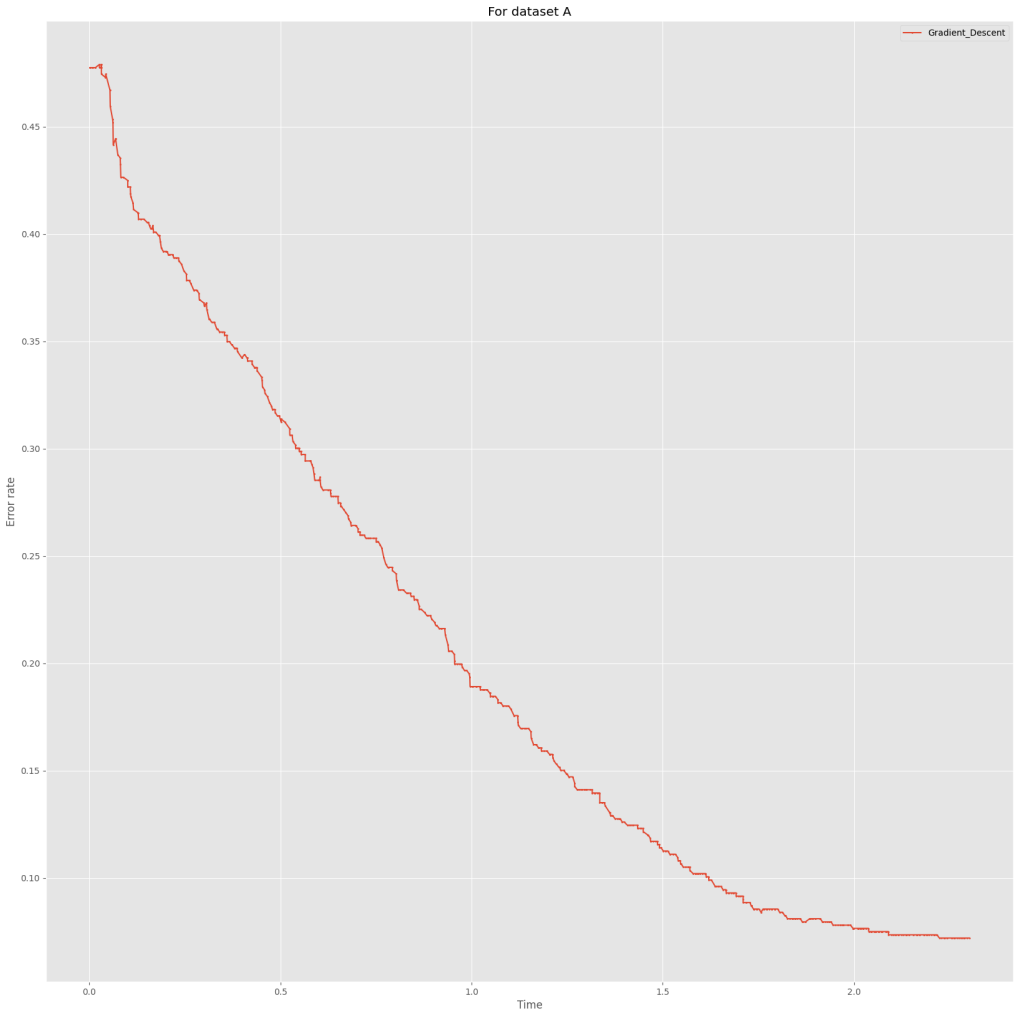


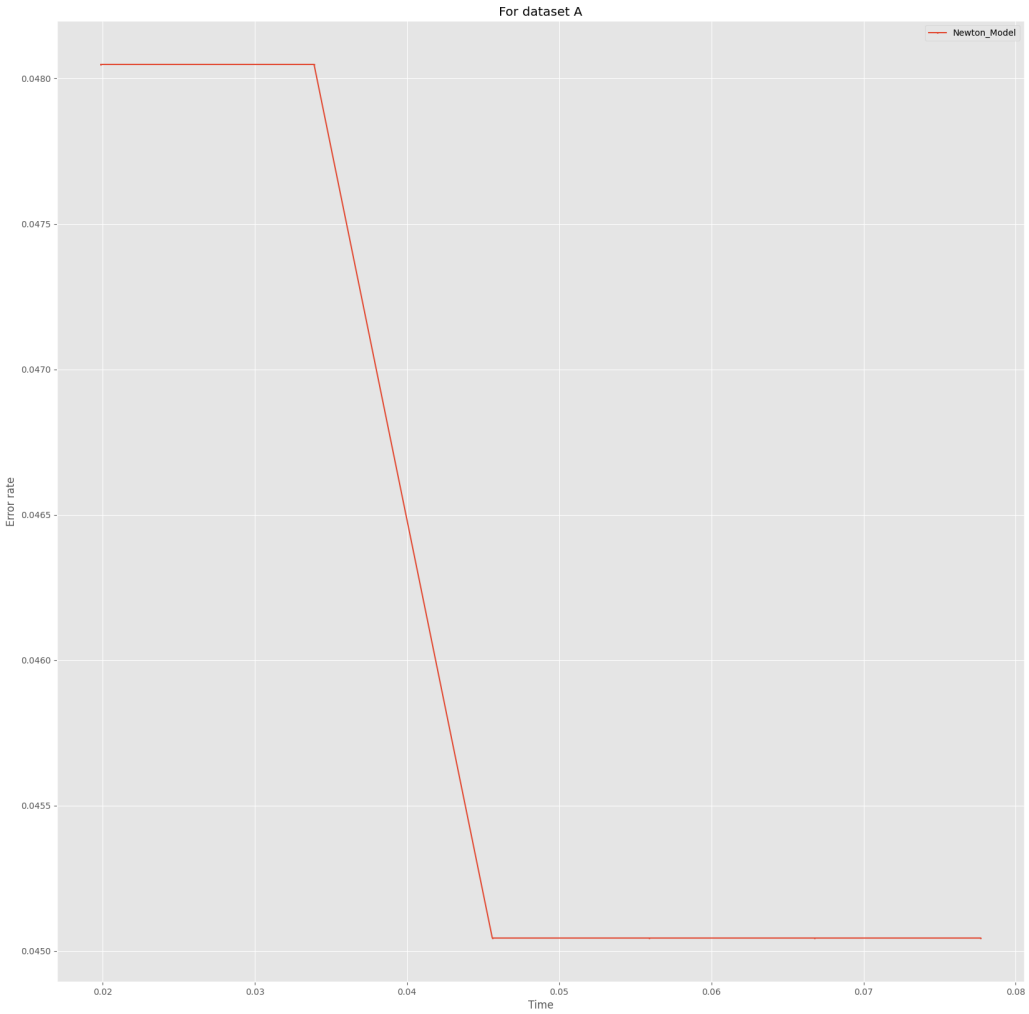

For dataset B
Generative Model
Bayesian Model
Error rate
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.2
0.4
0.6
0.8
1.0
Fraction of Data Set Size

