MACHINE LEARNING
COL-774

Assignment 1
Report file

Submitted by:
Anshu Bansal
2018MCS2142

Question. 1 (Linear Regression)

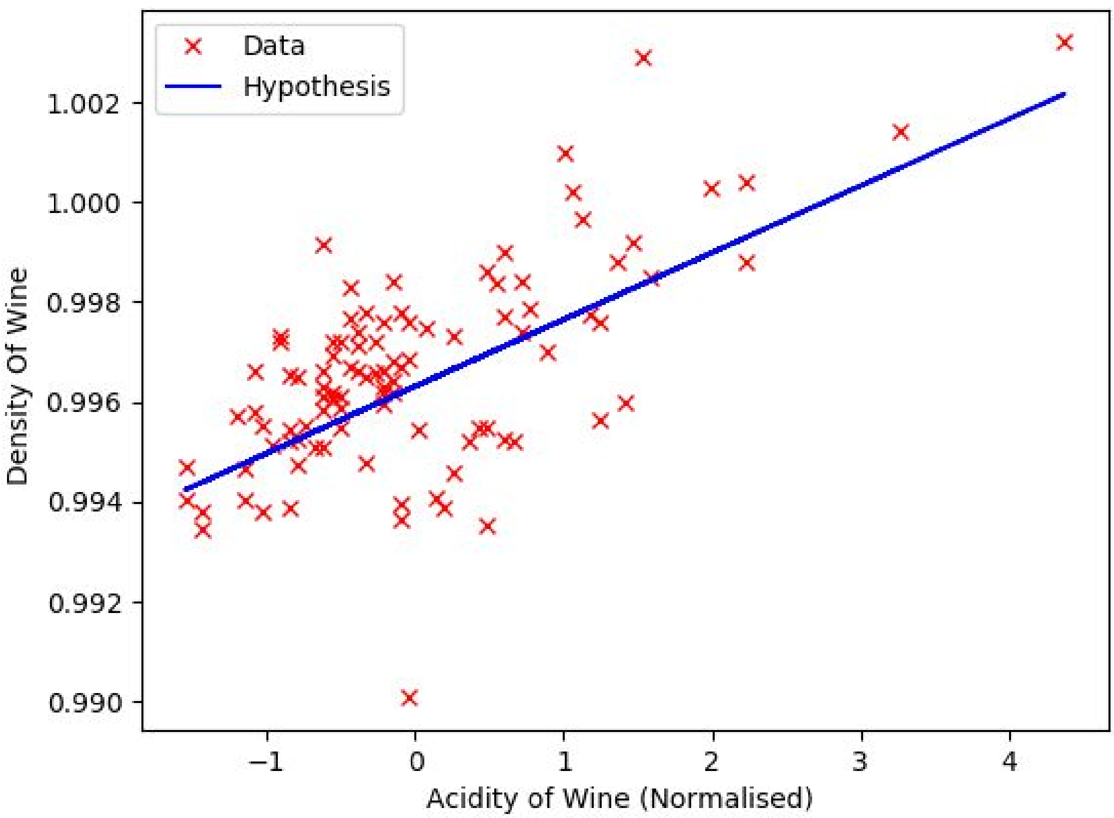
(a) Learning rate chosen = 0.001

theta Parameters Obtained = [0.9963043 0.00133977] ([θ_0 θ_1])

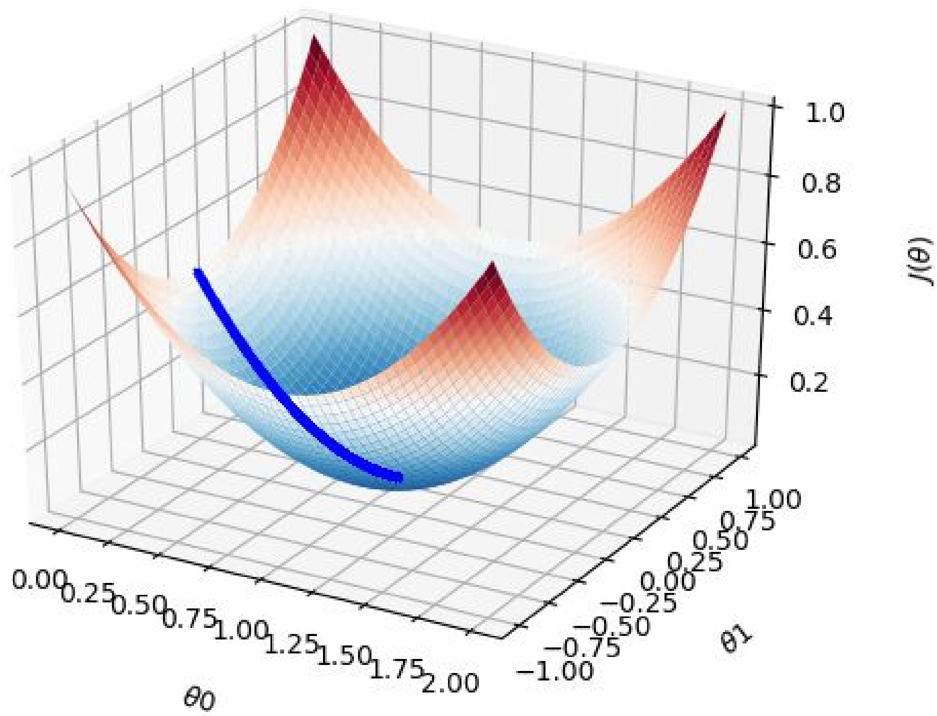
Stopping criteria for convergence : abs(prev error - new error) < 10 ** (-10)

Note : stopping the convergence process after taking 10000 iterations.

