COL774 - Machine Learning
Assignment 1

Ritesh Baldva
2012CS50711

February 4, 2017

1 Question 1

1.1 Batch Gradient Descent

I was very pessimistic while choosing the learning rate.

$$\eta = 0.00001$$
$$J(\Theta^{(t+1)}) - J(\Theta^{(t)}) = 10^{-12}$$
$$\Theta = [5.83888639, 4.59284554]^T$$

1.2 Data and Hypothesis Plot

The Data plot and hypothesis function can be seen in Q1 Figure (b)

1.3 3D Error function plot and Iteration Plot

The 3D error function plot can be seen in Q1 Figure (c1) and the iteration plot can be seen in Q1 Figure (c2).

1.4 Contour and Iteration Plot

The contour plot can be seen in Q1 Figure (d)

1.5 Different Learning Rates

The contour plots with different etas can be seen in Q1 Figure (e1=0.1, e2=0.3, e3=0.9, e4=1.3). You can see total iterations by seeing the number of dots on the plots. Note with higher learning rates (e5 = 2.1, e6 = 2.5), the gradient descent diverges quickly. And thus the it can't be plotted.

2 Question 2

2.1 Unweighted Linear Regression

The plot of Unweighted regression can be seen in Figure Q2 (a).

2.2 Locally Weighted Regression

The plot of weighted regression, with bandwidth = 0.8, can be seen in Figure Q2 (b).

2.3 Different Bandwidth Parameters

The plot of weighted regression, with bandwidth = 0.1, 0.3, 2, 10, can be seen in Figure Q2 (c1), Figure Q2 (c2), Figure Q2 (c2) and Figure Q2 (c4) respectively. It can be seen that bandwidth = 0.3 suits the best to data, this might also be attributed to the way intervals are chosen since twice the bandwidth is strictly reflected in the weights.

3 Question 3

3.1 Newton's Method Implementation & Logistic Parameters

The equations for Logisitic regression can also be solved using matrices making it easier to implement. The values of parameters obtained are:

$$\Theta = [-0.04717577, 1.46005896, 2.06586134]^T$$

3.2 Logistic Discriminant Plot

The data and discriminator can be seen in Figure Q3 (b).

4 Question 4

From the practice questions, we can see the values of the ML parameters, obtained by maximizing the log-likelihood equation:

$$LL(\phi, \mu_0, \mu_1, \Sigma) = \sum_{i=1}^{m} log(p(x^{(i)}|y^{(i)}; \mu_0, \mu_1, \Sigma)p(y^{(i)}; \phi))$$

4.1 Gaussian Discriminant Analysis

Note the $0, 1$ mapping to the outcomes $Alaska, Canada$ is maintained. Thus, the following values can be simply calculated from the values obtained:

$$\phi = \frac{1}{m}\sum_{i=1}^{m} 1(y^{(i)} = 1) = 0.5$$

$$\mu_0 = \frac{\sum_{i=1}^{m} 1(y^{(i)} = 0) * x^{(i)}}{\sum_{i=1}^{m} 1(y^{(i)} = 0)} = [-0.75529433, 0.68509431]^T$$

$$\mu_1 = \frac{\sum_{i=1}^{m} 1(y^{(i)} = 1) * x^{(i)}}{\sum_{i=1}^{m} 1(y^{(i)} = 1)} = [0.75529433, -0.68509431]^T$$

$$\Sigma = \begin{bmatrix} 0.42953048 & -0.02247228 \\ -0.02247228 & 0.53064579 \end{bmatrix}$$

4.2 Original Data Plot

The graph can be seen in Figure Q4(b).

4.3 Linear Discriminant Plot

The Linear Discriminator can be seen in Figure Q4(c).

