Assignment-01

Kapil verma
2018CS10348

How to run the programs:

1. **Put the data files in the respective question folder i.e the data files for question 1 should be in the folder 'Q1'**
2. **The name of data files should be the same as the name of files given in the assignment.**
3. **The program doesn't take any arguments so you can run them directly but iff you follow the above steps.**

Q(1).

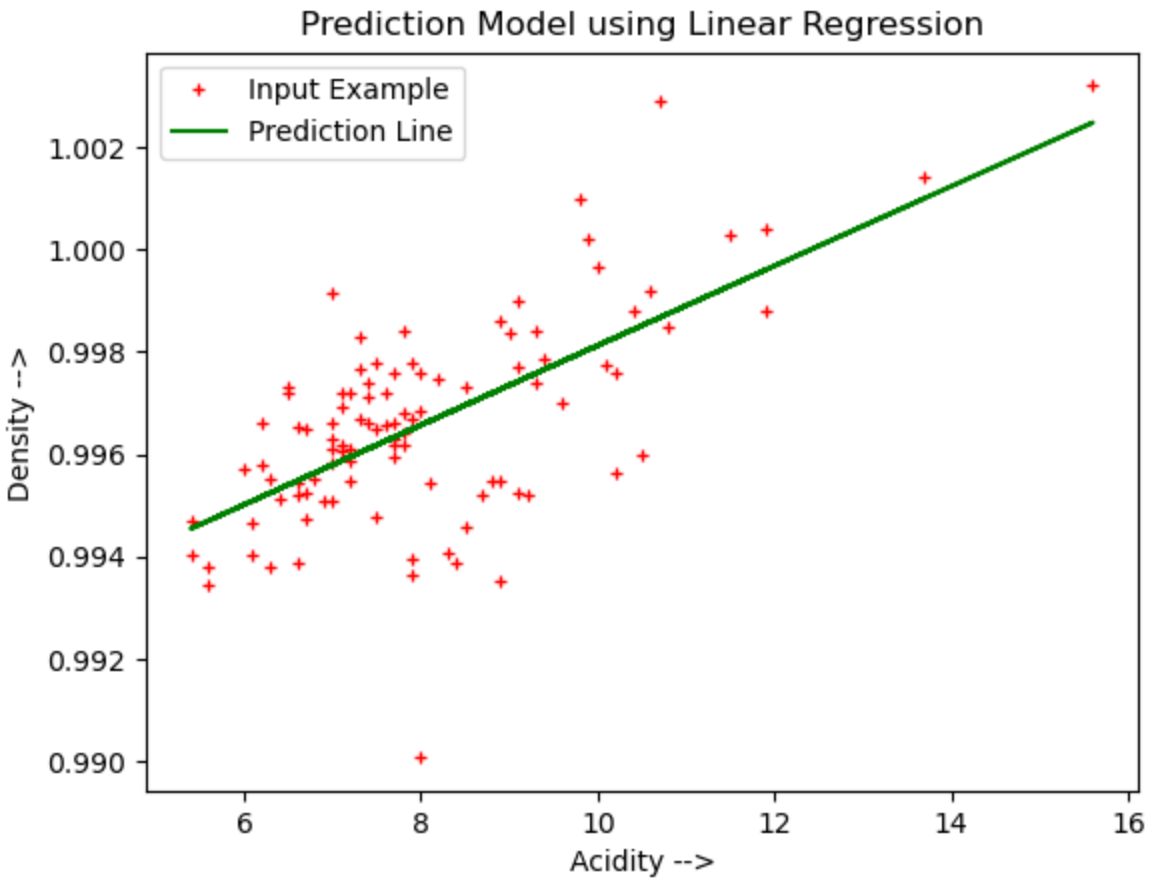
Part a.