(b)

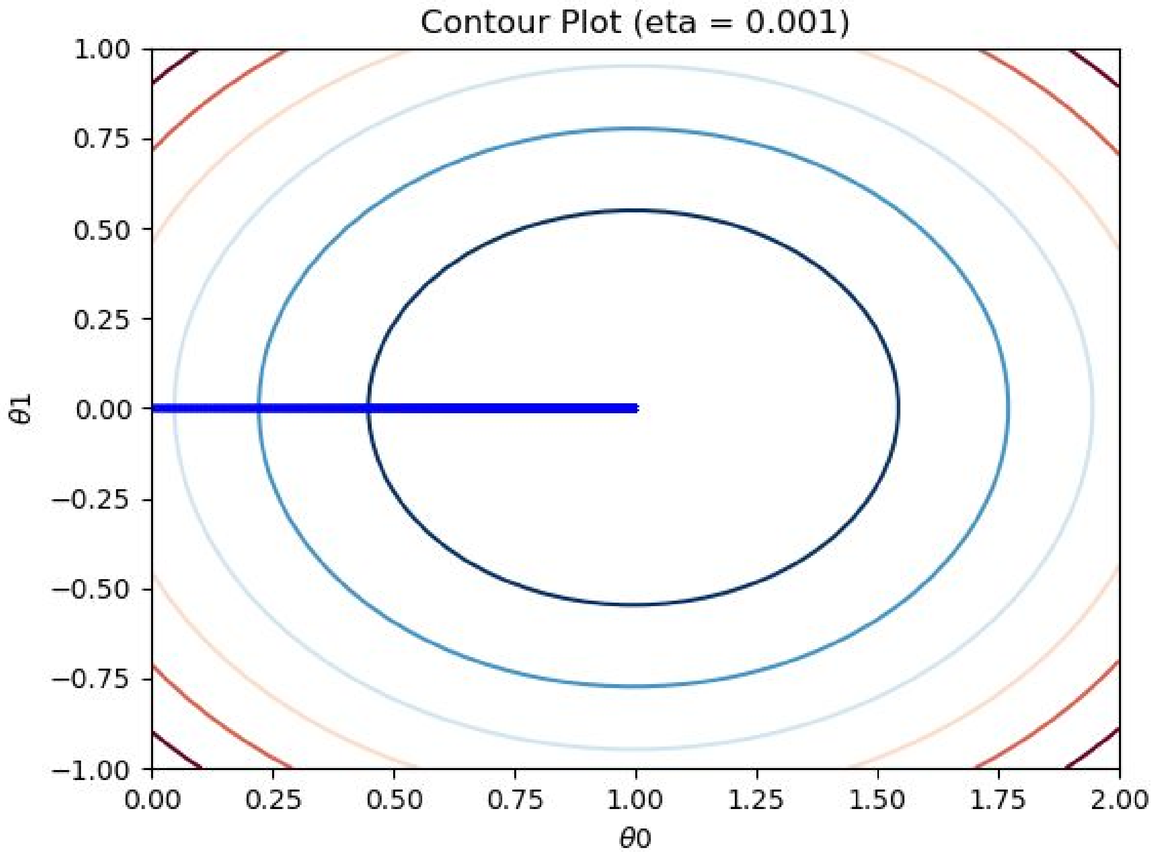


(c)

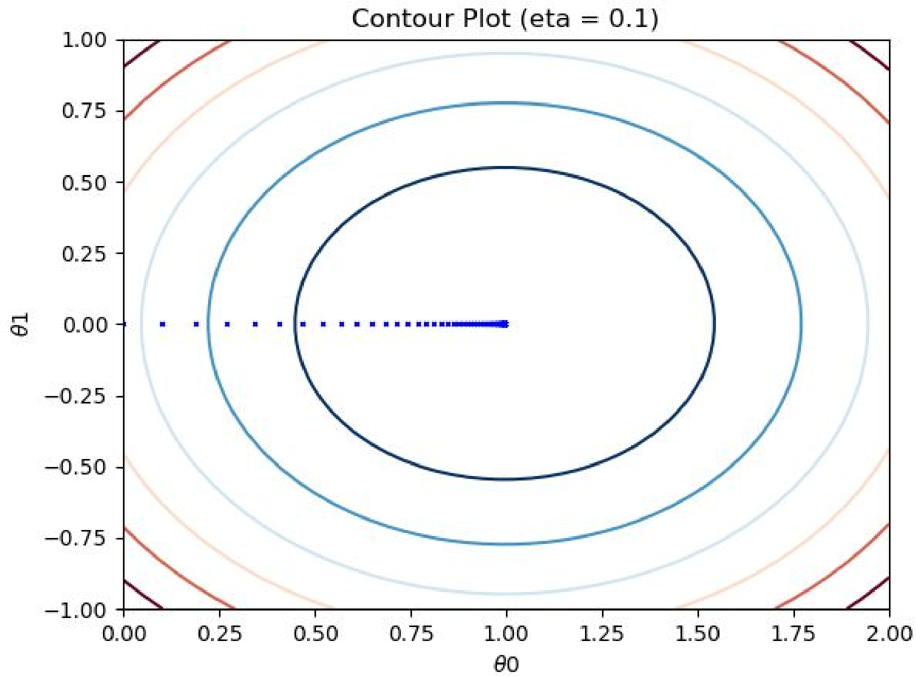


1.0
0.8
0.6
0.4
0.2
$J(\theta)$
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
$\theta 0$
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
$\theta 1$

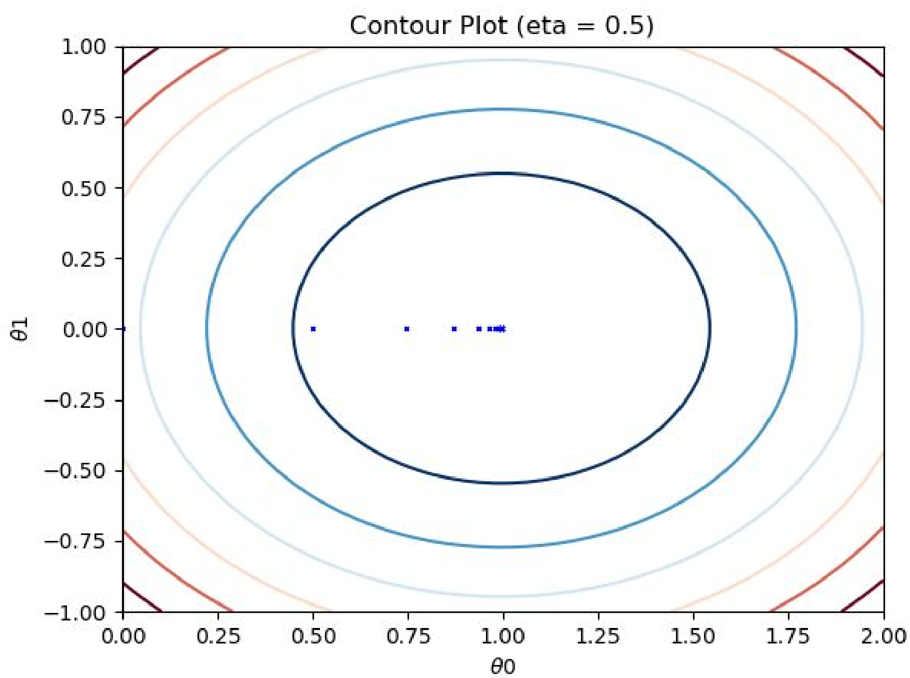
(d)


