Assignment 1 Report

Anoop (2015CS10265)

February 11, 2018

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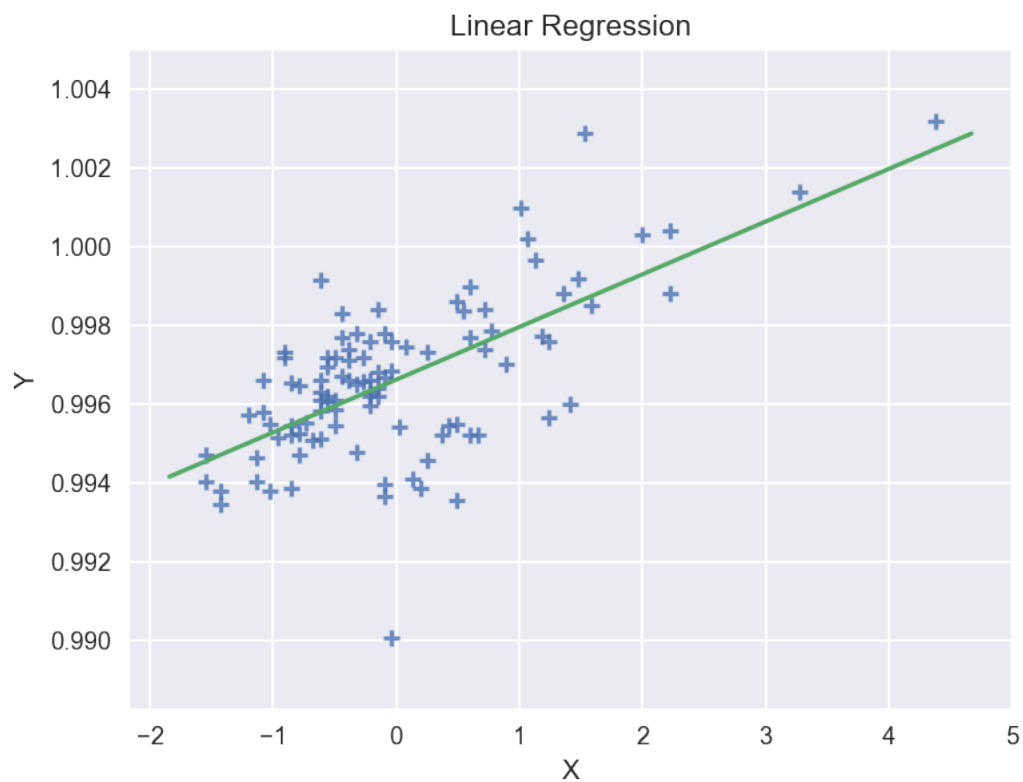
Part a
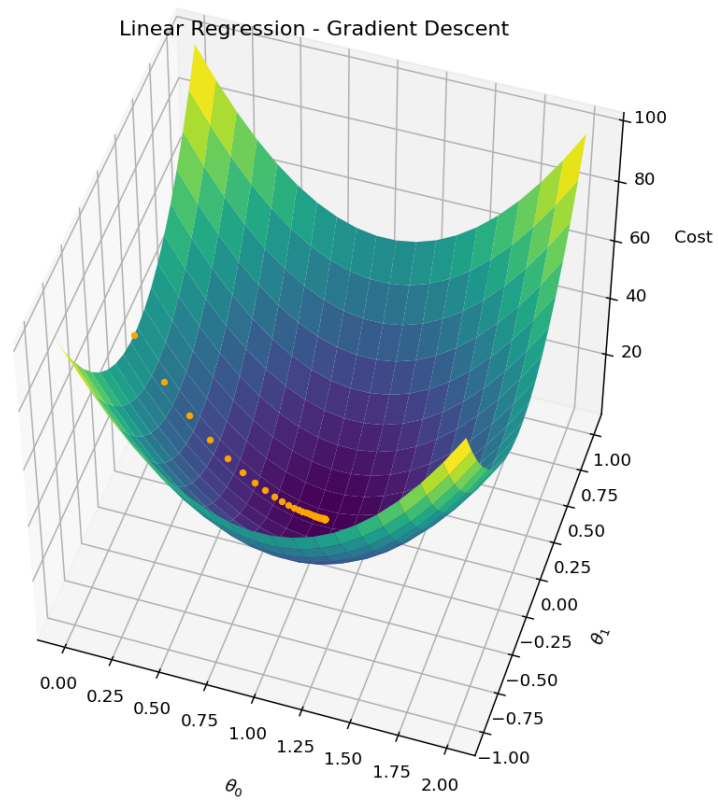
Learning Rate (η) - 0.0015
