Name – Samrajnee Ghosh

Entry No. – 2023CSY7547

COL774 Assignment 1

Q1. Linear Regression

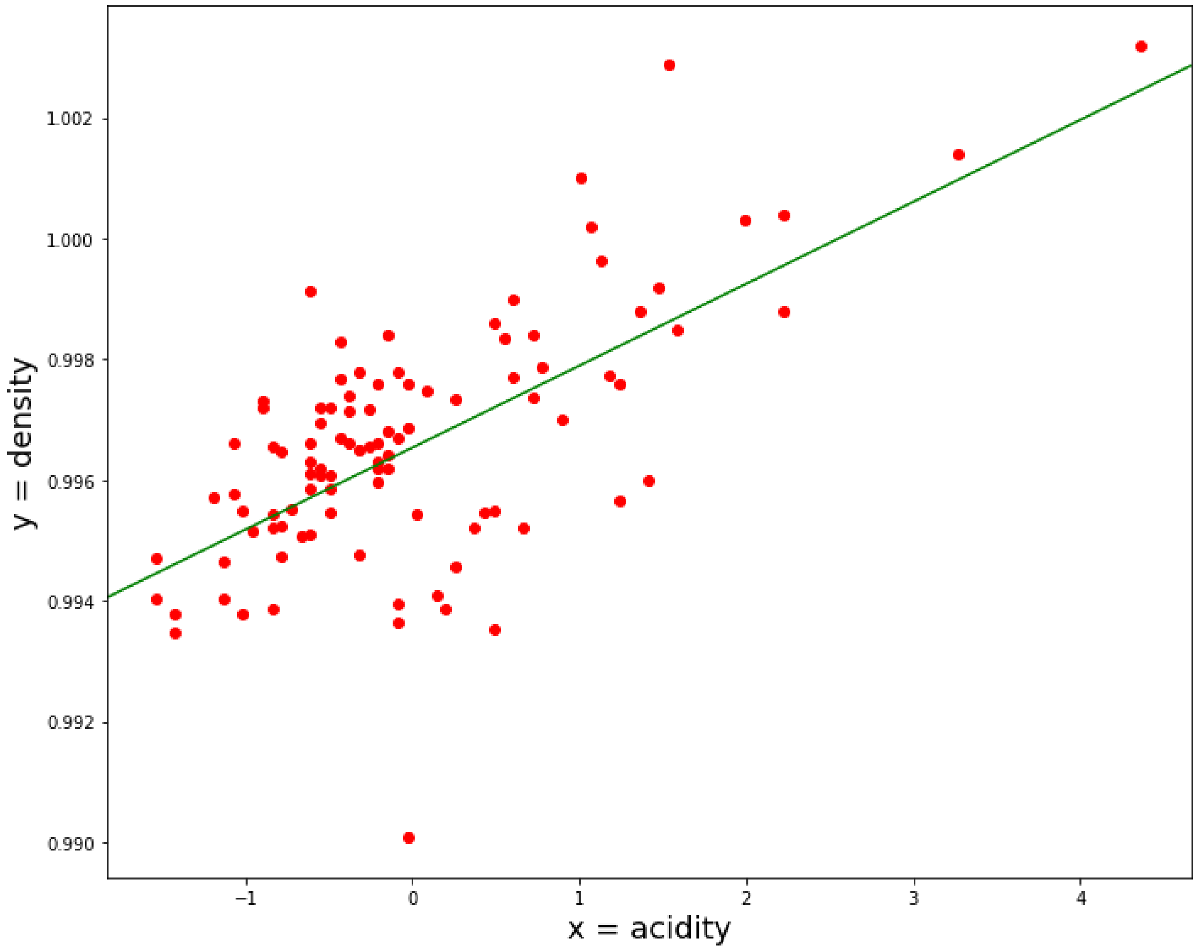
Part A

Learning rate = 0.001

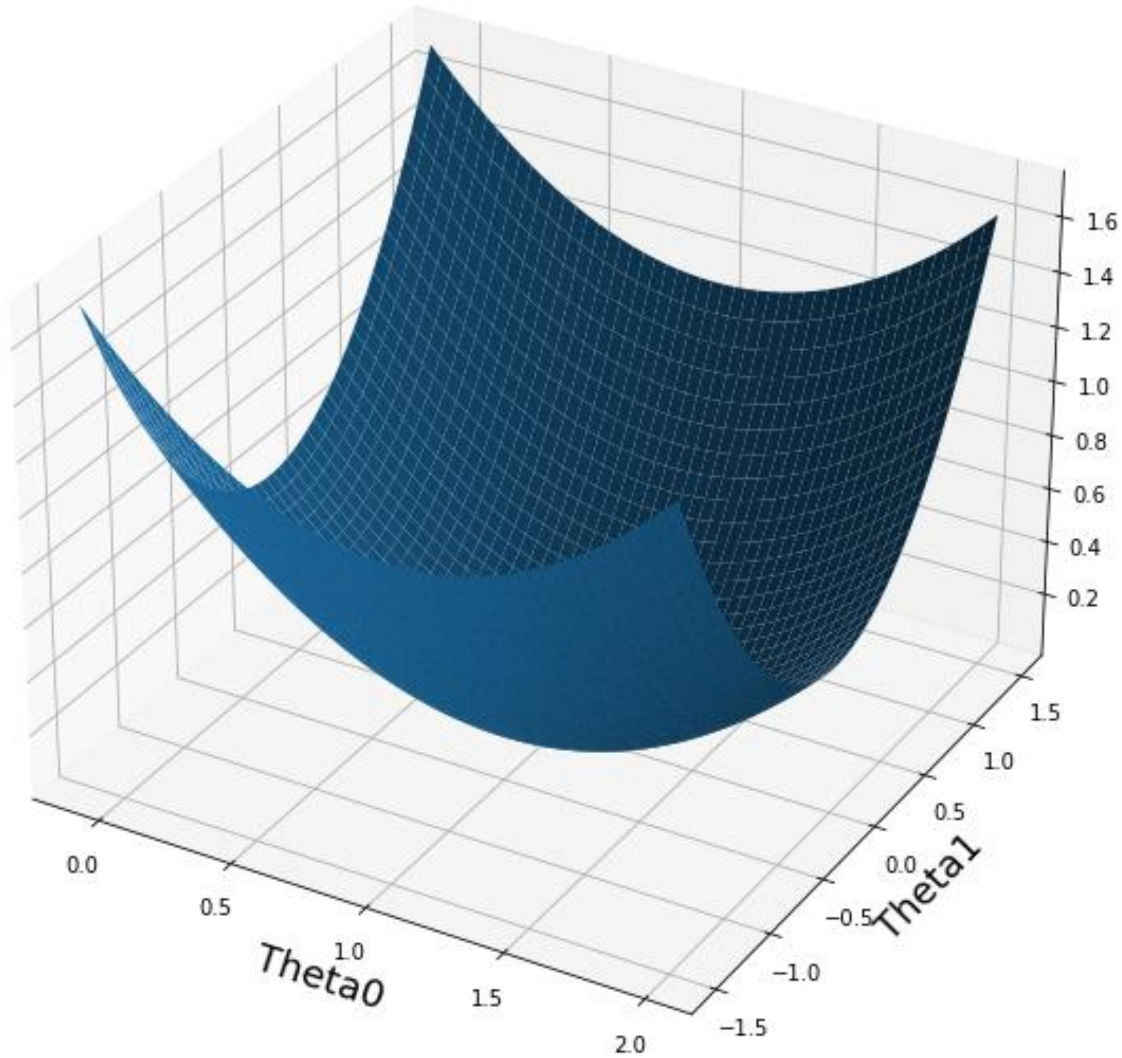
Stopping Criteria = 1-Norm of Delta_Theta_J_Theta < 0.0001

