COL 774: Assignment 1

Submitted By - Mahima Manik
Entry Number - 2017MCS2093

1. Linear Regression:

a. Learning rate: 0.01
   θ_0 = 0.996620099664
   θ_1 = 0.00134019601806
   J(θ) = 0.000119478981098
   Stopping Criteria: $J(\theta)_{i+1}$ >= $J(\theta)_i$

b.

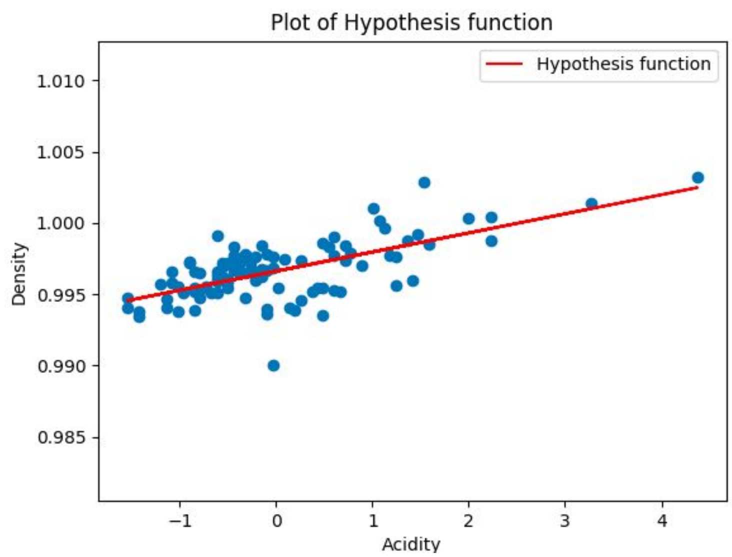


Figure 1

c.

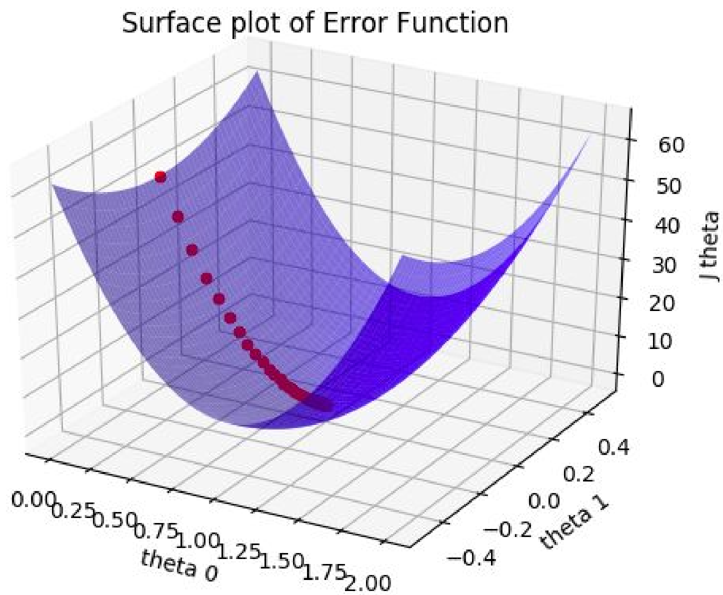


Figure 2

d.

