Machine Learning

Assignment 1

Shadab Zafar
2017MCS2076

Q1: Linear Regression

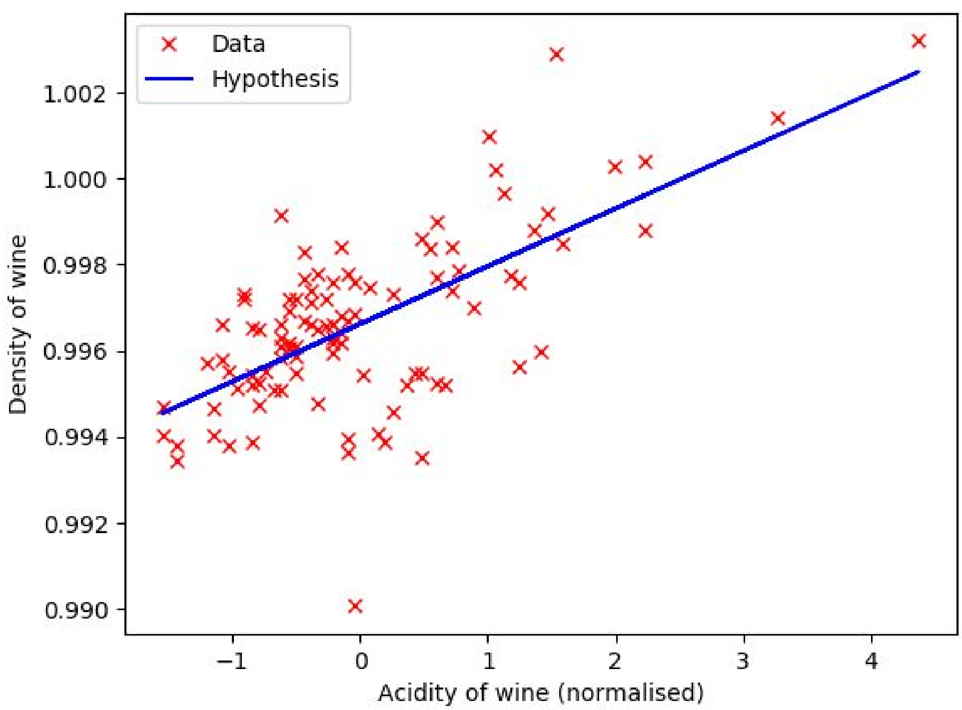
a.)

Batch Gradient Descent.

Learning Rate = 0.001
Stopping Criteria = abs(J_old - J_new) < 10^{-15}
Final Parameters = [0.99662009 0.0013402]
Number of iterations taken = 176

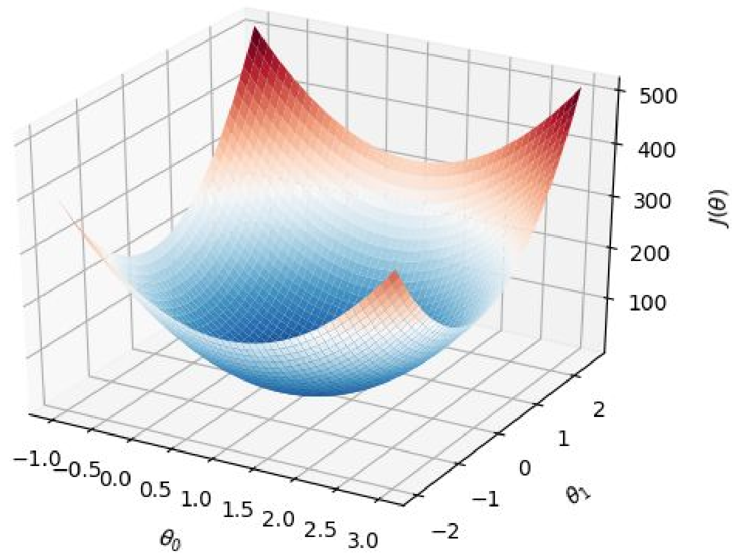
b.)

Plot of the hypothesis function learned:

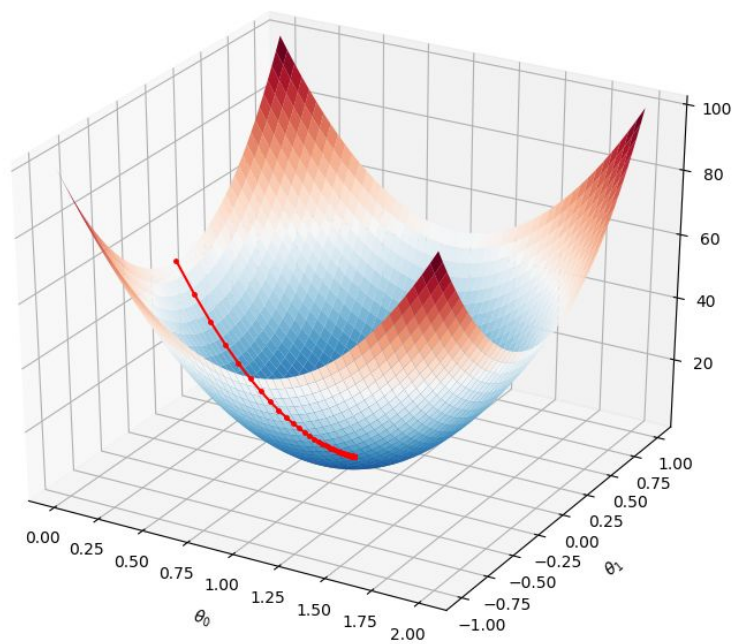


c.)

