COL-774: ASSIGNMENT-1

Question 1

Linear Regression
Learning rate chosen : 0.007
Stopping criteria : $max_j(\nabla_\Theta J(\Theta)) < \epsilon$ where $\epsilon = 10^{-8}$
Parameters :

$$\Theta = \begin{bmatrix} 0.99662009 \\ 0.0013402 \end{bmatrix}$$

It has been observed that for $\eta = 0.021$ and $\eta = 0.025$ the gradient descent algorithm diverges. For rest of the learning rate values the algorithm converges to the optimum value.

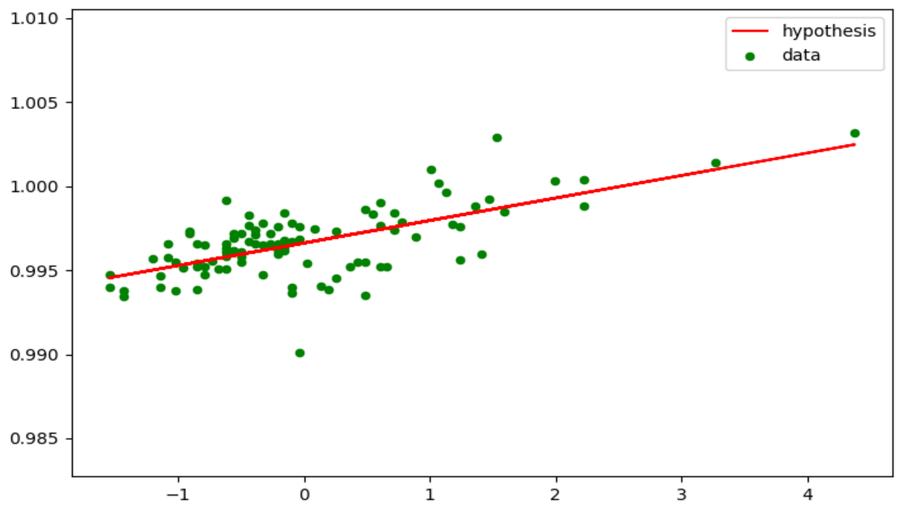

Figure 1.1 Linear regression using gradient descent.

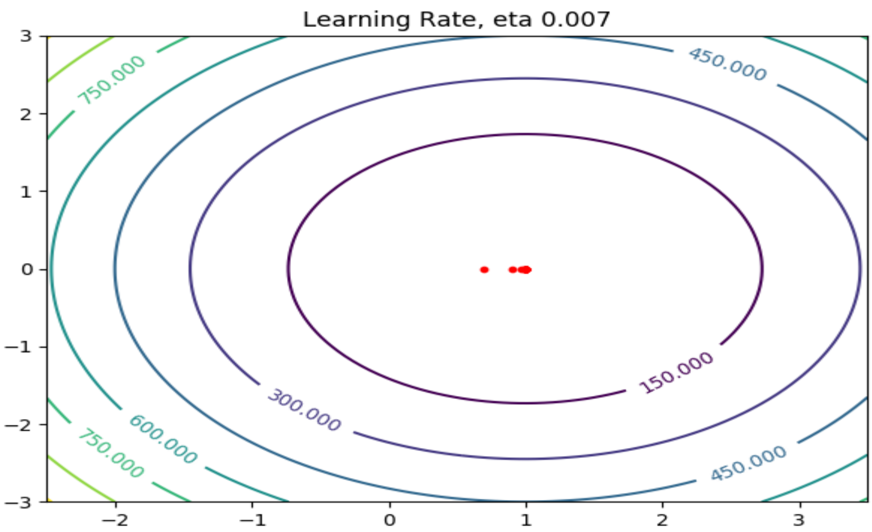

Figure 1.2 Contour plot for $\eta = 0.007$.

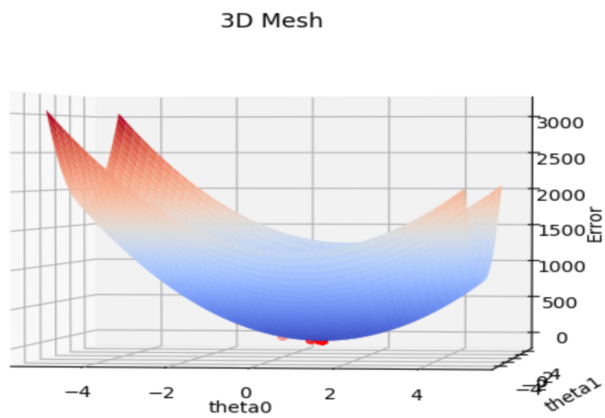

Figure 1.3 3D Surface plot.