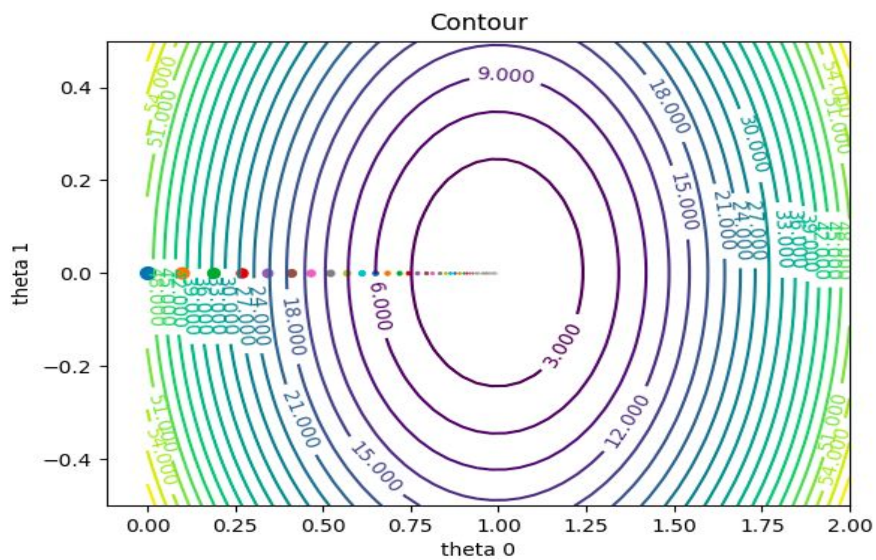


Figure 3: Contour for eta = 0.001

E.

Learning Rate	**Number of Iterations**	**Error Function (J theta)**
0.001	208	0.000119478981098
0.005	36	0.000119478981098
0.009	11	0.000119478981098

0.013	22	0.000119478981098
0.017	66	0.000119478981098
0.021	1	49.6627904715
0.025	1	49.6627904715

Learning rate decides that how well the function is converging to the desired point and guides the step size in the direction of the gradient.
If learning rate is chosen too small, then it takes very large number of iterations to converge, because the step size is very small.
If learning rate is chosen too high, then it might overshoots the desired point, because gradient changes at every stage. We observe that as the learning rate increases, number of iterations required to converge decreases. It oscillates before coming to the minima at learning rate of 0.013 and 0.017. It does not converge at all at 0.21 and 0.25.