Contour Plot (eta = 0.001)
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
θ1
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
θ0

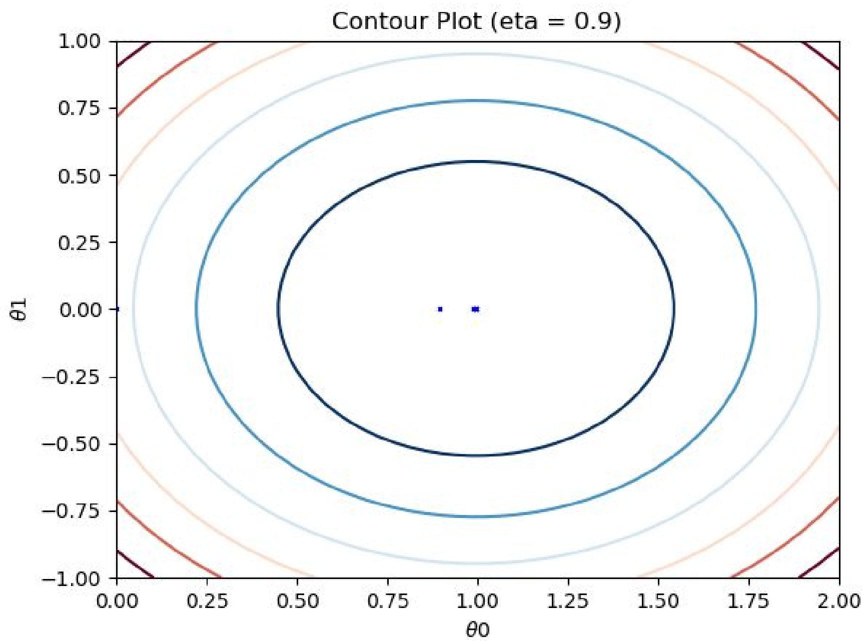
(e)

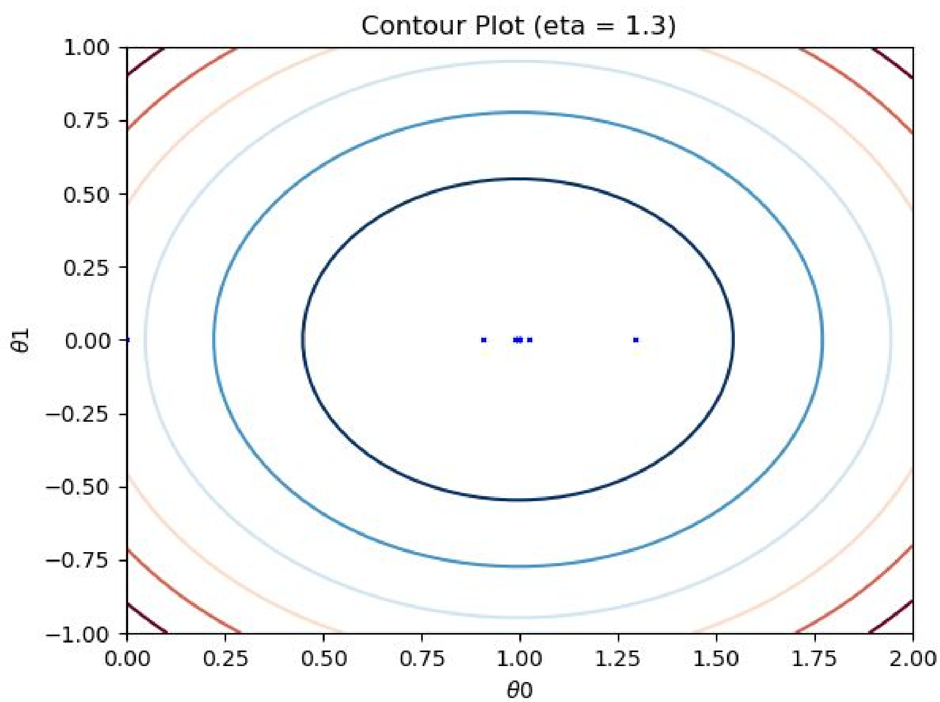


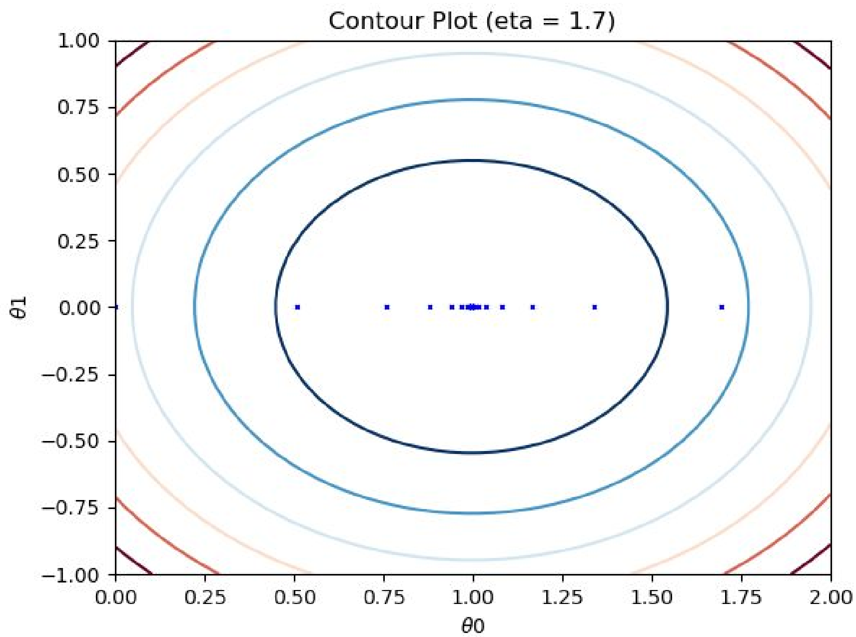
Contour Plot (eta = 0.1)
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
θ1
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
θ0