Stopping criteria - $cost < 0.0001$ or $iterations > 101$
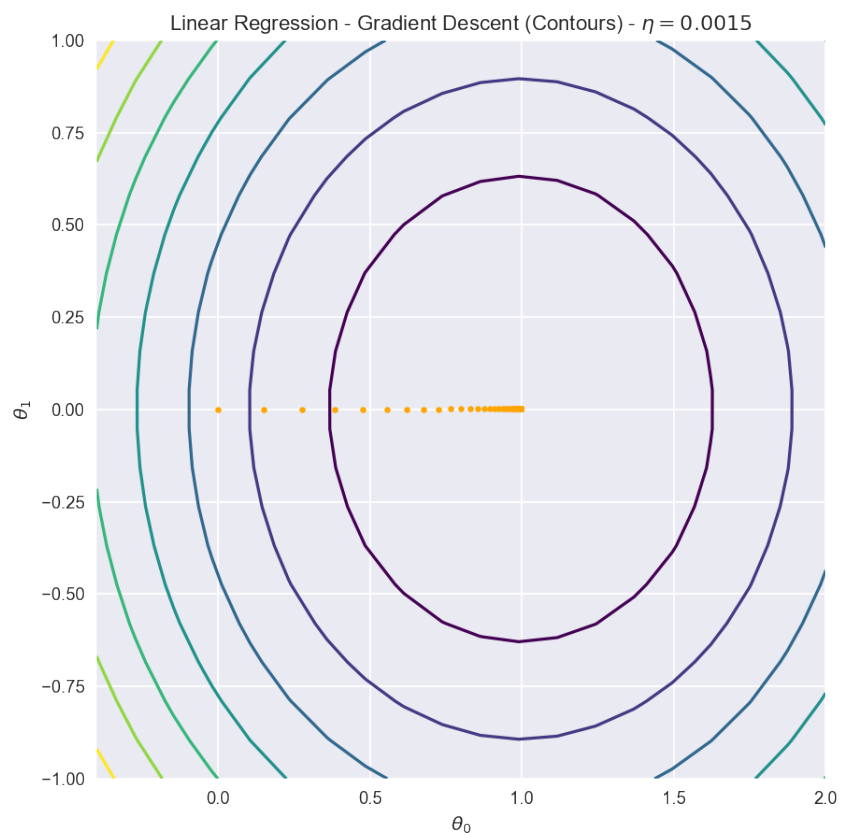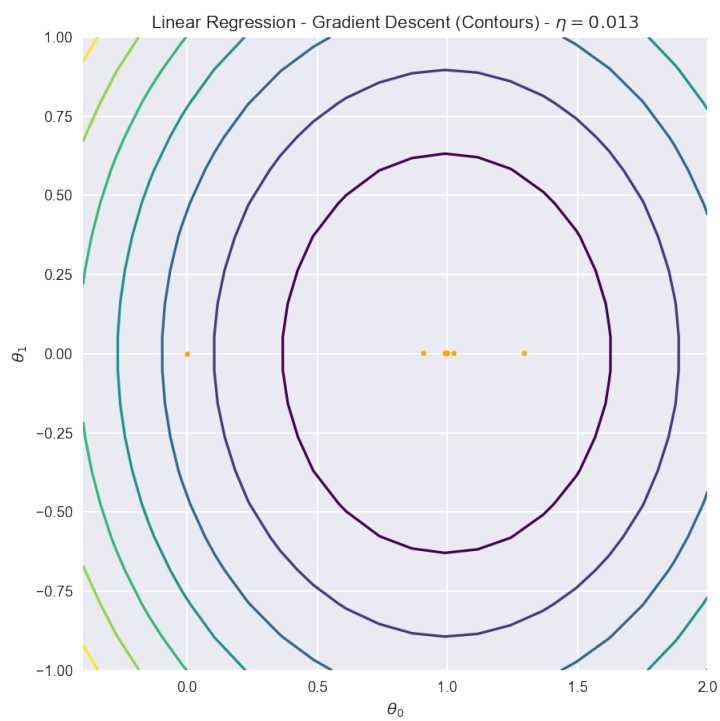Final Parameters - $\theta_0 = 0.9966201$ and $\theta_1 = 0.0013402$
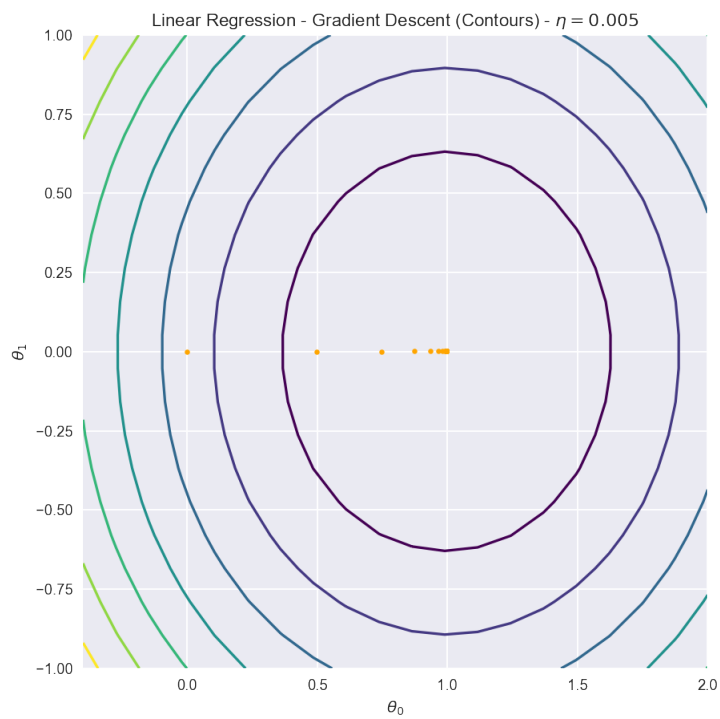
Part b