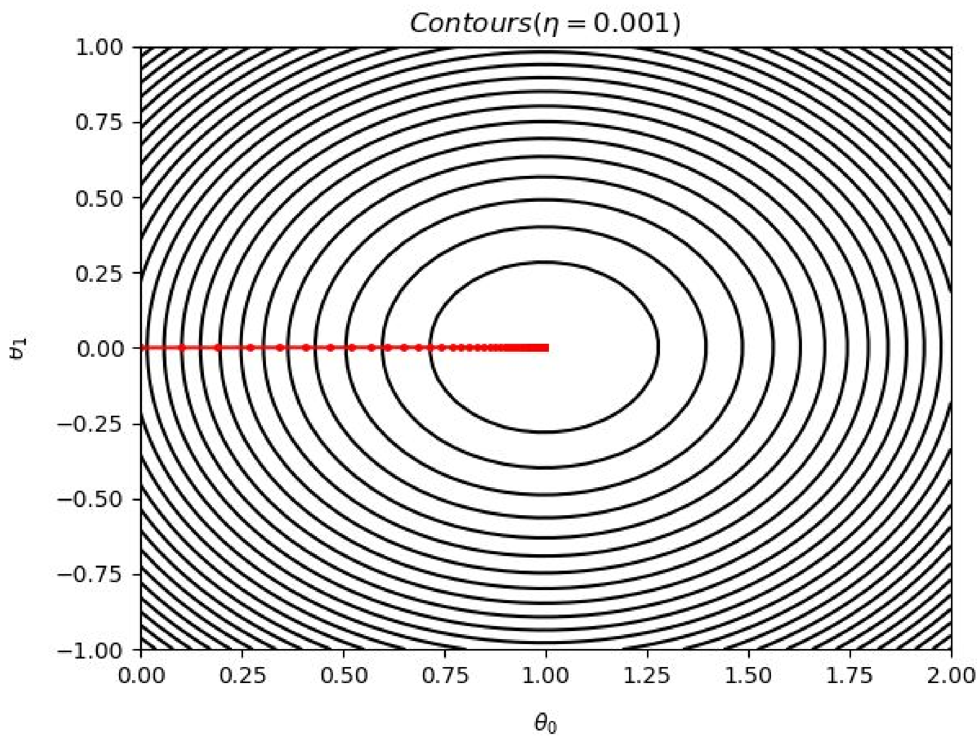
3D Plot of Error Function (J(θ))



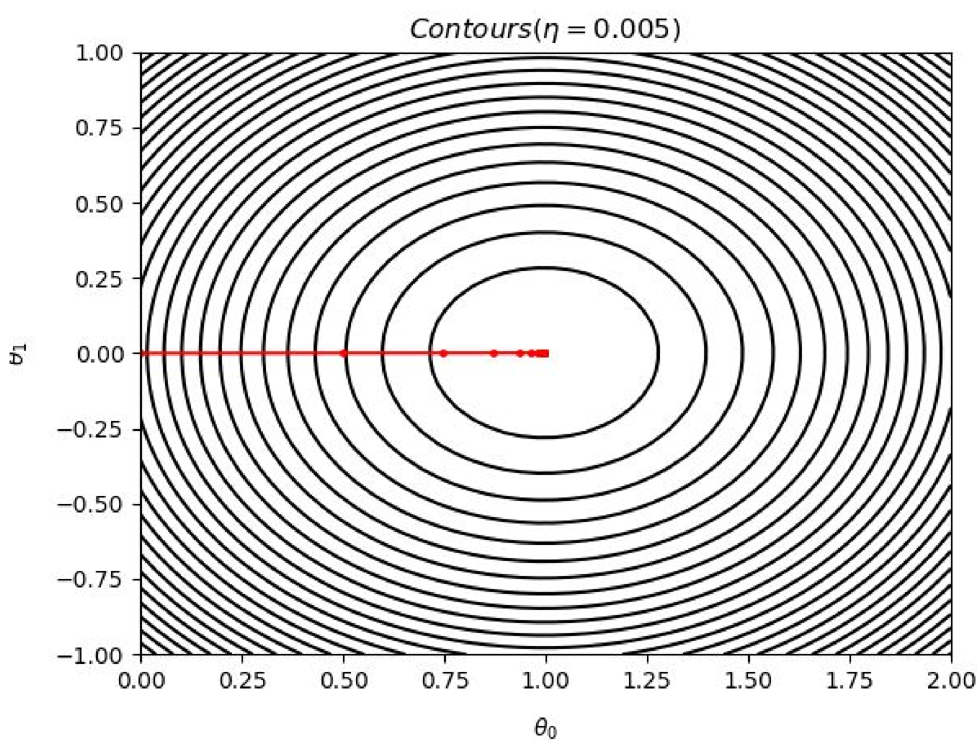
3D Plot with error values traced out:



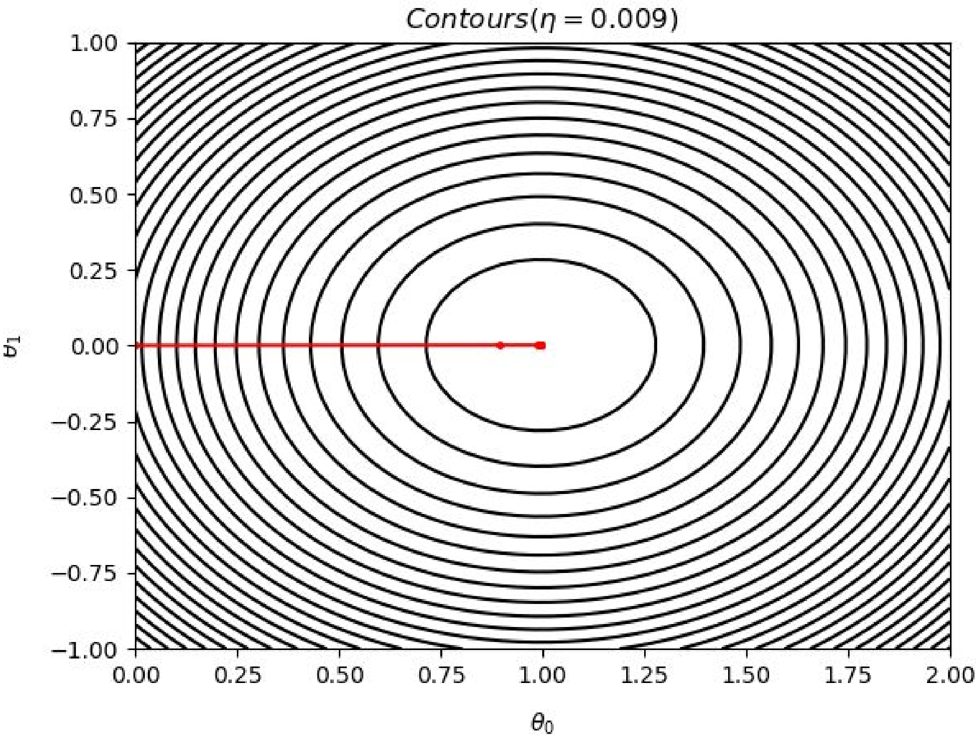
e.)