4.4 Different Covariance Matrices

Values of parameters obtained when covariance matrices are different.

$$\phi = 0.5$$

$$\mu_0 = [-0.75529433, 0.68509431]^T$$

$$\mu_1 = [0.75529433, -0.68509431]^T$$

$$\Sigma_0 = \begin{bmatrix} 0.38158978 & -0.15486516 \\ -0.15486516 & 0.64773717 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} 0.47747117 & 0.1099206 \\ 0.1099206 & 0.41355441 \end{bmatrix}$$

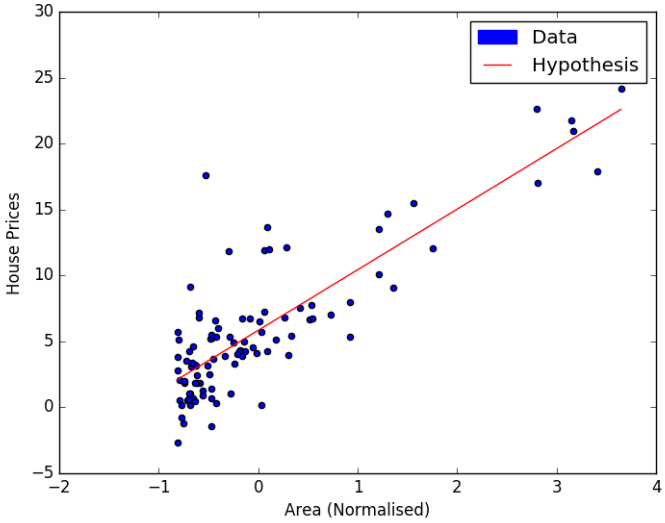

Figure 1: Q1(b) Data and Hypothesis

4.5 Quadratic and Linear Discriminant Plot

The Linear and Quadratic Discriminator in Figure Q4(e).

4.6 Observations

Following things were observed.