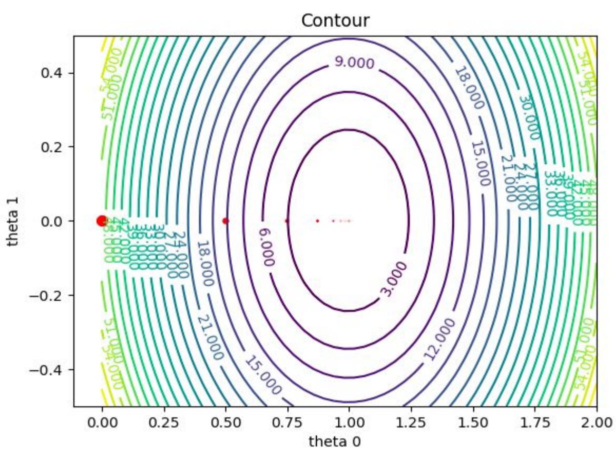


Figure 4: Contour for eta = 0.005

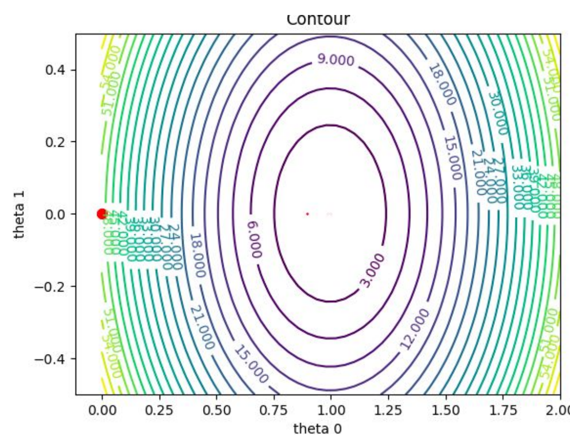


Figure 5: Contour for eta = 0.009

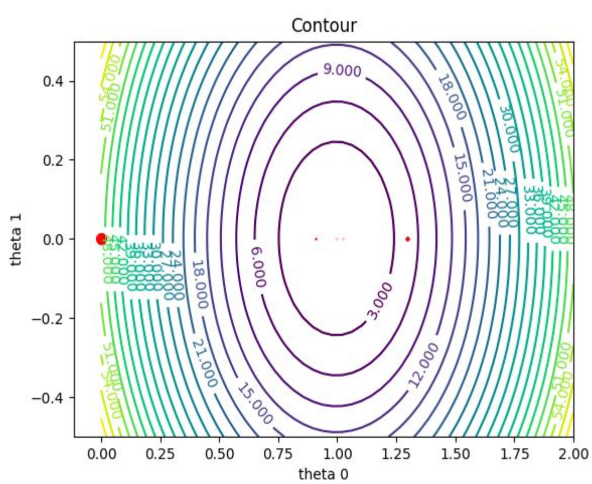


Figure 6: Contour for eta = 0.013

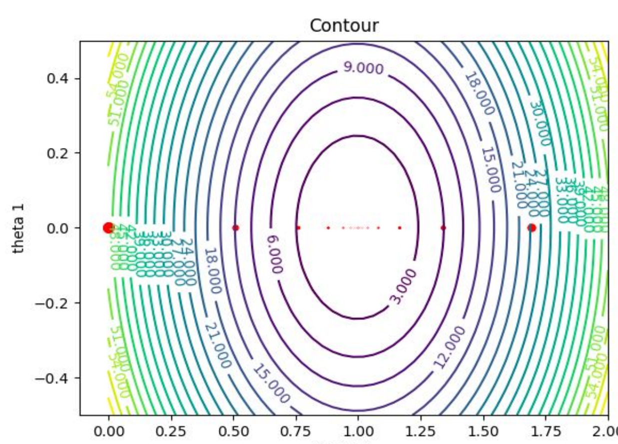


Figure 7: Contour for eta = 0.017

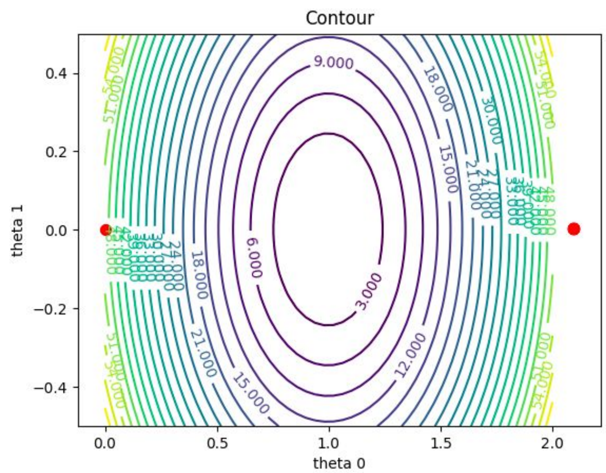


Figure 8: Contour for eta = 0.021

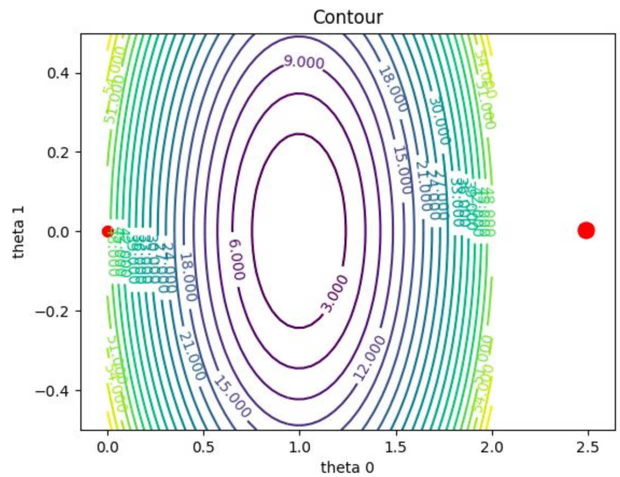


Figure 9: Contour for eta = 0.025

2. Locally Weighted Linear Regression

a. $\theta_0 = -1.87886923$
