Ashley Jain
2014CS50280
9 February 2017

COL774: Assignment 1

1.(a) Normalize:-

$x^{(i)} := (x^{(i)}\text{-mean}(x))/\text{std}(x)$ where std is standard deviation of column vector x

(We used normalized x variable for faster in convergence and circular contour graph)

In each iteration,Updating θ by $\theta := \theta - ((1/m)*\alpha*(y - X^{(i)}*\theta)' * X)'$ until converges.

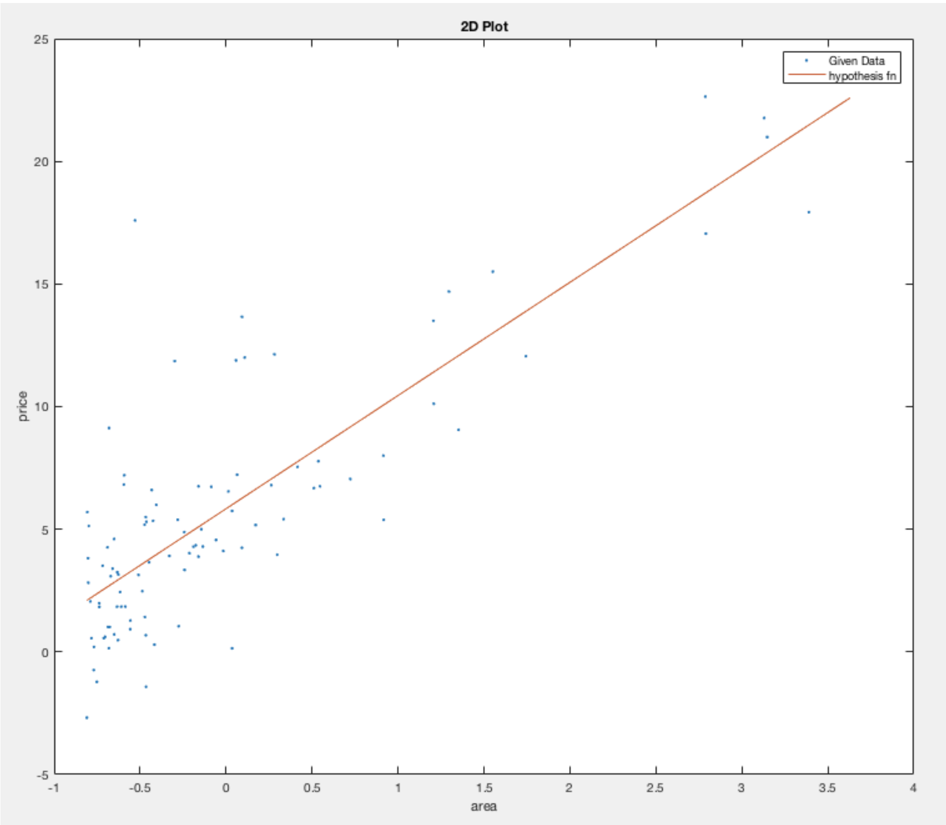
Stopping criteria :-

$|J(\theta_{t+1})\text{-}J(\theta_t)| < \epsilon$

$\alpha = 0.2$, $\epsilon = 0.00001$, $\theta = [5.8354;4.6137]$

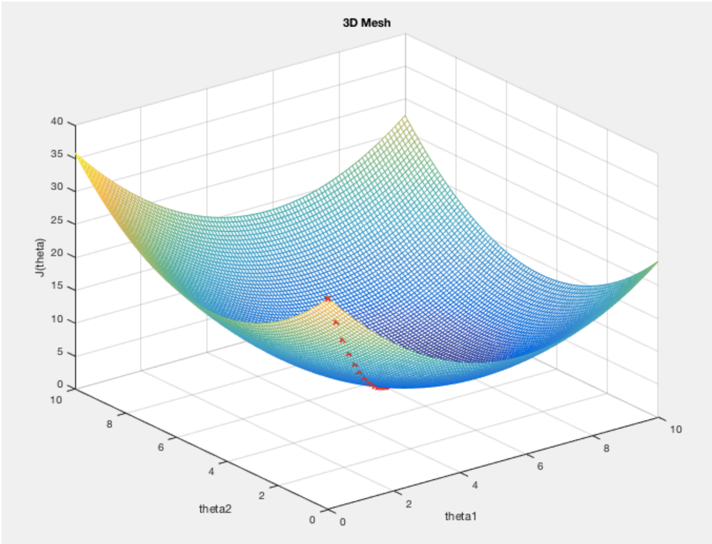
1.(b) Hypothesis function:- $y = \theta^T x$