Part c

Linear Regression - Gradient Descent

Part d

Linear Regression - Gradient Descent (Contours) - $\eta = 0.0015$

Part e

Linear Regression - Gradient Descent (Contours) - $\eta = 0.001$

Linear Regression - Gradient Descent (Contours) - $\eta = 0.005$

Linear Regression - Gradient Descent (Contours) - $\eta = 0.009$

Linear Regression - Gradient Descent (Contours) - $\eta = 0.013$

Linear Regression - Gradient Descent (Contours) - $\eta = 0.017$

Linear Regression - Gradient Descent (Contours) - $\eta = 0.021$

Linear Regression - Gradient Descent (Contours) - $\eta = 0.025$
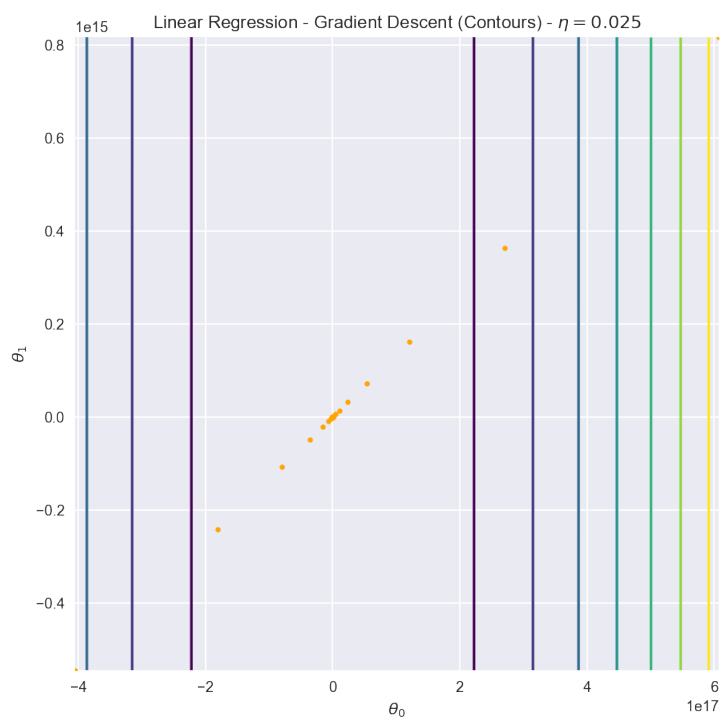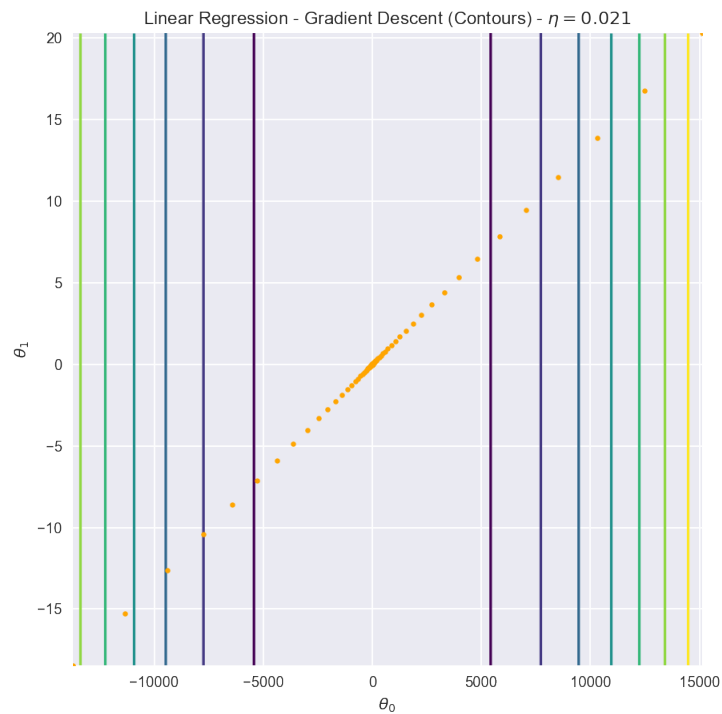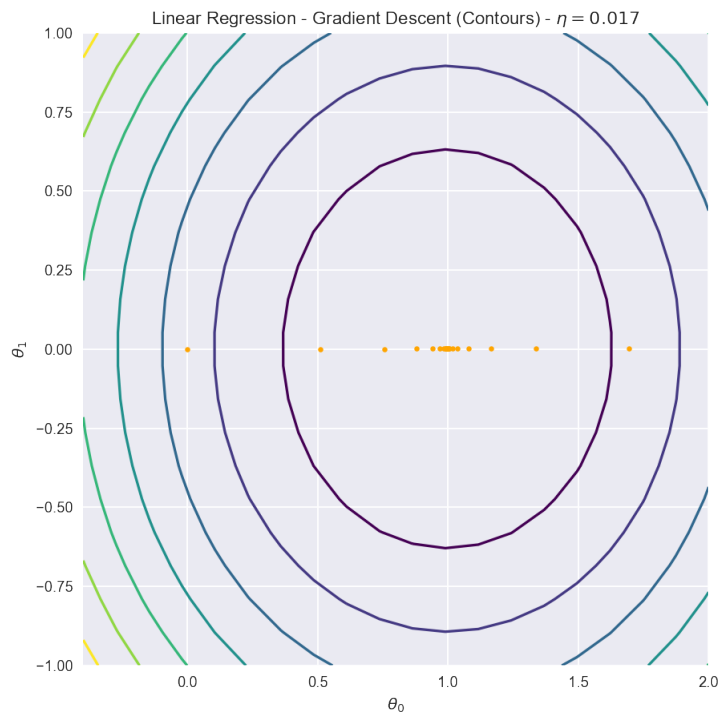
From the contours for various learning rates, it is clear that as the learning rate increases beyond some point, the cost function starts diverging. After initial overshooting, there are large oscillations around the minimum, each time overshooting even more.
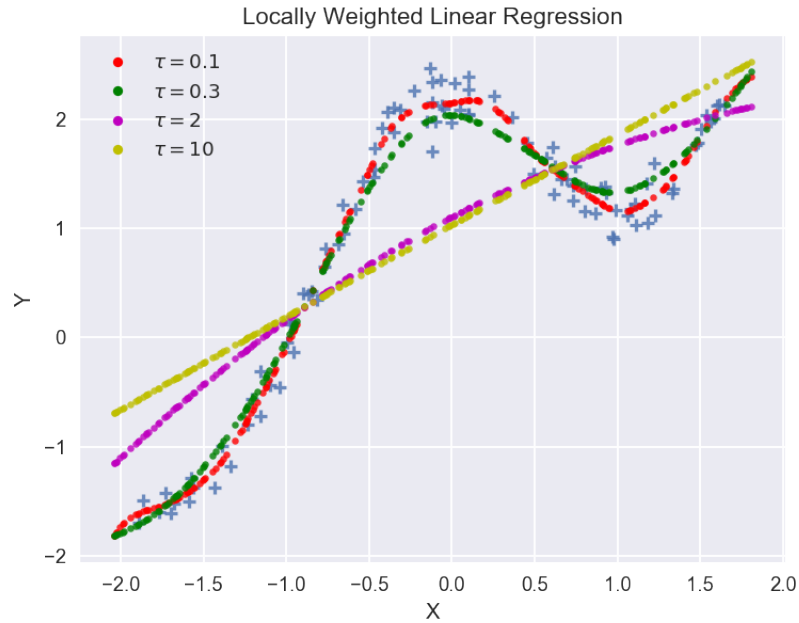
Locally Weighted Linear Regression