$\theta_1 = 0.64695273$

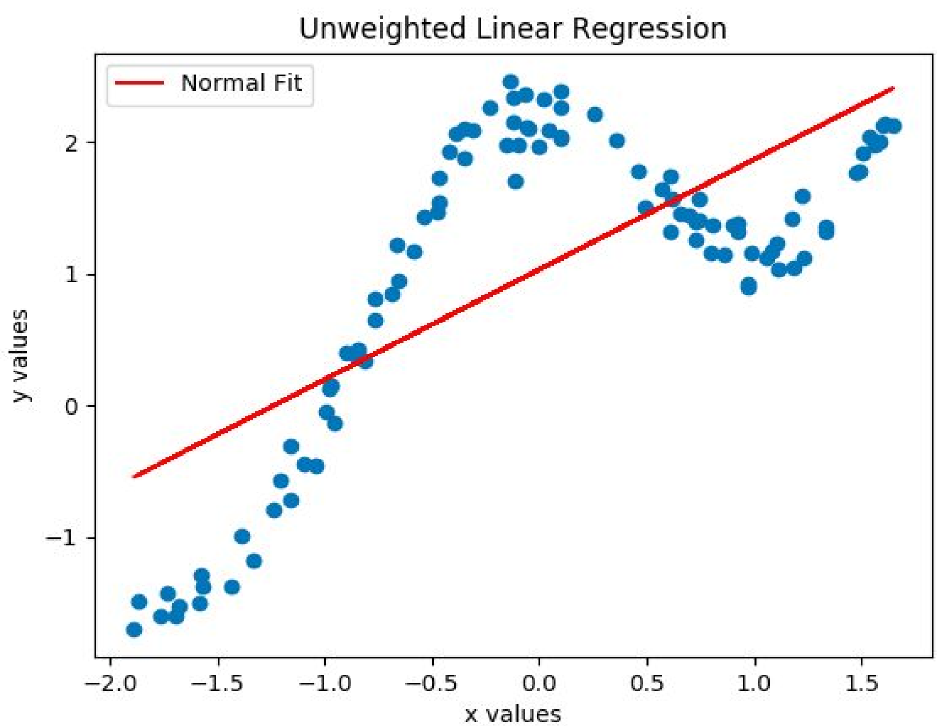


Figure 10

B. Equation for finding values of θ in locally weighted linear regression

$$\theta = (X^T X)^{-1} W (X^T Y)$$

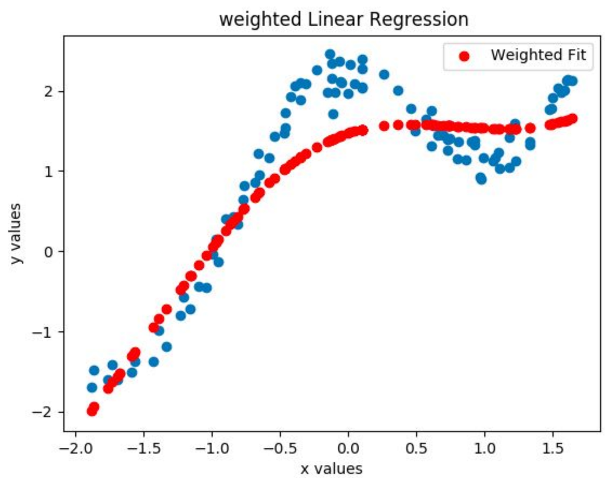


Figure 11

C. As bandwidth parameter increases, weight becomes 1 for all the m points in the example space and it starts behaving as linear regression. With very low bandwidth parameter, for all the points other than point of interest, their weight becomes 0. Function will pass through every points in the example space.