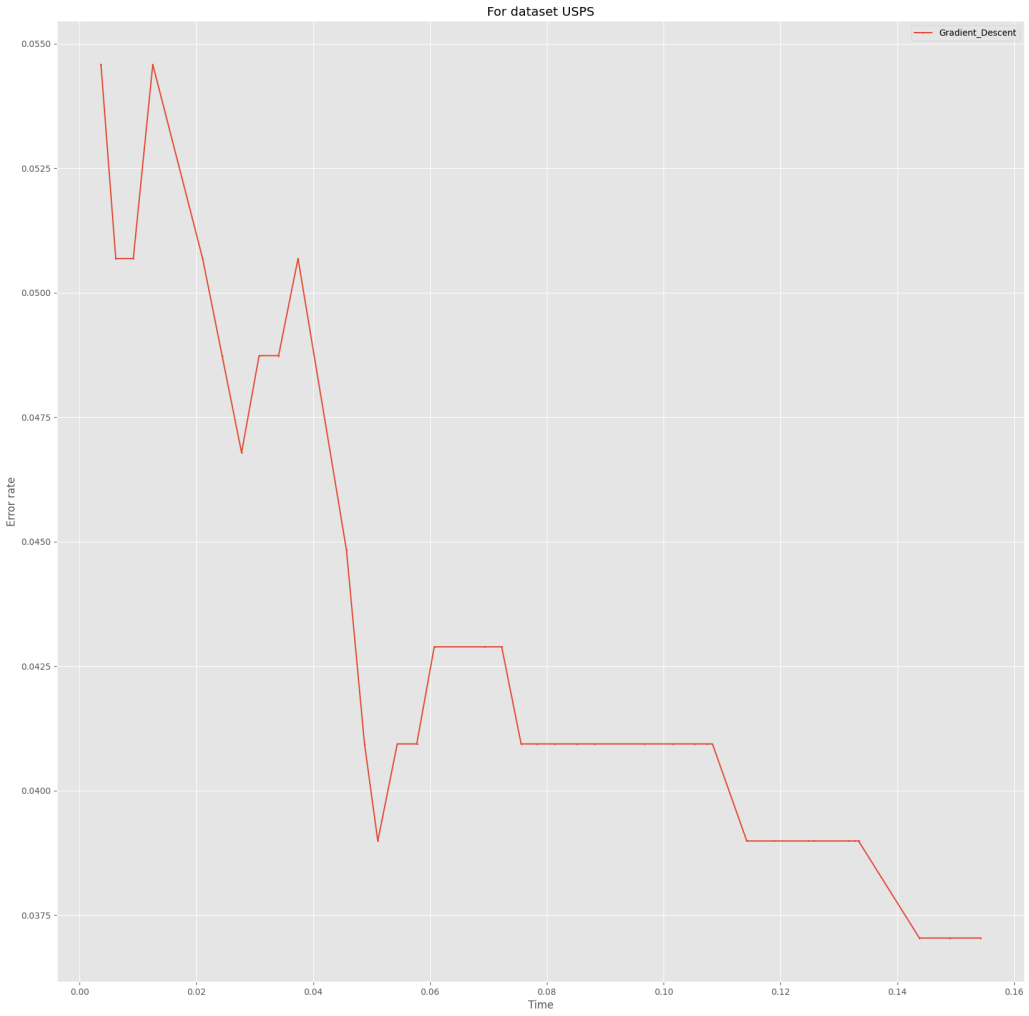


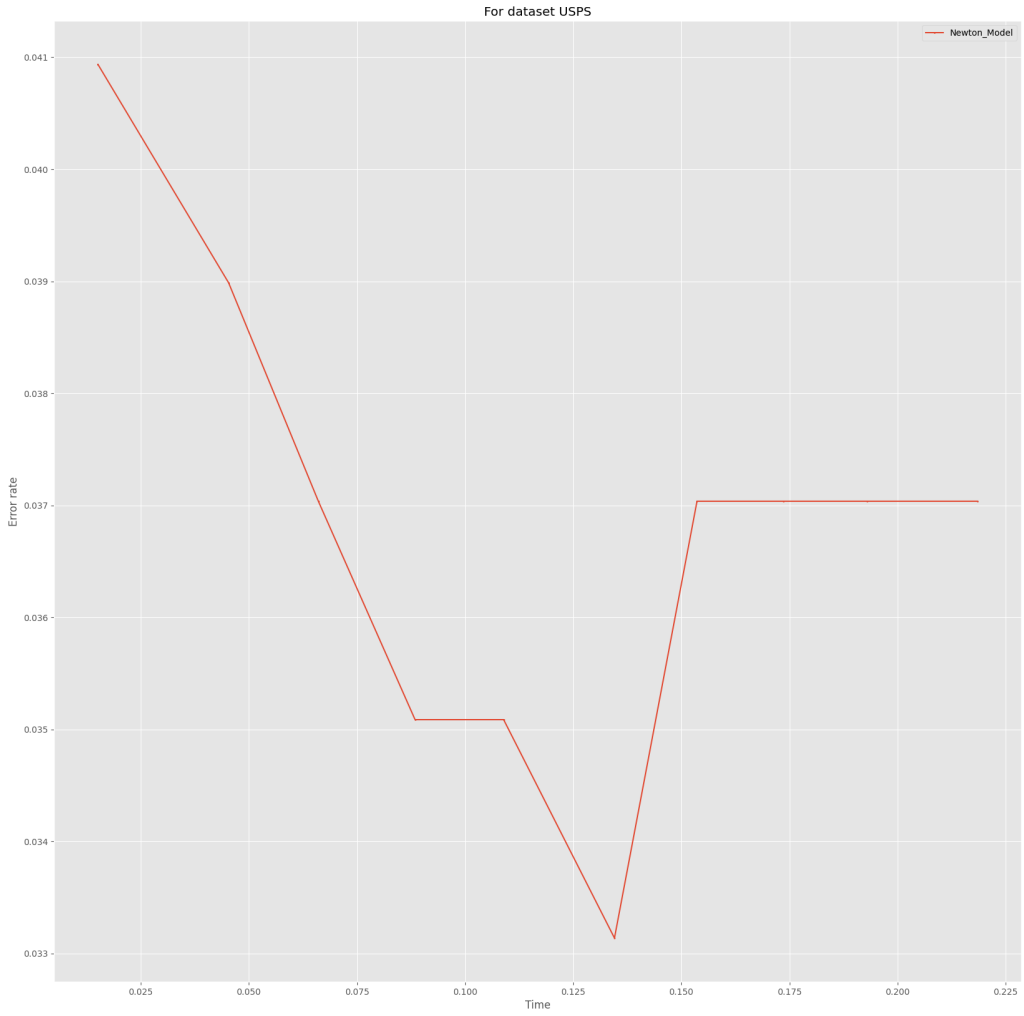
TASK-2

- For task 2, we got the following plots


For dataset A
Gradient_Descent
Error rate
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.0
0.5
1.0
1.5
2.0
Time