Final Parameters = [Theta0 Theta1] = [0.9965344 0.0013578]

Part B

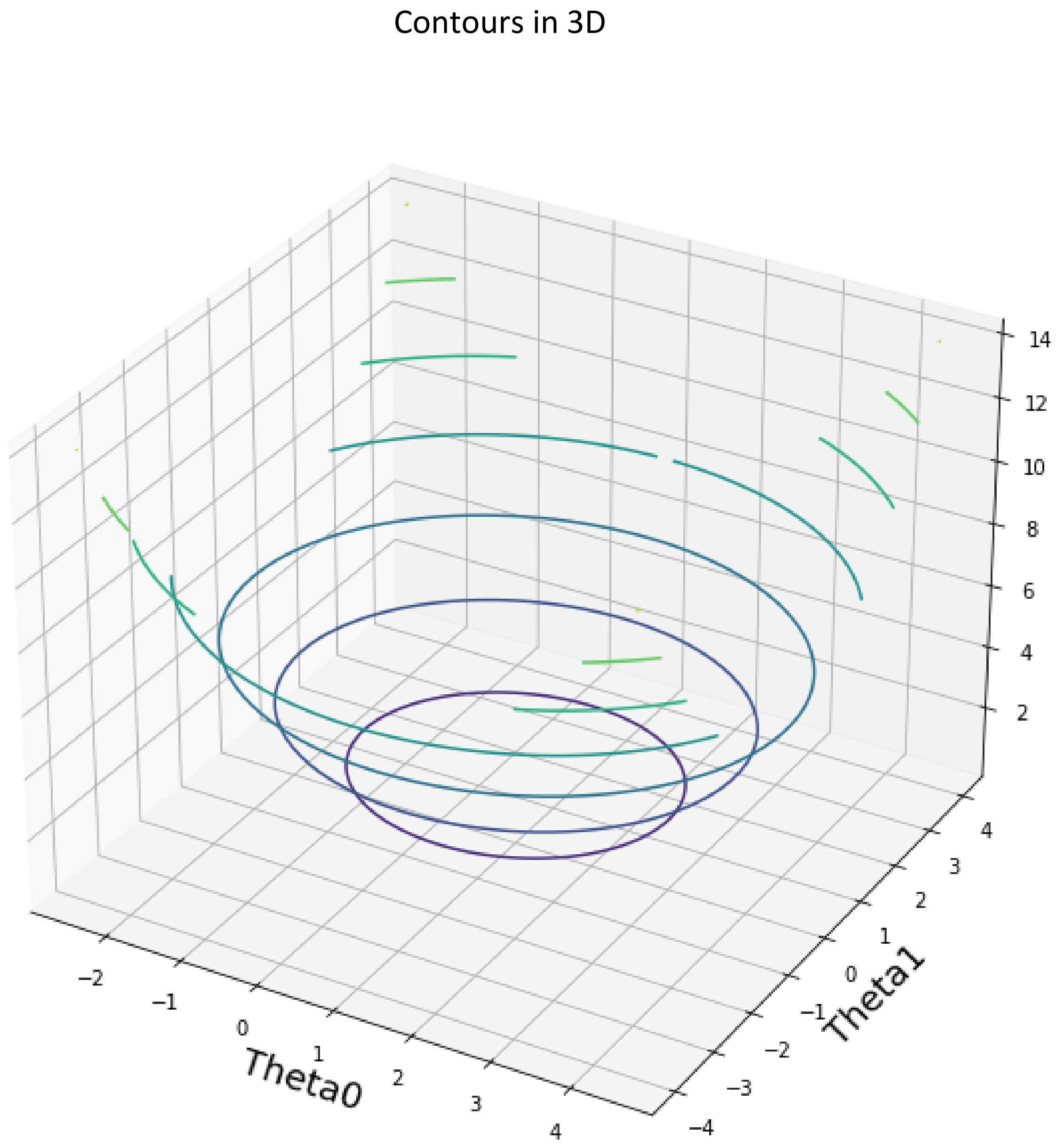


Part C

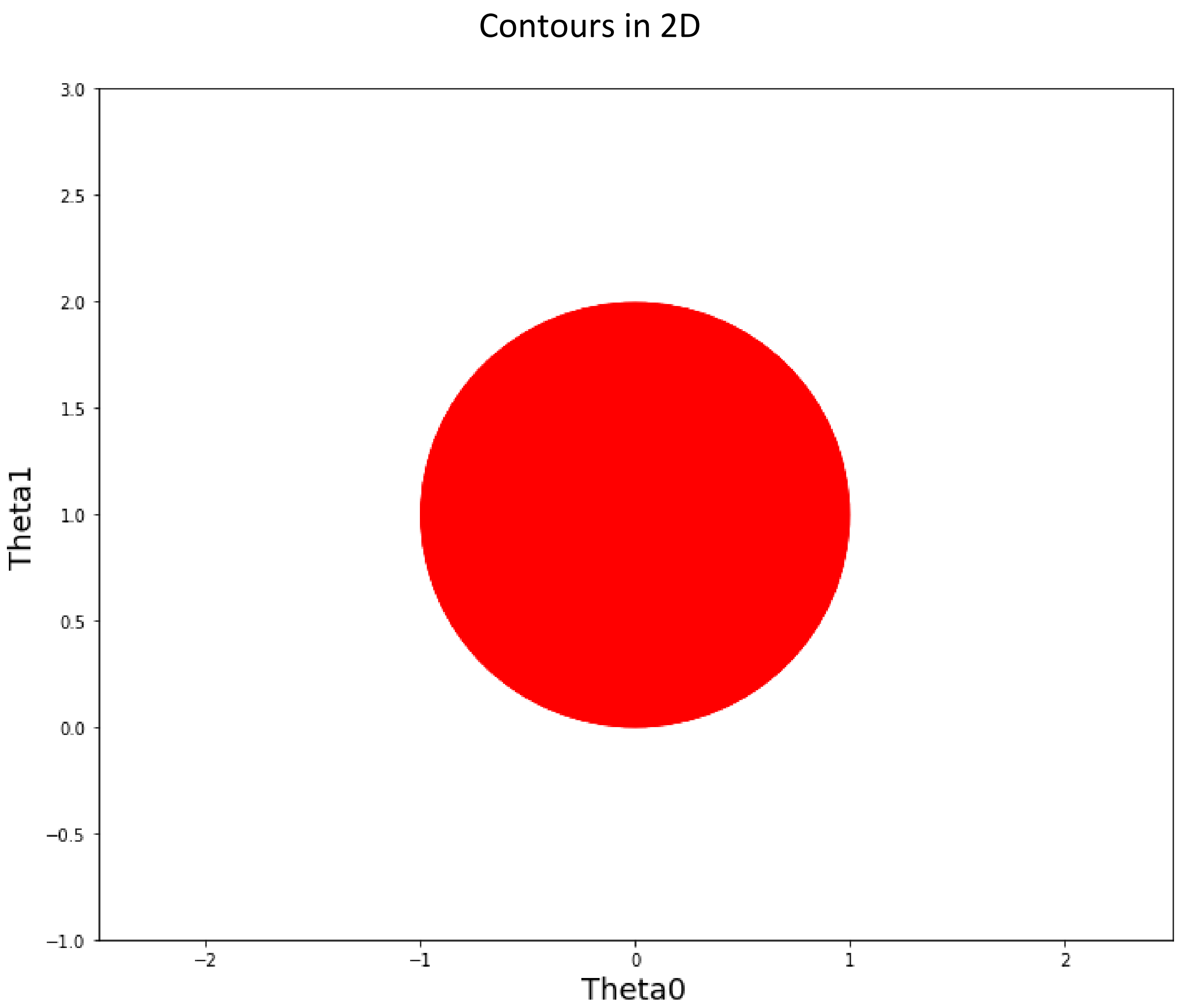


1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.5