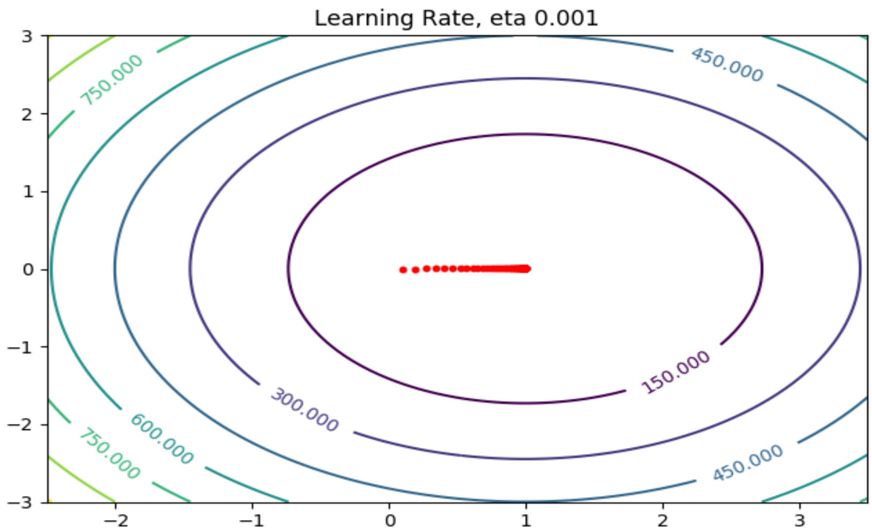

Figure 1.4 Contour plot for $\eta = 0.001$.

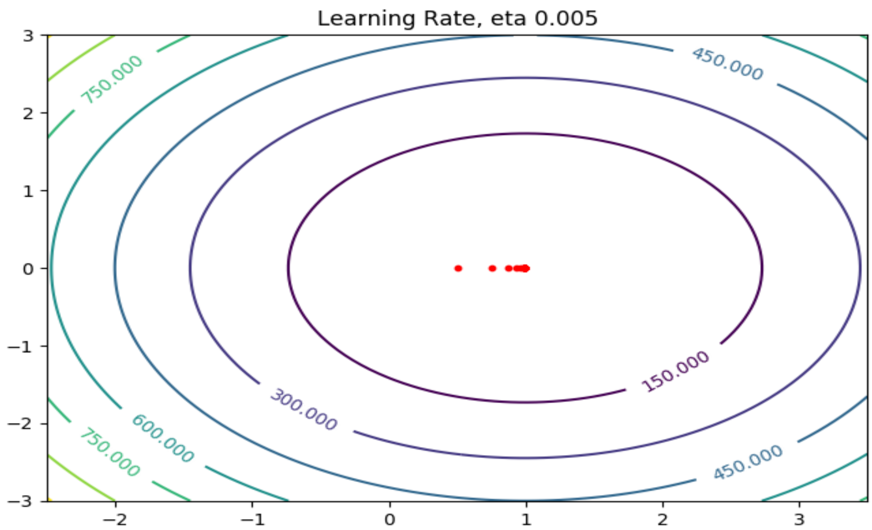

Figure 1.5 Contour plot for $\eta = 0.005$.