For dataset A
Newton_Model
0.0480
0.0475
0.0470
0.0465
0.0460
0.0455
0.0450
Error rate
0.02
0.03
0.04
0.05
0.06
0.07
0.08
Time


For dataset USPS
Gradient_Descent
Error rate
Time
0.0550
0.0525
0.0500
0.0475
0.0450
0.0425
0.0400
0.0375
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16

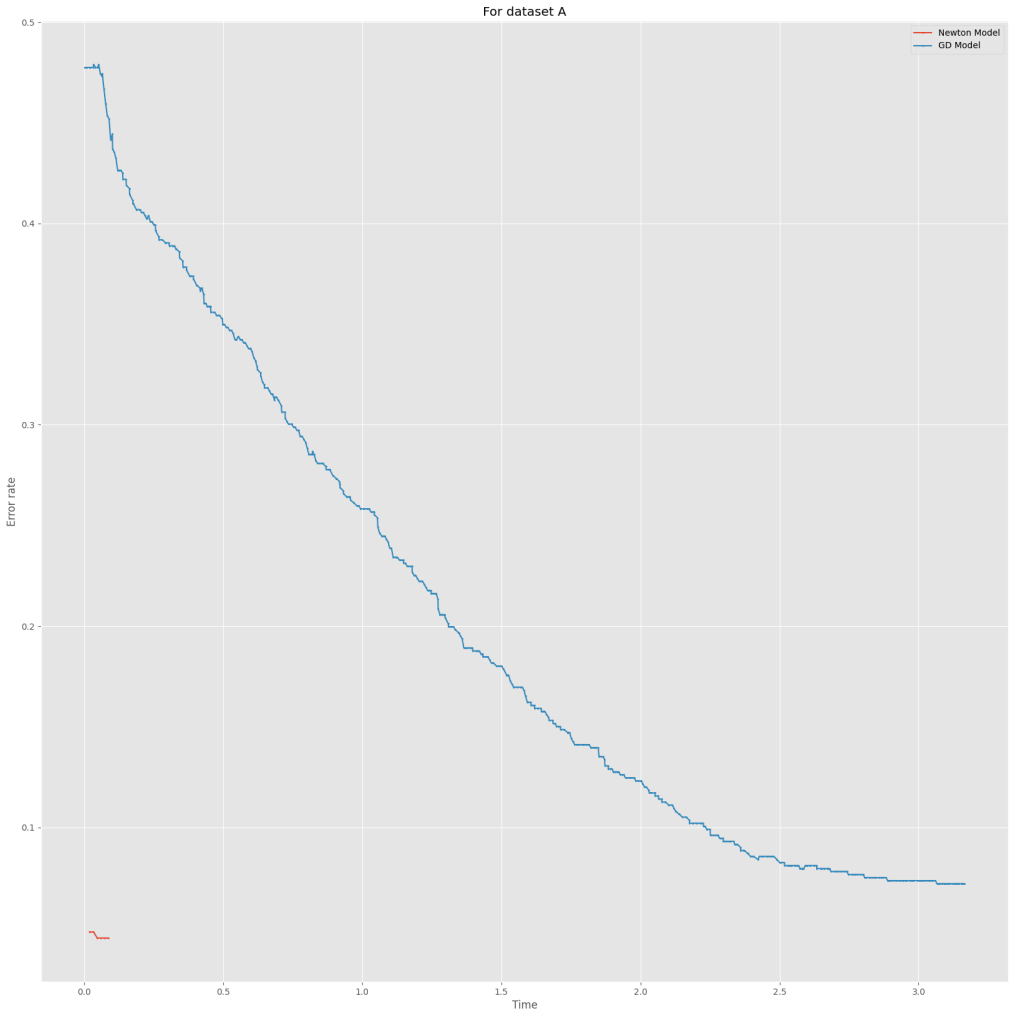


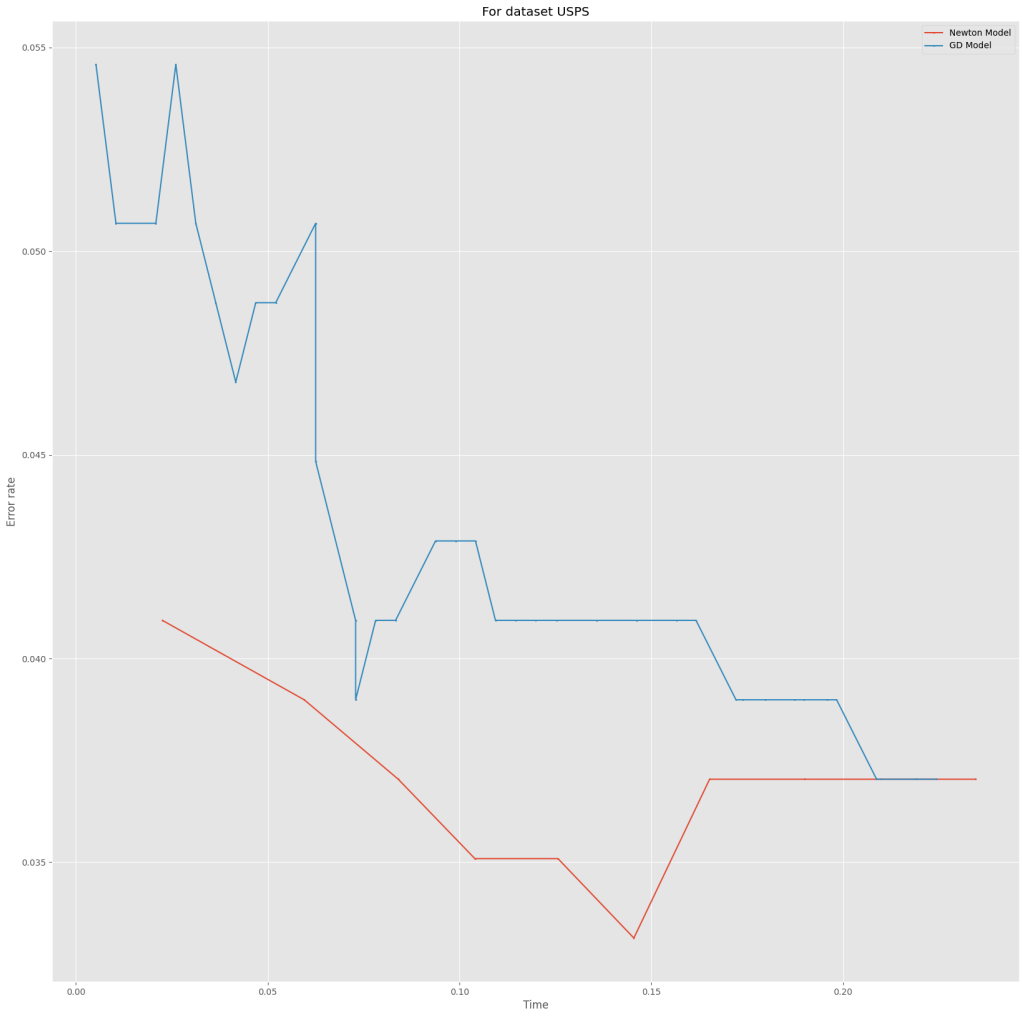
OBSERVATIONS:

- The Newton model takes a lot of time to update as the 10 iterations by gradient descent are comparable to a single update in Newton Model. So the later is computationally expensive.

- But the plots of Newton's model have rapid improvements as compared to gradient descent. Gradient descent takes many iterations to converge, but the Newton model converges in a few updates. That’s why we can see many ups-downs in the Newton plots.
- The matrix inversion of an NxN matrix in Newton Methods is O(N^3), hence the update equation takes a lot of time as compared to gradient descent. And, also the quality of the update in the Newton Model is better than the gradient descent with respect to the number of iterations. But as the number of iterations increases, these two methods converge.
- So GD performs many cheap updates and the Newton model performs less expensive updates.

COMPARISON PLOTS


For dataset A
Newton Model
GD Model
Error rate
Time
0.5
0.4
0.3
0.2
0.1
0.0
0.5
1.0
1.5
2.0
2.5
3.0



For dataset A, we can see that the Newton Model converges very fast as we have less number of features (60). But as the number of features is high in the USPS dataset, the Newton and GD method both struggle initially, but their performance converges after some time. And, the time taken is

nearly similar, which suggests that with a higher number of features, we can also do many cheap updates and get equivalent results as per Newton Method.