

Assignment 1

Suraj S, 2018MCS2024

February 12, 2019

The Assignments were made on python2

1 Linear Regression

1.1 Implementing the batch gradient descent

The batch gradient descent was implemented in the following section used the following equations in order to implement the batch gradient descent.

- the cost function

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m \left(h_{\theta}(x^{(i)}) - y^{(i)} \right)^2$$

- gradient descent for the j^{th} theta parameters

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \text{ for } j=0,1,\dots,n$$

for $j=1$ can be further rewritten as :

$$\frac{\partial}{\partial \theta_1} \left(\frac{1}{2m} \sum_{i=1}^m \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right)^2 \right) = \frac{1}{m} \sum_{i=1}^m \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right) \cdot x^{(i)} =$$
$$\frac{1}{m} \sum_{i=1}^m \left(h_{\theta}(x^{(i)}) - y^{(i)} \right) \cdot x^{(i)}$$

now implementing the gradient descent, we have first normalized the data in order to have mean 0 and variance 1

using the gradient descent we have used the convergence condition as the difference between the initial and final cost should be less than a given threshold (i.e about 10^{-9})

the following output was given by the program upon convergence

Converged

Following are the details of the gradient descent

Learning Rate: 0.1

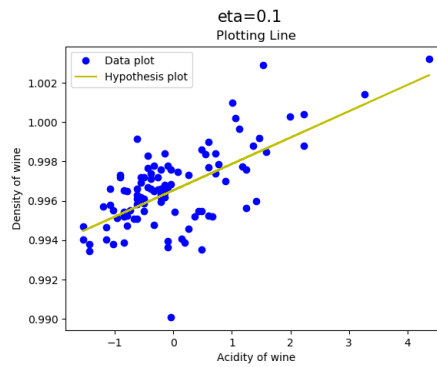
Stopping Criteria: absolute value of difference between the initial and final error

The number of iterations took in order to converge is 29

Theta Values (Final Parameters) [[0.99653574 0.00134008]]