Learning rate	No. of iterations	Cost of training-data
0.1	209	1.1947898109837145e-06
0.01	2033	1.1947898109845757e-06
0.001	3960	1.1947898109865624e-06
0.0030	6539	1.194789810987968e-06
0.005	18973	1.1947898109998543e-06

1. **Stopping criteria:**
   a. Max no. of iterations: 20,000
   b. Precision: 1e-20 (consecutive difference in cost)
2. **Final parameter for part a:**
   a. Learning- rate: 0.015
   b. Theta =[[0.9966201], [0.0013402]]
   c. No of iterations : 1365
   d. Cost : 1.1947898109842884e-06

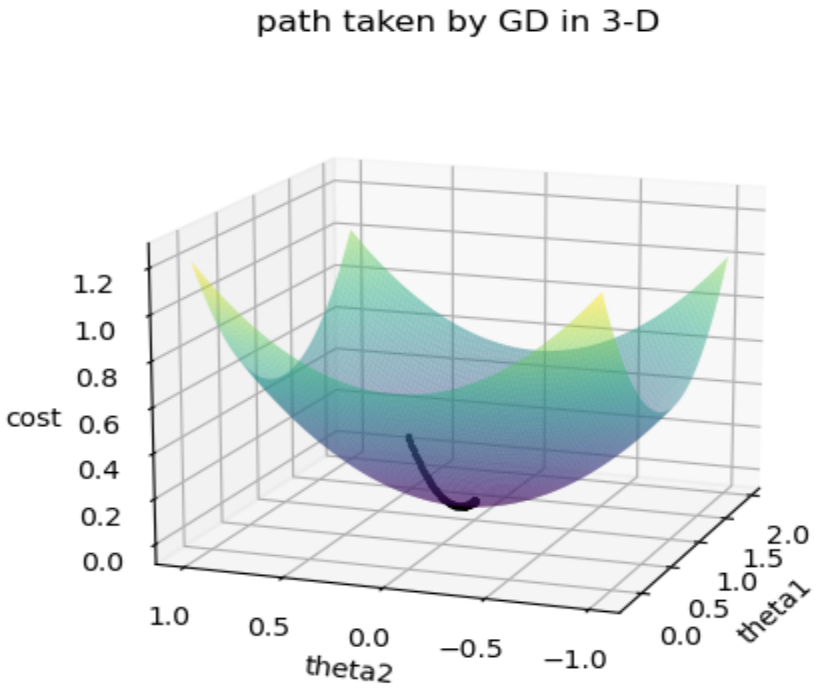
Part b.

The below figure is with learning rate: 0.015



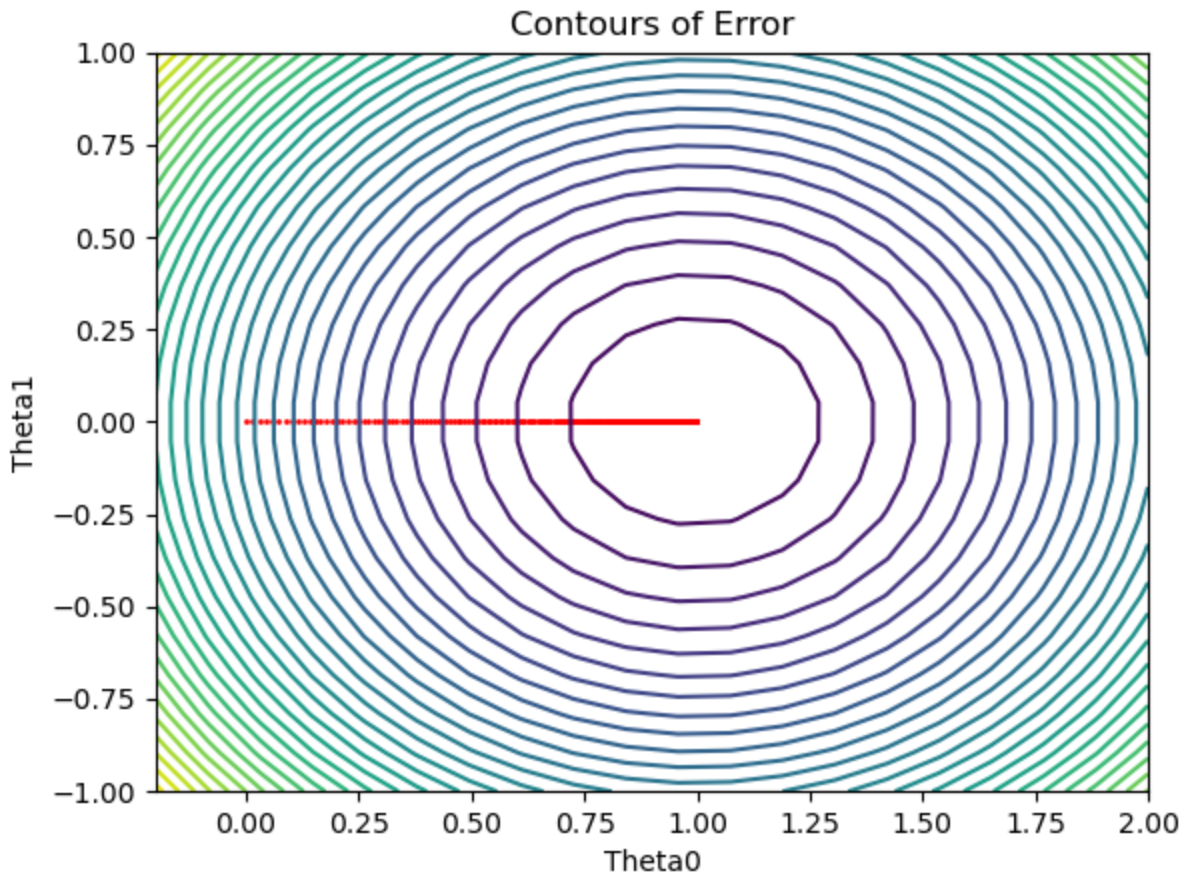
Part c:

The below 3-D mesh plot is for a learning rate of 0.015



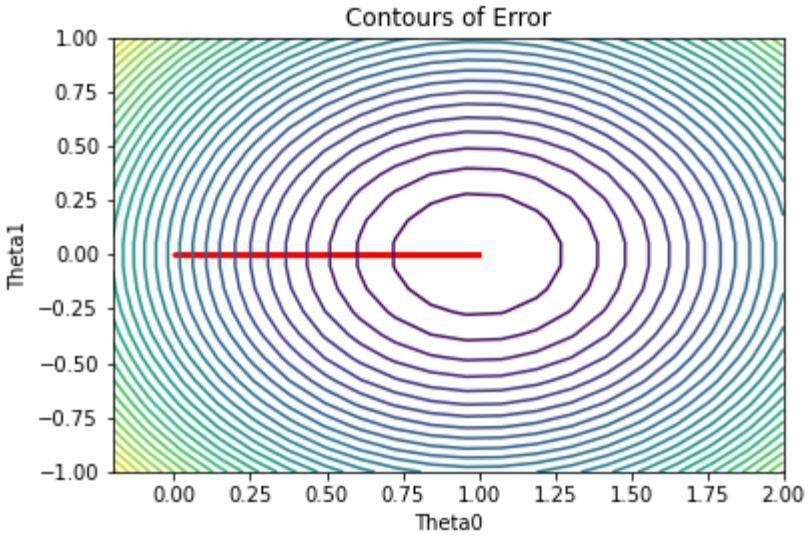
Part d:

The below contours is of 0.015 learning rate.

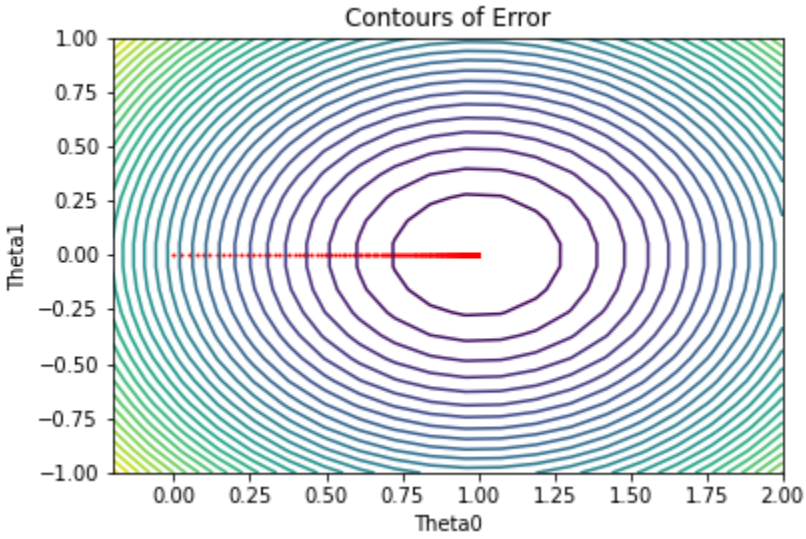


Part e:

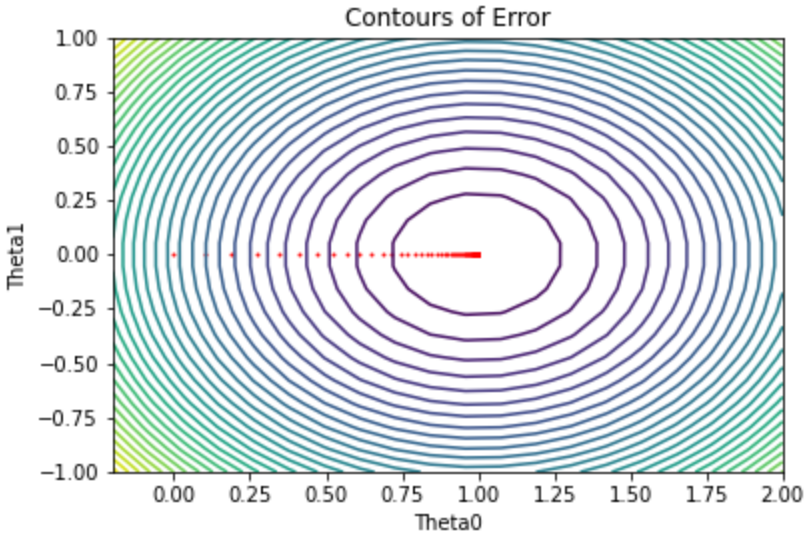
The below contours is of learning rate 0.001



The below contours is of learning rate 0.025



The below contours is of learning rate 0.1



We notice that when we increase the learning rate then gradient descent converges faster but it starts oscillating close to minima.

The higher the learning rate, the more oscillation but if we increase it more than the threshold then it diverges instead of convergence.

Question 2.

Final parameters:

- Learning rate: 0.001
- Stopping conditions:
  - Max-iterations: 20,000
  - Precision: 1e-20 (consecutive difference in cost)

Part b.

Batch size	iterations	theta	error
1	2	[3.02923194] [1.00997357] [1.99602322]	1.002851314768300
100	12	[3.00960376] [0.99926772] [2.0056786]	1.000602387131553

10000	698	[3.01008645] [0.99722328] [2.00070725]	1.000545904135258
1000000	20000(did not converge yet)	[2.99778176] [0.99989644] [1.99990016]	1.000567510452209

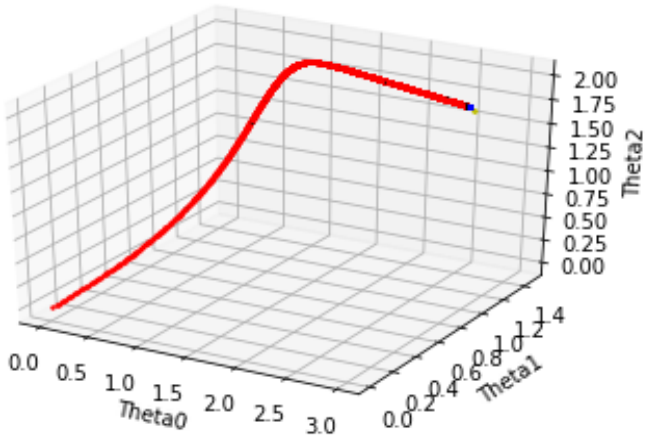
Part c.

Different algorithms converge different parameters but all of them are very close. Cost with original parameters is: 0.9982150452078077

This table shows the errors calculated over the qtest.csv test case and parameters generated from different models like batch sizes 1, 100, 10000, 1000000.

A model with batch size	error
1	`1.2436811879124732`
100	`0.9837191677054679`
10000	`0.9828872813676782`
1000000	`0.9835786686494543`

Part d.



Question 3.

Part a.

