Machine Leaning Assignment-1

Lalith Satya Srinivas Kaladi

February 2020

1 Question 1

1.1 Part 1

(a) **Stopping Criteria**: The Iteration stops when the difference in error between previous iteration and pesent iteration is less than 10^{-10}. That is change in $J(\theta)$ is less than 10^{-10}.

(b) **Learning Rate:** Learning rate is 0.05

(c) **Obtained Parameters:**

$$\theta_0 = 0.9966$$
$$\theta_1 = 0.0013$$

1.2 Part 2

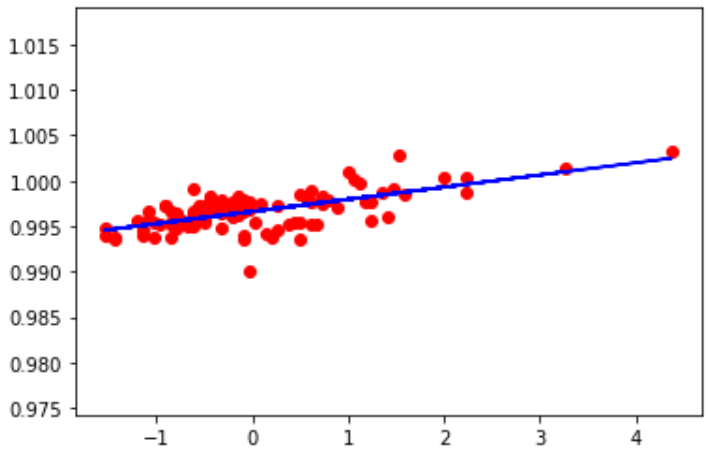

Figure 1: Plot showing datapoints and Hypothesis function

1.3 Part 3

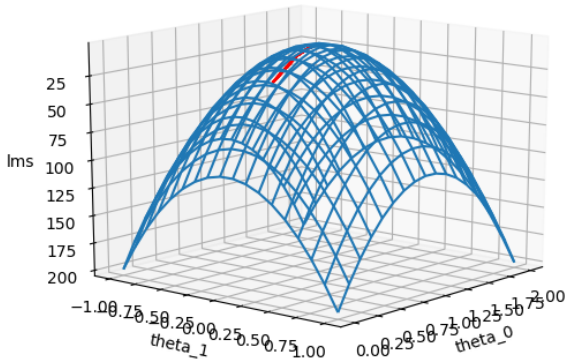

Figure 2: 3D Mesh showing J(θ) with inverted axis, parameters.