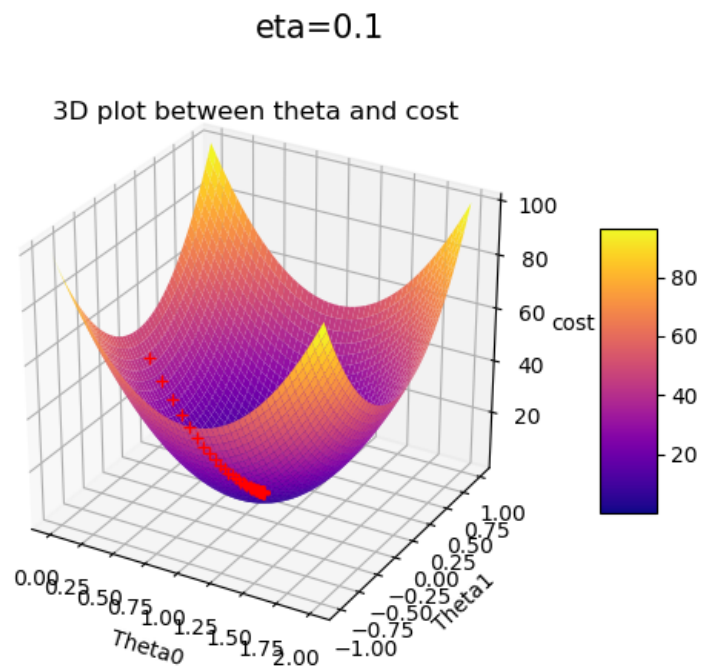
1.2 Plotting the line

the lines were plotted using the value of $\eta = 0.1$



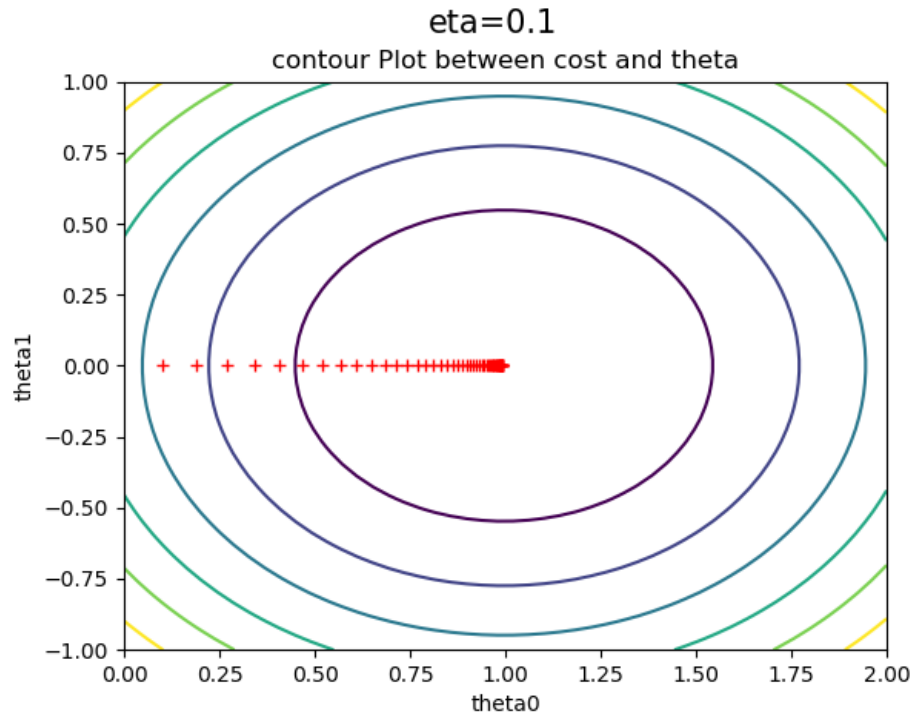
1.3 Plotting the mesh

the mesh were plotted using the value of $\eta = 0.1$



1.4 Plotting the contours

the contour were plotted using the value of $\eta = 0.1$



1.5 trying with different values of η

In this section we have used the various values of $\eta = \{0.1, 0.5, 0.9, 1.3, 1.7, 2.1, 2.4, 2.9\}$ and ran the above procedure with these values we have noticed that the gradient descent didn't converge in the value from 2.1 also we have noticed that the number of iterations that was required for the convergence was reduced as the value was increased.

2 Locally Weighted Linear regression

The data was first normalized and then the data was processed further

The normal equations in order to find the parameters used in the following sections are as follows:

For unweighted linear regression:

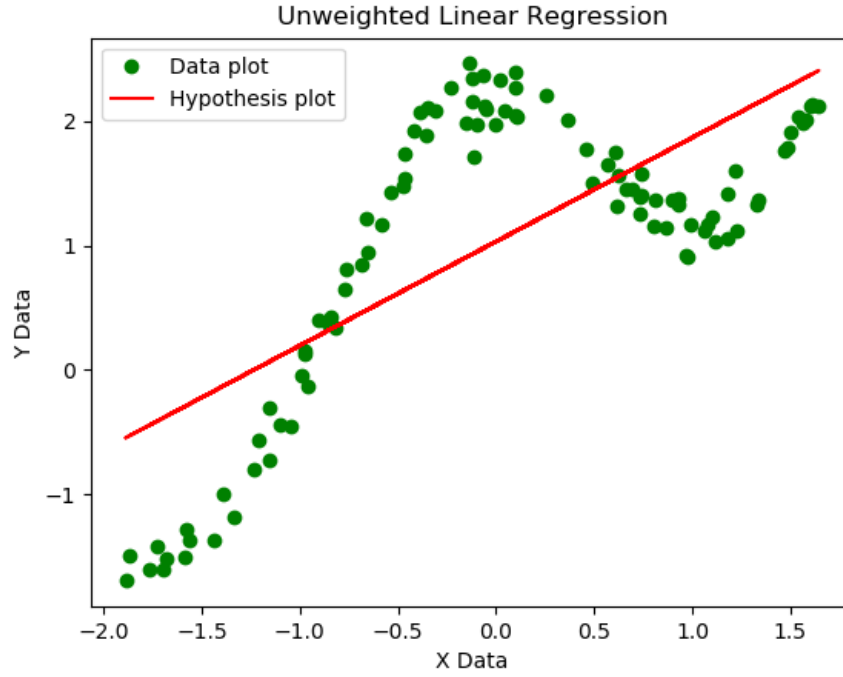
$$\theta = (X^T X)^{-1} X^T Y$$

For locally weighted linear regression:

$$\theta = (X^T W X)^{-1} X^T W Y$$

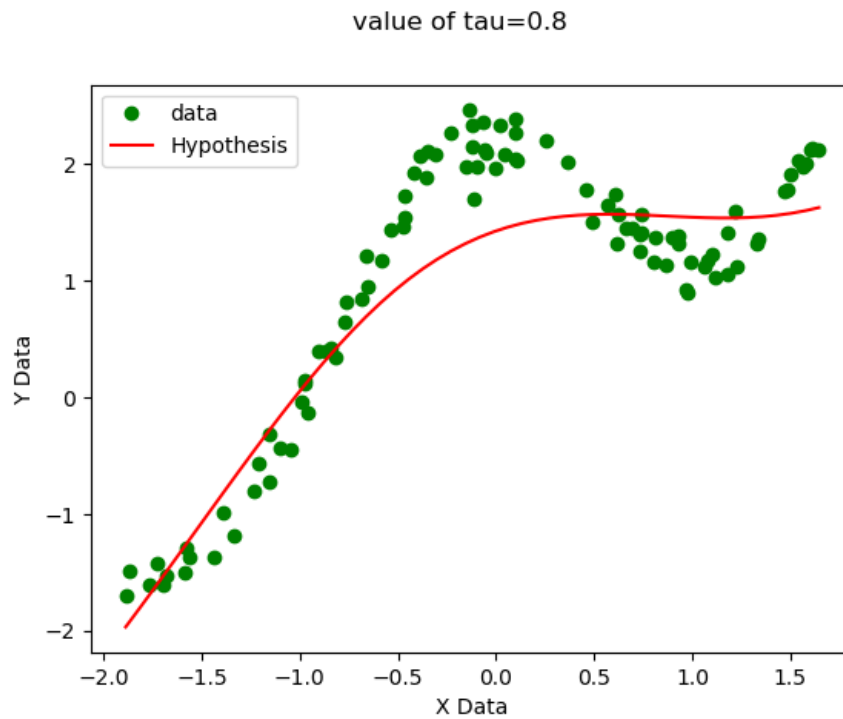
2.1 Unweighted Linear Regression

In this section we have implemented the unweighted linear regression for the distribution.



2.2 Locally Weighted Linear Regression

In this section we have implemented the weighted linear regression for the distribution using the value of $\tau = 0.8$.