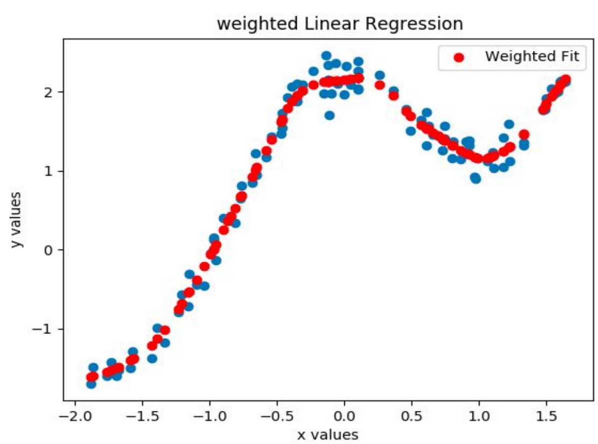


Figure 12: For eta=0.1

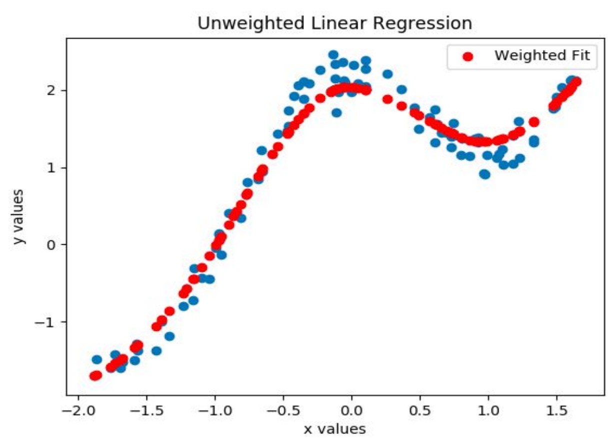


Figure 13: For eta=0.3

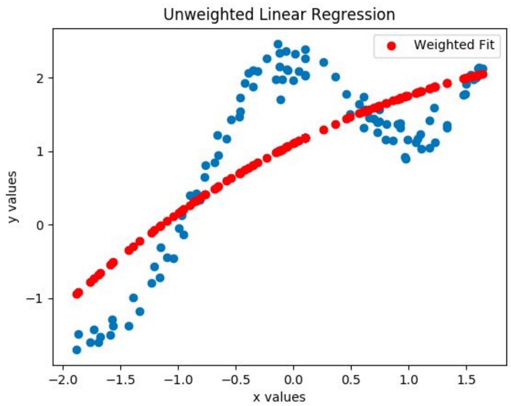


Figure 14: For eta=2

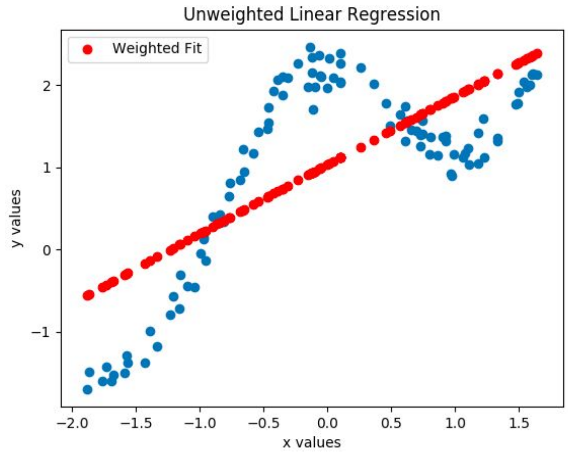


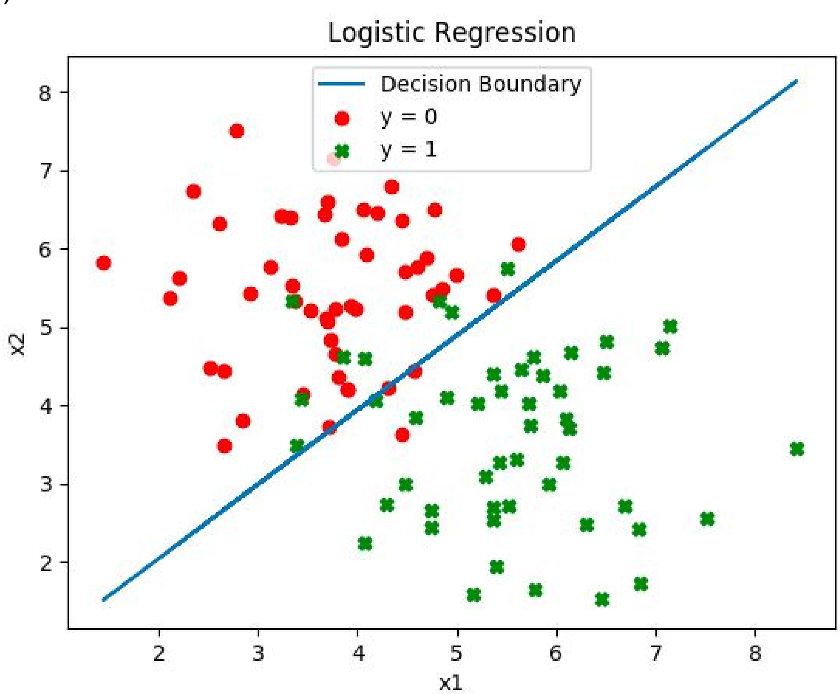
Figure 13: For eta=10

Q. 3.

a)