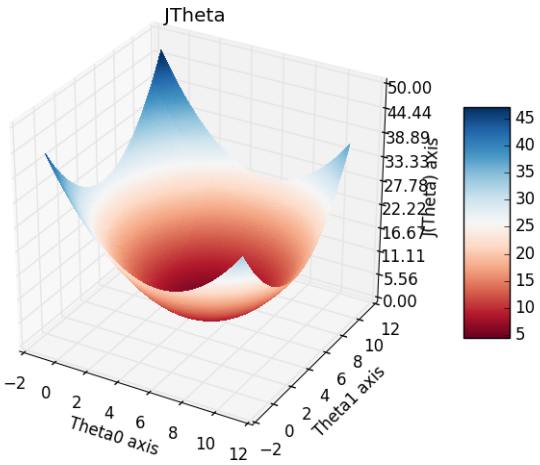

Figure 2: Q1(c1) 3D Error Function Plot

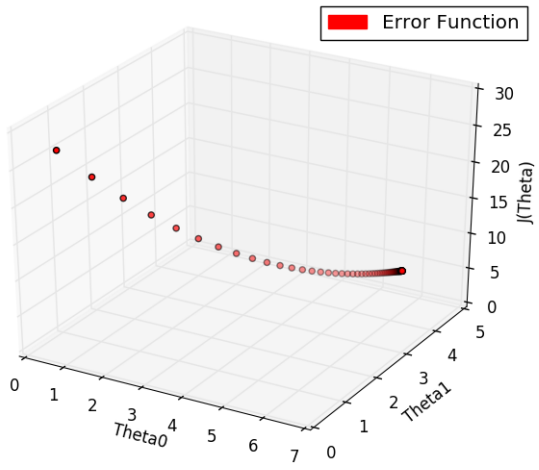

Figure 3: Q1(c2) 3D Iteration Plot

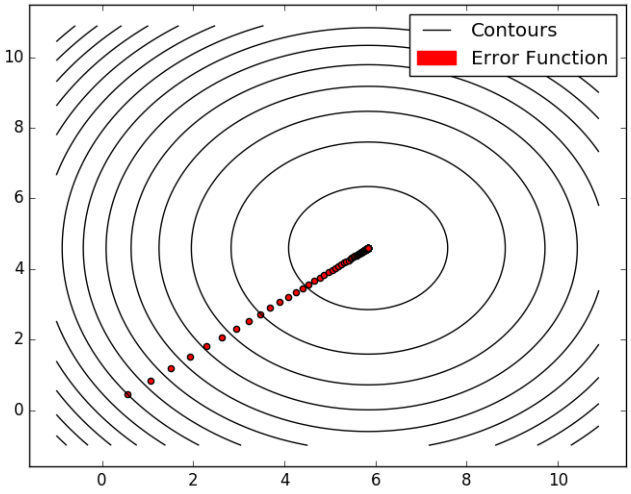

Figure 4: Q1(d) 3D Iteration Plot

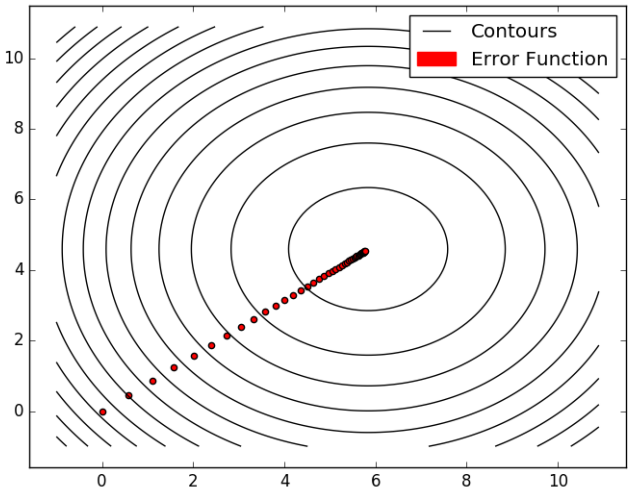

Figure 5: Q1(e1) Eta = 0.1

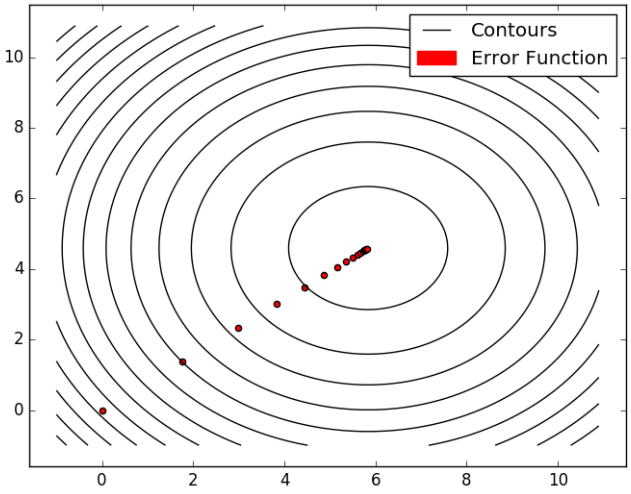

Figure 6: Q1(e2) Eta = 0.5

