Assignment 1 Report

Anoop (2015CS10265)

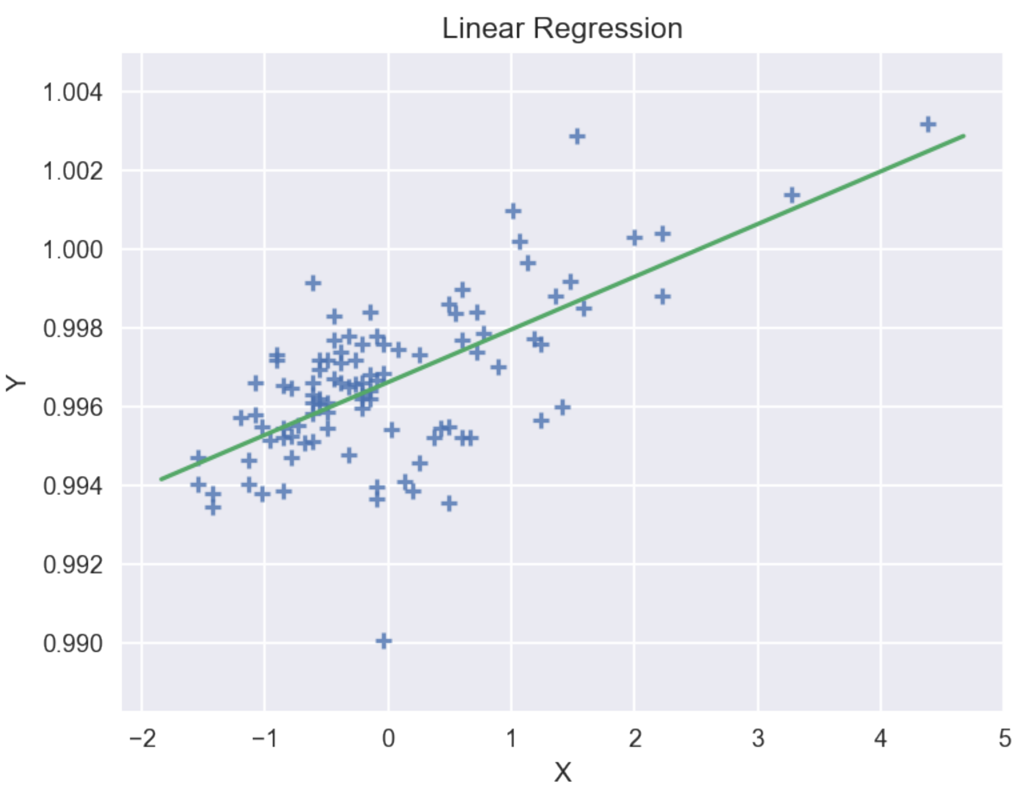
February 12, 2018

Linear Regression

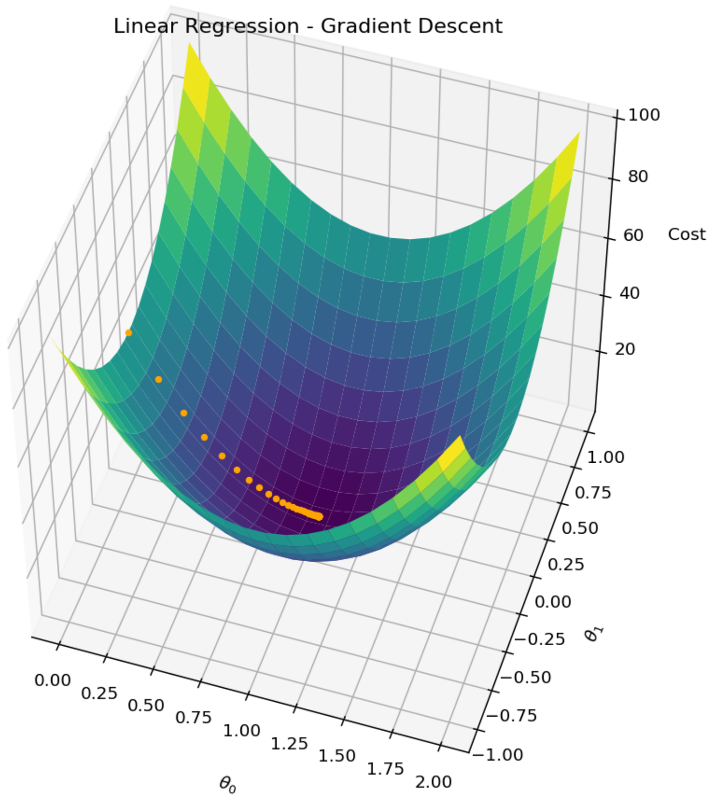
Part a

Learning Rate (η) - 0.0015
Stopping criteria - $cost < 0.0001$ or $iterations > 101$
Final Parameters - $\theta_0 = 0.9966201$ and $\theta_1 = 0.0013402$

