COL774 Assignment 1 Report

By Khyateeswar Naidu Nalla

2019CS10376

Question 1.

A)

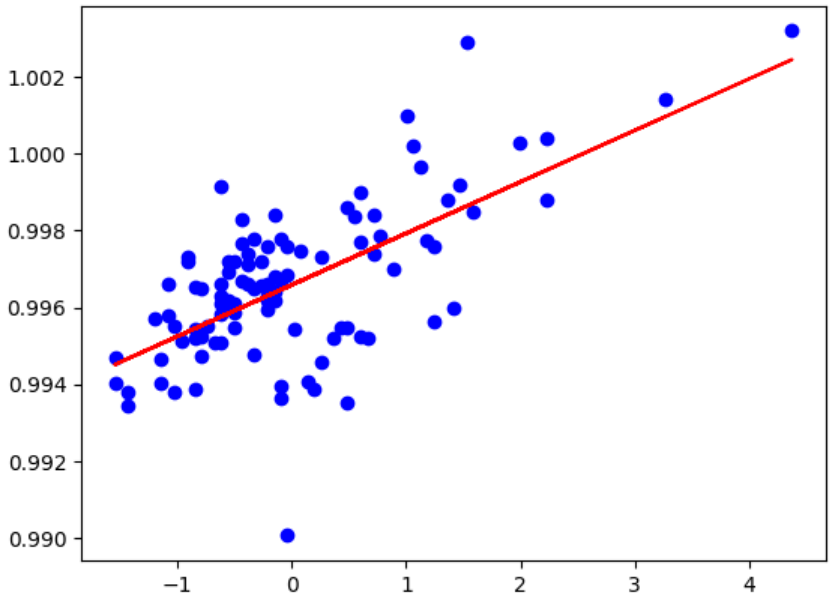
Learning rate = 0.001

Convergence criteria: $J(\theta)$ - prev_$J(\theta) \leq 10^{-9}$

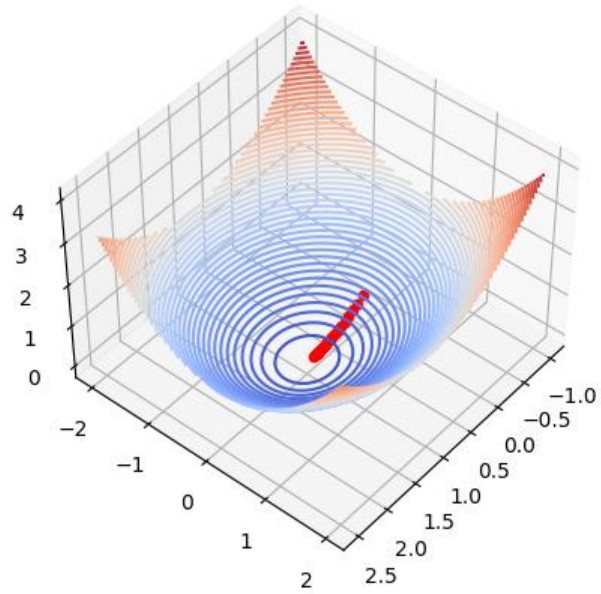
Converged Theta: [[0.9966139804],[0.0013401877]]

No of iterations = 474

B)