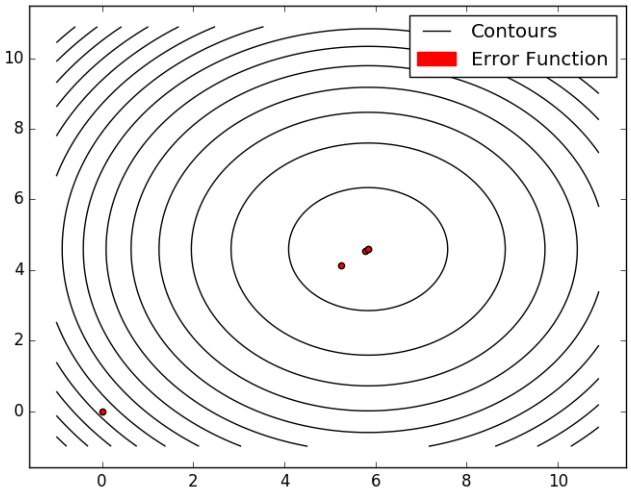

Figure 7: Q1(e3) Eta = 0.9

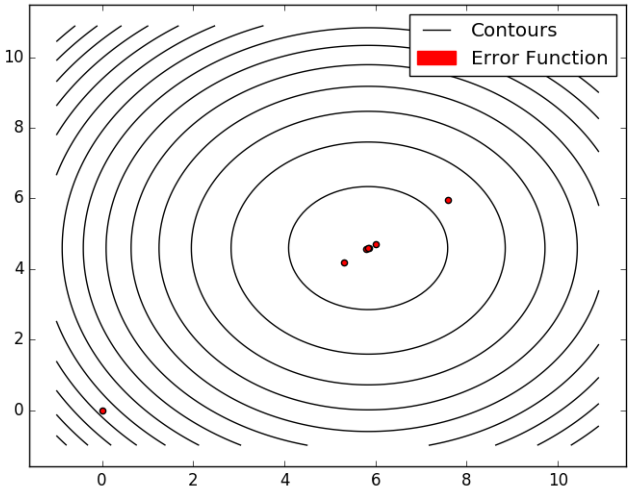

Figure 8: Q1(e4) Eta = 1.3

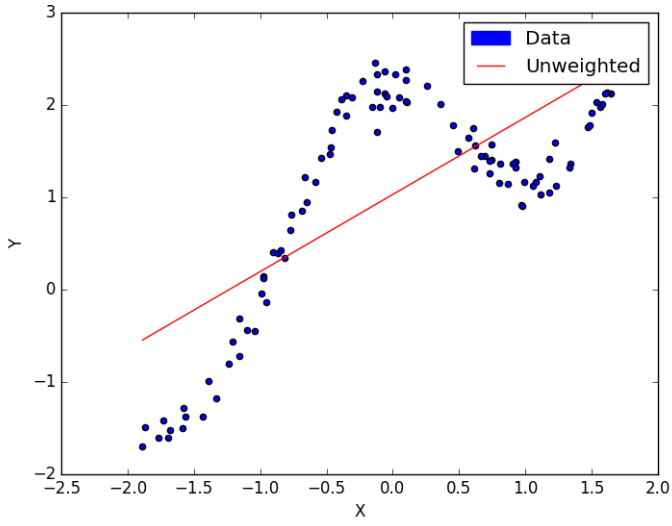

Figure 9: Q2(a) Unweighted linear regression

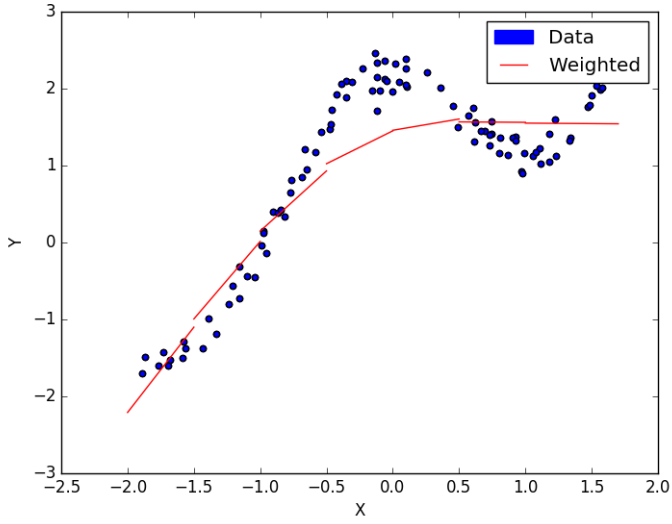

Figure 10: Q2(b) Weighted (Bandwidth = 0.8) linear regression

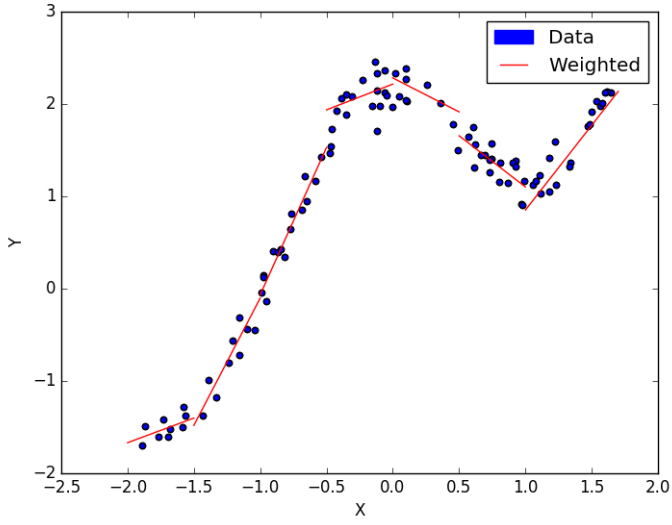