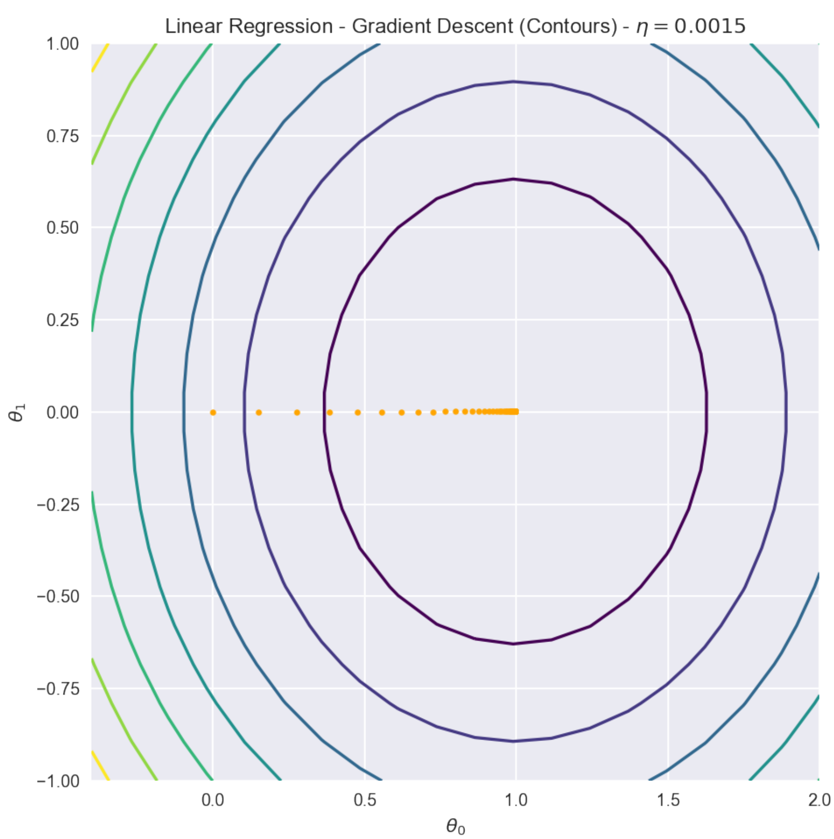
Part b



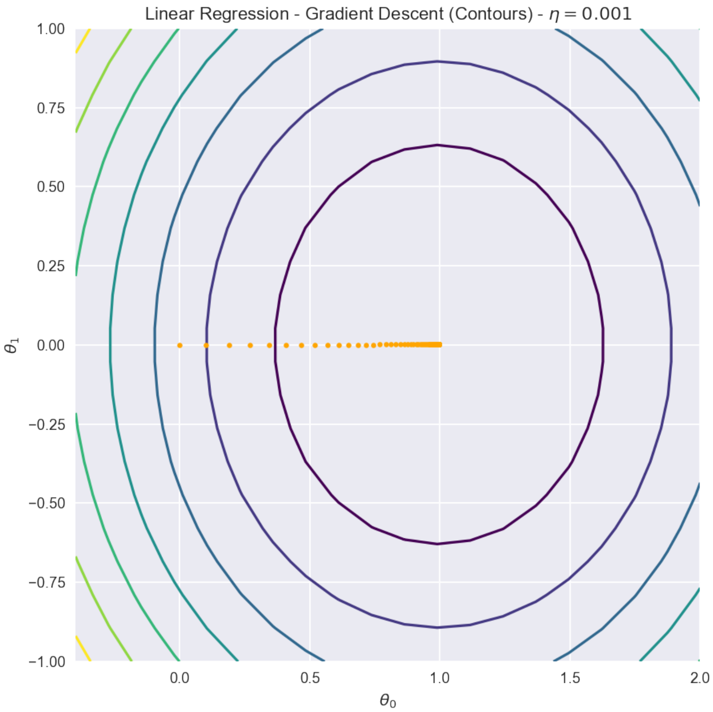
Part c

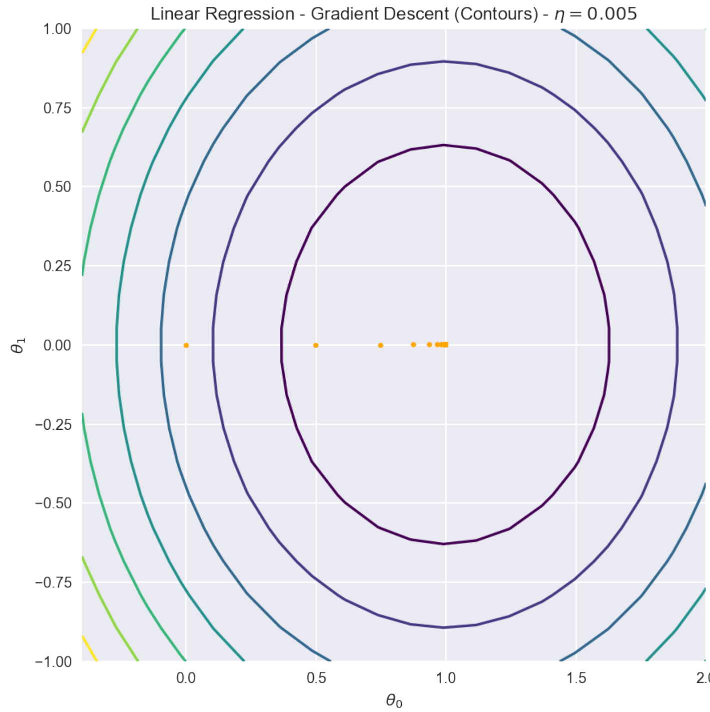

Linear Regression - Gradient Descent
Cost
100
80
60
40
20
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
θ_1
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
θ_0