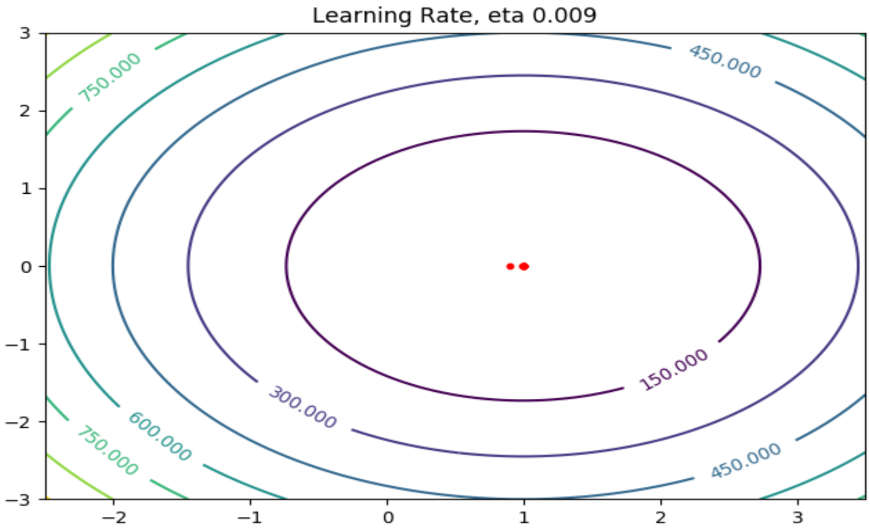

Figure 1.6 Contour plot for $\eta = 0.009$.

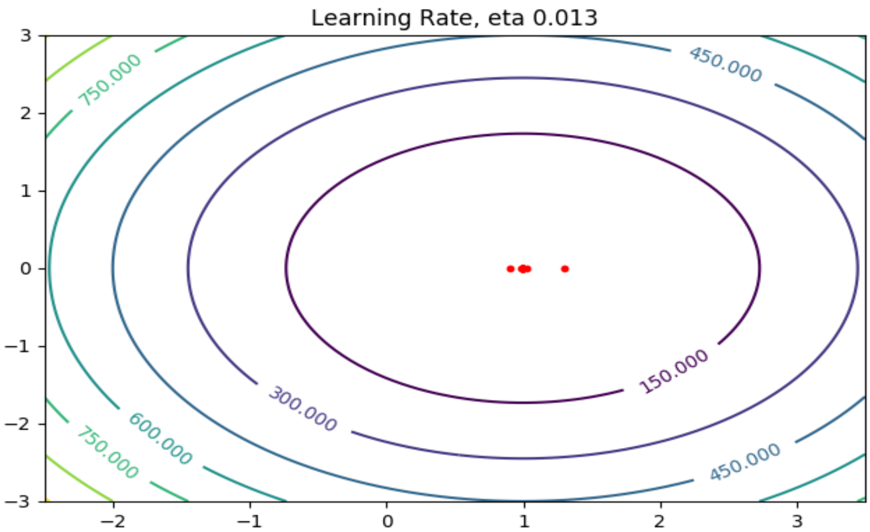

Figure 1.7 Contour plot for $\eta = 0.013$.

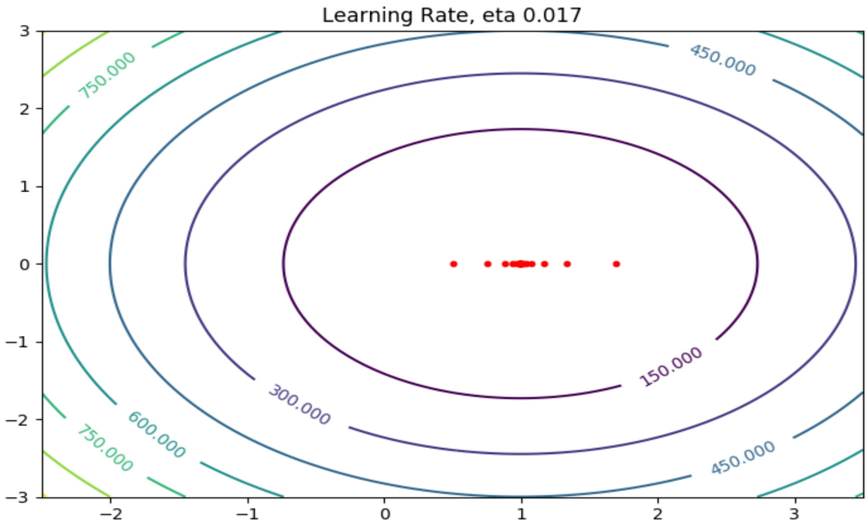

Figure 1.8 Contour plot for $\eta = 0.017$.

