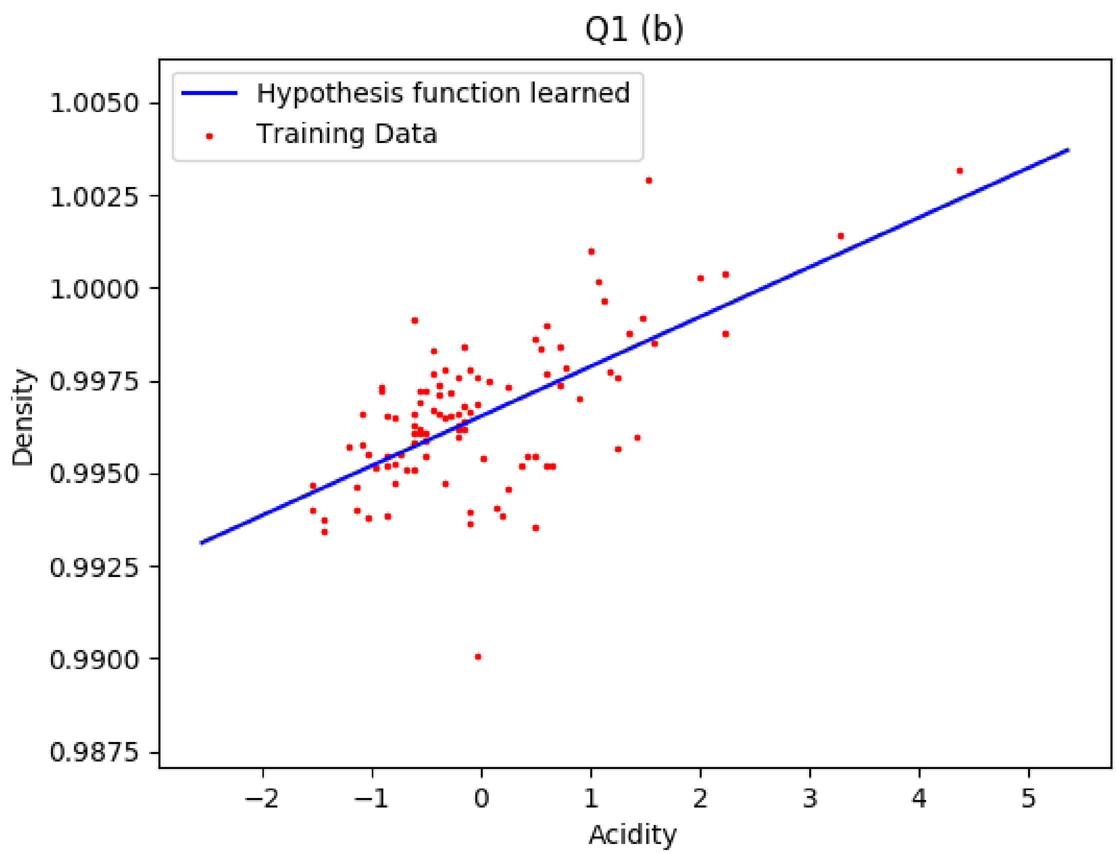
COL774 Assignment-1

Q1
Part(a)

Learning rate: tried many values and chose η=**0.001**
Stopping criteria: If value of cost in 2 consecutive iterations changes by less than ε=**0.0000001** then stop.
I takes **89** iterations to converge at η=**0.001**

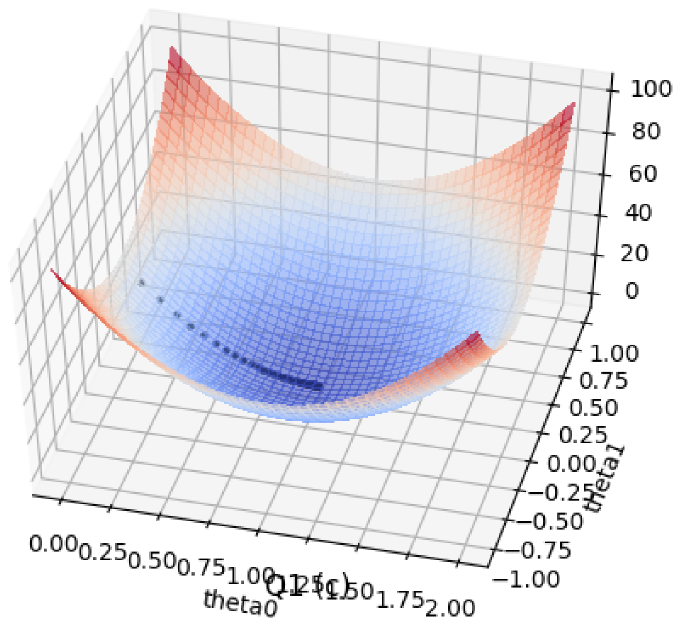
θ comes out to be [0.99653574 0.00134008]

Part(b)
Plotted training data and hypothesis function learned.