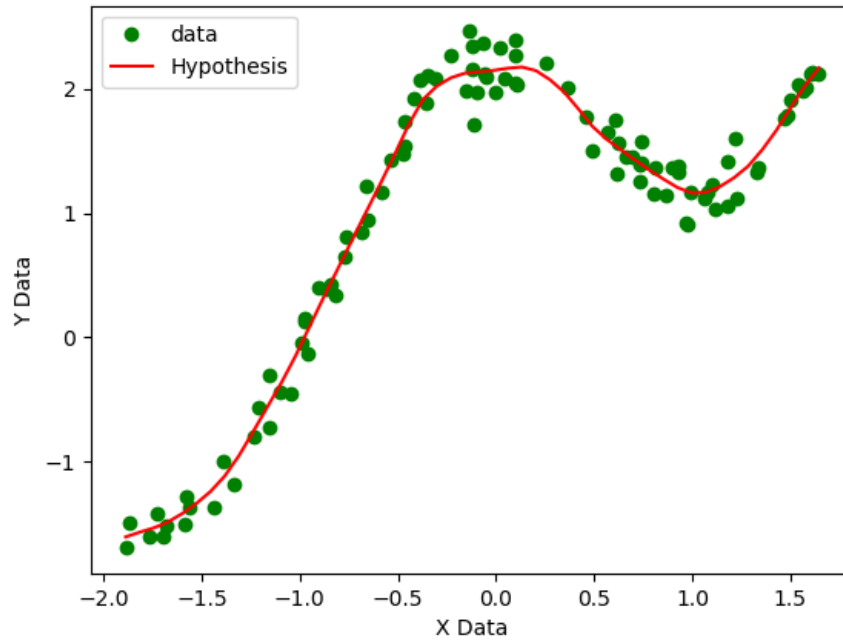


2.3 Weighted Linear Regression using different τ

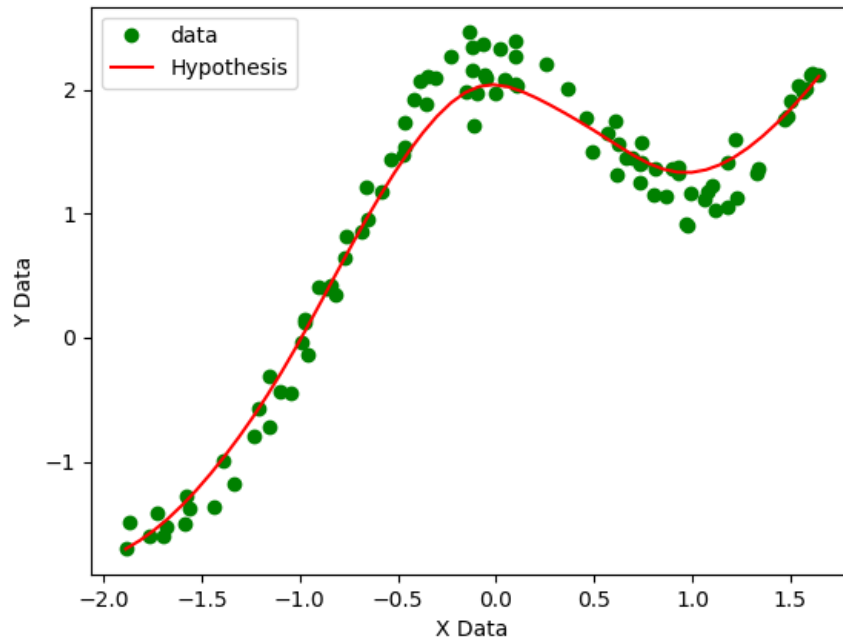
In this section we have implemented the weighted linear regression for the distribution using different values of τ , the values of τ used are $\{0.1, 0.3, 2, 10\}$ we can see that as we increase the value of τ the locally weighted linear regression tends to become linear regression.

Also if we take the value of τ too large it might underfit the curve and might can act like the linear regression and if we take the value of τ to be small it is likely to overfit the data