Part a

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

Part b

Part c

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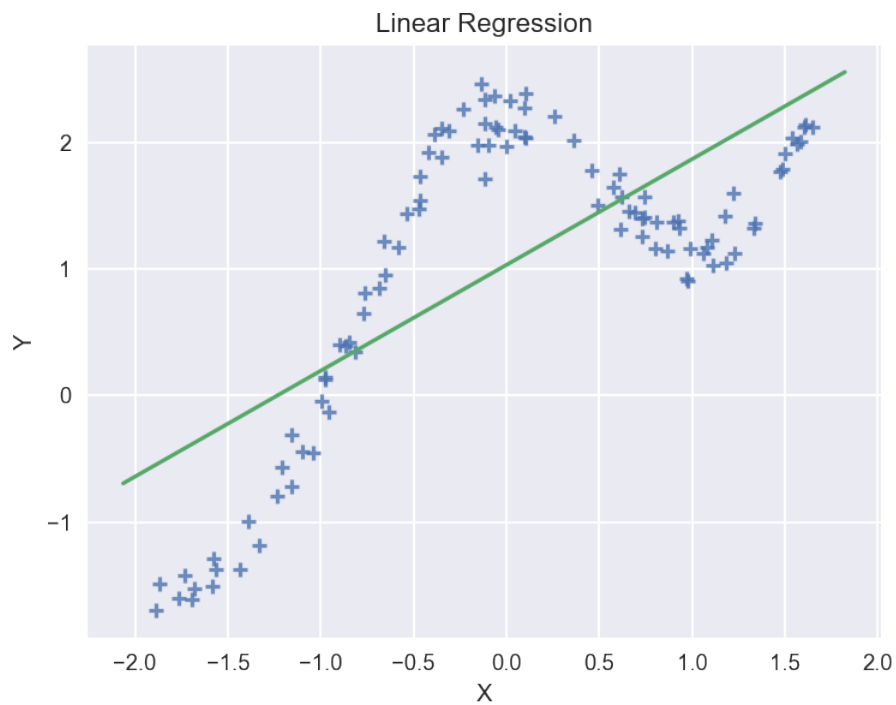
As τ becomes too large, the model starts underfitting (tends to be linear). As τ becomes too small, the model starts overfitting (tends to pass through all points).