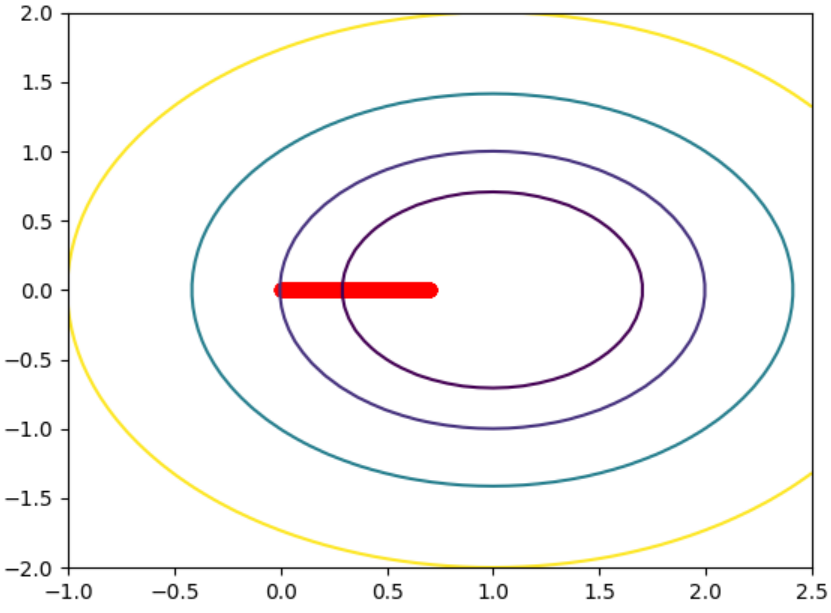


C)

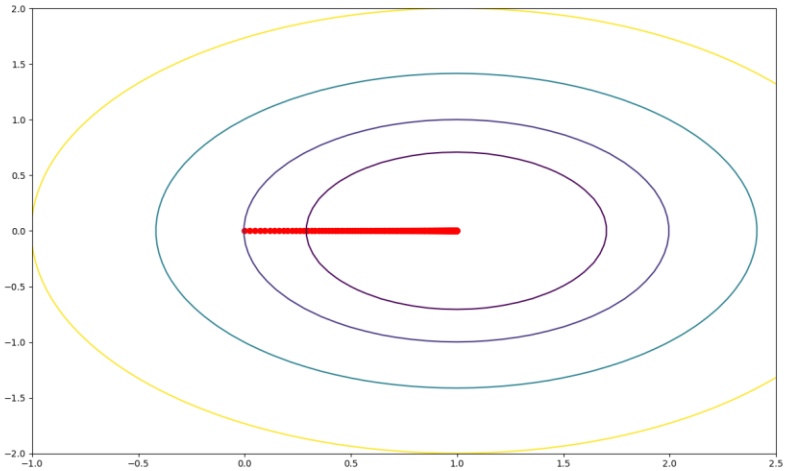


E)

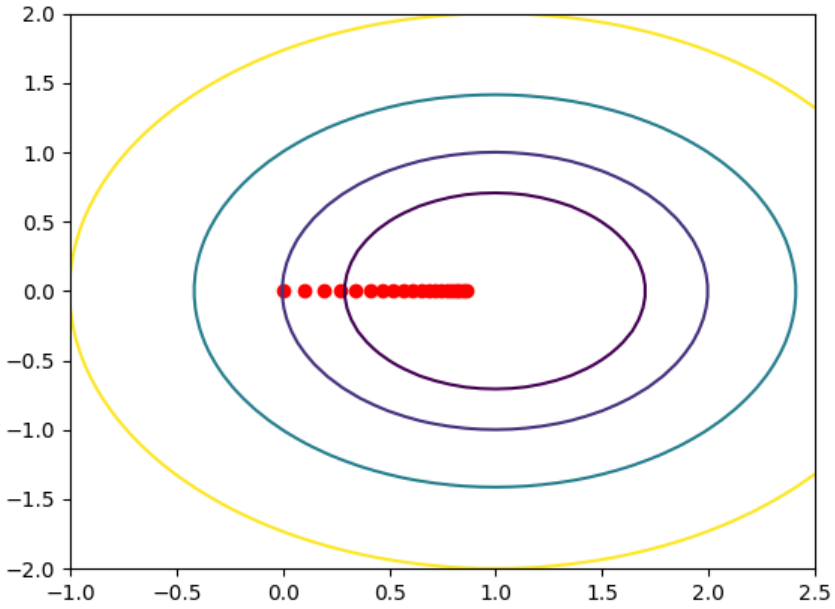
Learning rate = 0.001



Learning rate = 0.025



Learning rate = 0.1



Observation: It converges faster for 0.1 compared to 0.025 and 0.01. So if we increase the learning rate we get the minima in less time but if we increase it higher it might cause overfitting and we might miss the minima.