1.4 Part 3

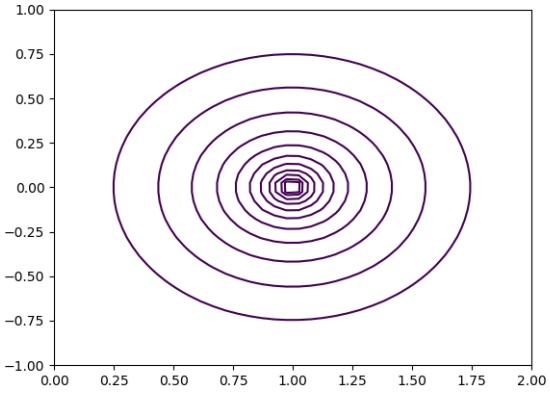

Figure 3: Contour plot of J(θ) with parameters at η=0.05

1.5 Part 4

1.5.1 η=0.001

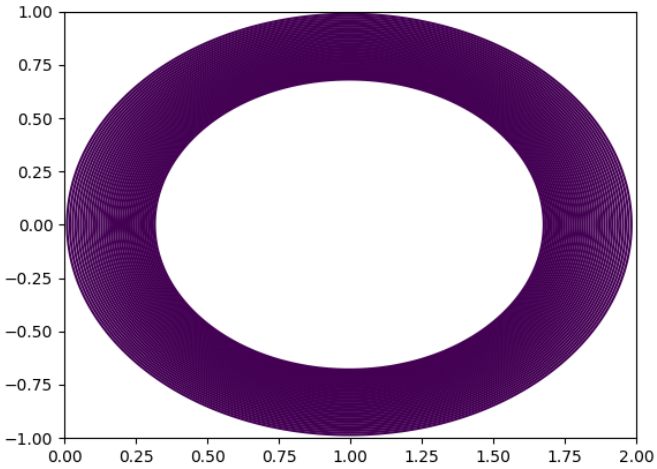

Figure 4: Contour plot of J(θ) with parameters at η=0.05

1.5.2 η=0.025

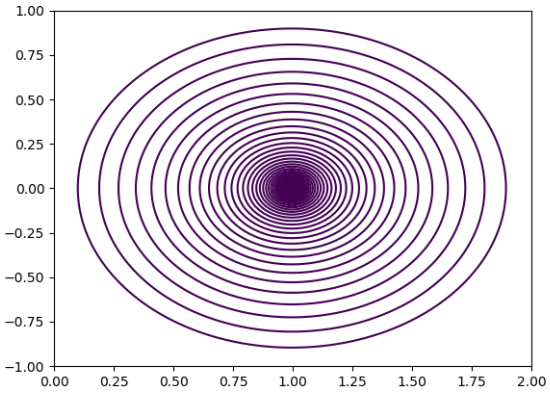

Figure 5: Contour plot of J(θ) with parameters at η=0.05

1.5.3 η=0.1

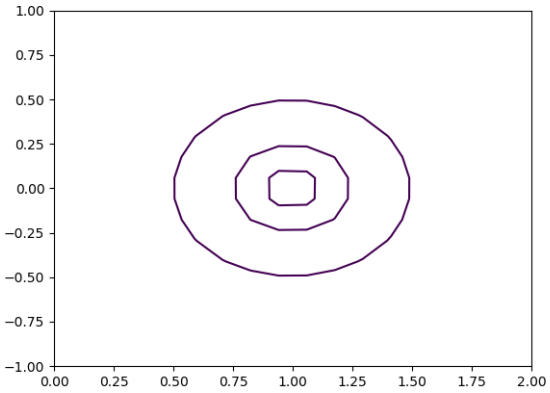

Figure 6: Contour plot of J(θ) with parameters at η=0.05

1.5.4 Inference

The contours for lower η value are closer. Hence as η increases the distance between them is increasing.

2 Question 2

2.1 Part 1

Sampled 1 Million data points

2.2 Part 2 & 3

Batch Size	θ_0	θ_1	θ_2	**Error on q2test**
1	2.870	1.208	2.017	62931.003
100	2.971	1.027	2.048	22531.489
10000	2.983	1.004	2.004	19684.212
1000000	1.996	1.220	1.927	77961.753
NA	3	1	2	19658.938

Convergence Criteria: Two convergence criterias are considered. One for batch convergence and one for total convergence. Change in J(θ) for each batch should be less than a particular value (different for different batch sizes)

and the change average error for the whole 1 Million data points should be less than 10^{-2}.
Observations

1. The initial hypothesis gives minimum error among all of the above mentioned hypothesis obtained by varieing batch sizes.
2. The one with batch size equal to 1,000,000 took 3 complete iteration over all 1 Million data points and rest took only one iteration.

2.3 Part 4

For all the batches theta is converging on the first batch so the plots are a single point plots.

3 Question 3

3.1 Part 1

Parameters Obtained:

$$\theta_0 = 2.583929e6$$
$$\theta_1 = -2.093480e7$$
$$\theta_2 = 2.0325298e7$$

3.2 Part 2

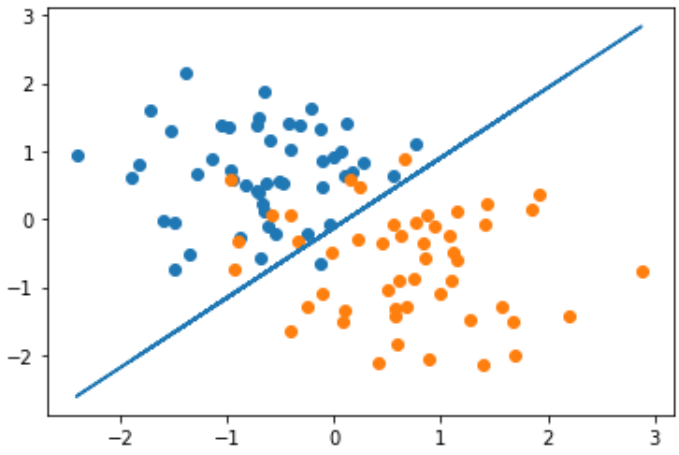

Figure 7: Plot showing datapoints and decision boundary

4 Question 4

4.1 Part 1

1. $\mu_0 = \begin{bmatrix} 0.755 \\ -0.685 \end{bmatrix}$
2. $\mu_1 = \begin{bmatrix} -0.755 \\ 0.685 \end{bmatrix}$
3. $\Sigma = \begin{bmatrix} 0.42953048 & -0.02247228 \\ -0.02247228 & 0.53064579 \end{bmatrix}$