1.0
1.5
2.0
Theta0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
Theta1

Part D

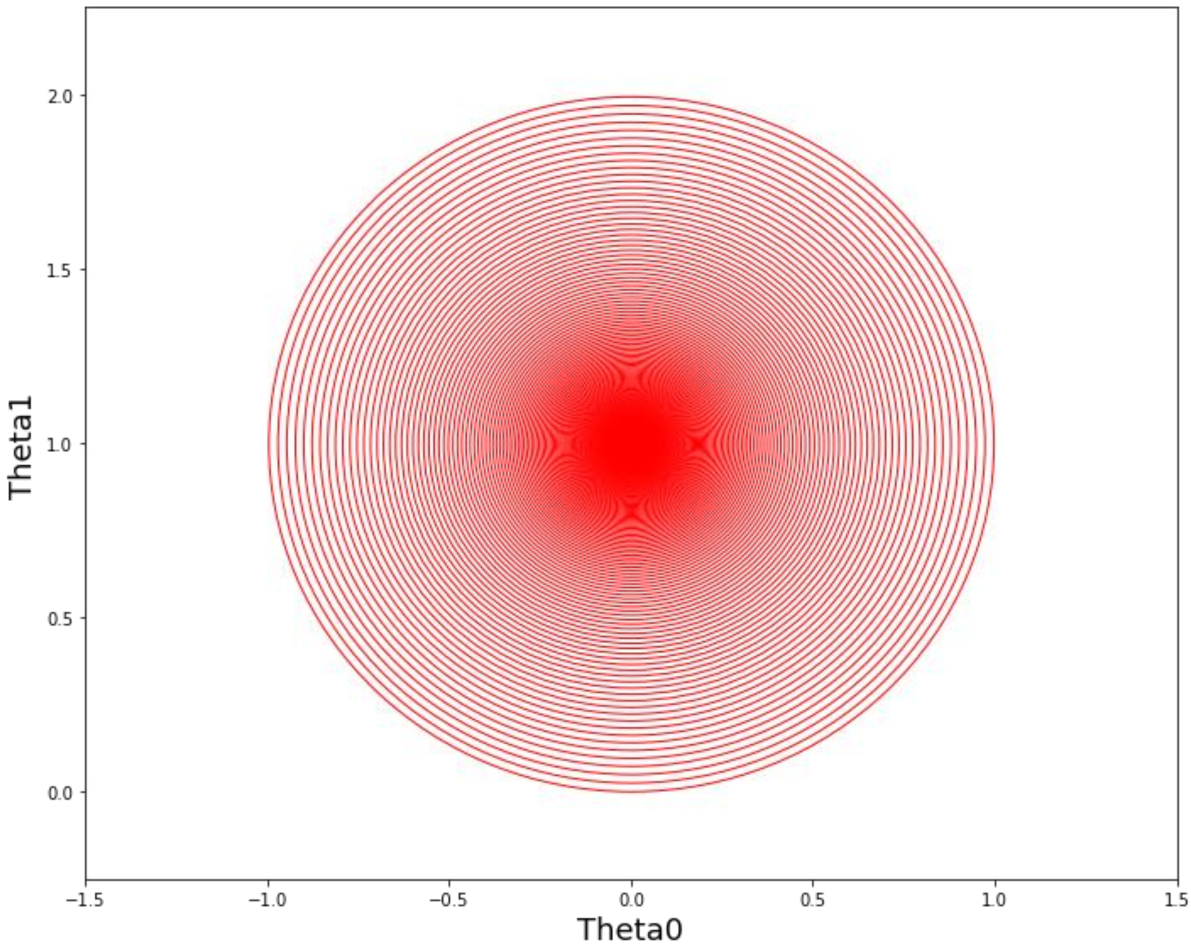


Contours in 3D
14
12
10
8
6
4
2
−2
−1
0
1
2
3
4
Theta0
4
3
2
1
0
−1
−2
−3
−4
Theta1