Question 2.

B)

Convergence Criteria: For each epoch $J(\theta)$ - prev_$J(\theta) \leq 10^{-9}$

Batch size	no of epochs	Convergence criteria	theta
1	2	10^{-9}	
100	9	10^{-9}	
10000	634	10^{-9}	
100000	5494	10^{-9}	

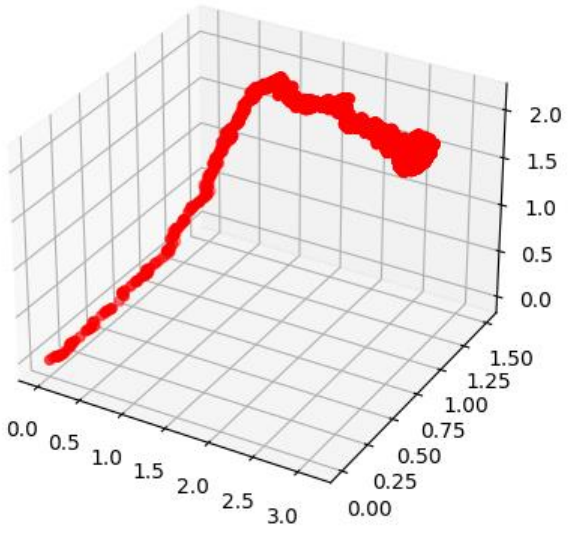
C)

Batch Size	MSE
1	1.5597
100	1.8887
10000	1.997
100000	1.7306

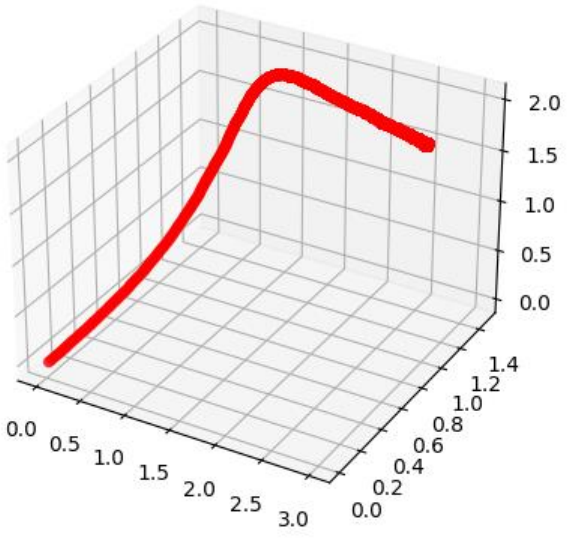
Error depends on the convergence criteria

D)