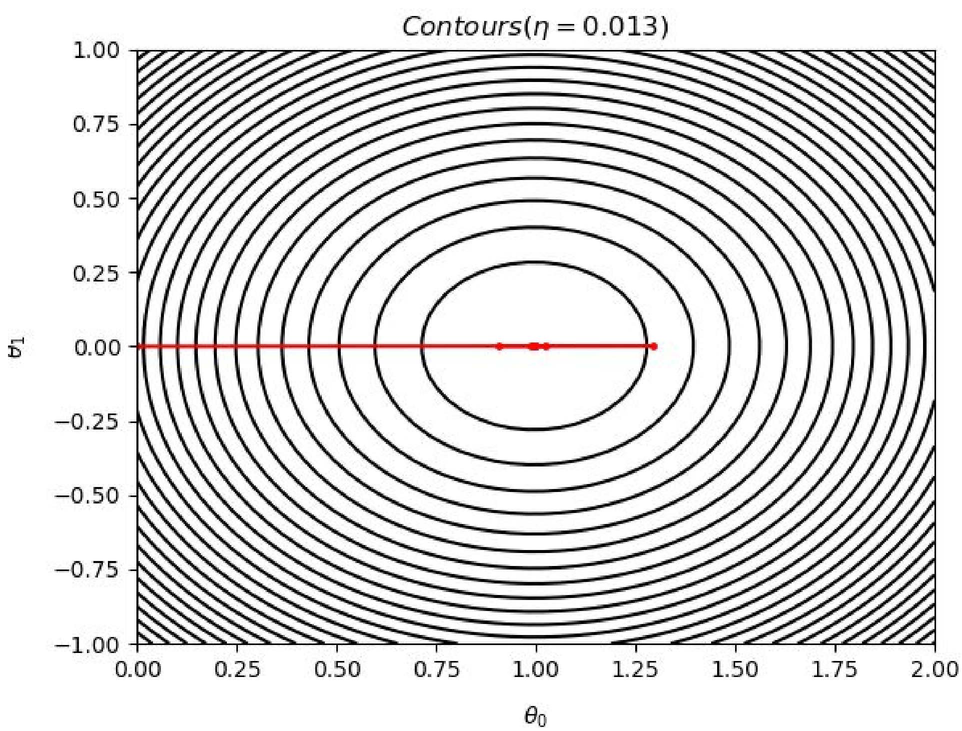
$Contours(\eta = 0.001)$
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
θ_1
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
θ_0



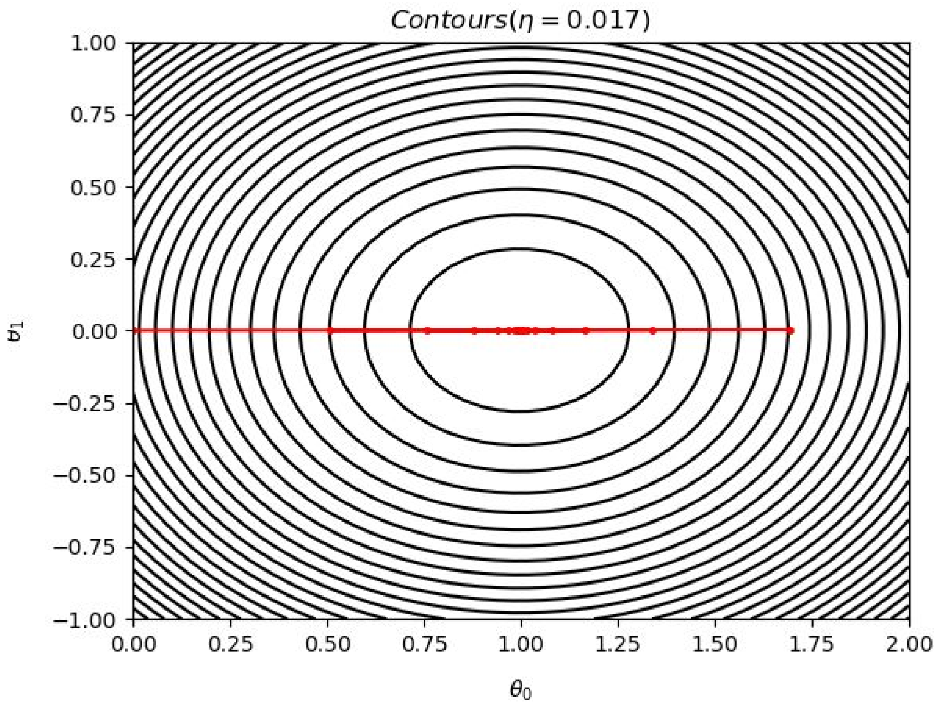
$Contours(\eta = 0.005)$
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
θ_1
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
θ_0