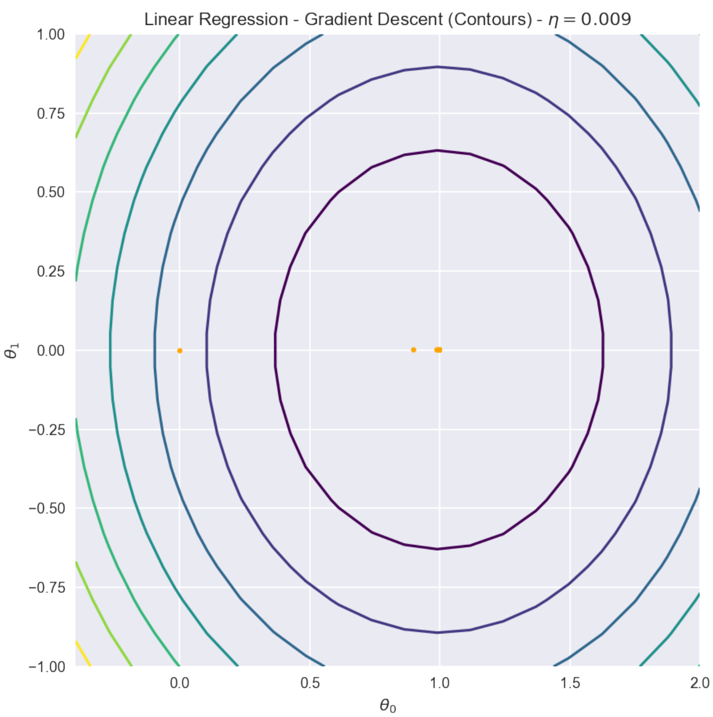
Part d

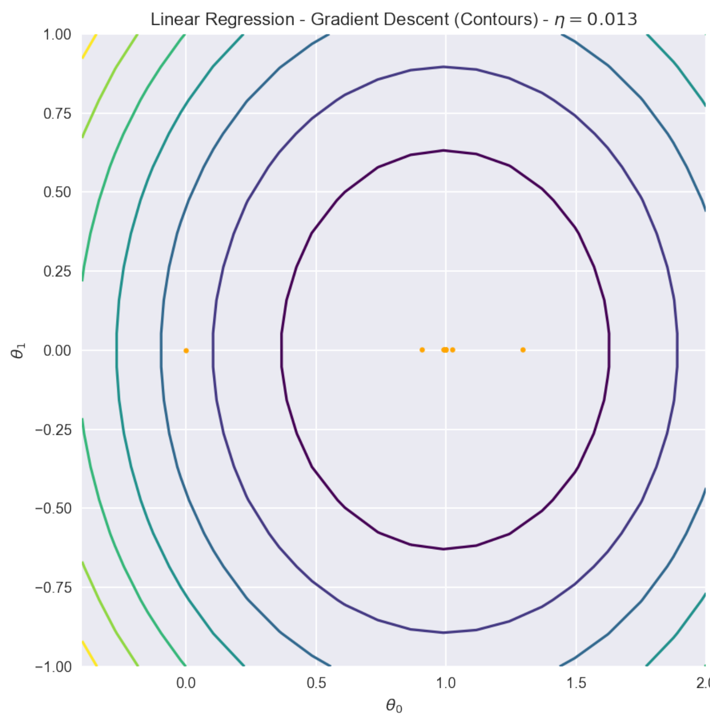

Linear Regression - Gradient Descent (Contours) - $\eta = 0.0015$
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
θ_1
0.0
0.5
1.0
1.5
2.0
θ_0