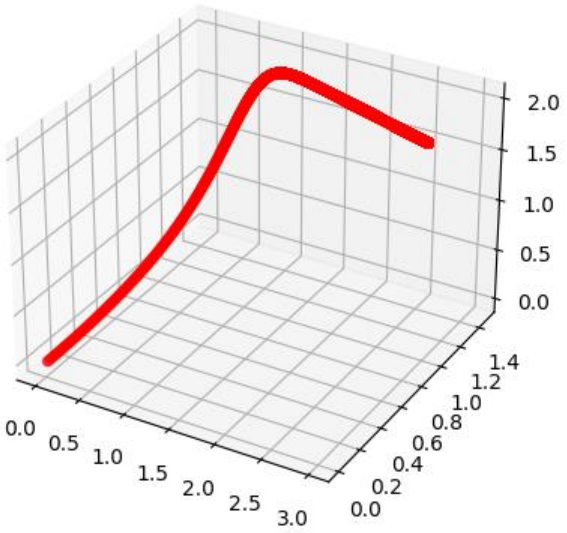
Batch size = 1



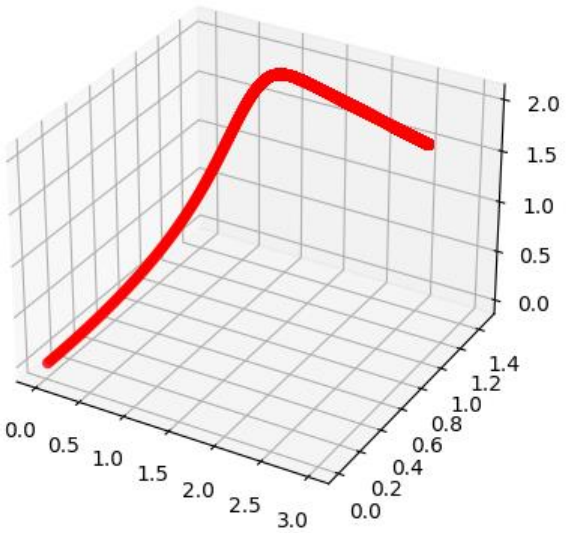
Batch size = 100



Batch size = 10000



Batch size = 100000

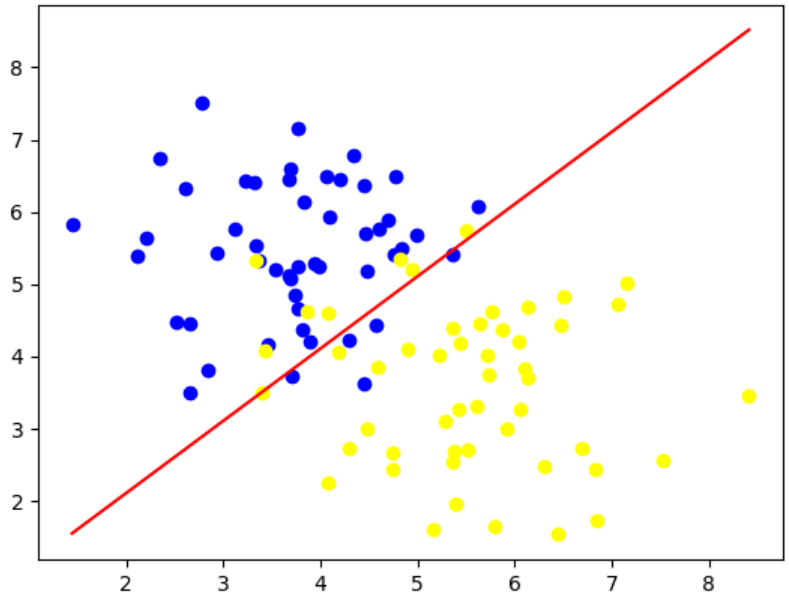


Observation: Initially it is not smooth for r=1 but increase in r it gets smoother. This is because for small r it only looks for small dataset and so it doesn't know how to behave so it causes the sudden changes in the theta

Question 3.

A) final parameters = [[0.223295367],[1.9626155],-1.9648612]]

B)



Question 4.

A) Assumption Σ1 = Σ2 = Σ

ϕ: 0.5

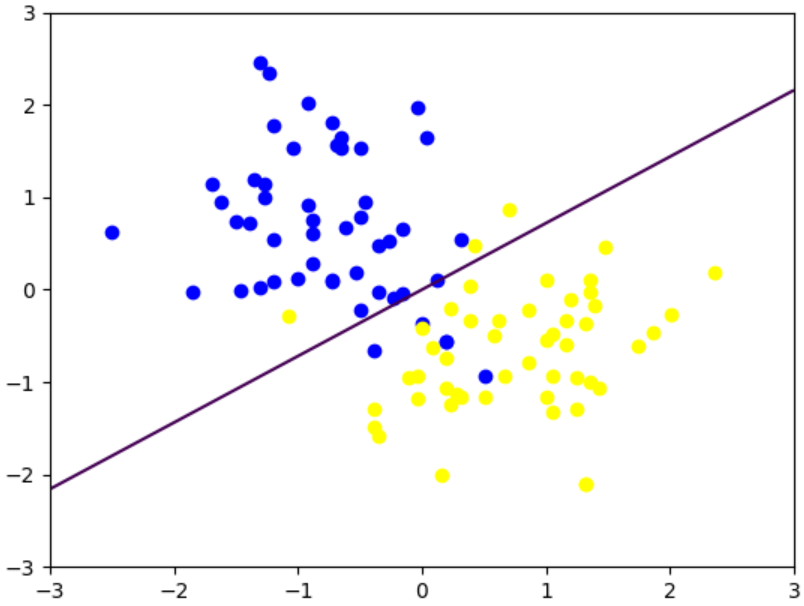
μ_0: [[-0.75529433] [0.68509431]]