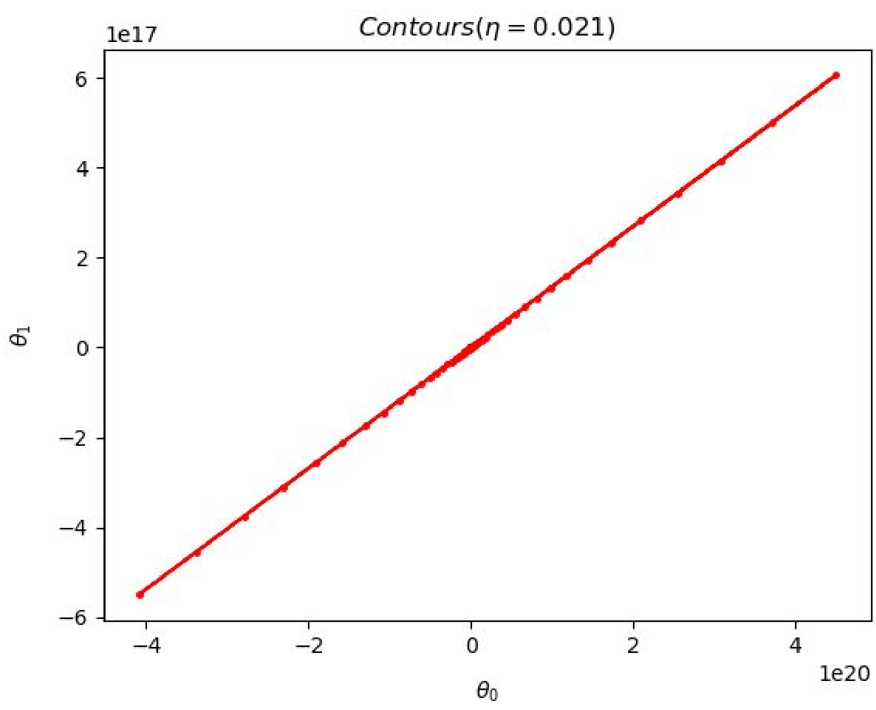


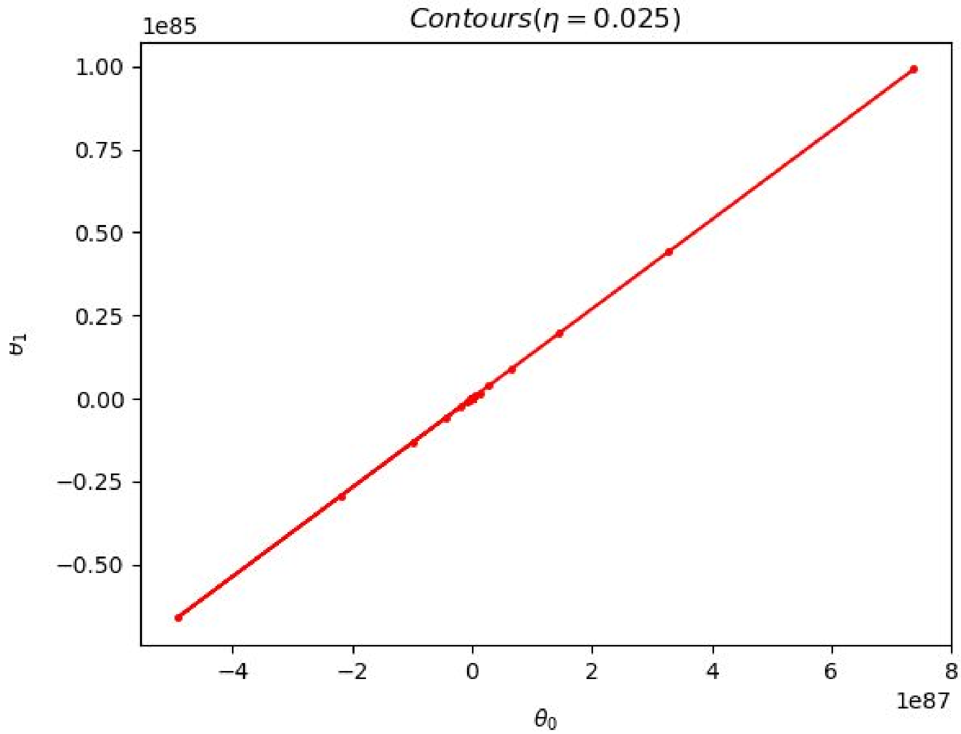
$Contours(\eta = 0.009)$
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
θ_1
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
θ_0



$Contours(\eta = 0.013)$
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
θ_1
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
θ_0


$Contours(\eta = 0.017)$
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
θ_0
θ_1


$Contours(\eta = 0.021)$
1e17
6
4
2
0
−2
−4
−6
−4
−2
0
2
4
1e20
θ_0
θ_1



As eta increases the number of iterations required by the algorithm decrease so it converges faster (which is shown by lesser number of points on the graph - around eta = 0.09) this seems to be the best of the given values.

But then, when eta is increased further the error function oscillates around the minima (can be observed for eta 0.013 & 0.017)

Further increasing eta values results in divergence so the contours hide from view.

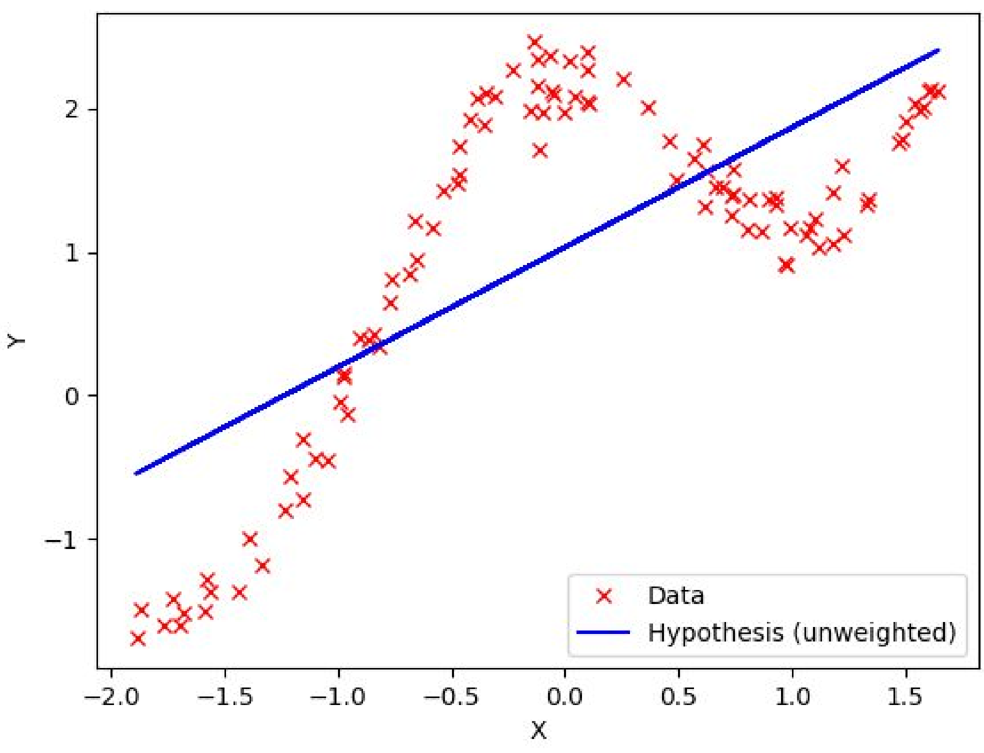
Q2: Locally Weighted Linear Regression

a.)

Unweighted Linear Regression using Normal Equations

Final Parameters = [1.03128116, 0.83519315]

Plot of the Hypothesis function learnt:

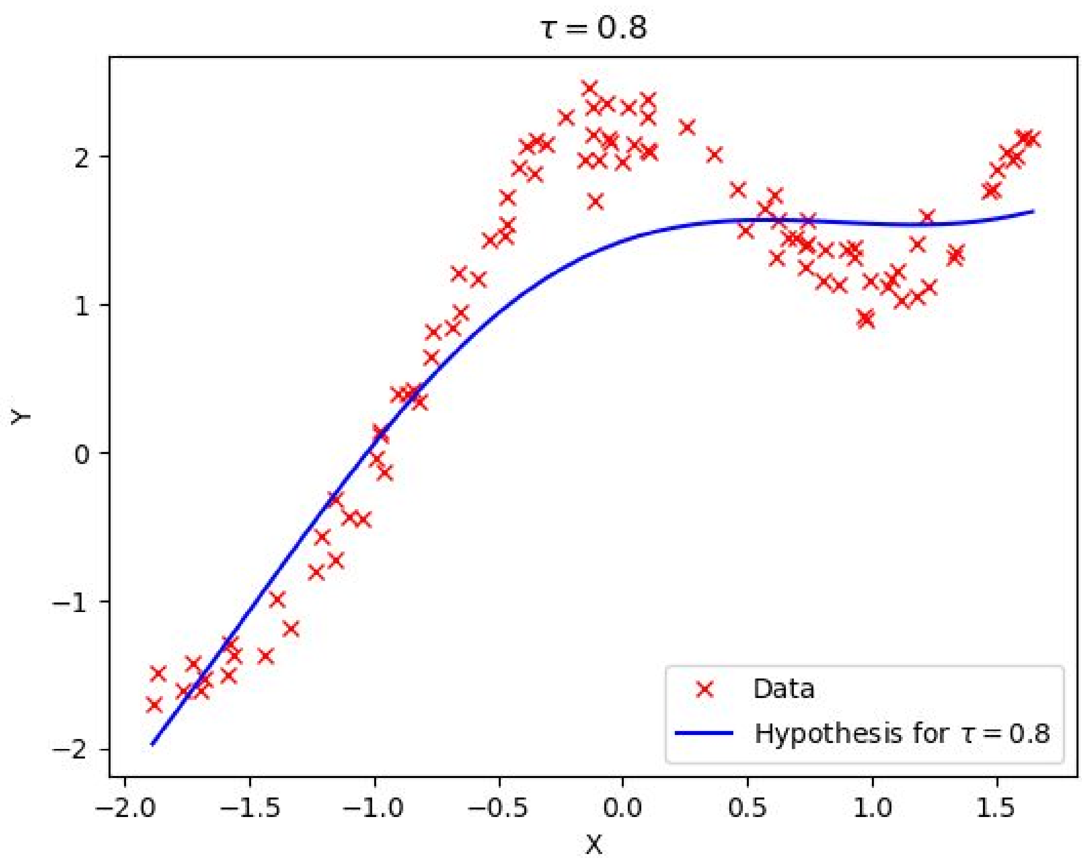


As can be observed, a simple unweighted linear regression is not able to the fit the data properly, because the data is inherently non linear.

b.)

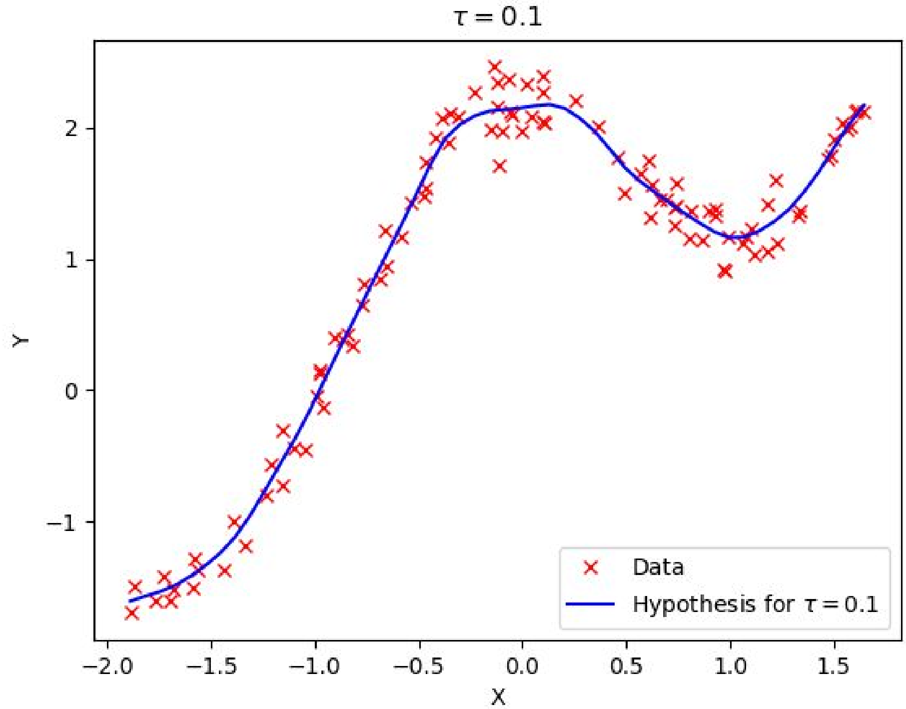
Weighted Linear Regression using Normal Equations

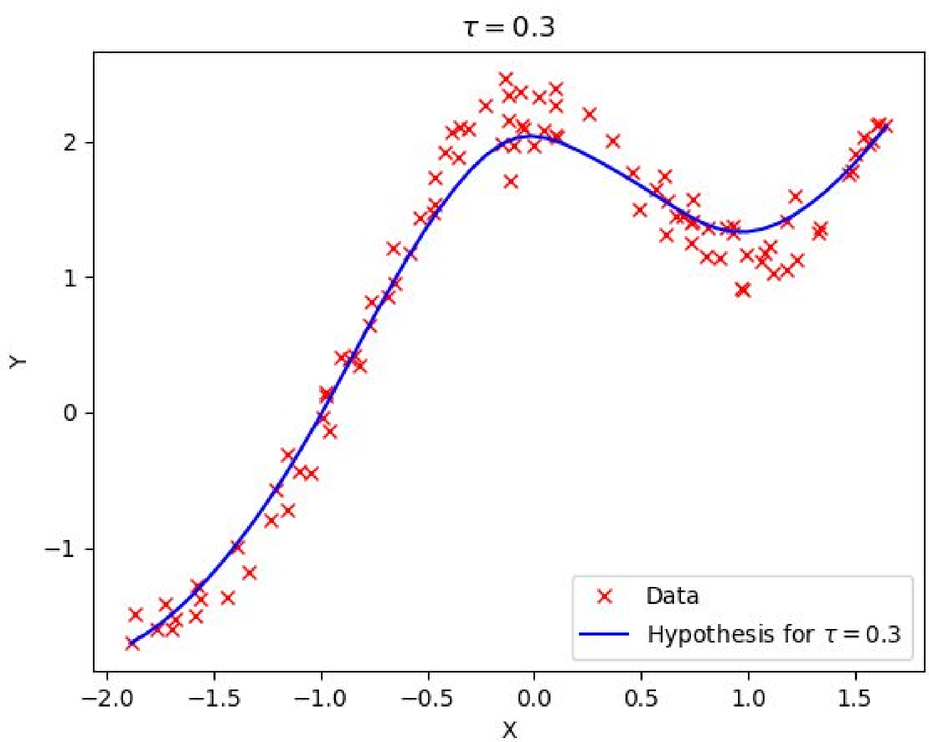
Plot of the Hypothesis function learnt when Tau = 0.8:

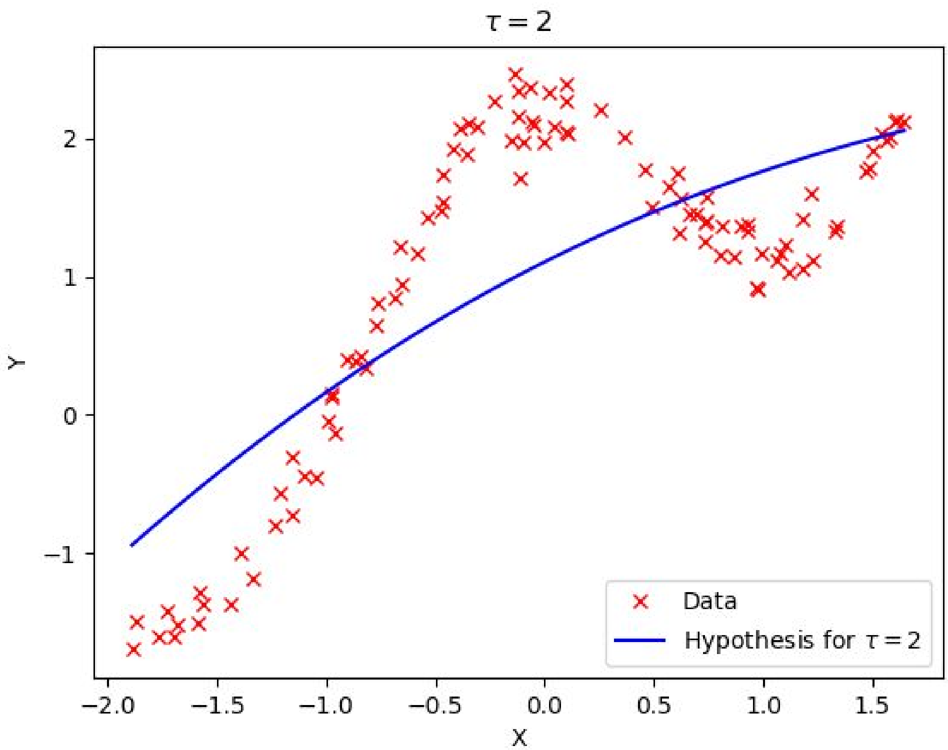


c.)

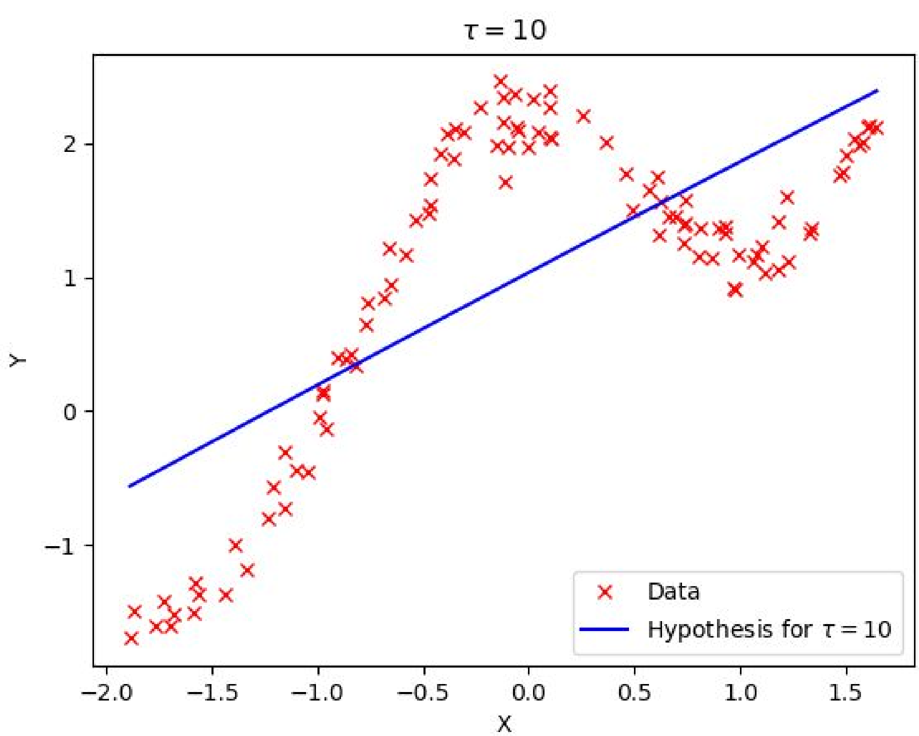
Effect of values of Tau







$\tau = 2$
Data
Hypothesis for $\tau = 2$
Y
X



$\tau = 10$
Data
Hypothesis for $\tau = 10$
Y
X

As can be observed from the plots the value of Tau = 0.3 works the best.

As Tau gets larger, the curve becomes a straight line as all weights get close to 1 and the process reduces to that of the unweighted regression.

As Tau gets smaller, the curve tries to pass through every data point. Since this is overfitting the data, it won’t work well on new test data.

Q3: Logistic Regression

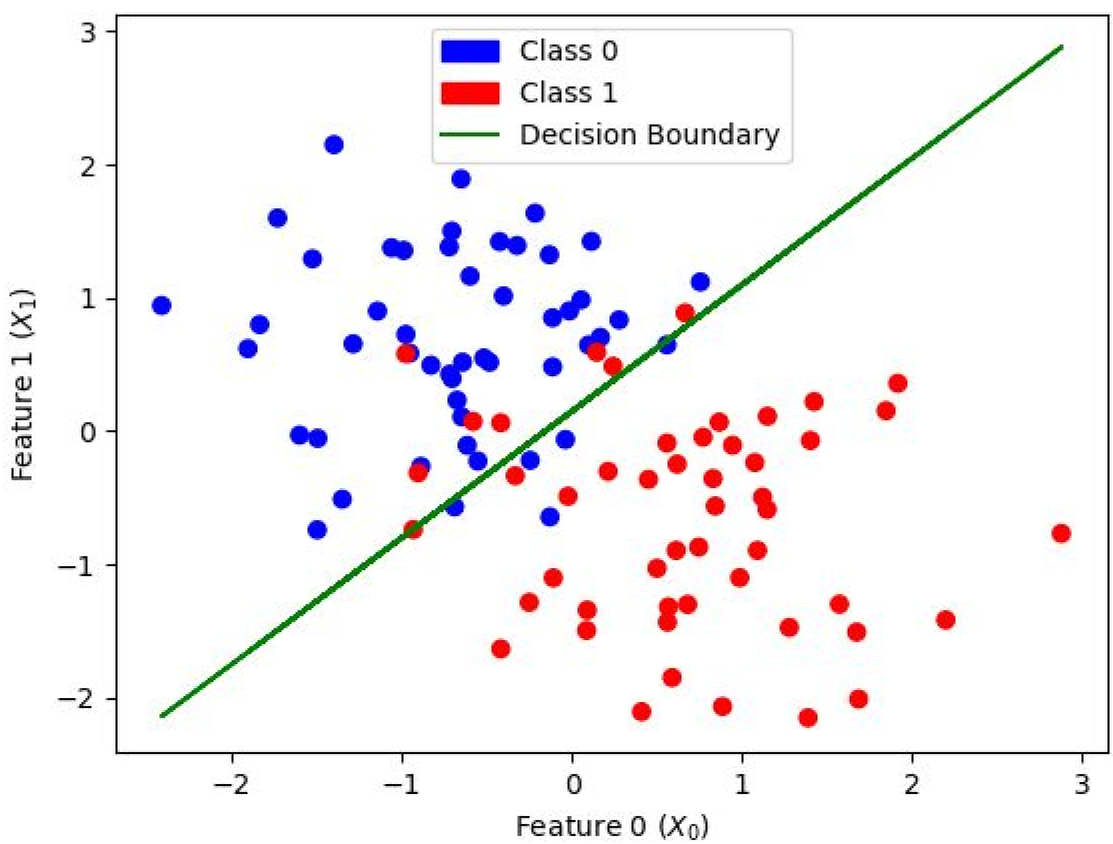
a.)