Graph:-

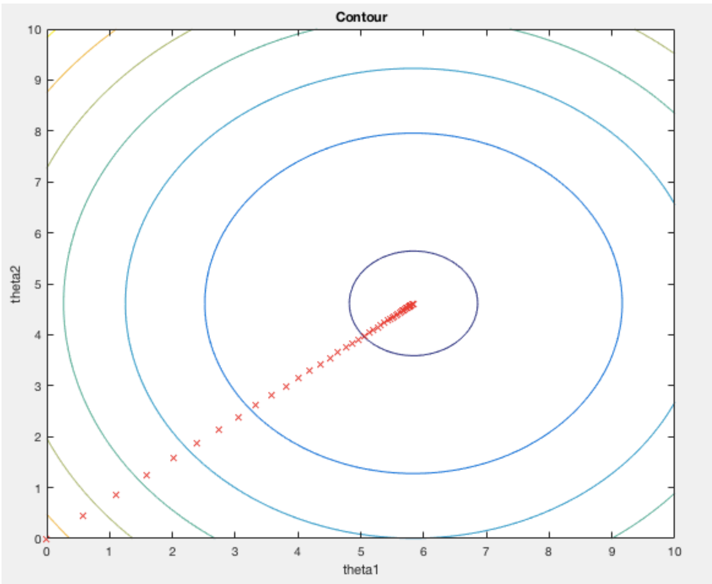


1.(c) #iterations=34

Graph(at the end of iteration):-

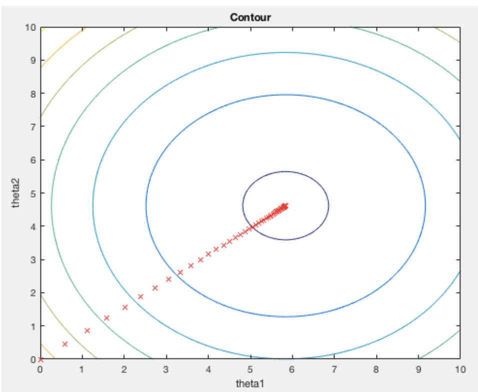


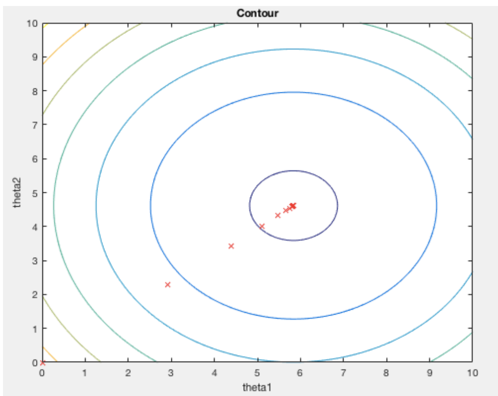
1.(d) Graph:-



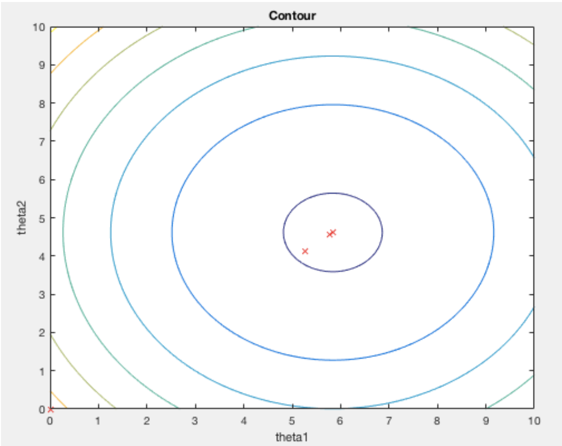
1.(e) α = 0.1,#iterations=65

α = 0.5,#iterations=13

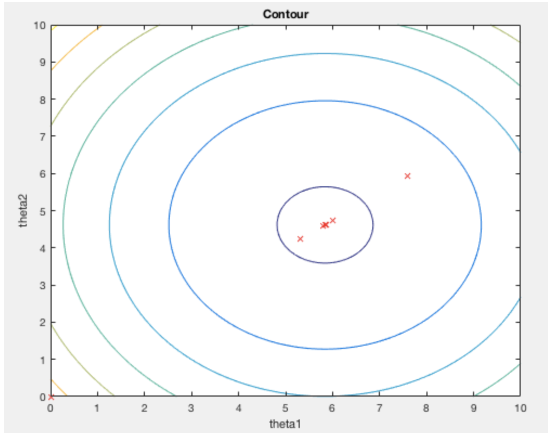




α = 0.9,#iterations=6



α=1.3,#iterations=9



for α>2.0 , #iterations would be greater than 3500.

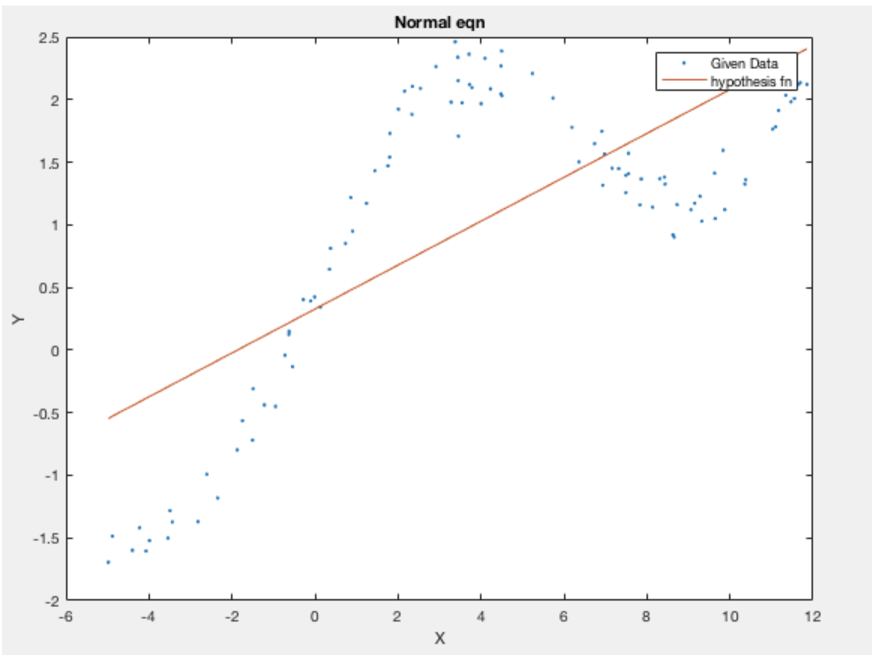
<u>Observation and conclusion</u>:-

- α = 0.9 has minimum no. of iterations. $\alpha < 0.9$ will change θ in small amount toward minima, hence has more no. of iterations. $\alpha > 0.9$, θ will jump to opposite side of minima and then comes back to the minima, hence oscillating around minima.
- Increase the α value after 0.9, no. of iterations would be larger due to overshooting of the parameters and then coming back to minima.
- For α = 2.1 and 2.5, θ never reach minima due to distance between the new θ point and optimal point is greater than that of previous θ point.Hence Error value keeps on increasing in each iteration.