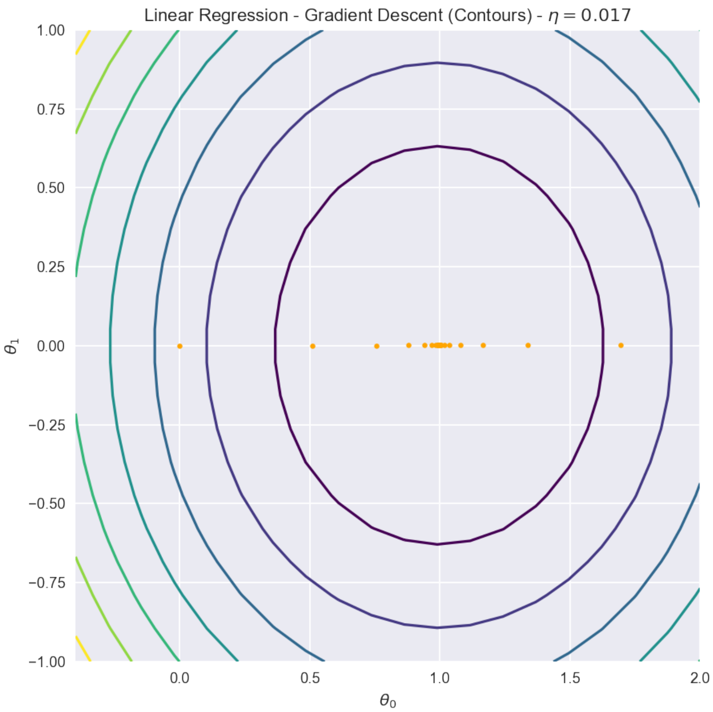
Part e

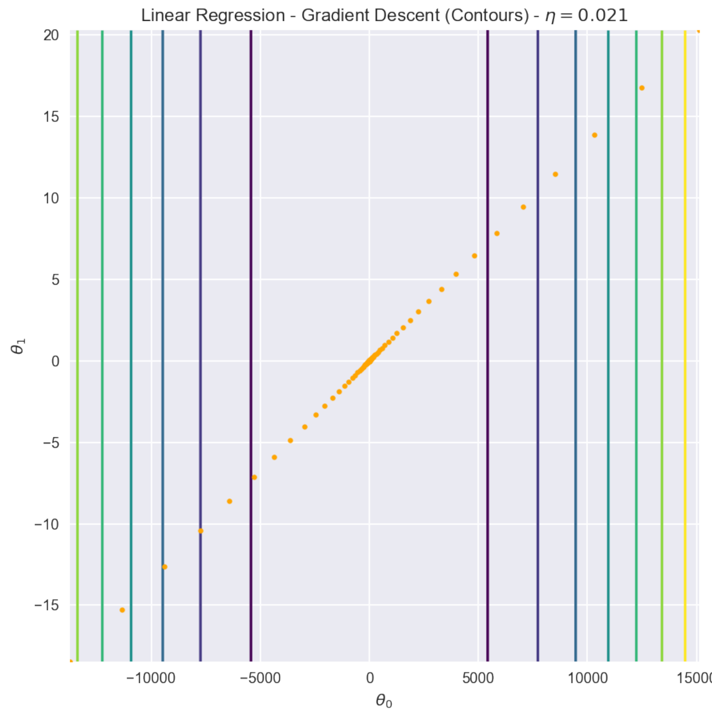

Linear Regression - Gradient Descent (Contours) - $\eta = 0.001$
θ_1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
0.5
1.0
1.5
2.0
θ_0

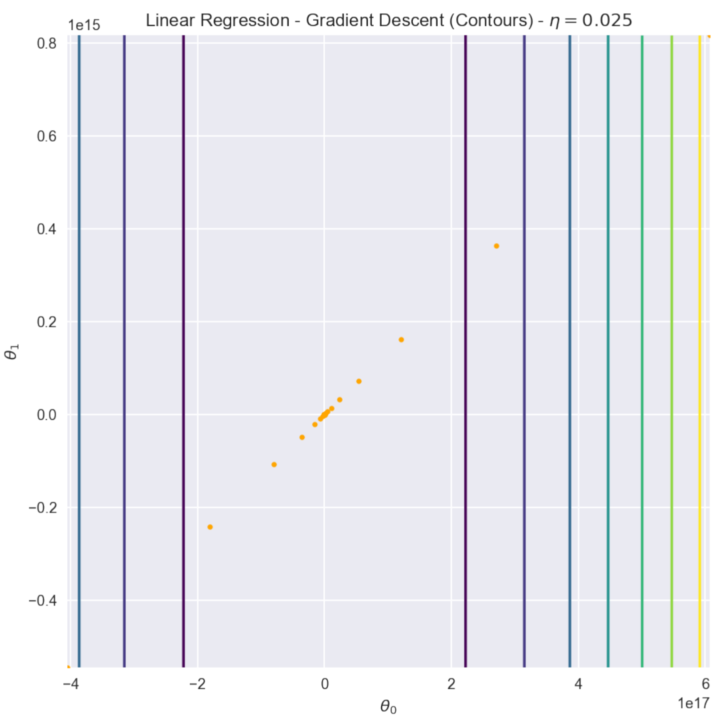

Linear Regression - Gradient Descent (Contours) - $\eta = 0.005$
θ_1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
0.5
1.0
1.5
2.0
θ_0


Linear Regression - Gradient Descent (Contours) - $\eta = 0.009$
θ_1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
0.5
1.0
1.5
2.0
θ_0


Linear Regression - Gradient Descent (Contours) - $\eta = 0.013$
θ_1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
0.5
1.0
1.5
2.0
θ_0