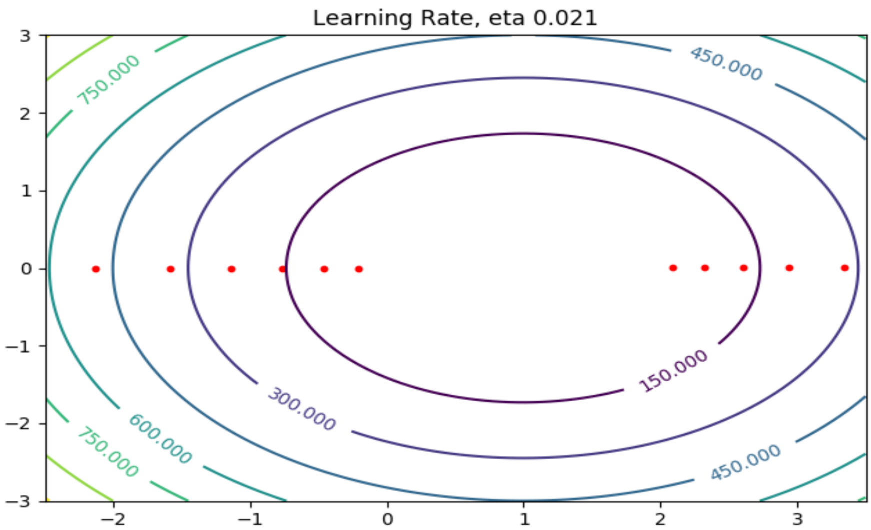

Figure 1.9 Contour plot for $\eta = 0.021$.

Question 2

Weighted Linear regression implementation

It has been observed that too small value of τ results in over-fitting whereas too large value results in under-fitting. I think among all the values that have been tested, 0.3 works best for the given dataset.

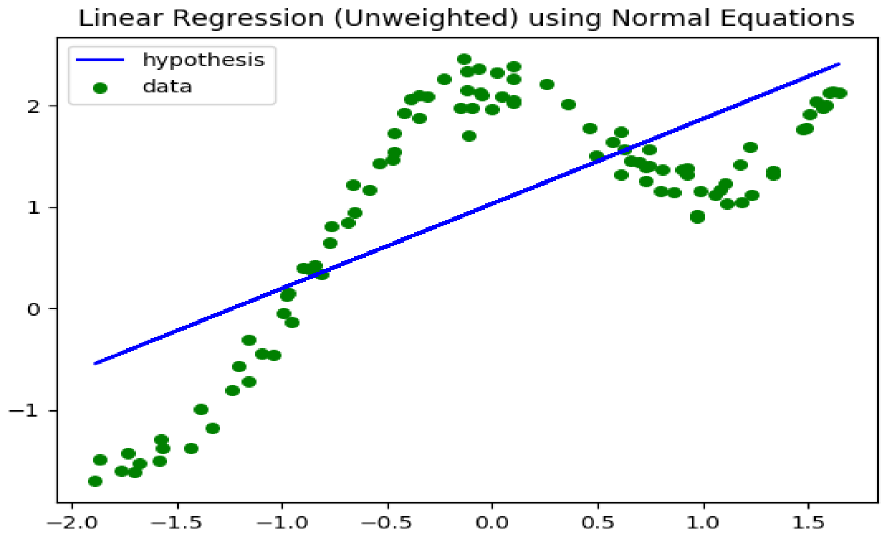

Figure 2.1 Linear regression (Unweighted) using normal equation.

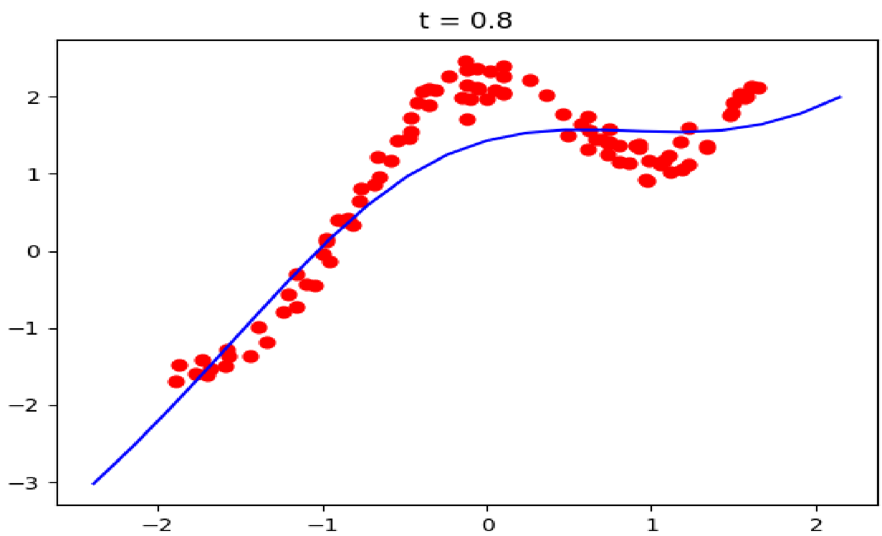

Figure 2.2 Linear regression (Weighted) with bandwidth = 0.8.