2. (a) Given X and y, we can compute theta by using following equation:-

$$\theta = (X^TX)^{-1}X^Ty \quad \text{(derived from min } J(\theta)\text{)}$$

And finally plot with function $y = \theta^T x$



2. (b) For each query point:-

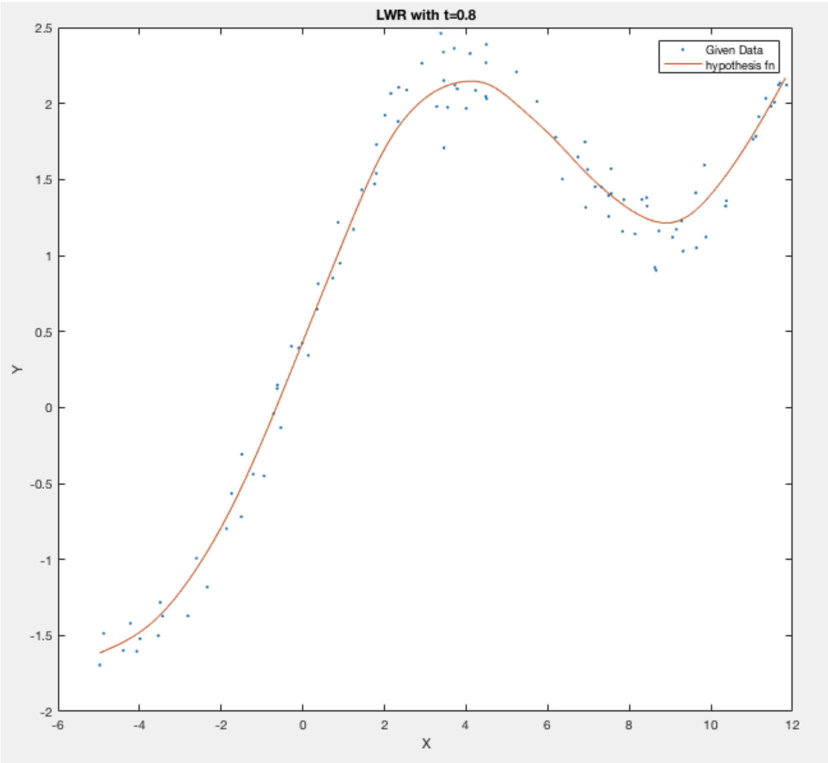
$\theta = (X^TWX)^{-1}X^TWY$

where W is diagonal matrix with m*m entries, each entry is

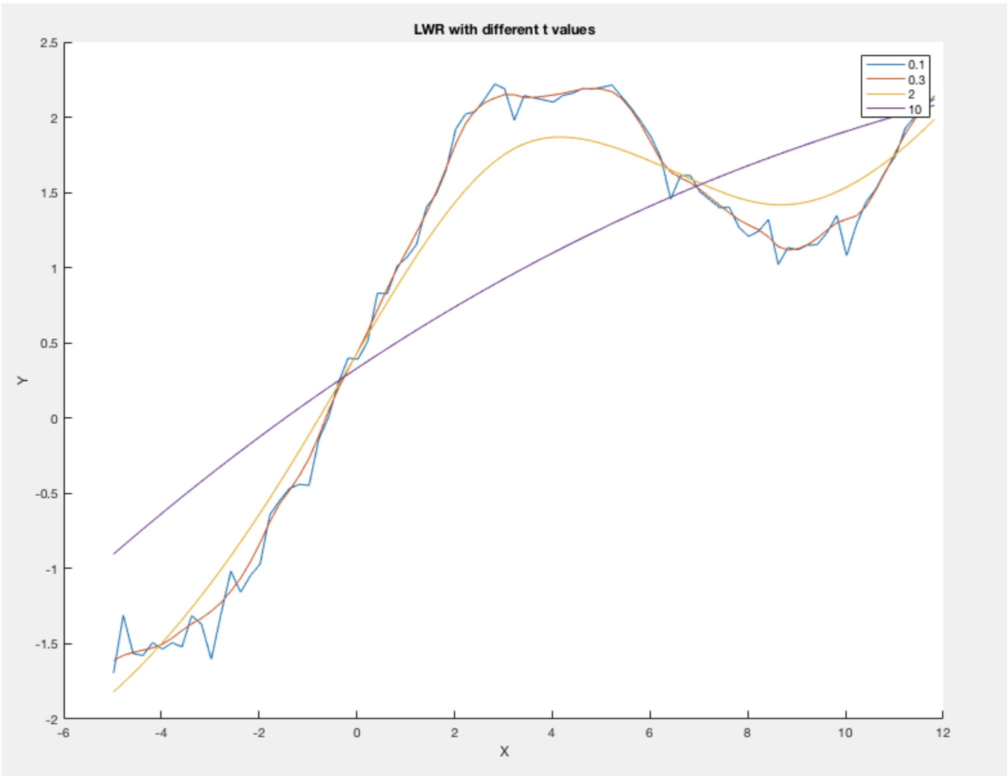
$w_{i,i} = e^{-((x-xi)^\wedge 2)/2*(\tau 2)}$