Theta 0	0.302258111805
Theta 1	2.29958184979
Theta 2	-2.40859570745

b)



Q. 4.

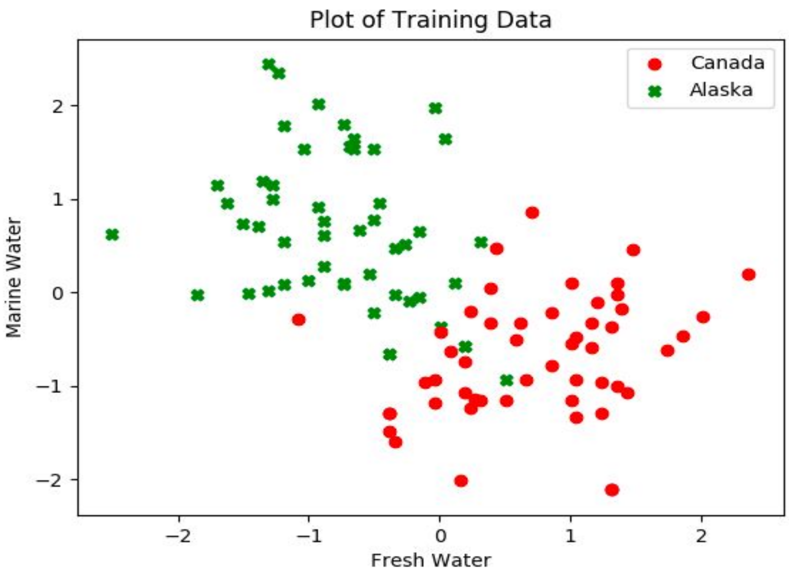
A.

$\mu_0 = [0.75529433 \ -0.68509431 \ ]$

$\mu_1 = [-0.75529433 \ 0.68509431 \ ]$

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

B. Decision variable is taken as 1 for Alaska and 0 for Canada in the Vector Y.



C. Linear equations are expressed in the following form of:

$$Ax_1 + Bx_2 + C = 0$$

Here we express x_2 as the function of x_1 and obtain:

$$x_2 = \frac{-C-Ax_1}{B}$$

Solving for $\Sigma_1 = \Sigma_0$, we get the following form of the equation:

$$2 * ((\mu_0^T - \mu_1^T) * \Sigma^{-1}) * x_1 + (\mu_1^T * \Sigma^{-1} * \mu_1 - \mu_0^T * \Sigma^{-1} * \mu_0) - log\ \frac{\varnothing}{1-\varnothing} = 0$$