Logistic Regression using Newton's Method

Final Parameters = [0.40125316, 2.5885477 , -2.72558849]
Number of iterations = 8

b.)

Plot of the training data and decision boundary learnt:



Q4: Gaussian Discriminant Analysis

a.)

GDA using Closed form equations.

ϕ = 0.5

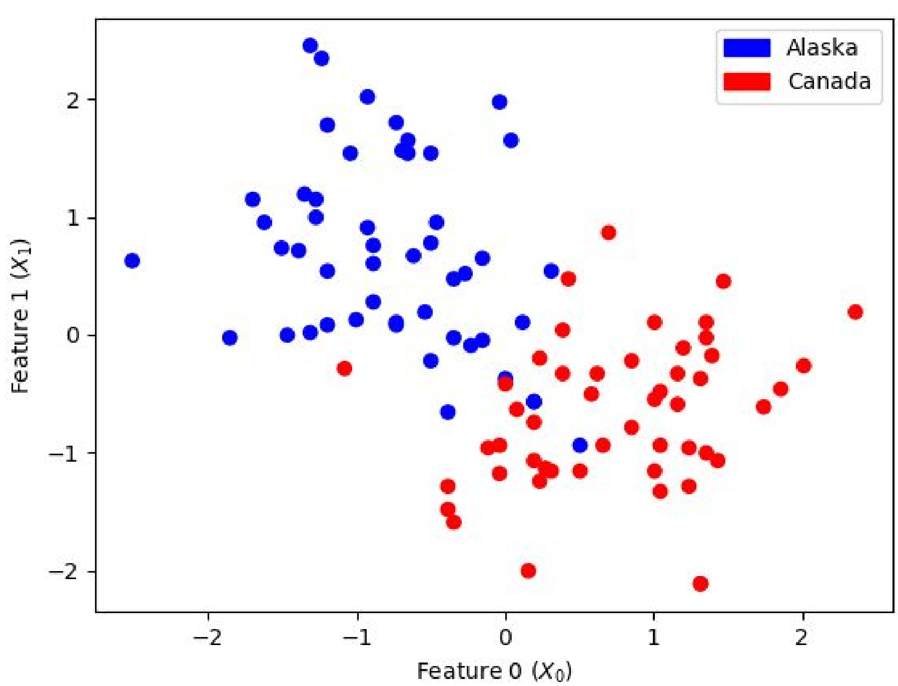
μ_0 = [-0.75529433 0.68509431]

μ_1 = [0.75529433 -0.68509431]

Σ = [[0.42953048, -0.02247228],
[-0.02247228, 0.53064579]]

b.)

Plot of Training Data:



c.)

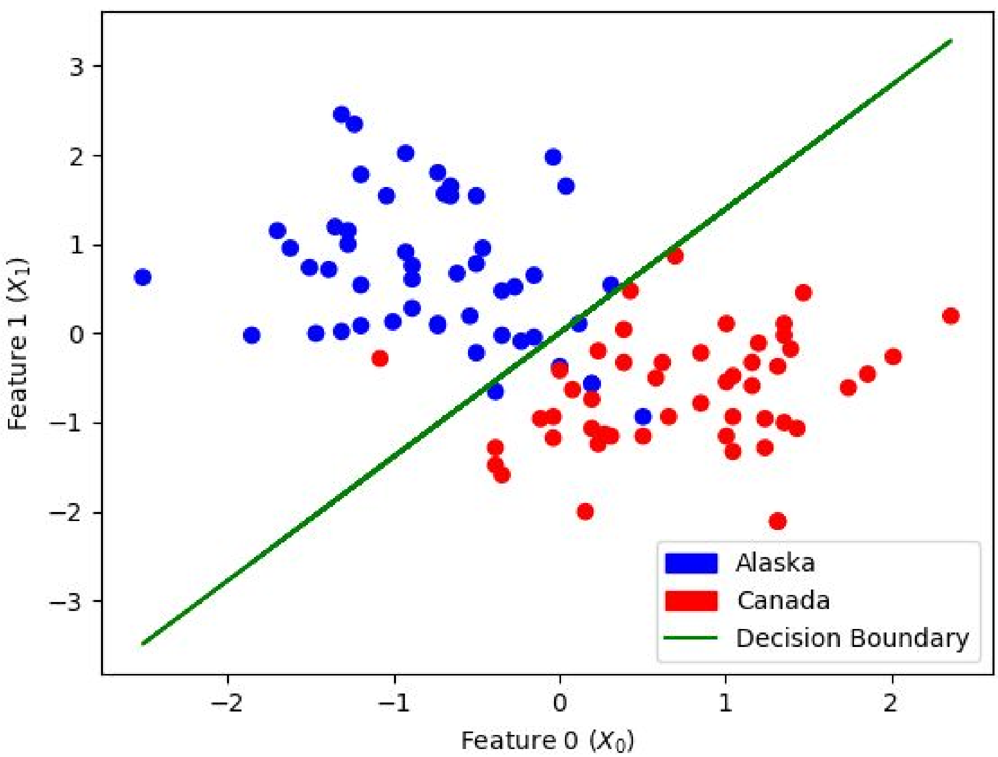
Equation of Linear Boundary:

Linear boundary can be expressed as : A X = B

Where $A = 2 * (\mu_0^T \Sigma^{-1} - \mu_1^T \Sigma^{-1})$

$B = (\mu_0^T \Sigma_0^{-1} \mu_0 - \mu_1^T \Sigma_1^{-1} \mu_1 - 2 * \log (((1/\phi) - 1))$

Plot of Decision Boundary:



d.)