Finally plot with set of pair points $(x\ ,\ \theta_1+\theta_2 x)$ with $\boldsymbol{\tau} = 0.8$

Graph:-



2. (c) Graph:-



<u>Observation and conclusion</u>:-

- $\tau = 2$ works best to fit.
- For smaller value of τ,predicted values is almost same as measured value ===> overfitting
- For larger value of τ,predicted value would be closer to the points on curve (obtained by unweighted case) ===> underfitting

3.(a) Updating θ by using rule for Newton's Method:-

$$\theta := \theta - H^{-1}\nabla_\theta J(\theta) \quad \text{until convergence.}$$

where $\nabla_\theta J(\theta) = (1/m)[(y^{(i)} - h_\theta(x^{(i)}))^T * \theta]$ and

$H = (1/m)\Sigma[\, h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))x^{(i)}(x^{(i)})^T\,]$ $(\Sigma \longrightarrow i=1:m)$

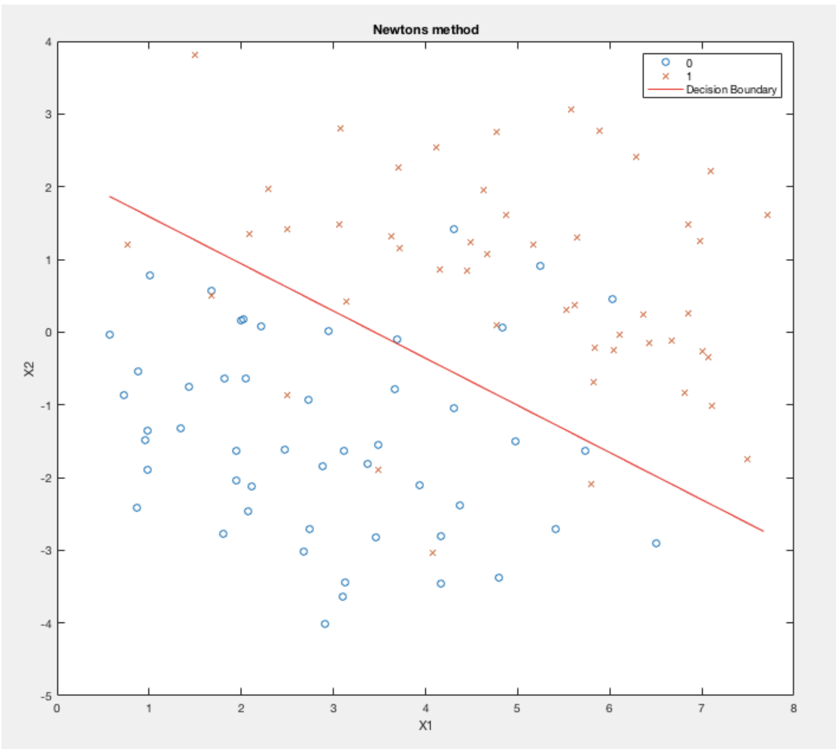
<u>Stopping criteria</u> :-

$$|J(\theta_{t+1}) - J(\theta_t)| < \epsilon \qquad , \epsilon = 0.00001$$

<u>Final θ</u> = [-2.6205 ; 0.7604 ; 1.1719]