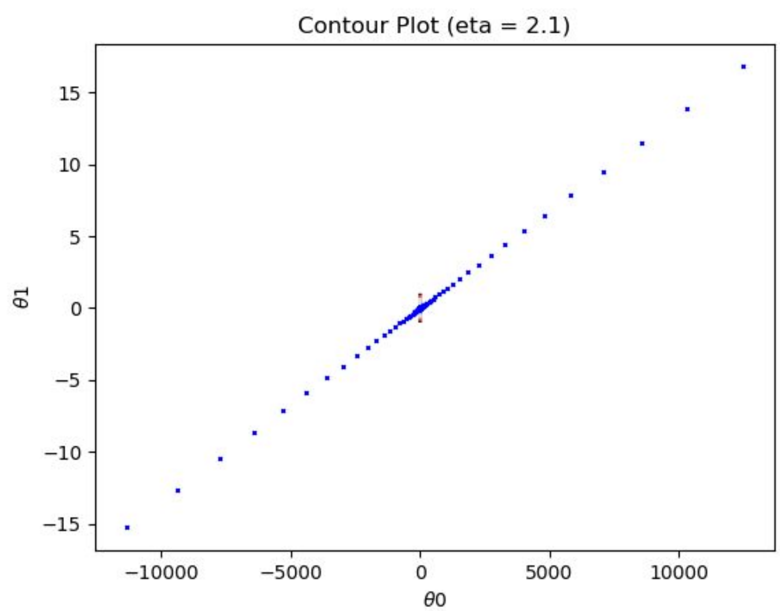
Contour Plot (eta = 0.5)
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
θ1
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
θ0


Contour Plot (eta = 0.9)
θ1
θ0
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

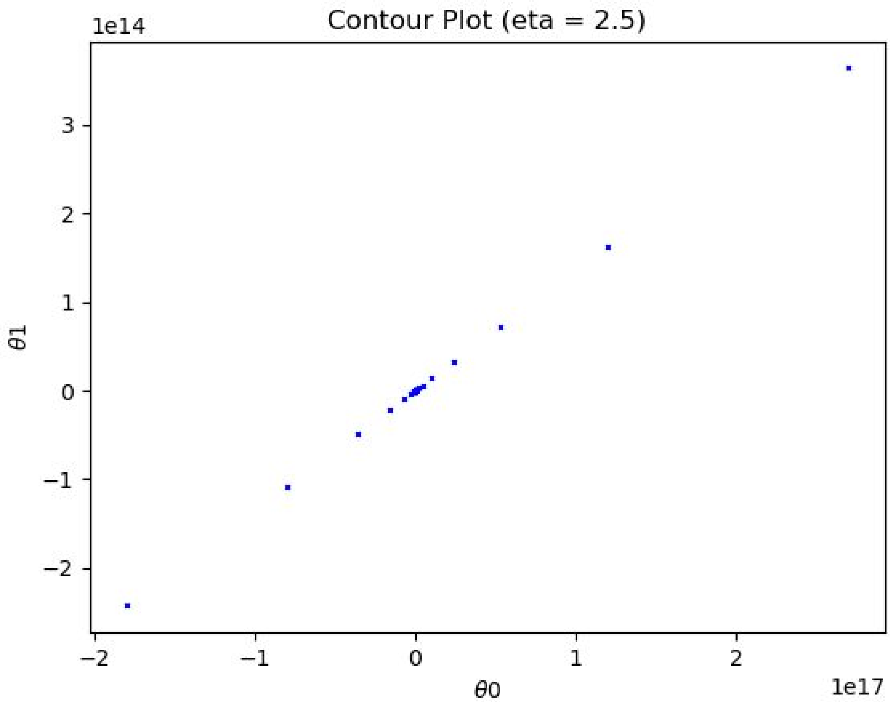

Contour Plot (eta = 1.3)
θ1
θ0
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00



Contour Plot (eta = 1.7)
θ1
θ0



Contour Plot (eta = 2.1)
θ1
θ0



Observations :

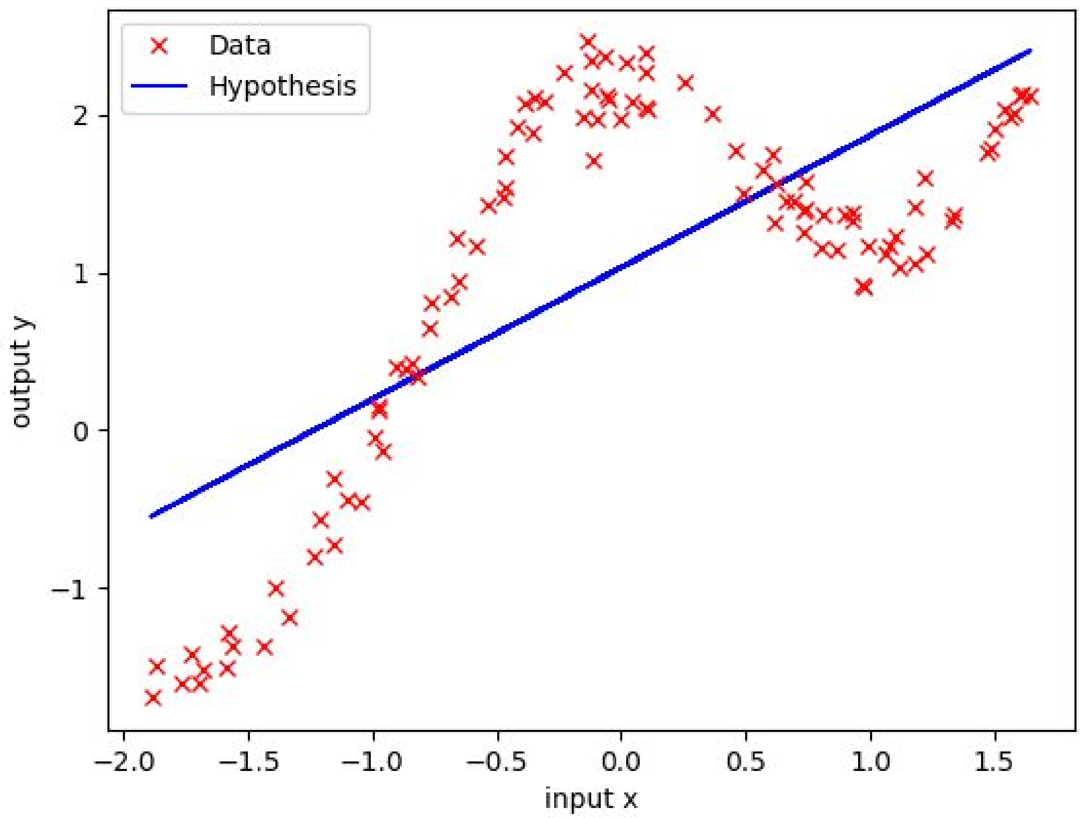
When value of learning rate (η) is very small as in 0.03 to 0.1 , it is taking too many iteration for converging to local minima because gradient takes very small steps but eventually it does converge to the local minima after some large number of iterations.