4.2 Part 2

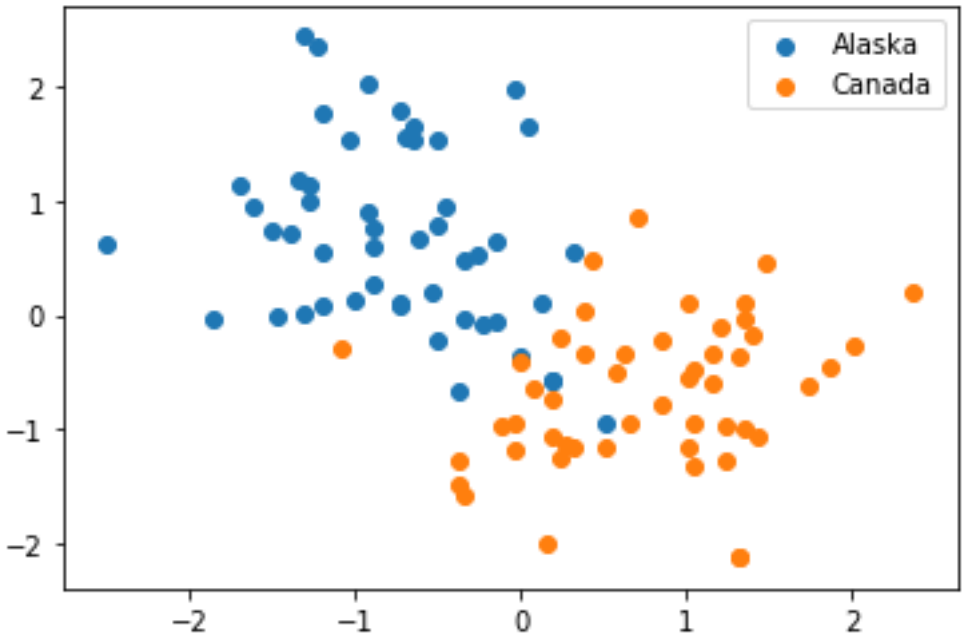

Figure 8: Plot showing datapoints

4.3 Part 3

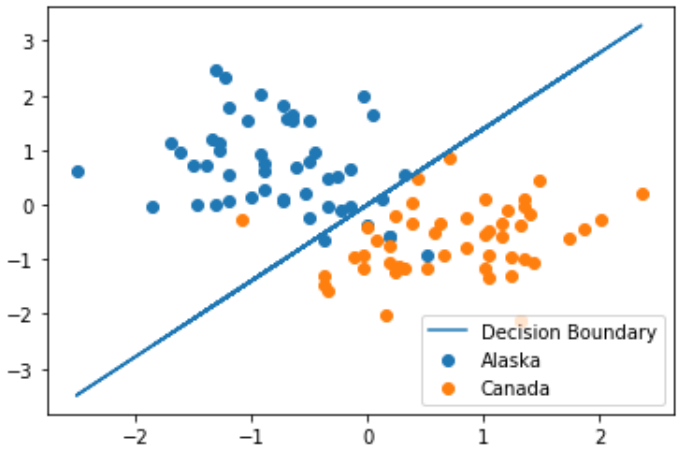

Figure 9: Plot showing datapoints, decision boundary

The equation of the decision boundary is

$$-1.11022302 \times 10^{-15} - 3.38925452x_1 + 2.43858399x_2 = 0$$

Parameters:

$$\theta = [-1.11022302 \times 10^{-15}, -3.38925452, 2.43858399]$$

4.4 Part 4

1. $\mu_0 = \begin{bmatrix} 0.75529433 \\ -0.68509431 \end{bmatrix}$
2. $\mu_1 = \begin{bmatrix} -0.75529433 \\ 0.68509431 \end{bmatrix}$
3. $\Sigma_0 = \begin{bmatrix} 2.75934926 & -1.95987077 \\ -1.95987077 & 2.29097124 \end{bmatrix}$
4. $\Sigma_1 = \begin{bmatrix} 2.66346787 & -2.22465653 \\ -2.22465653 & 2.525154 \end{bmatrix}$

4.5 Part 5

The equation of the decision boundary is

$$-0.00636171 - 0.3719437x_1 + 0.24665587x_2 - 0.24888549x_1^2 - 0.19340056x_2^2 - 0.46210777x_1x_2 = 0$$

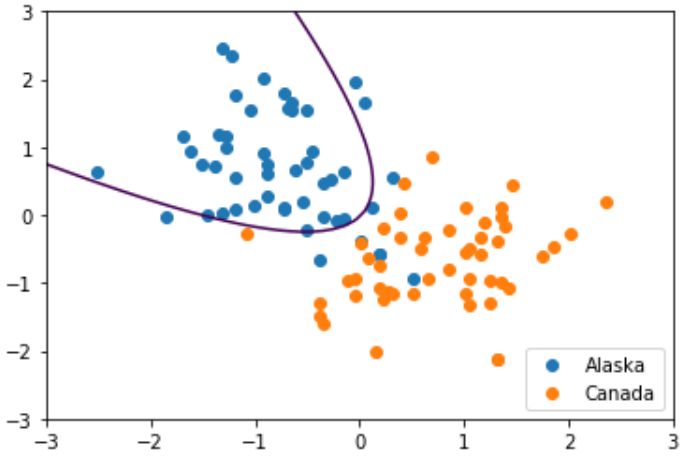

Figure 10: Plot showing datapoints, decision boundary

Parameters:

$\theta = [-0.00636171, -0.3719437, 0.24665587, -0.24888549, 0.19340056, -0.46210777]$

4.6 Part 6

Both take good care about the spread of points but for the given data quadratic decision boundary is considering more precisely when "Canada" points are considered, than the linear boundary.