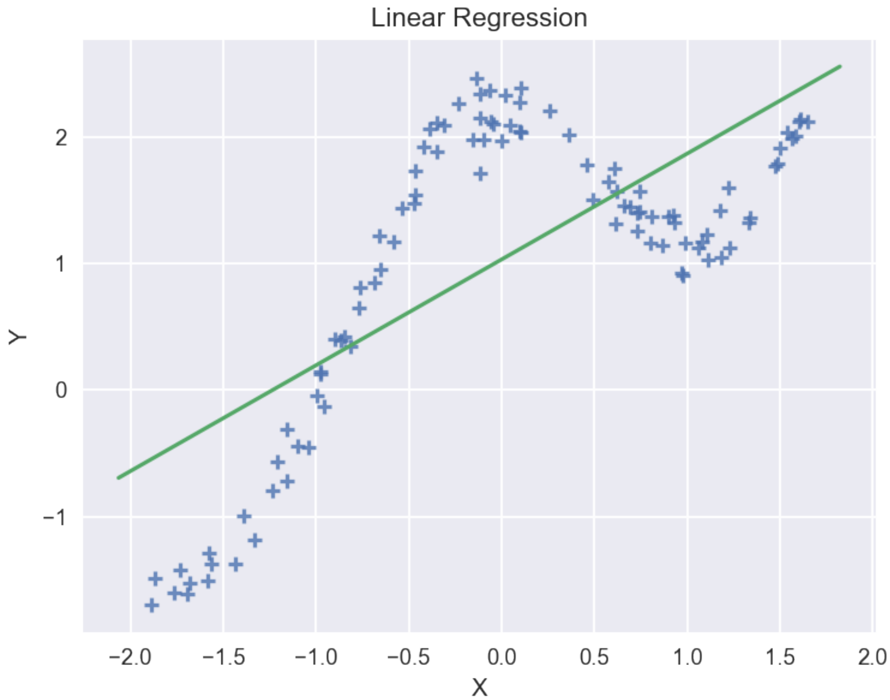


From the contours for various learning rates, it is clear that as the learning rate increases beyond some point, the cost function starts diverging. Initially there are small oscillations and the function still converges. As learnging rate increases more, there are large oscillations around the minimum, each time overshooting even more.