Logistic Regression

Part a

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

Part b

Logistic Regression
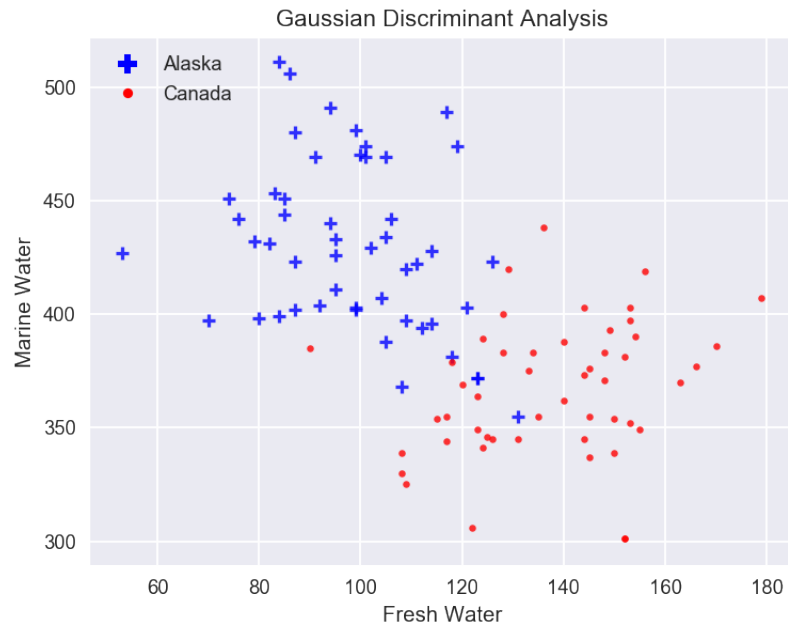
Gaussian Discriminant Analysis