As η increases to 0.9 or 1.7 , it is taking less iterations to converge as gradient takes a bigger step than the previous one. So it gives us the local minima at much faster pace.

As η increases to 2.1 or 2.3 , it is taking very large steps and will start oscillating resulting the contours to be hidden. The function will never converge to minima and keep oscillating around the minima.

Question. 2 (Locally Weighted Linear Regression)

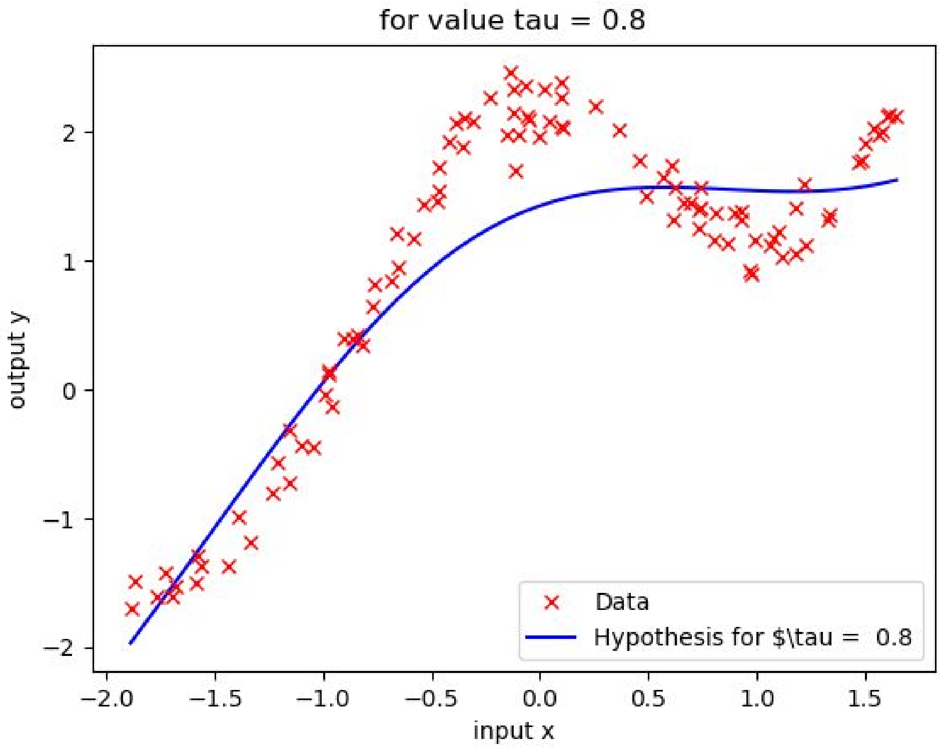
(a)



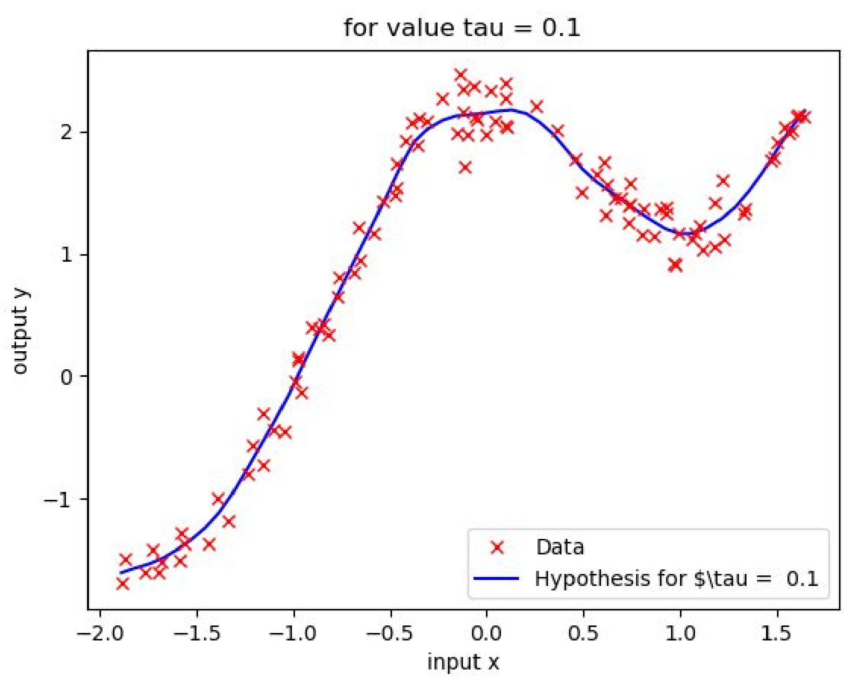
Linear Regression plotting without considering weights using normal equations.

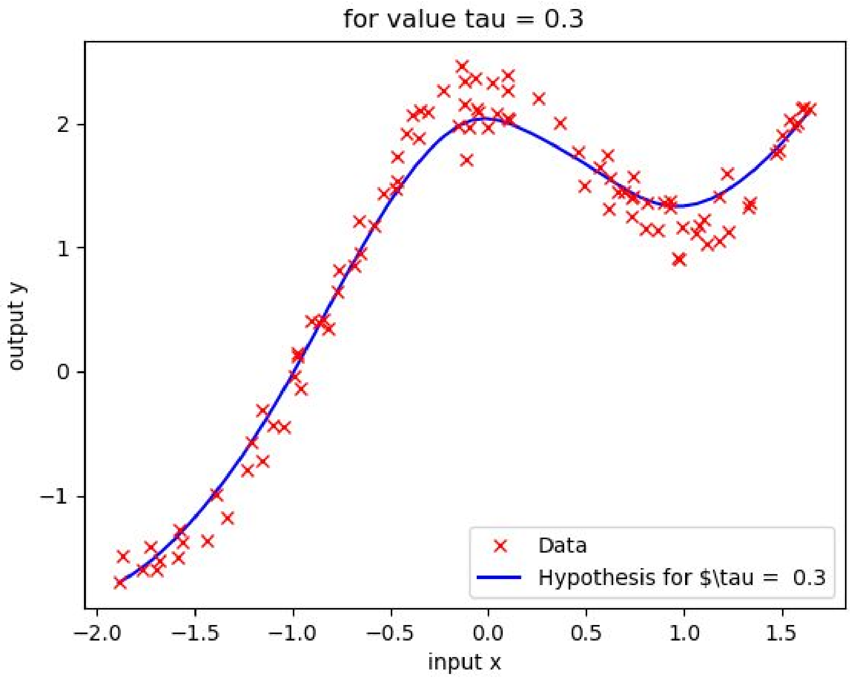
Since we can see that the linear regression doesn't fit well for the given training data using linear regression, we will use the locally weighted linear regression for this dataset.