$$H = \begin{matrix} -0.1001 & -0.3882 & 0.0259 \\ -0.3882 & -1.8401 & 0.1742 \\ 0.0259 & 0.1742 & -0.1850 \end{matrix} \quad , \qquad \#iterations=7$$

3.(b) Plotting Decision Boundary by using $\theta^T x=0$;



4.(a) $\mu_1 = (\Sigma\ 1\{y^{(i)}=1\}x^{(i)})\ /\ (\Sigma\ 1\{y^{(i)}=1\})$

$\mu_0 = (\Sigma\ 1\{y^{(i)}=0\}x^{(i)})\ /\ (\Sigma\ 1\{y^{(i)}=0\})$

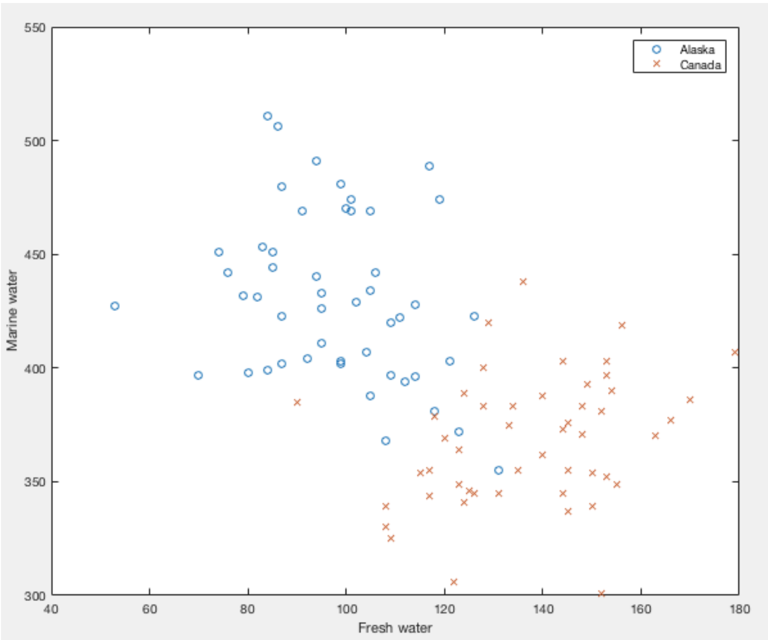
$\boldsymbol{\Sigma} = (1/m) * (\Sigma\ (x^{(i)} - \mu_{yi}) * (x^{(i)} - \mu_{yi})^T)$

Ans:-

$\mu_1 = [137.4600\ ;\ 366.6200]$, $\mu_0 = [98.3800\ ;\ 429.6600]$,

$$\boldsymbol{\Sigma} = \begin{matrix} 287.5 & -26.7 \\ -26.7 & 112.33 \end{matrix}$$