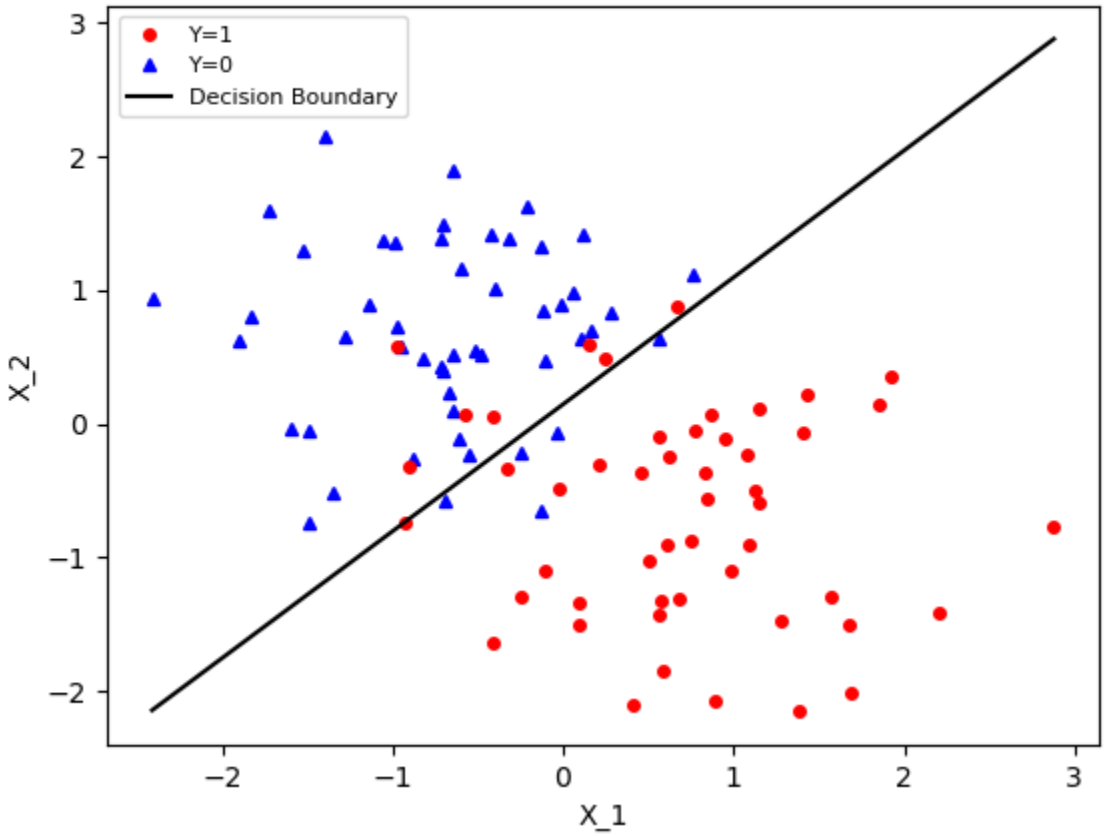
The final parameters found using Newton's method are:

Theta = [0.40125316054552523, 2.588547696944718 5, -2.7255884878403958]

Cost = -22.834144984472392

Part b.

The below graph shows boundary lines for logistic regression.



Question 4.

Part a.

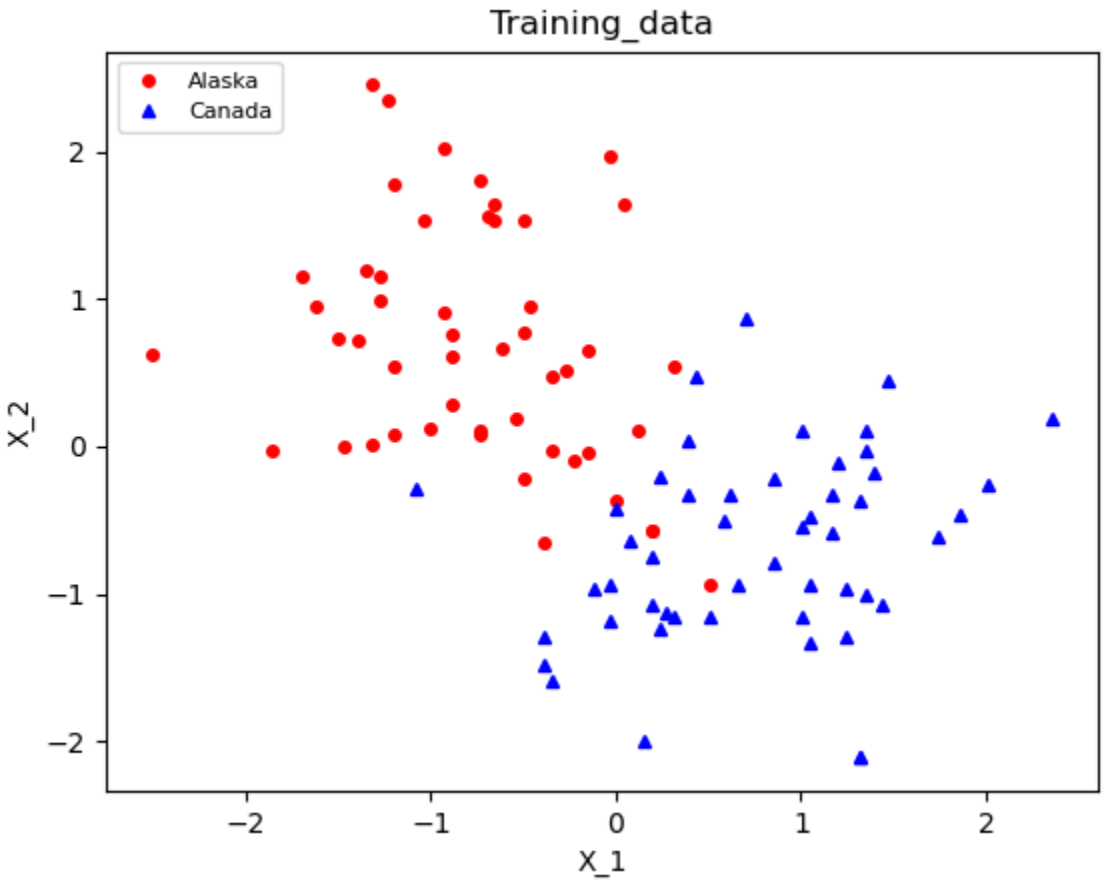
Final parameters found using GDA assuming sigma_0 = sigma_1