μ_1: [[0.75529433] [-0.68509431]]

Σ: [[0.42953048 -0.02247228]

[-0.02247228 0.53064579]]

B)



C) Equation:

(x- μ1)* $\boldsymbol{\Sigma}$ 1−1*(x- μ1).T – (x- μ0)* $\boldsymbol{\Sigma}$ 0−1*(x- μ0).T + 2* log (1−ϕ /ϕ)

+ log(|$\boldsymbol{\Sigma}$ 1|/| $\boldsymbol{\Sigma}$ 0|) = 0. where Σ1 = Σ2 = Σ

D) Σ1 ≠ Σ2 ≠ Σ

ϕ: 0.5

μ_0: [[-0.75529433] [0.68509431]]

μ_1: [[0.75529433] [-0.68509431]]

$\boldsymbol{\Sigma}$: [[0.42953048 -0.02247228]

[-0.02247228 0.53064579]]

$\boldsymbol{\Sigma}_0$: [[0.38158978 -0.15486516]

[-0.15486516 0.64773717]]

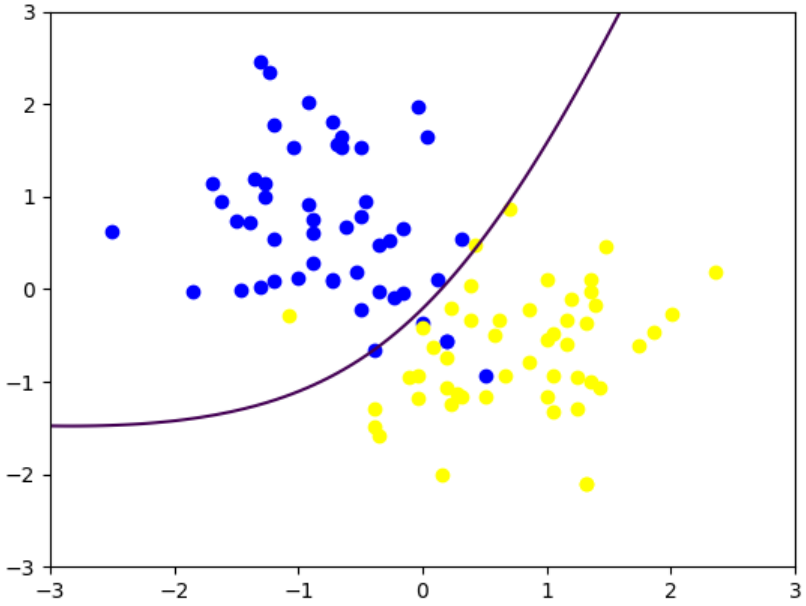
$\boldsymbol{\Sigma}_1$: [[0.47747117 0.1099206]

[0.1099206 0.41355441]]

E) Equation:

(x- μ1)* $\boldsymbol{\Sigma}$ 1−1*(x- μ1).T – (x- μ0)* $\boldsymbol{\Sigma}$ 0−1*(x- μ0).T + 2* log (1−ϕ /ϕ)

+ log(|$\boldsymbol{\Sigma}$ 1|/| $\boldsymbol{\Sigma}$ 0|) = 0.



F).

we can observe that the quadratic decision boundary is separating the points more accurately than the linear decision boundary. In general, both the classes have their own covariance matrix. Linear decision boundary is assuming that both have the same covariance matrix, so it is not able to separate accurately