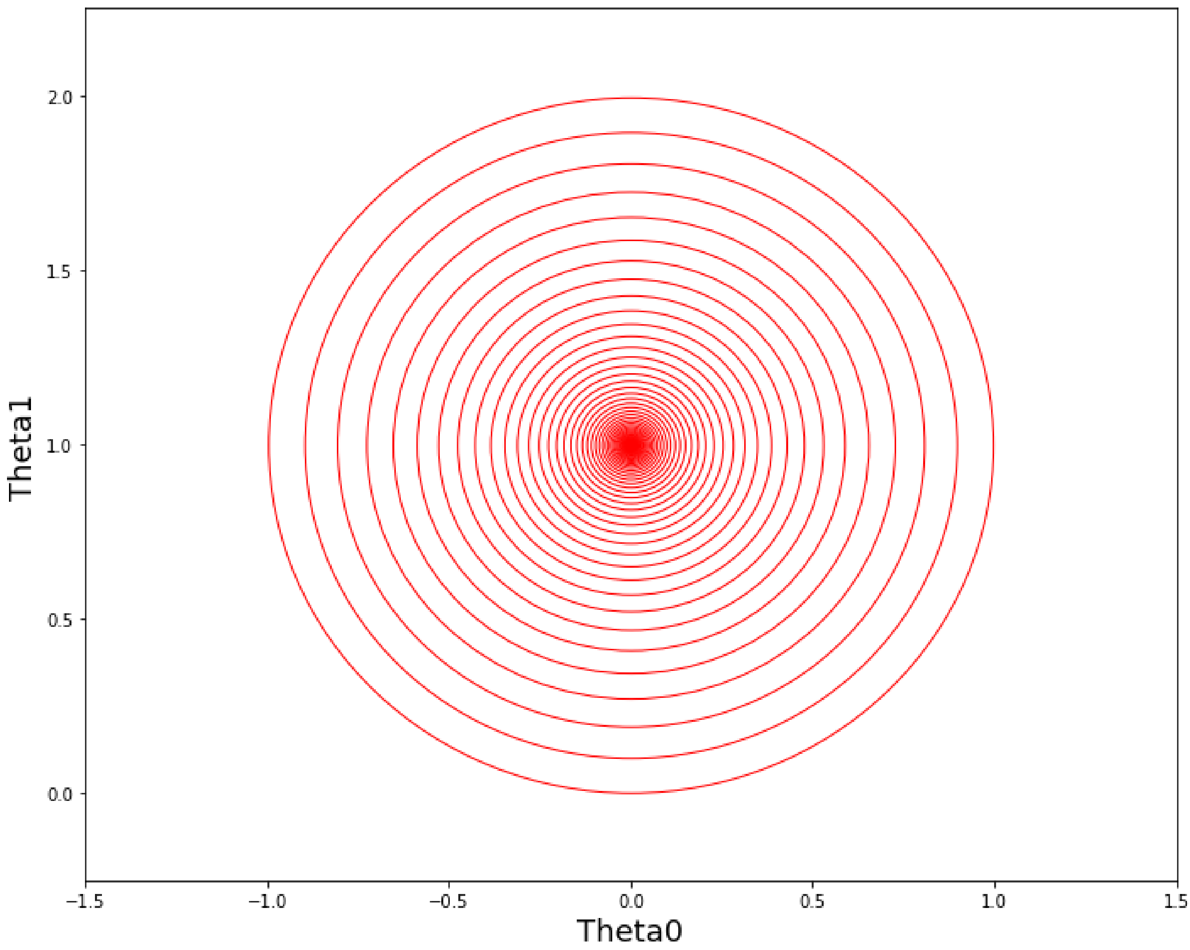

Contours in 2D
Theta1
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−2
−1
0
1
2
Theta0

Part E

Contours for eta = 0.025



Contours for eta = 0.1



With the increase in the learning rate, we see that –

1. The jumps in theta values i.e., the change in the theta values for every iteration becomes larger and so, each of the concentric circles are farther apart from the previous one.
2. Since, the theta values take larger leaps, the convergence criteria is satisfied faster and so the no. of iterations become lesser with the increase in learning rate.

But the learning rate should be increased carefully because if the change in theta is too large, the theta values might diverge instead of converging.

Q2. Stochastic Gradient Descent

Part C

Batch Size	Gamma	Eta	Theta0	Theta1	Theta2
1	3	0.001	1.96780807	1.00657195	1.96820701
100	3	0.001	1.9978498	1.00464364	1.99767346
10000	3	0.001	0.21964044	0.9081632	0.71607144
1000000	3	0.001	0.10503717	0.45217961	0.23461861

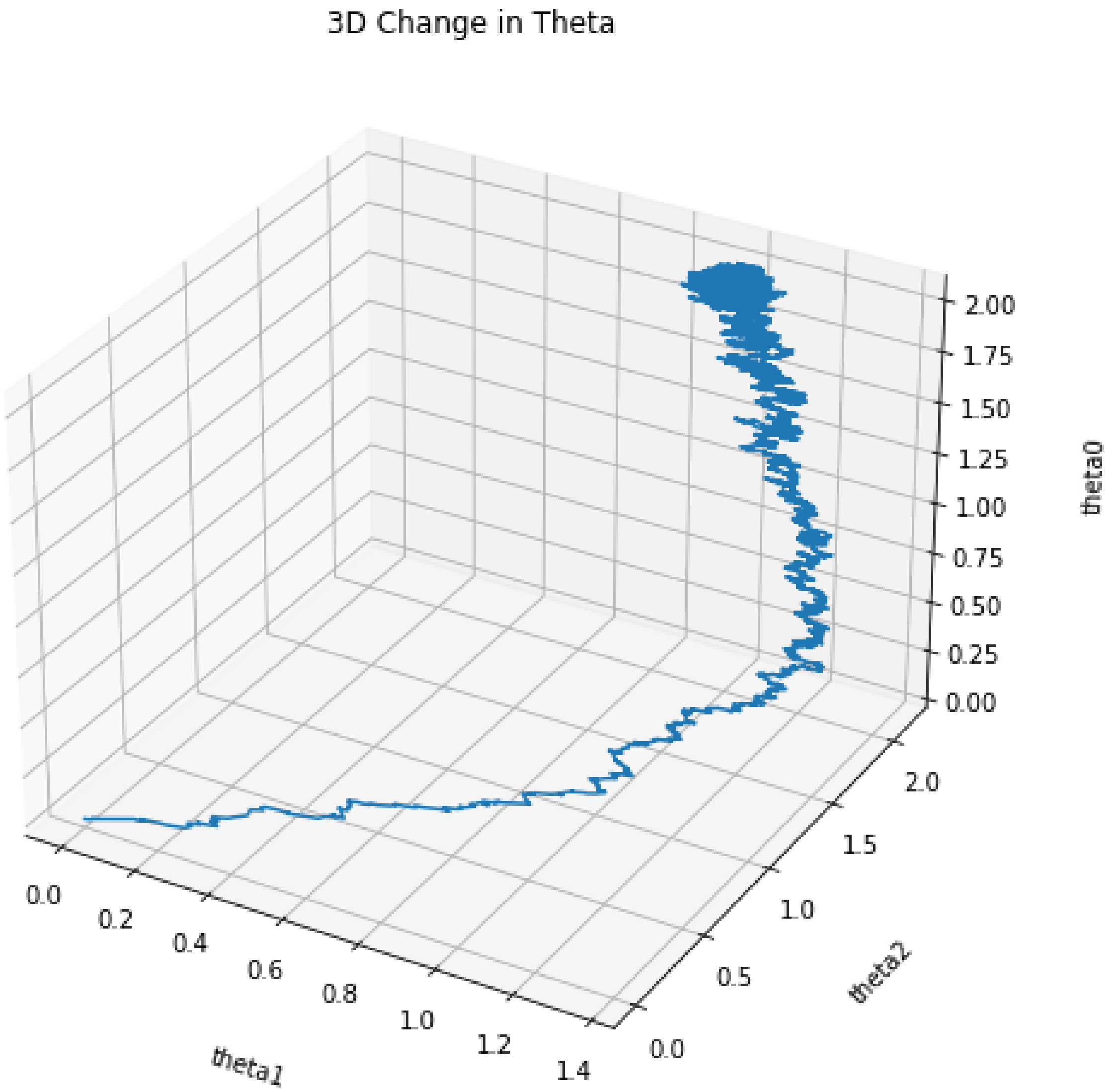
We see that, the values of theta are not exactly the same for the different batch sizes in SGD. With the increase in the batch size, the theta values start becoming less and less accurate.