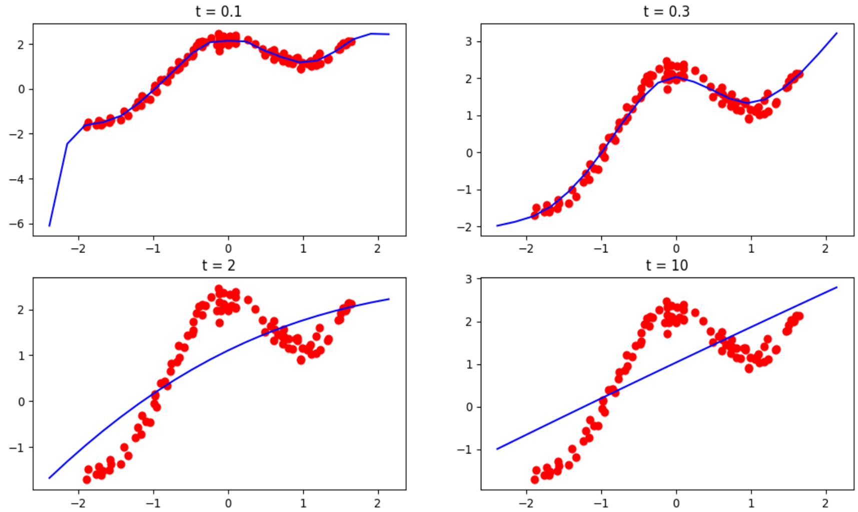

Figure 2.3 Linear regression (Weighted) with different values of bandwidth.

Question 3

Newton's Method implementation
The values of parameters obtained are:

$$\Theta = \begin{bmatrix} 0.40125316 \\ 2.5885477 \\ -2.72558849 \end{bmatrix}$$

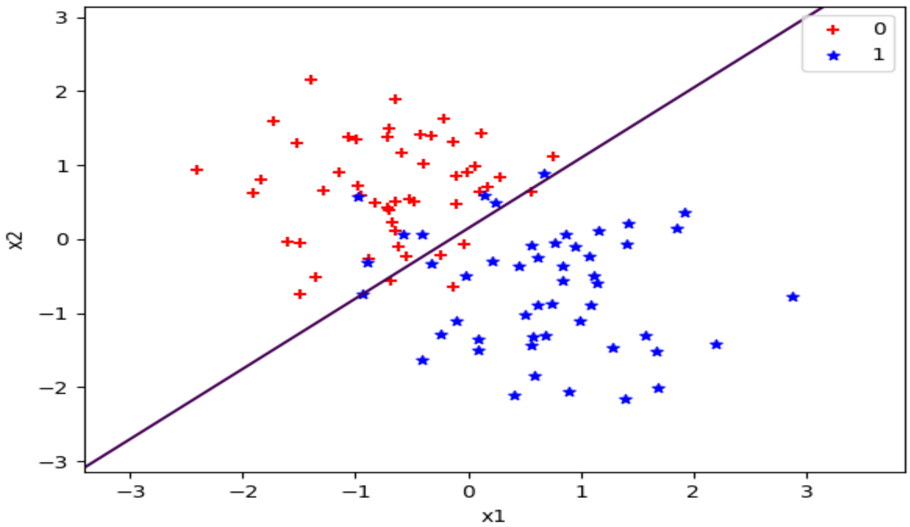

Figure 3.1 Logistic regression optimized using Newton's Method.

Question 4

Gaussian Discriminant Analysis

Phi

$$\phi = 0.5$$

Mean of distribution corresponding to Canada class

$$\mu_0 = \begin{bmatrix} 0.75529433 \\ -0.68509431 \end{bmatrix}$$

Mean of distribution corresponding to Alaska class

$$\mu_1 = \begin{bmatrix} -0.75529433 \\ 0.68509431 \end{bmatrix}$$

Covariance Matrix

$$\Sigma = \begin{bmatrix} 0.42953048 & -0.02247228 \\ -0.02247228 & 0.53064579 \end{bmatrix}$$

Covariance Matrix corresponding to Canada class

$$\Sigma_0 = \begin{bmatrix} 0.47747117 & 0.1099206 \\ 0.1099206 & 0.41355441 \end{bmatrix}$$