GDA with different covariance matrices.

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.)

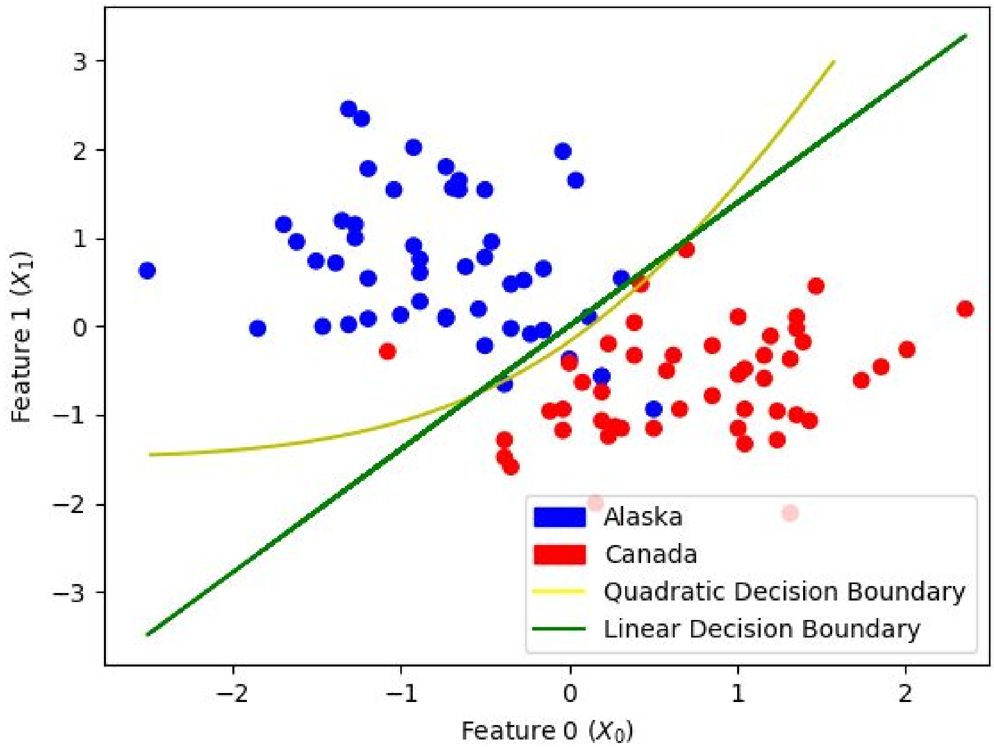
Equation for Quadratic Decision Boundary:

The general decision boundary can be written as:

$X^T A X + B X + C = 0$

Where $A = \Sigma_0^{-1} - \Sigma_1^{-1}$
$B = -2 * (\mu_0^T \Sigma_0^{-1} - \mu_1^T \Sigma_1^{-1})$
$C = (\mu_0^T \Sigma_0^{-1} \mu_0 - \mu_1^T \Sigma_1^{-1} \mu_1 - 2 * \log (((1/\phi) - 1) * (|\Sigma_1| / \Sigma_0))$

Plot of Quadratic & Linear Decision Boundaries:

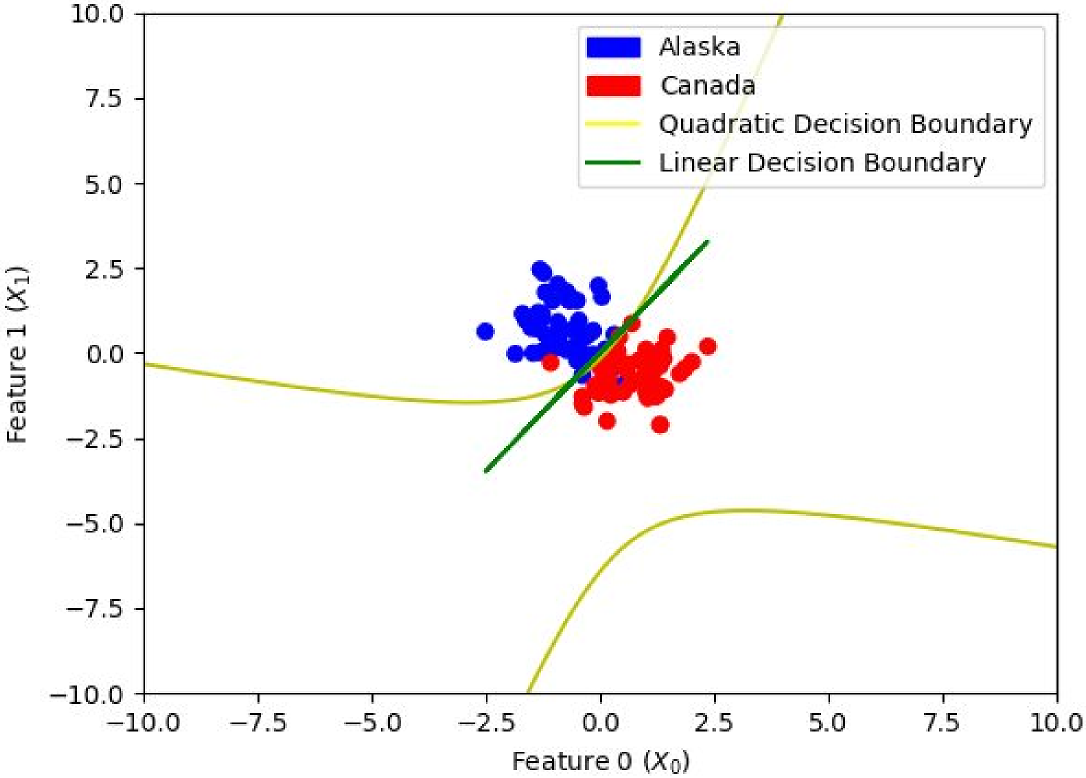


f.)

Analysis/Comment on Decision Boundaries

As we can see from the last figure, the quadratic classifier separates the classes a bit better. Some of the Alaskan fishes that were being incorrectly classified are now classified correctly.

Zooming out on the same graph, we see that the curve is actually a hyperbola!



Alaska
Canada
Quadratic Decision Boundary
Linear Decision Boundary
Feature 1 (X_1)
Feature 0 (X_0)
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0