Part a

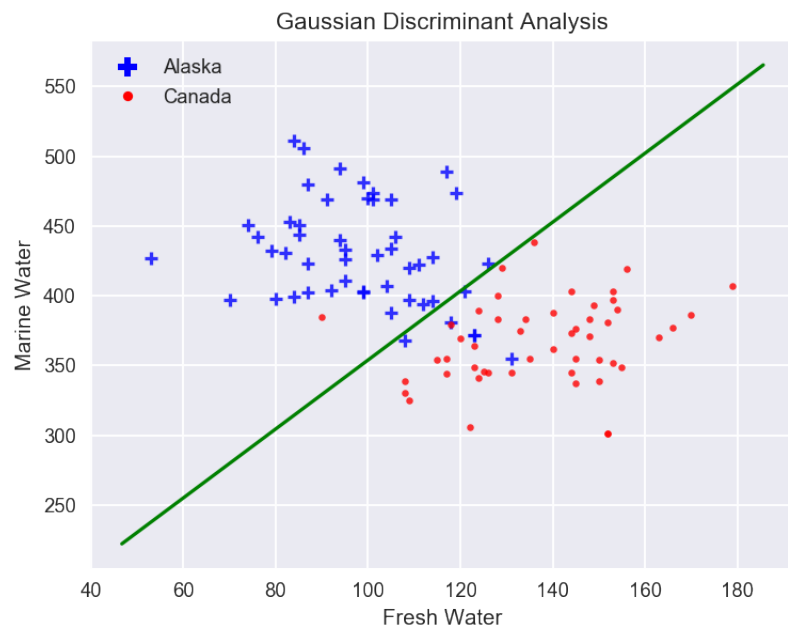$$\mu_0 = \begin{bmatrix} 137.46 & 366.62 \end{bmatrix}^T$$
$$\mu_1 = \begin{bmatrix} 98.38 & 429.66 \end{bmatrix}^T$$
$$\Sigma_0 = \Sigma_1 = \Sigma = \begin{bmatrix} 287.482 & -26.748 \\ -26.748 & 1123.25 \end{bmatrix}$$
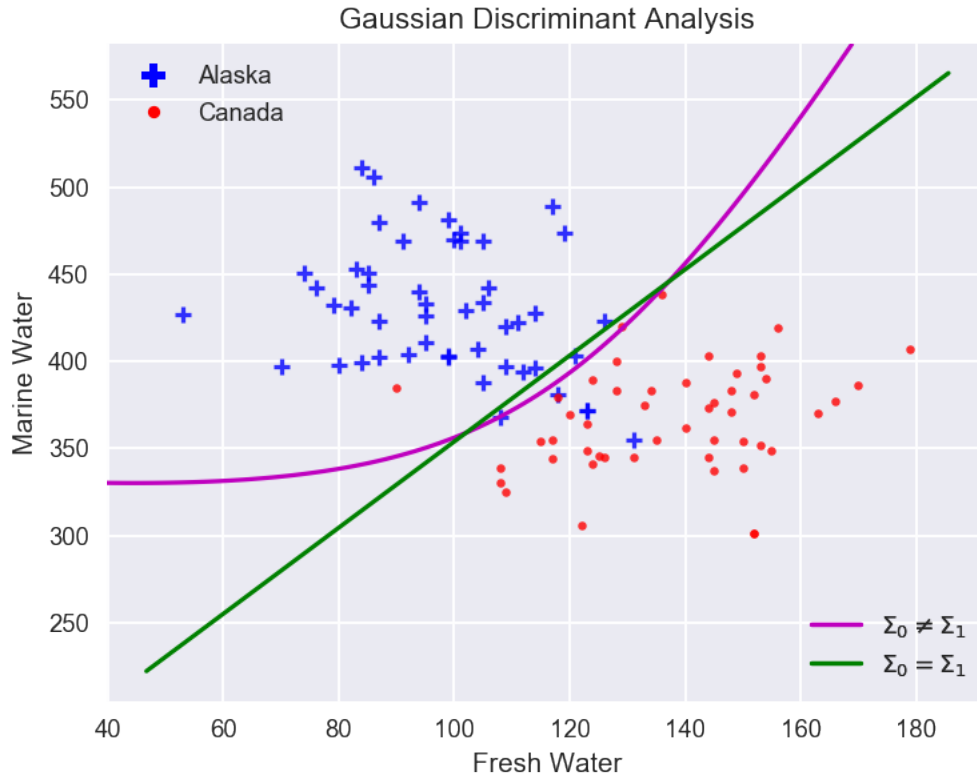
Part b

Part c

Part d

$$\mu_0 = \begin{bmatrix} 137.46 & 366.62 \end{bmatrix}^T$$
$$\mu_1 = \begin{bmatrix} 98.38 & 429.66 \end{bmatrix}^T$$
$$\Sigma_0 = \begin{bmatrix} 319.5684 & 130.8348 \\ 130.8348 & 875.3956 \end{bmatrix}$$
$$\Sigma_1 = \begin{bmatrix} 255.3956 & -184.3308 \\ -184.3308 & 1371.1044 \end{bmatrix}$$

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.