Where,

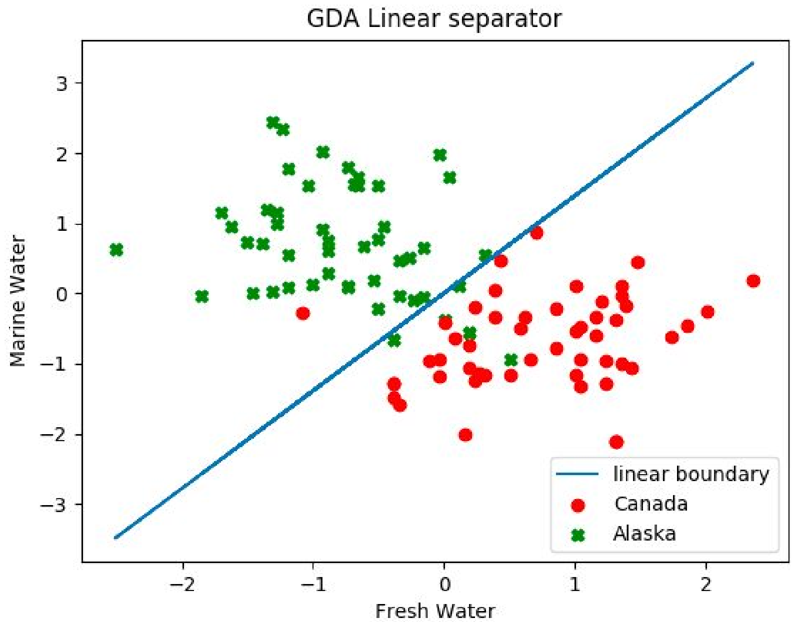
X = Input Features

μ_0 = Mean of input features where the target variable is 0

μ_1 = Mean of input features where the target variable is 1

Σ = Sample Covariance

$\varnothing$ = Probability of the target variable being 1



D.

μ_0 = [0.75529433 -0.68509431]

μ_1 = [-0.75529433 0.68509431]

Σ_0 = [[0.47747117 0.1099206]
 [0.1099206   0.41355441]]

Σ_1 = [[0.38158978 -0.15486516]
 [-0.15486516 0.64773717]]

E. Quadratic equations can be expressed in the following form:

$$A * x_1^2 + B * x_1 + C = 0$$

Equation for GDA quadratic separator becomes:

$$X^T(\Sigma_1^{-1} - \Sigma_0^{-1}) * X + 2 * (\mu_0^T * \Sigma_1^{-1} - \mu_1^T * \Sigma_0^{-1}) * X + (\mu_1^T * \Sigma_1^{-1} * \mu_1 - \mu_0^T * \Sigma_0^{-1} * \mu_0) - 2 * log \frac{\varnothing}{1-\varnothing} + 2 * log \frac{|\Sigma_1|}{|\Sigma_0|} = 0$$

where,

X = Input Features

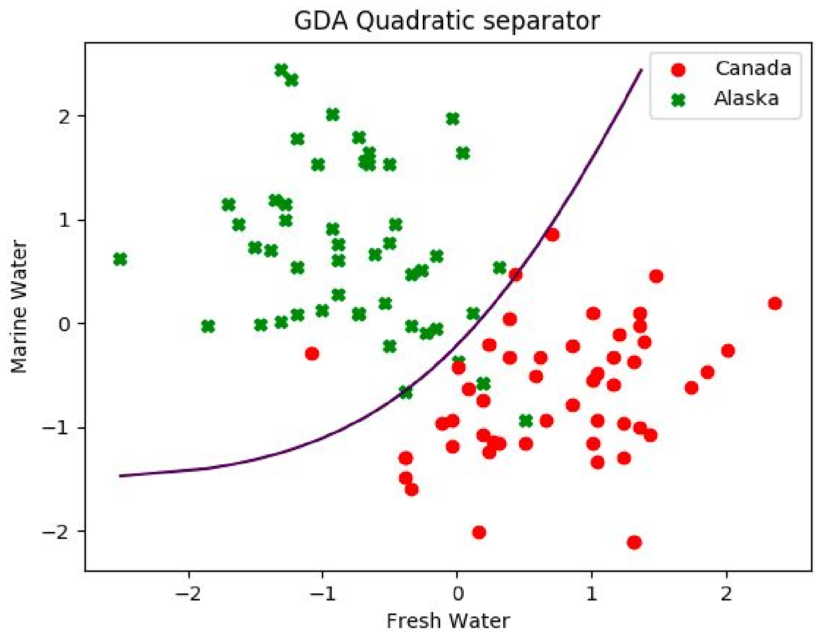
μ_0 = Mean of input features where the target variable is 0

μ_1 = Mean of input features where the target variable is 1

Σ_0 = Sample Covariance where the target variable is 0

Σ_1 = Sample Covariance where the target variable is 1

$\varnothing$ = Probability of the target variable being 1



F. Quadratic curve gives better fit than the linear as the points gets better separated.