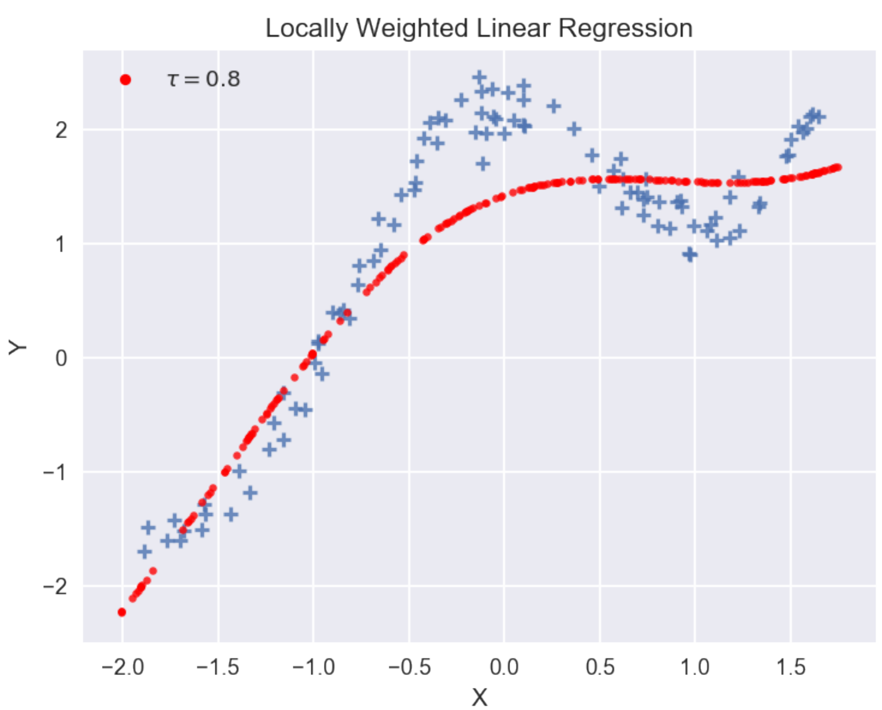
Locally Weighted Linear Regression

Part a

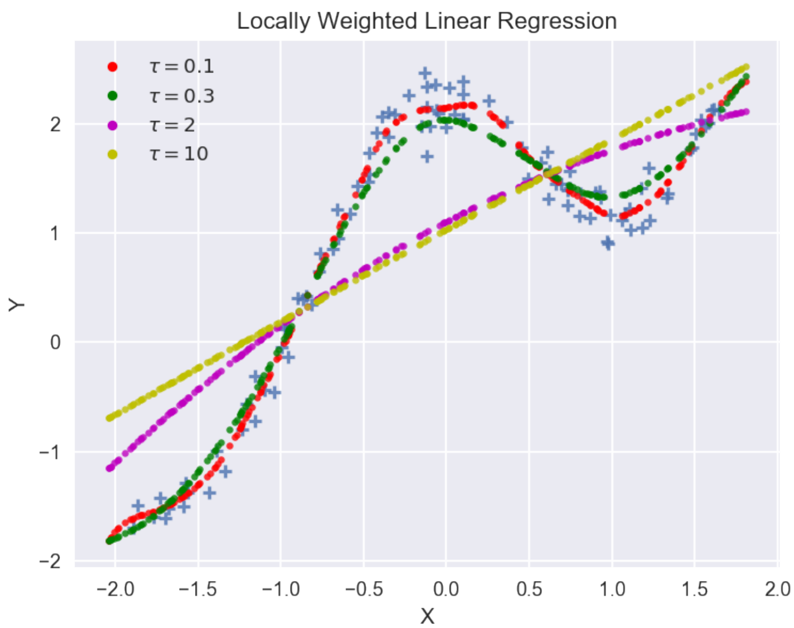
Final Parameters - $\theta_0 = 1.03128116$ and $\theta_1 = 0.83519315$



Part b



Part c



As τ becomes too large, the model starts underfitting (tends to be linear). As τ becomes too small, the model starts overfitting (tends to pass through all points).

Logistic Regression

Part a

Final Parameters - $\theta_0 = -0.2130308$, $\theta_1 = -2.65801937$ and $\theta_2 = 2.66106075$