Mu_0 = [0.7552943279913608, -0.6850943055489274]

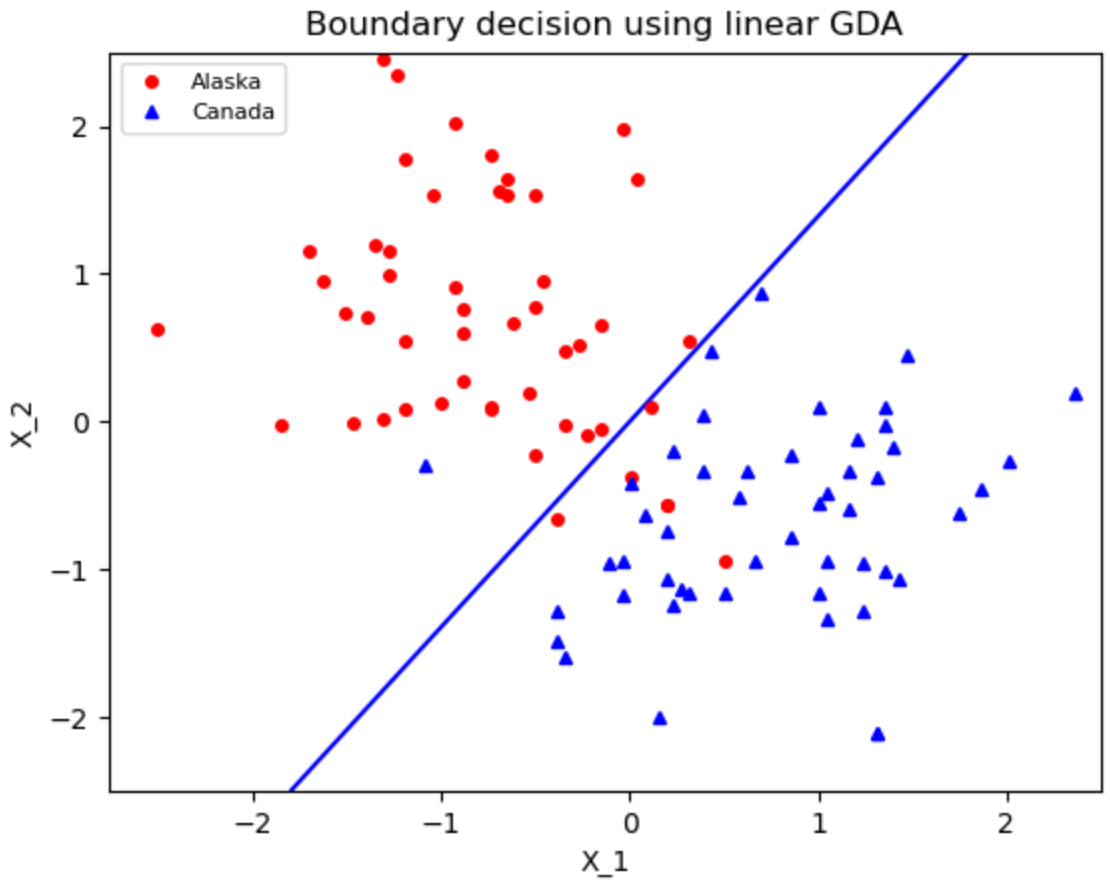
Mu_1 = [-0.7552943279913609, 0.685094305548928]