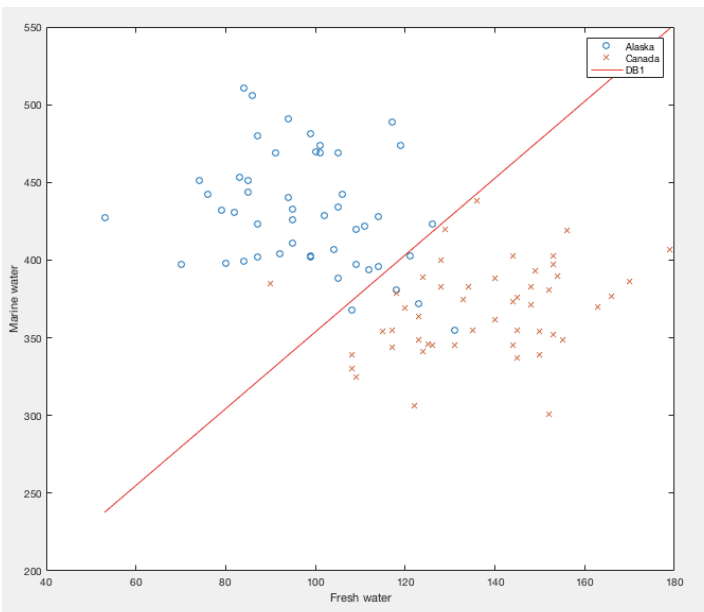
4.(b) Graph (training datas) :-



4.(c) Linear Decision Boundary equation:-

$$\log((1-\Phi)/\Phi) + 0.5[(x^{(i)} - \mu_1)^T \boldsymbol{\Sigma}^{-1} (x^{(i)} - \mu_1) - (x^{(i)} - \mu_0)^T \boldsymbol{\Sigma}^{-1} (x^{(i)} - \mu_0)] = 0;$$

Graph:-



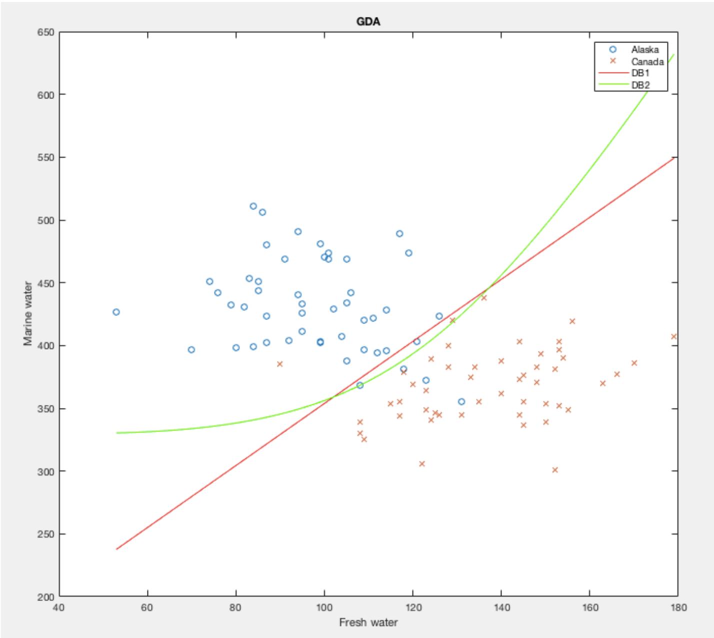
4.(d) μ_1 = [137.4600 ; 366.6200] , μ_0 = [98.3800 ; 429.6600] ,

$$\boldsymbol{\Sigma}_1 = \begin{matrix} 319.5684 & 130.8348 \\ 130.8348 & 875.3956 \end{matrix} \quad , \quad \boldsymbol{\Sigma}_0 = \begin{matrix} 255.4 & -184.3 \\ -184.3 & 137.11 \end{matrix}$$

4.(e) Quadratic Decision Boundary equation:-

$$\log\left((1-\Phi)/\Phi \right) \left(|\, \boldsymbol{\Sigma}_1^{\;1/2} | \,/\, |\, \boldsymbol{\Sigma}_0^{\;1/2}) \right) + 0.5[(x^{(i)} - \mu_1)^T \boldsymbol{\Sigma}_1^{-1} (x^{(i)} - \mu_1)-(x^{(i)} - \mu_0)^T \boldsymbol{\Sigma}_0^{-1} (x^{(i)} - \mu_0)] = 0$$

Graph:-



4.(f) Two real circle Alaska Points(between two Decision Boundary lines) are in Alaska classified by Quadratic Boundary.But Linear Boundary classifies these two points as Canada. Thus, Quadratic boundary better classifies the data as compared to linear boundary classification due to quadratic boundary computing the difference in covariance matrix of the two classes and hence providing more accurate estimate.