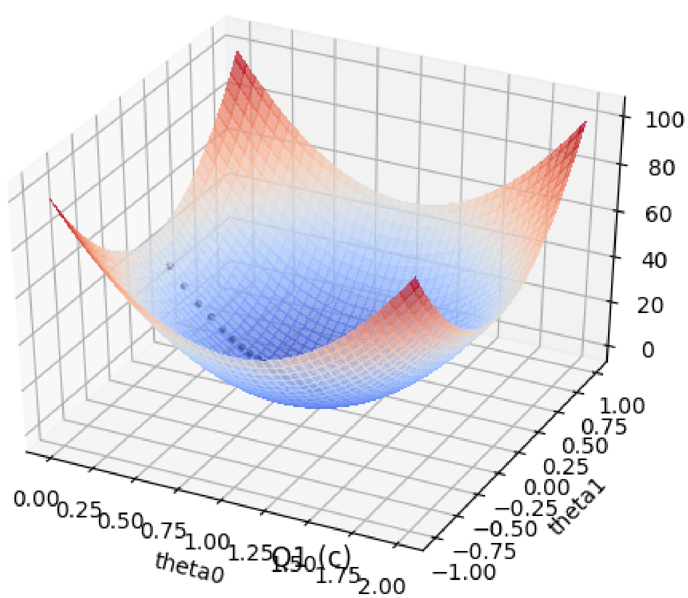


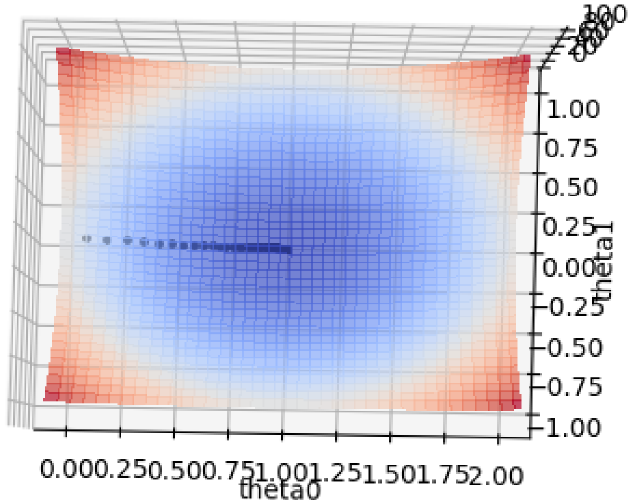
Part(c)

Plotted J(θ) and then plotted points at each iteration of gradient decent.