Figure 11: Q2(c1) Weighted (Bandwidth = 0.1) linear regression

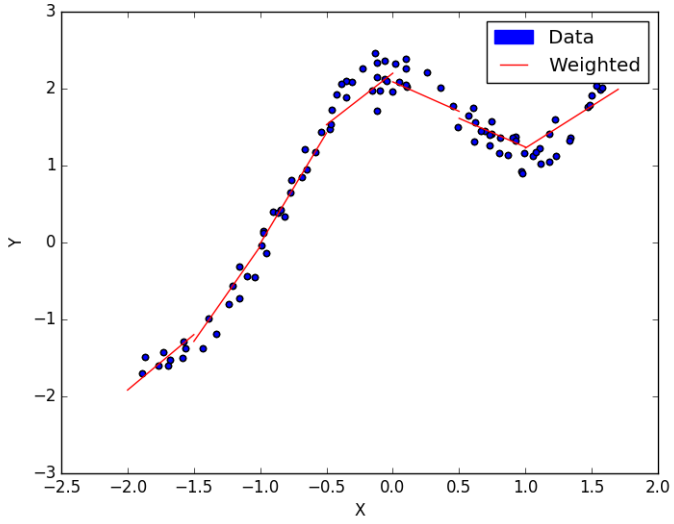

Figure 12: Q2(c2) Weighted (Bandwidth = 0.3) linear regression

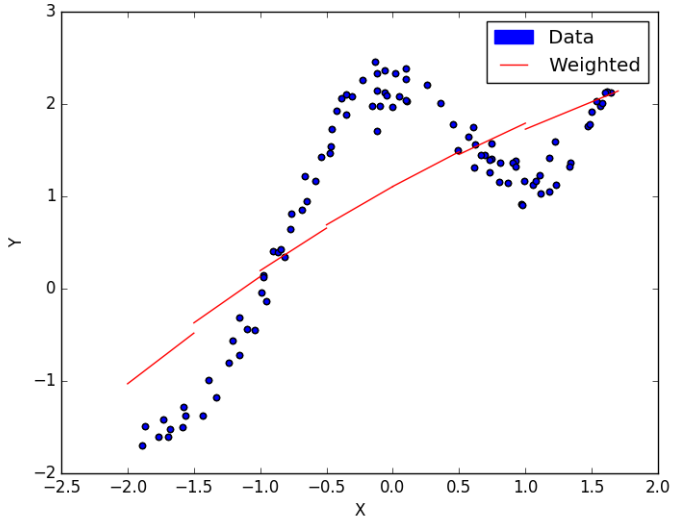

Figure 13: Q2(c3) Weighted (Bandwidth = 2) linear regression

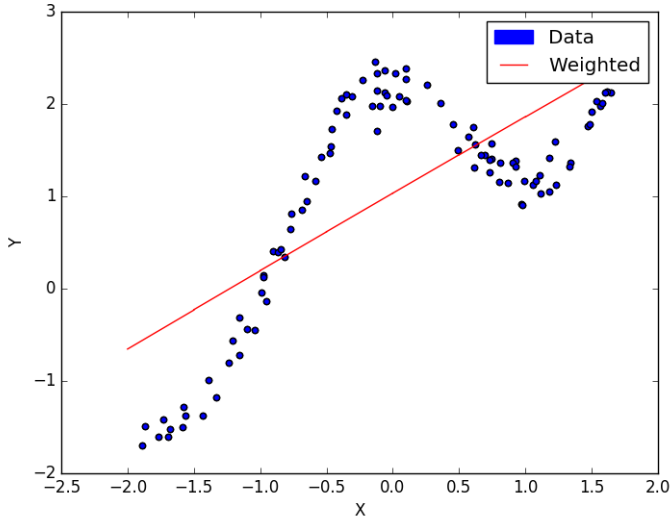

Figure 14: Q2(c4) Weighted (Bandwidth = 10) linear regression