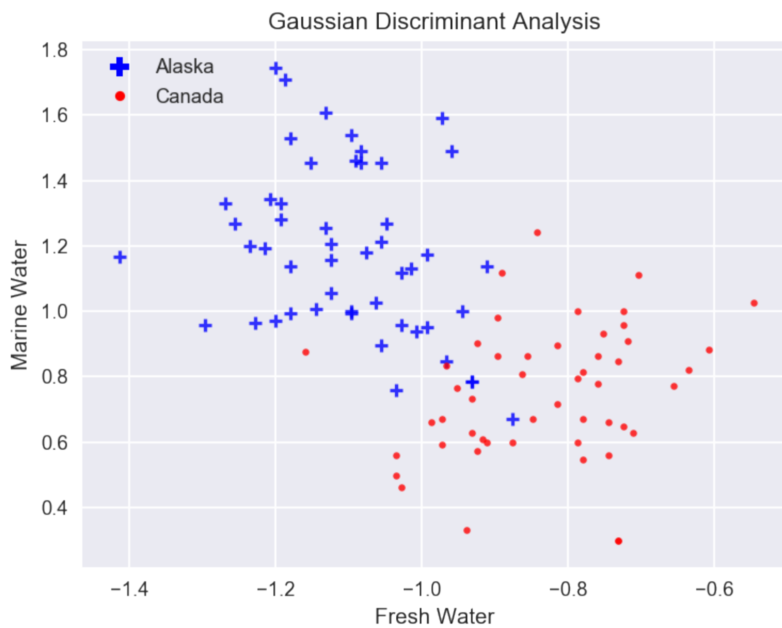
Part b

Gaussian Discriminant Analysis

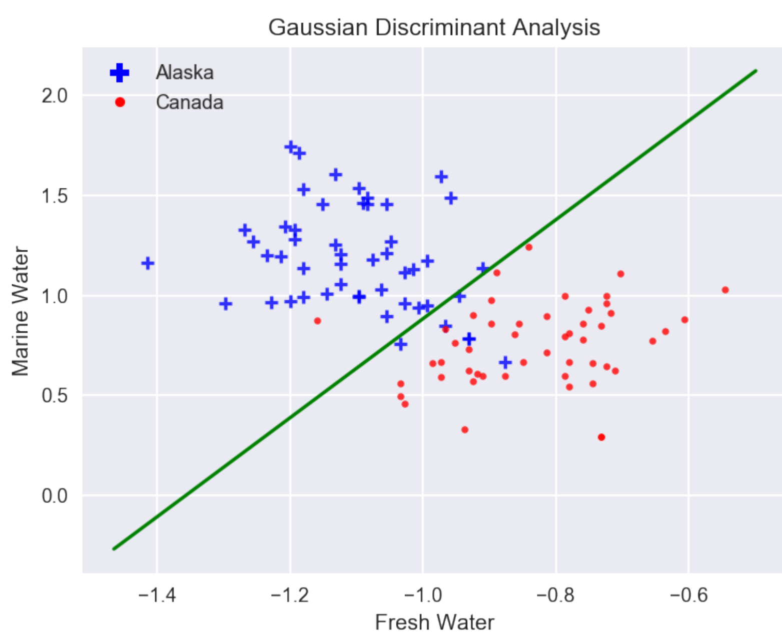
Part a

$$\mu_0 = \begin{bmatrix} -0.8315402 & 0.74891723 \end{bmatrix}^T$$
$$\mu_1 = \begin{bmatrix} -1.10106488 & 1.18368785 \end{bmatrix}^T$$
$$\Sigma_0 = \Sigma_1 = \Sigma = \begin{bmatrix} 0.0136741 & -0.00127227 \\ -0.00127227 & 0.05342744 \end{bmatrix}$$

Part b



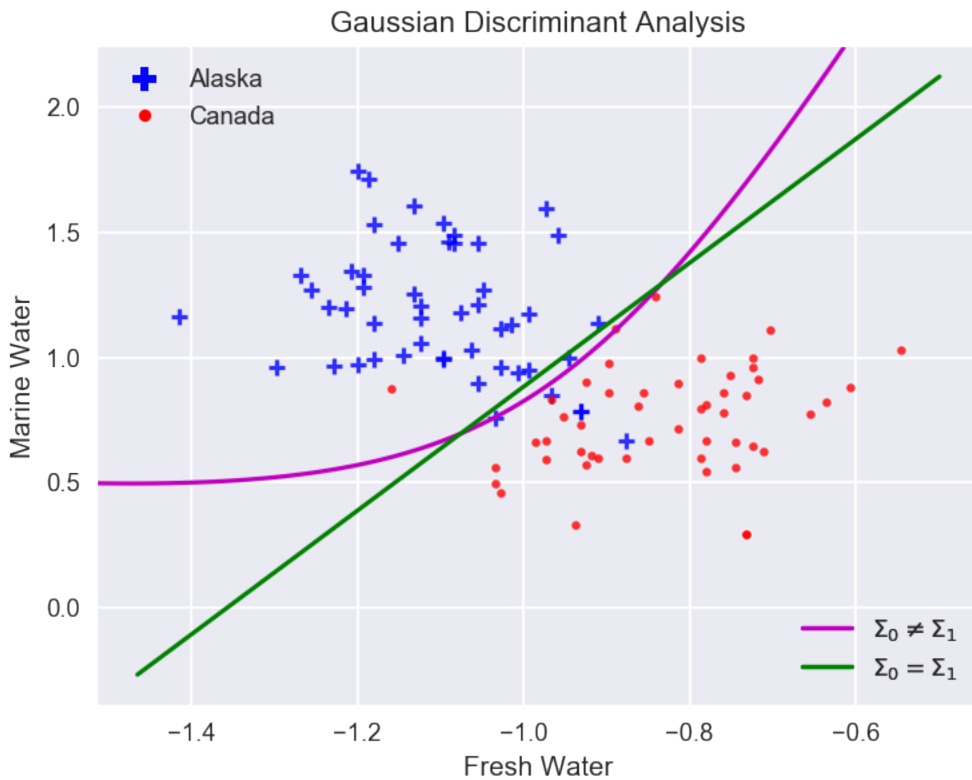
Part c



Part d

$$\mu_0 = \begin{bmatrix} -0.8315402 & 0.74891723 \end{bmatrix}^T$$
$$\mu_1 = \begin{bmatrix} -1.10106488 & 1.18368785 \end{bmatrix}^T$$
$$\Sigma_0 = \begin{bmatrix} 0.01520029 & 0.00622316 \\ 0.00622316 & 0.04163824 \end{bmatrix}$$
$$\Sigma_1 = \begin{bmatrix} 0.0121479 & -0.0087677 \\ -0.0087677 & 0.06521665 \end{bmatrix}$$

Part e



Part f

From the plot (in Part e) we can see that the quadratic decision boundary does better job of classification than the linear decision boundary (equivalent to logistic regression). This means that the assumption that the underlying data belongs to Gaussian distribution is valid to a great extent.