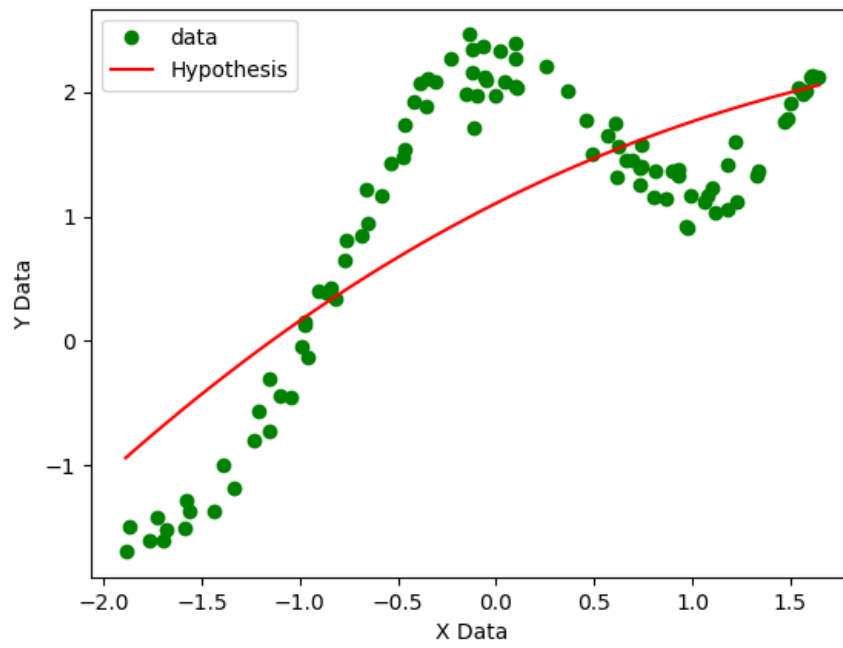
value of tau=0.1



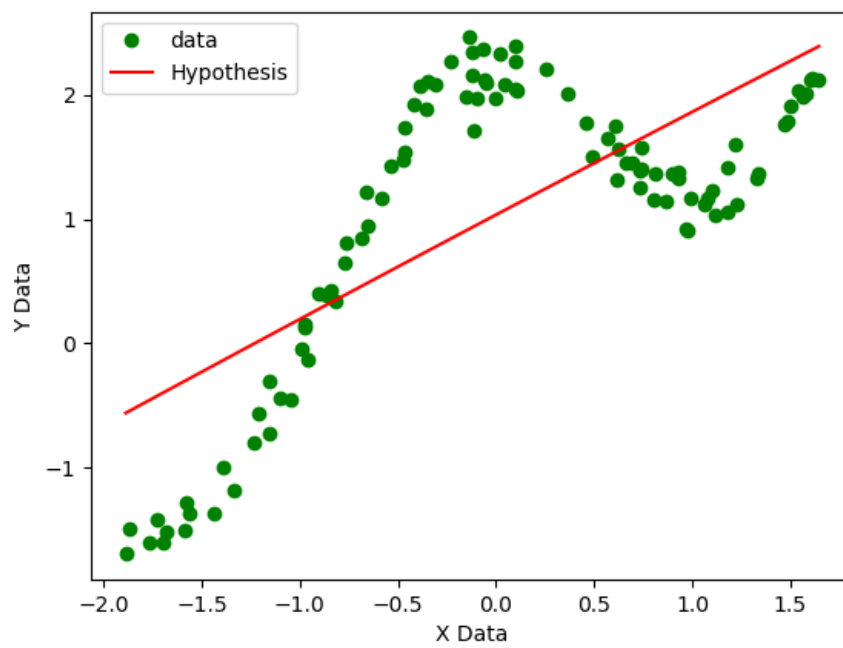
value of tau=0.3



value of tau=2



value of tau=10



3 Logistic Regression

The data was first normalized and then the data was processed further

3.1 Newton's method

In this section we have implemented the logistic regression for the distribution for classifying the two element using the newton's method to find the parameters.

Following are the details of the Newton's method

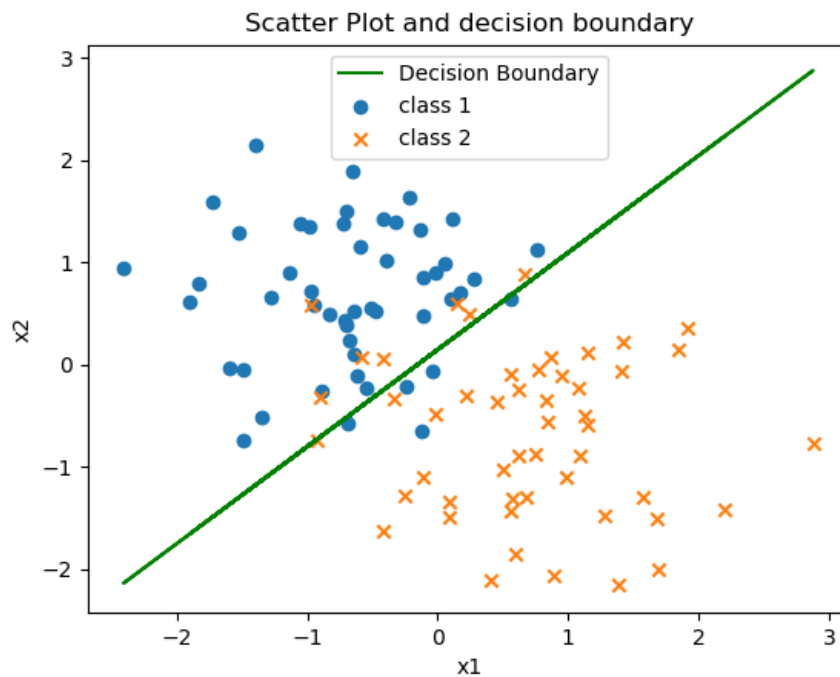
Stopping Criteria: absolute value of difference between the initial and final error