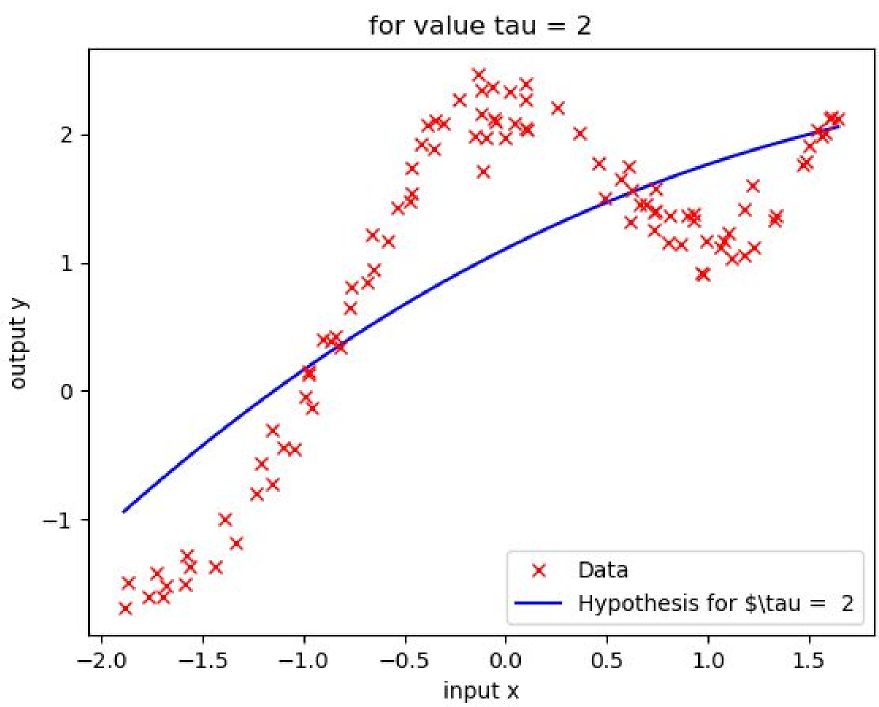
(b)

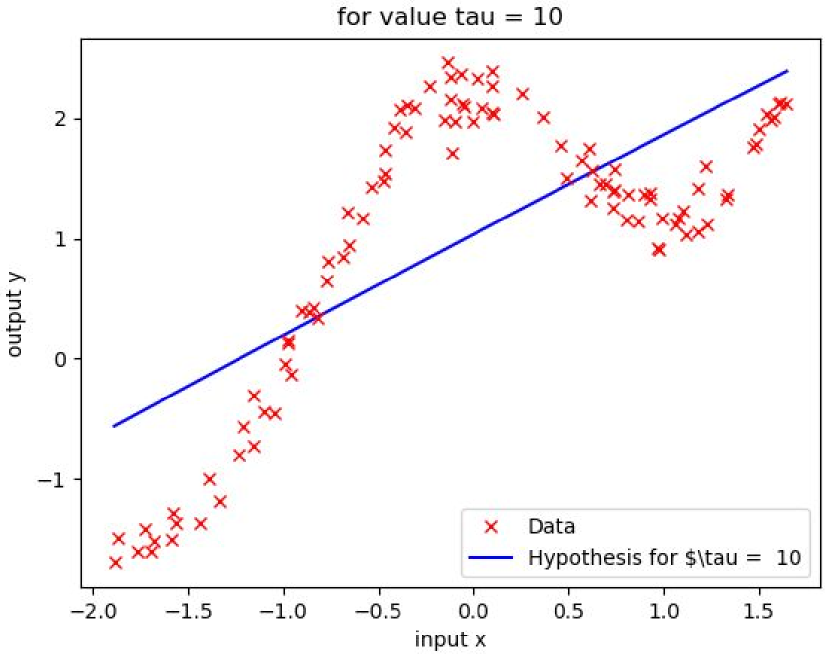

for value tau = 0.8
output y
input x
Data
Hypothesis for $\tau = 0.8
2
1
0
−1
−2
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

(c)


for value tau = 0.1
output y
input x
Data
Hypothesis for $\tau = 0.1
2
1
0
−1
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5


for value tau = 0.3
Data
Hypothesis for $\tau = 0.3
output y
input x
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2
1
0
−1


for value tau = 2
Data
Hypothesis for $\tau = 2
output y
input x
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2
1
0
−1



Observations:

The best fit is obtained when τ = 0.3

When τ is very small then the algorithm is trying to fit each and every data as with value 0.1. This will lead to overfitting.