Covariance Matrix corresponding to Alaska class

$$\Sigma_1 = \begin{bmatrix} 0.38158978 & -0.15486516 \\ -0.15486516 & 0.64773717 \end{bmatrix}$$

Observation

The decision boundary satisfies the following equation:

$$\frac{p(y=1|x;\theta)}{p(y=0|x;\theta)} = 1$$

On simplifying the above equation, we get,

$$\frac{1}{2}x^T(\Sigma_1^{-1} - \Sigma_0^{-1})x - x^T(\Sigma_1^{-1}\mu_1 - \Sigma_0^{-1}\mu_0) + \frac{1}{2}(\mu_1^T\Sigma_1^{-1}\mu_1 - \mu_0^T\Sigma_0^{-1}\mu_0) - log(\frac{\phi}{1-\phi}) + \frac{1}{2}log(\frac{|\Sigma_1|}{|\Sigma_0|}) = 0$$

This is the quadratic equation of the form $Ax_0^2 + Bx_1^2 + Cx_0 + Dx_1 + E = 0$ which is the equation of hyperbola in this case as the coefficient of degree 2 term that is either A or B is negative.
From the plot, it is observed that the quadratic boundary separates the two classes better.

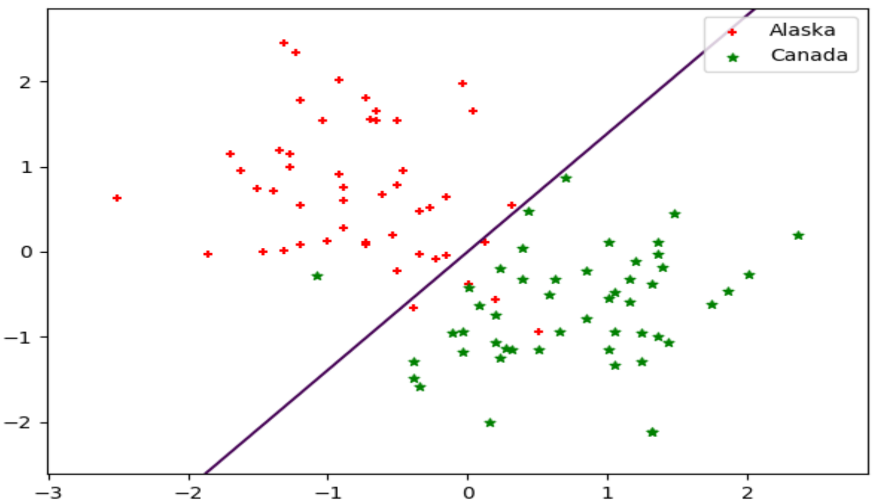

Figure 4.1 Linear separator

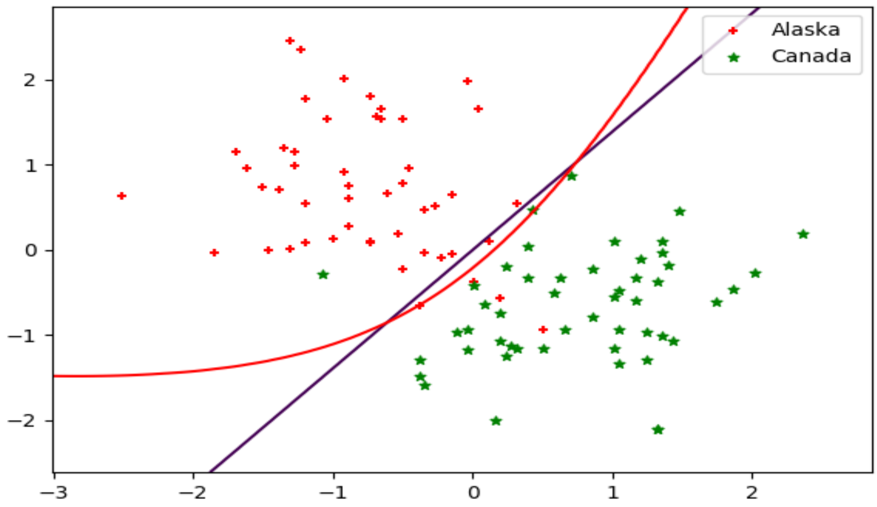

Figure 4.2 Linear and Quadratic separator