Batch Size	Iterations	Speed	Error
1	131990	Very high	1.536483823743384
100	66171	High	1.4846843625098576
10000	143	Mediocre	85.70842460748858
1000000	45	Low	174.1605386847608

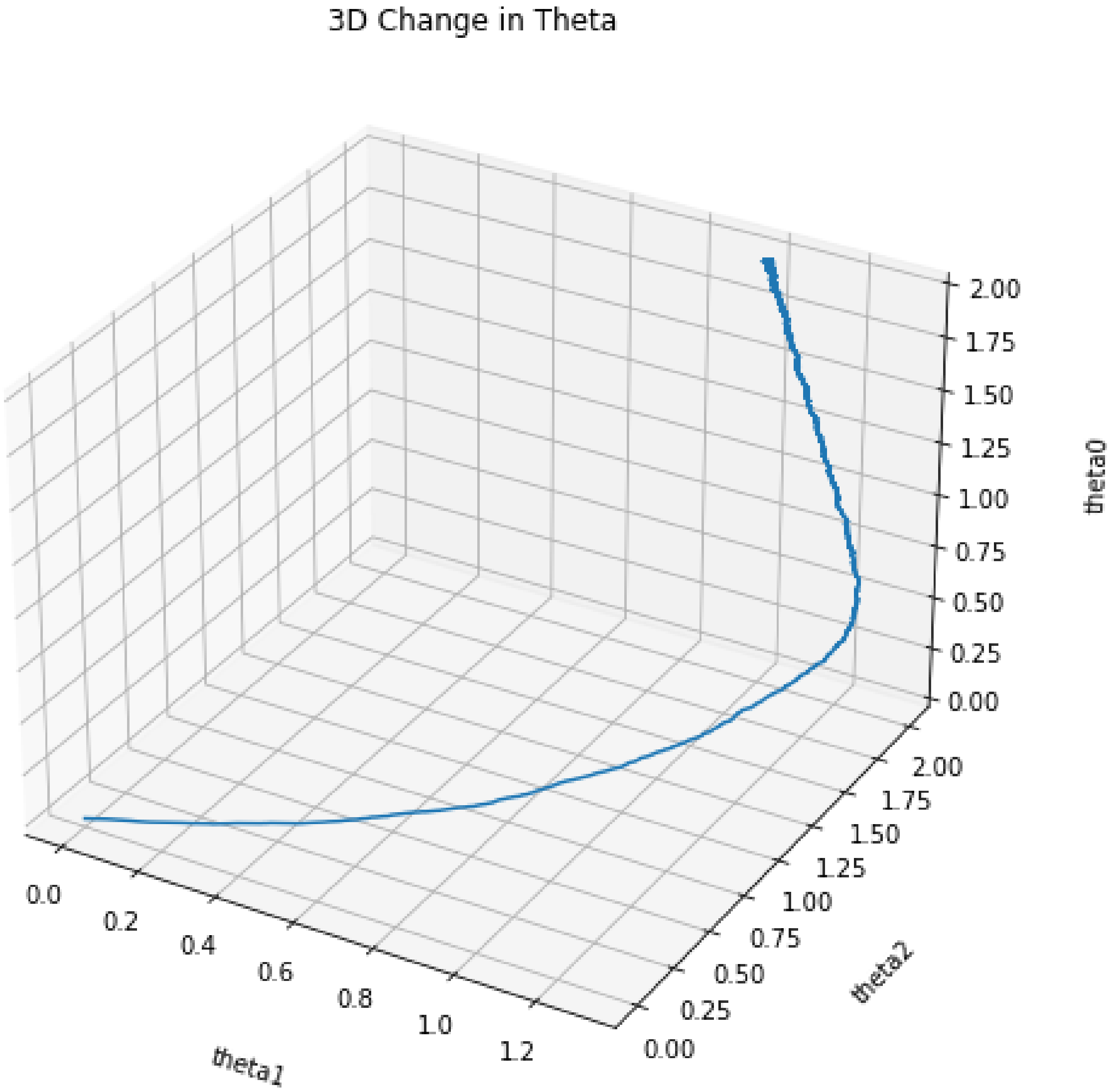
We see that, although the no. of iterations required to converge decreases with the increase in the batch size but each of the iterations for a large batch is very time-taking and so the overall speed decreases for larger batches. Also, the error values are more for larger batches because their thetas are less accurate.