Sigma = [[0.42953048 -0.02247228]
[-0.02247228 0.53064579]]

Part b.
Training data:



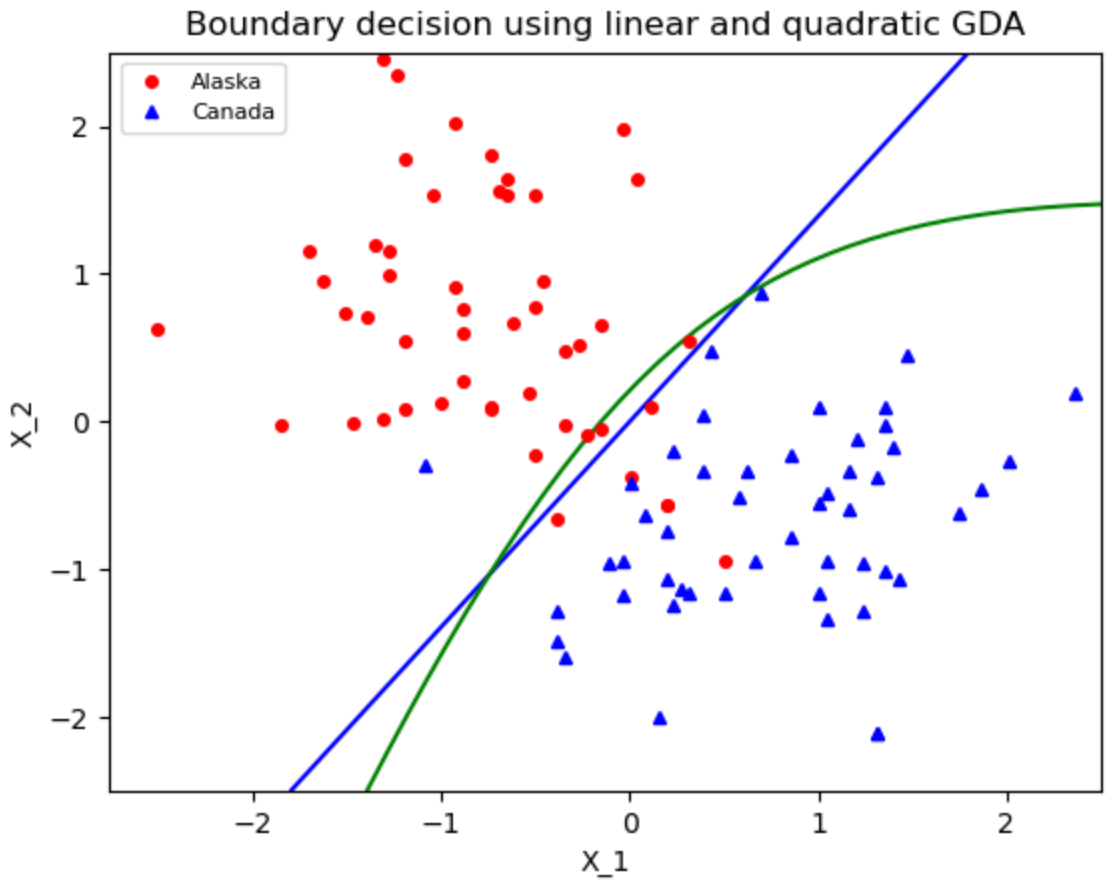
Part c.



Part d.

Mu_0 = [0.7552943279913608, -0.6850943055489274]
Mu_1 = [-0.7552943279913609, 0.685094305548928]
Sigma_0 = [[0.38158978 -0.15486516]
[-0.15486516 0.64773717]]
Sigma_1 = [[0.47747117 0.1099206]
[0.1099206 0.41355441]]

Part e.



Part f.

If we assume the spread of both the data then the decision boundary will be straight which we saw in the derivation of the GDA and we got the straight line in data also.

But in reality, the spread might not be the same that is why when we assume covariance of both the data are different then we get quadratics which is more accurate than a straight line

this shows that GDA classifies more variety of data.