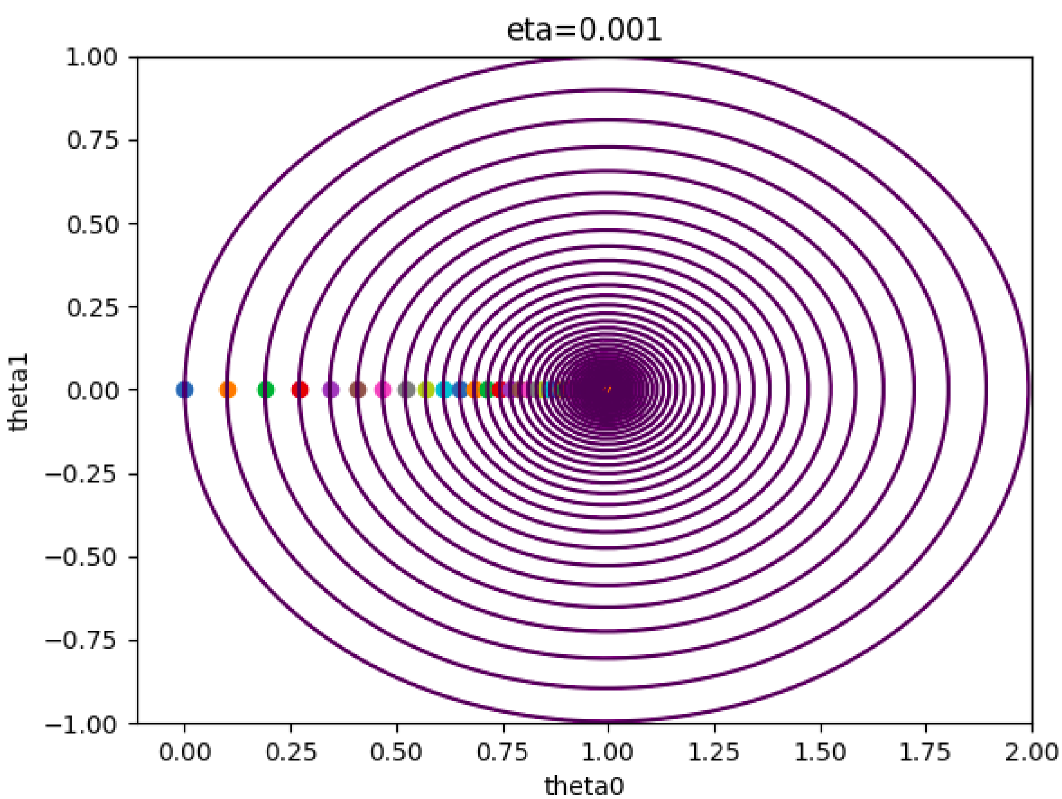
Plots of J(θ) vs θ

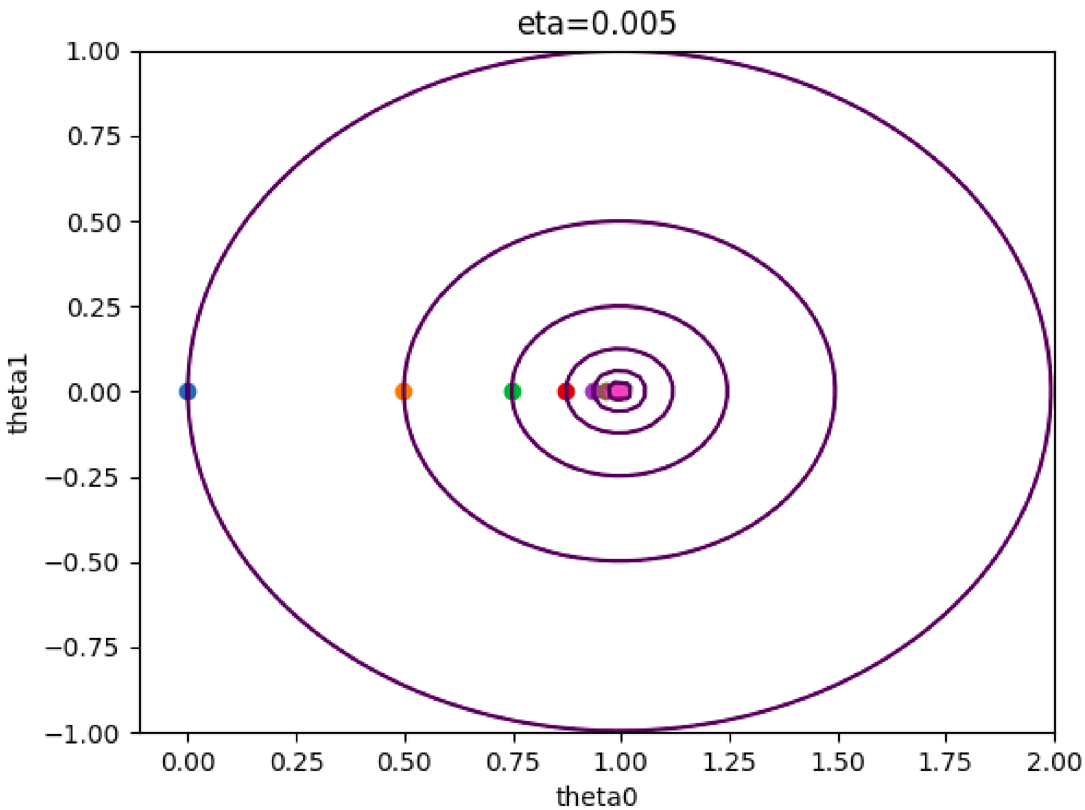