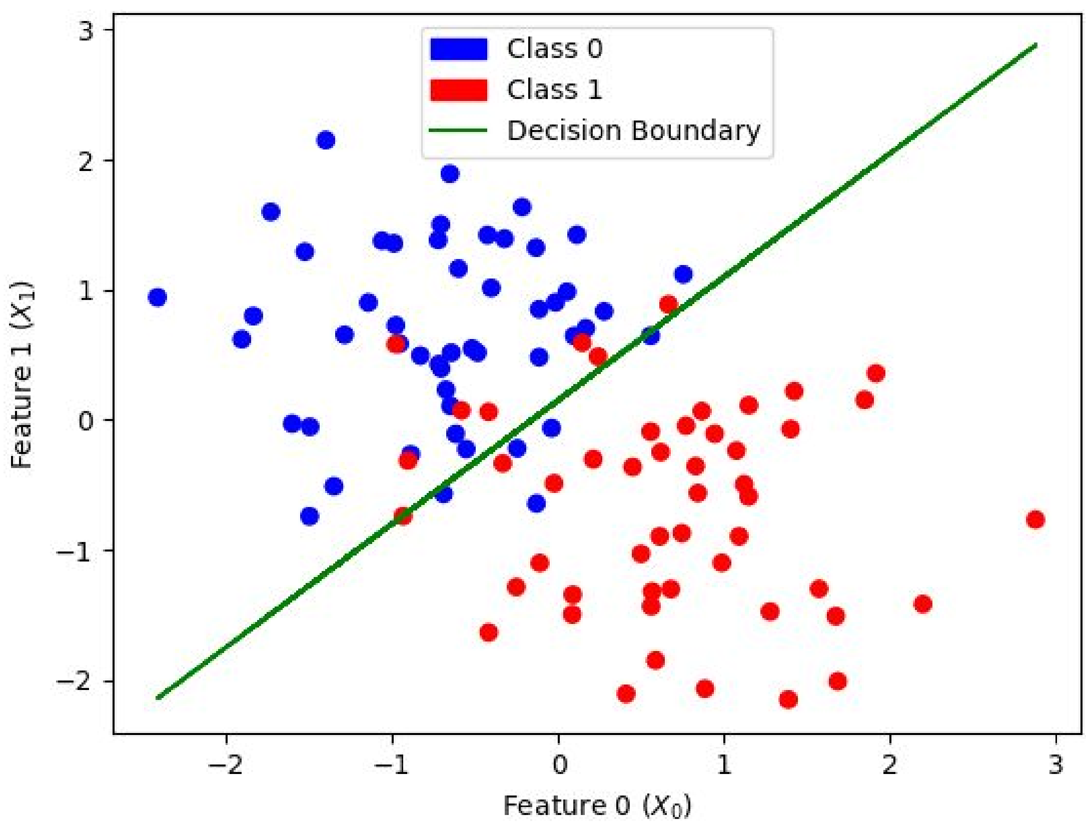
When τ is too large as 0.8, or 2 or 10 , algorithm is not doing well and since the training set accuracy is very poor, which will result in underfitting. (even with value 10, it almost similar to the unweighted linear regression). Which means that when τreaches to infinity , all neighbour points get weight close to 1 and the algo will be reduced to the unweighted linear regression.

Question. 3 (Logistic Regression)

(a) Theta parameters obtained :

[0.40125316 2.5885477 -2.72558849] ([θ_0 θ_1 θ_2])

(b)



Decision Boundary by Logistic regression

Question. 4 (Gaussian Discriminant Analysis)

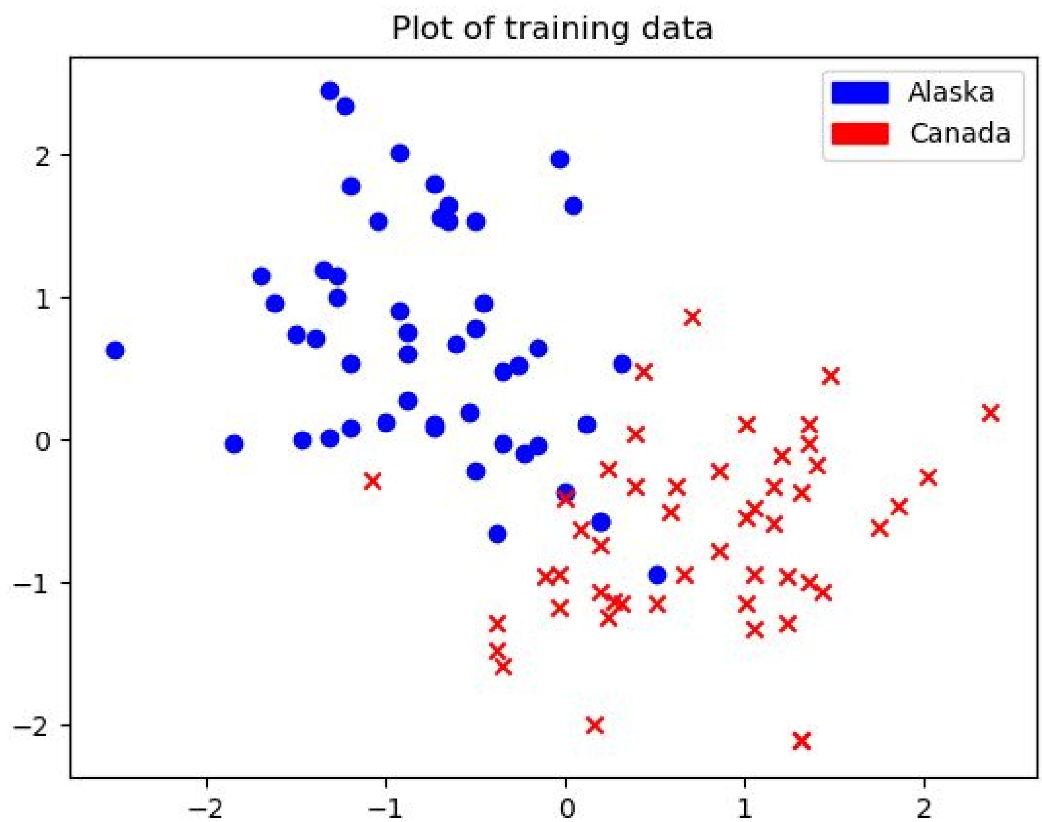
(a) μ_0= [-0.75529433 0.68509431]

μ_1 = [0.75529433 -0.68509431]

$\sum$= [[0.42953048 -0.02247228]

[-0.02247228 0.53064579]]