Part D

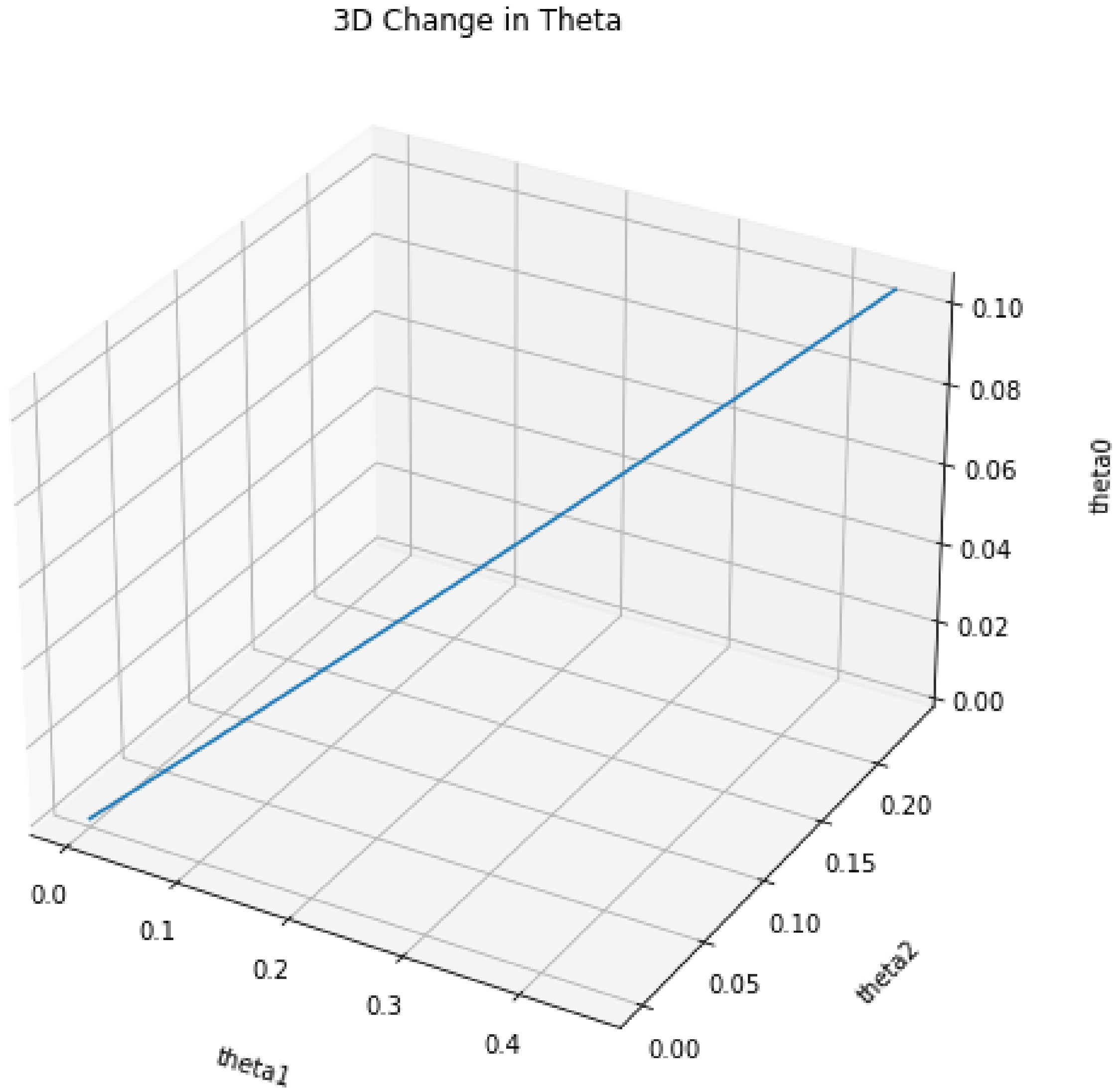
For batch_size = 1



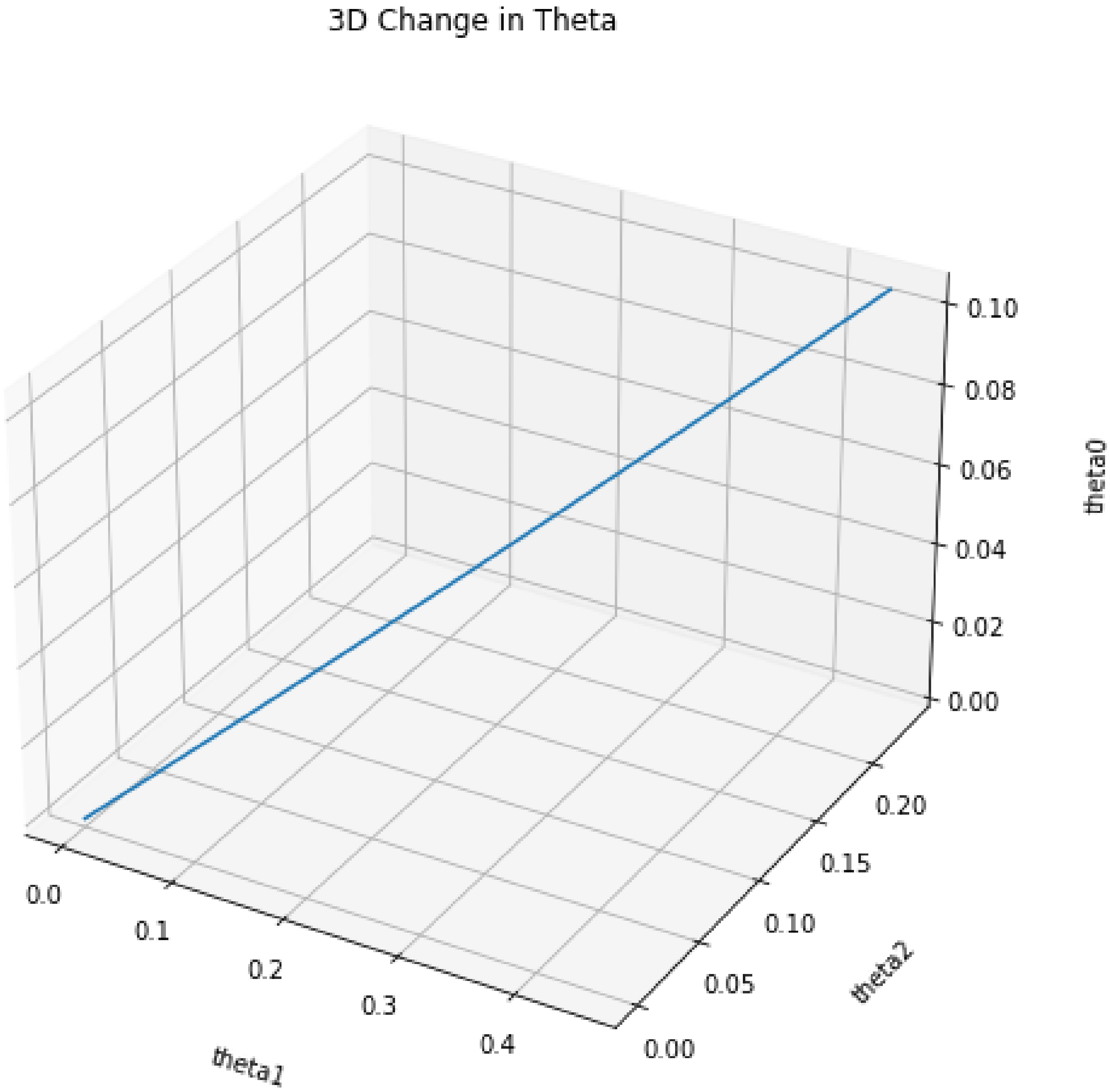
For batch_size = 100



For batch_size = 10000



For batch_size = 1000000



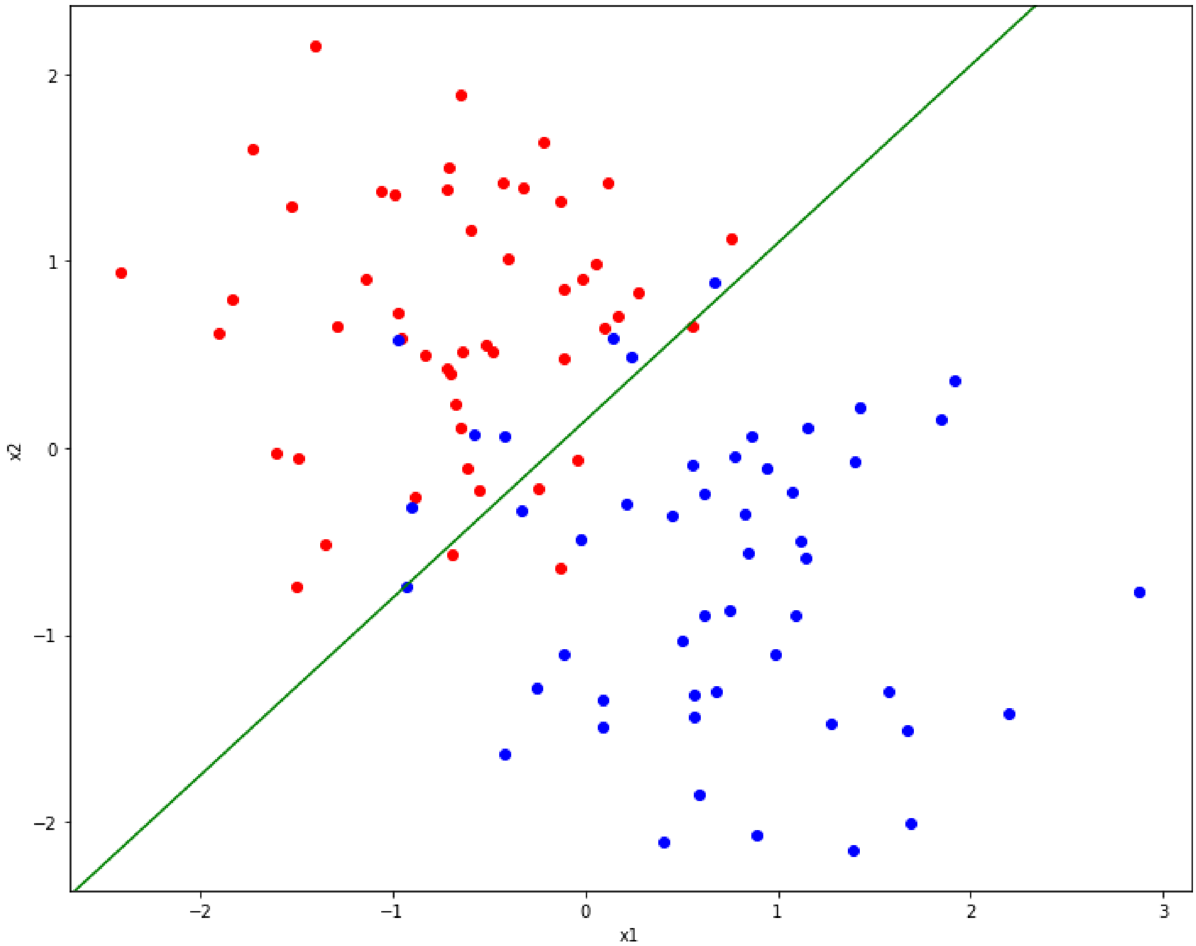
We see that, with the decrease in the batch size, the plotted curve becomes more zig-zag before converging. This justifies the fact that for smaller batch sizes, the theta hovers around the minima before converging due to the sampling of smaller batches from the original batch. However, for