Theta Values(Final Parameters)

```
[[ 0.40125316]
 [ 2.5885477 ]
 [-2.72558849]]
```

No. of iterations:7

3.2 Graph plot



4 Gaussian Discriminant Analysis

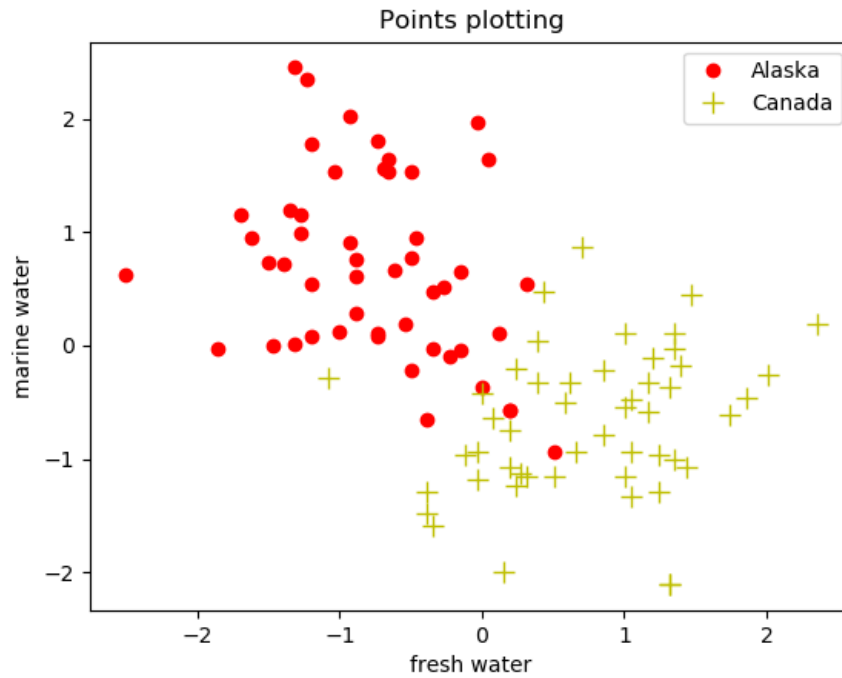
The data was first normalized and then the data was processed further

4.1 various values of the parameters:

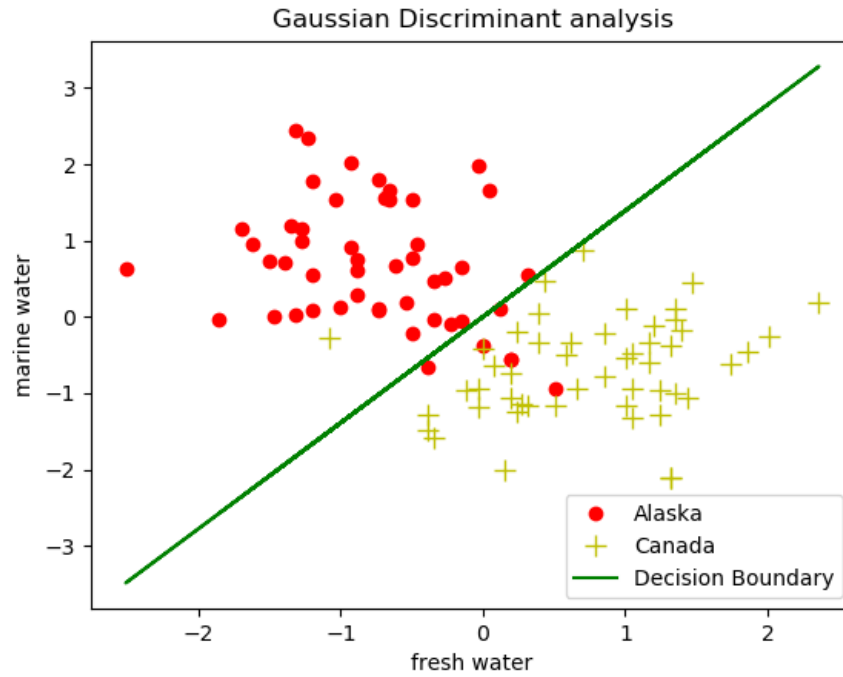
The values of the various parameters are given below:

- $\mu_0 = [-0.75529433 \ 0.68509431]$
- $\mu_1 = [0.75529433 \ -0.68509431]$
- $\Sigma = \begin{bmatrix} 0.42953048 & -0.02247228 \\ -0.02247228 & 0.53064579 \end{bmatrix}$
- $\phi = 0.5$

4.2 data plot for points



4.3 data plot for linear speration using same covariance matrix

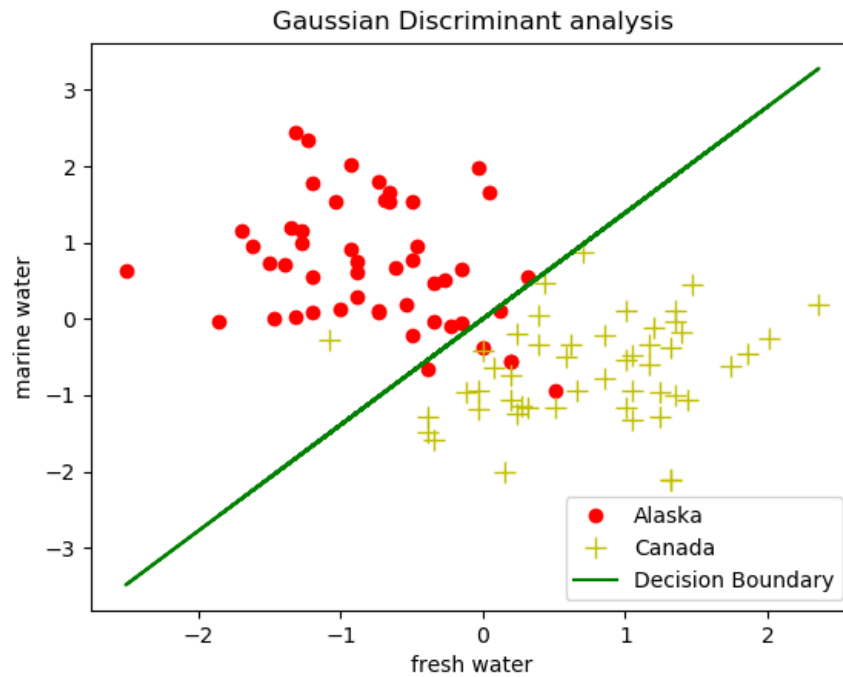


4.4 various values for different covariance matrix

The values of the various parameters are given below:

- $\mu_0 = [-0.75529433 \ 0.68509431]$
- $\mu_1 = [0.75529433 \ -0.68509431]$
- $\Sigma_0 = \begin{bmatrix} 0.38158978 & -0.15486516 \\ -0.15486516 & 0.64773717 \end{bmatrix}$
- $\Sigma_1 = \begin{bmatrix} 0.47747117 & 0.1099206 \\ 0.1099206 & 0.41355441 \end{bmatrix}$
- $\phi = 0.5$

4.5 data plot for linear separation using same covariance matrix



4.6 linear along with quadratic boundary