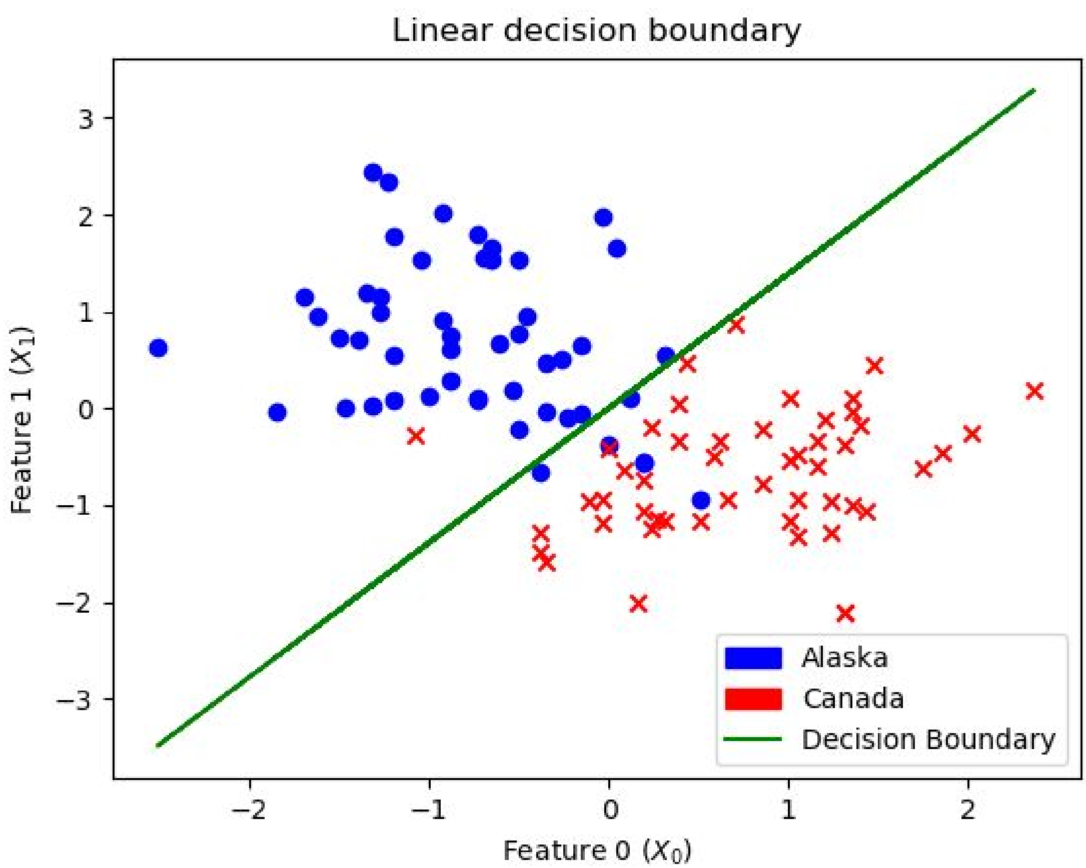
(b)



(c) Equation of Line for Linear Gaussian Discriminant analysis: (AX = B) [reference from http://www-scf.usc.edu/~csci567/05-LDA.pdf --> page 19] Where A = 2 * ($\mu_0^T\Sigma^{-1}$- $\mu_1^T\Sigma^{-1}$) and

$B = (\mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1 - 2 * \log(P(y=1) / P(y=0))$

I used this equation of line and for every input feature X1, I calculated X2 using this equation. It resulted in a line which is as shown below.



(d)

Parameters obtained using Quadratic Gaussian Discriminant Analysis :

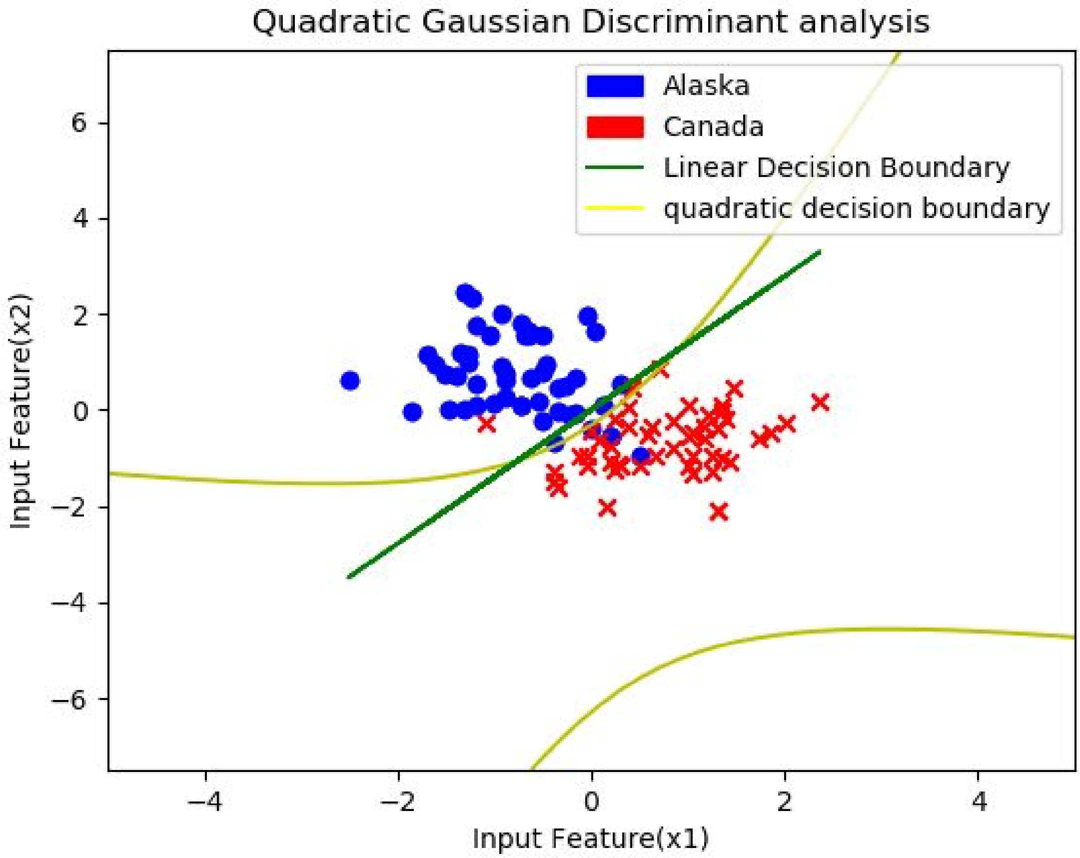
μ_0 = [-0.75529433 0.68509431]

μ_1 = [0.75529433 -0.68509431]

Σ_0 = [[0.38158978 -0.15486516]
[-0.15486516 0.64773717]]
Σ_1 = [[0.47747117 0.1099206]
[0.1099206   0.41355441]]

(e)
Equation for quadratic gaussian discriminant analysis :
(When Σ_0 != Σ_1)
$X^T AX + BX + C = 0$
Where $A = \Sigma_0^{-1} - \Sigma_1^{-1}$
$B = - 2 * (\mu_o^T \Sigma_0^{-1} - \mu_1^T \Sigma_1^{-1})$
And
$C = \mu_o^T \Sigma_0^{-1} \mu_o - \mu_1^T \Sigma_1^{-1} \mu_1 - 2 * \log((P(y=1) / P(y=0)) * (|\Sigma 0| / |\Sigma 1|))$

These are the coefficients of the quadratic equation.



(f) The quadratic boundary is better seperator as it classifies the salmons more correctly unlike linear boundary. The quadratic boundary is hyperbolic curve as shown in above figure.