larger batch sizes, especially where the batch size is equal to the original batch size, SGD behaves like Gradient Descent and so, the curve is smooth.

Q3. Logistic Regression

Part A

Theta = [Theta0, Theta1, Theta2] = [0.39170362 2.55715793 -2.69153177]

Part B



Q4. Gaussian Discriminant Analysis

Part A

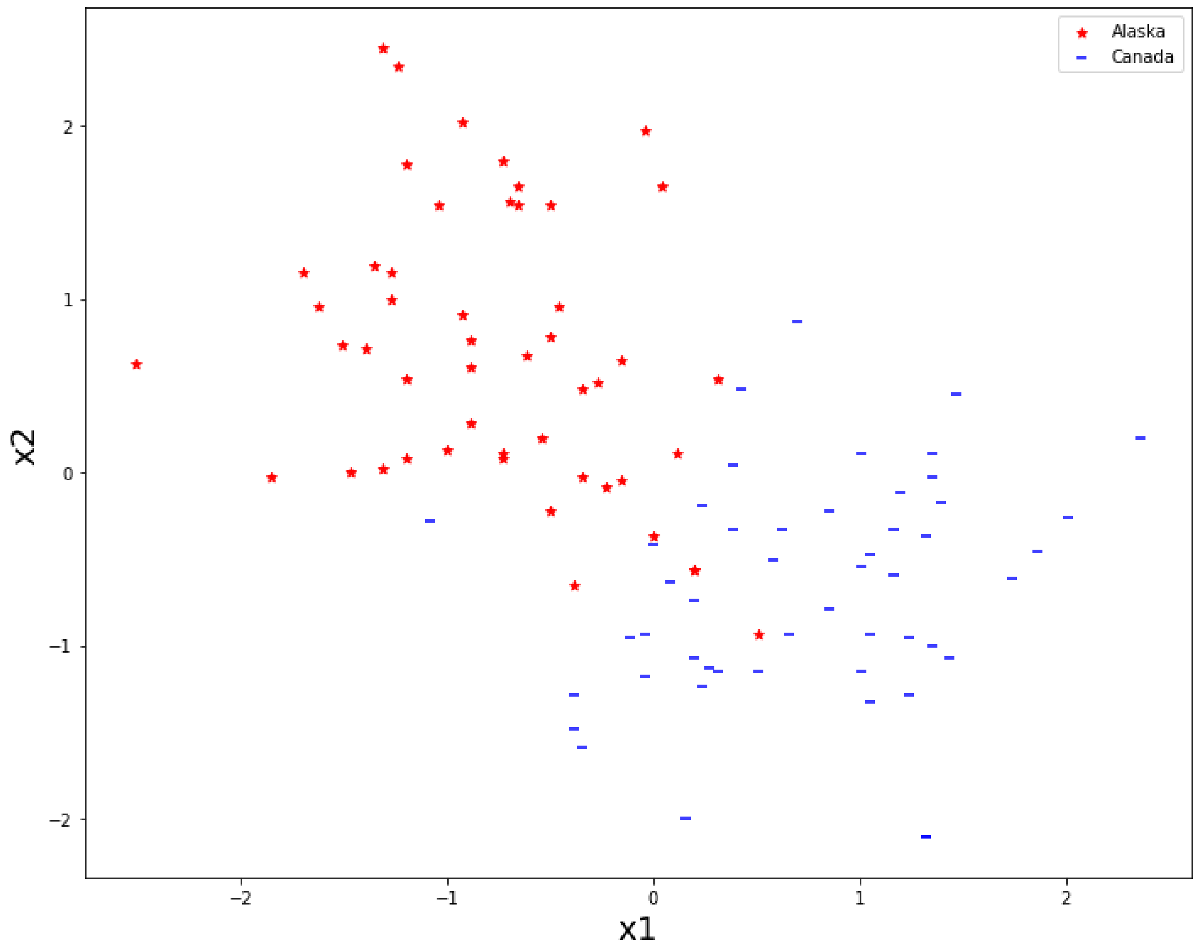
Mu0 = [-0.7552943279913609 0.685094305548928]

Mu1 = [0.7552943279913608 -0.6850943055489274]

Sigma = [[0.42953048 -0.02247228]

[-0.02247228 0.53064579]]