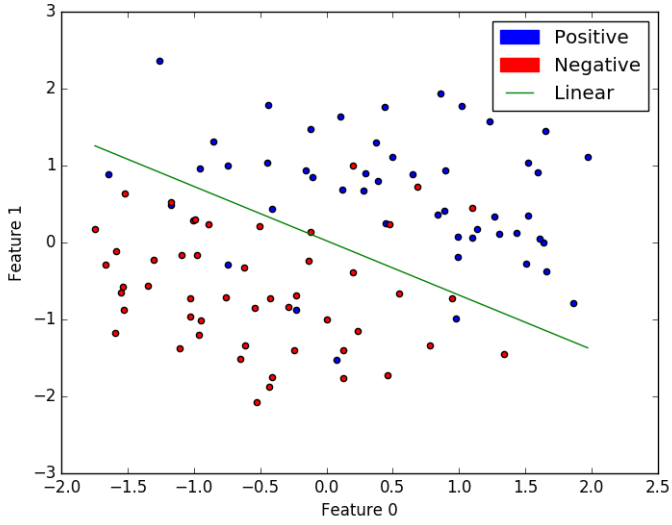

Figure 15: Q3(b) Logistic Discriminator and Data in the Feature Space

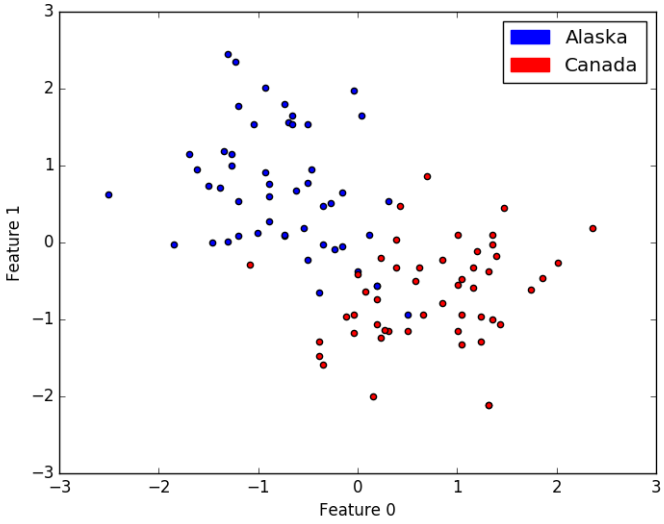

Figure 16: Q4(b) Data in the Feature Space

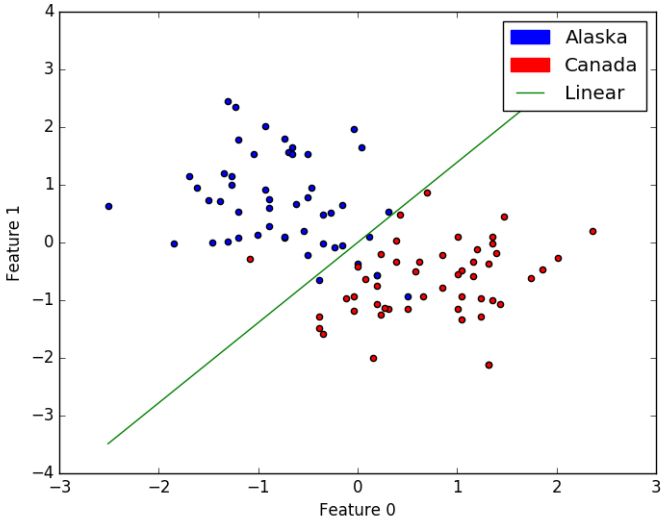

Figure 17: Q4(c) Linear Separator

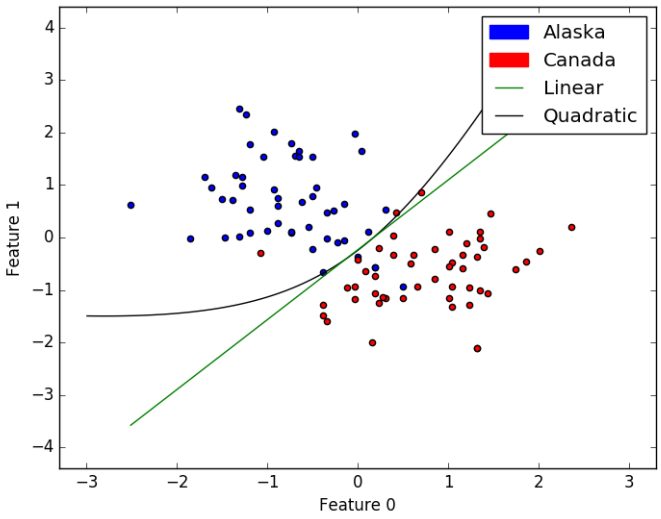

Figure 18: Q4(e) Quadratic and Linear Separator