Part B

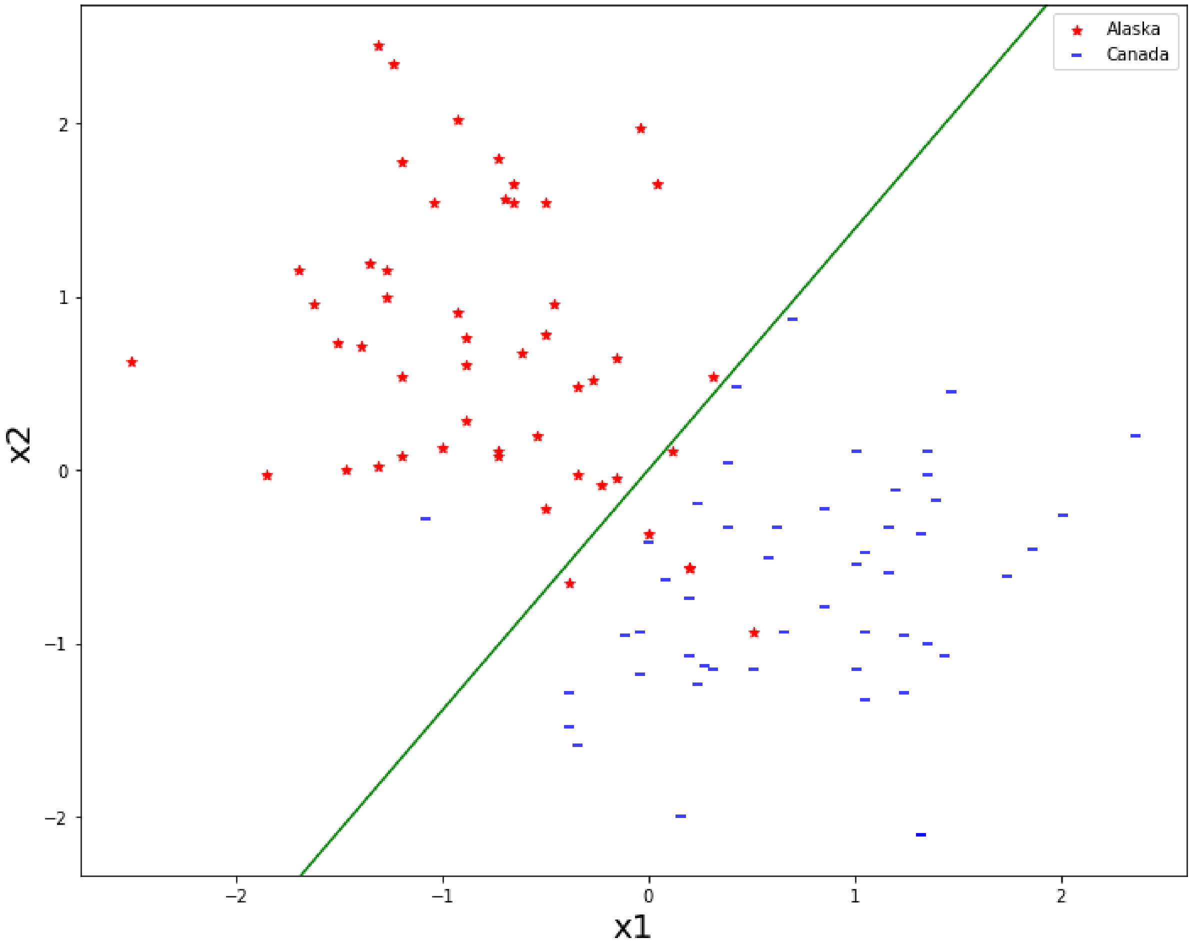


Part C

The equation for the linear decision boundary is –

$$2(\mu_1-\mu_0)^T\Sigma^{-1}x + \mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1 = 0$$

The plot of the data separated by the linear boundary is –



Part D

Mu0 = [-0.7552943279913609 0.685094305548928]

Mu1 = [0.7552943279913608 -0.6850943055489274]

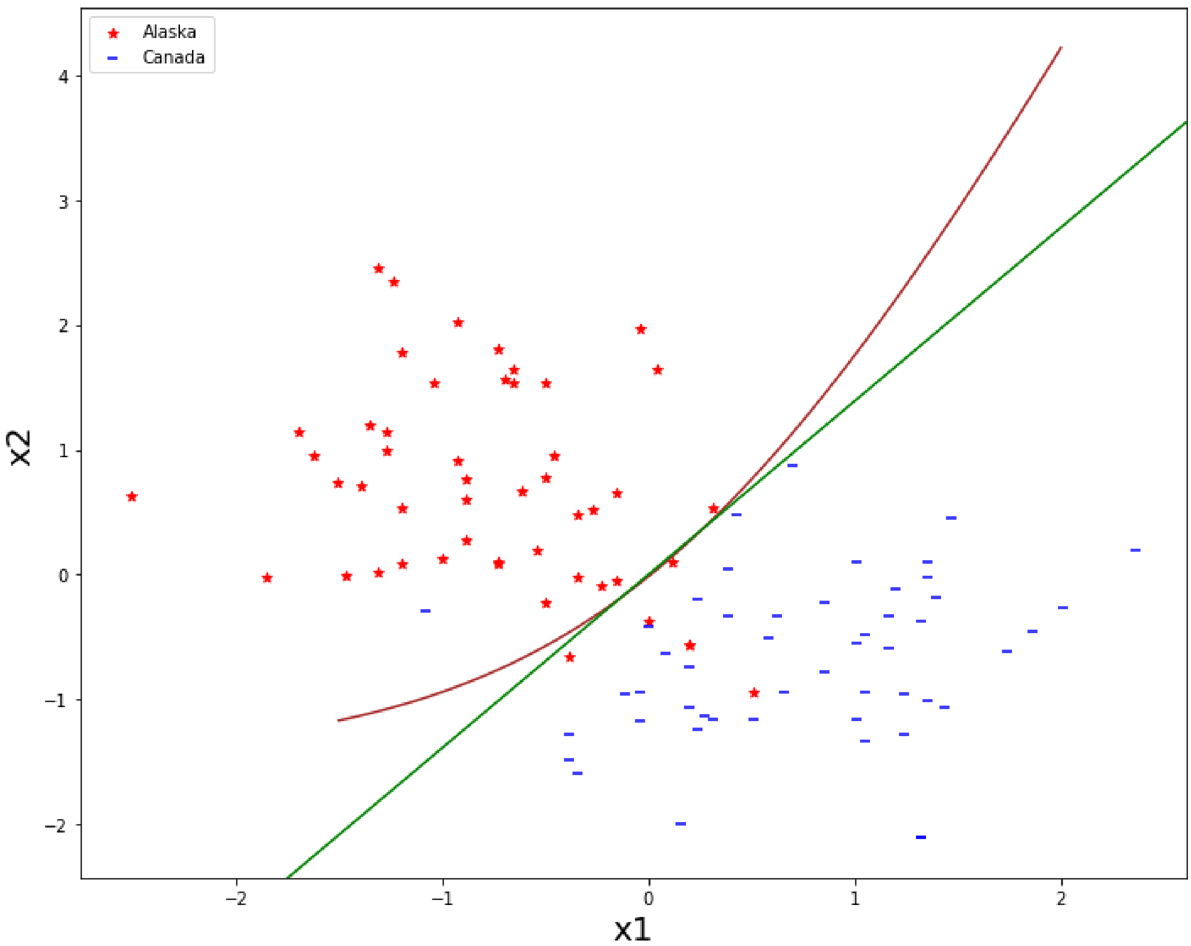
Sigma0 = [[0.38158978 -0.15486516]

[-0.15486516 0.64773717]]

Sigma1 = [[0.47747117 0.1099206]

[0.1099206 0.41355441]]

Part E



Part F

From the plot it is visible that the quadratic boundary classifies the data slightly better than the linear boundary. Although the improvement is not large but for very large dataset, this slight change may cause significant improvement.