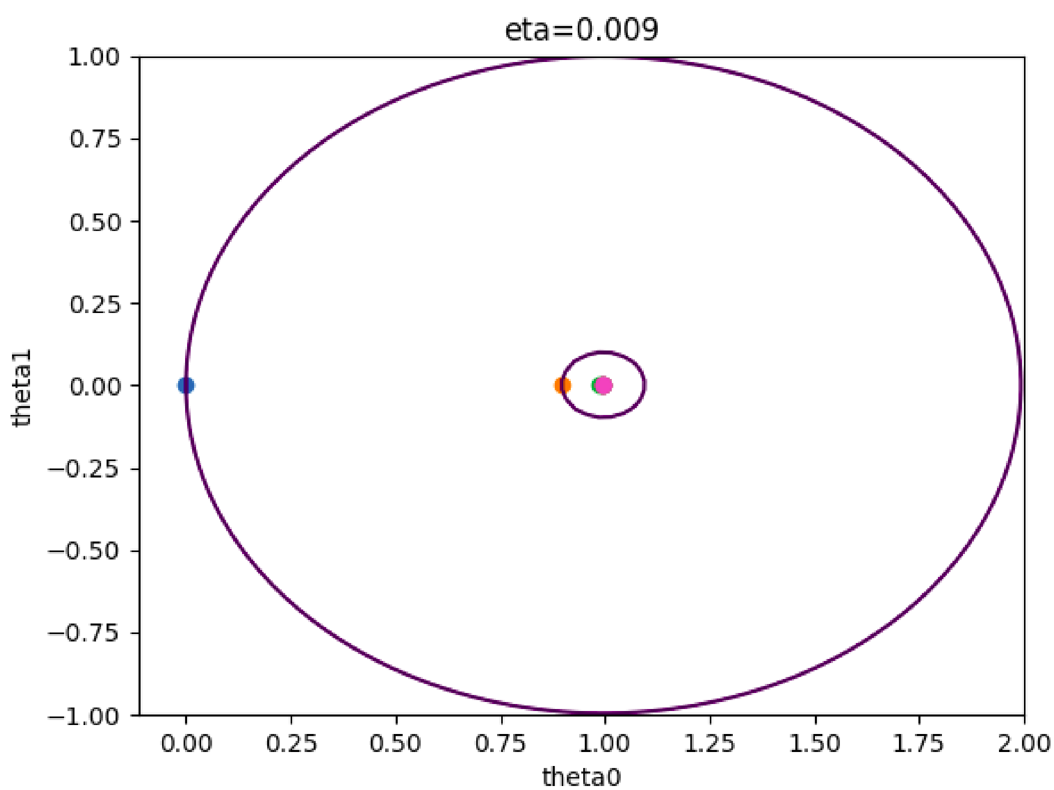


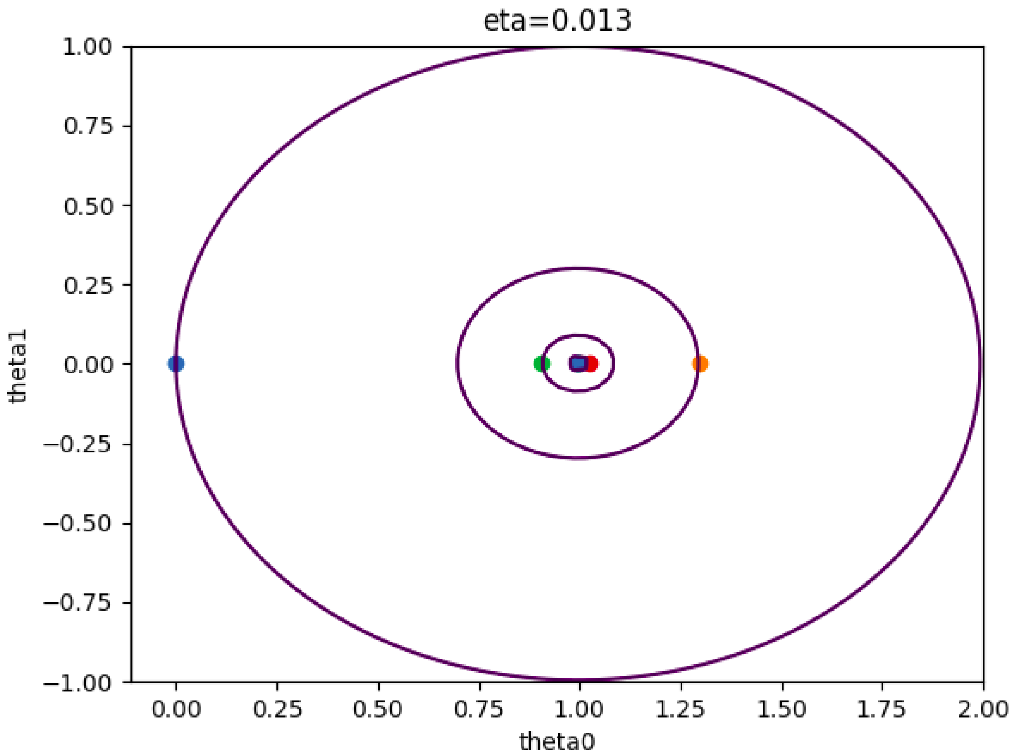
Q1 (c)

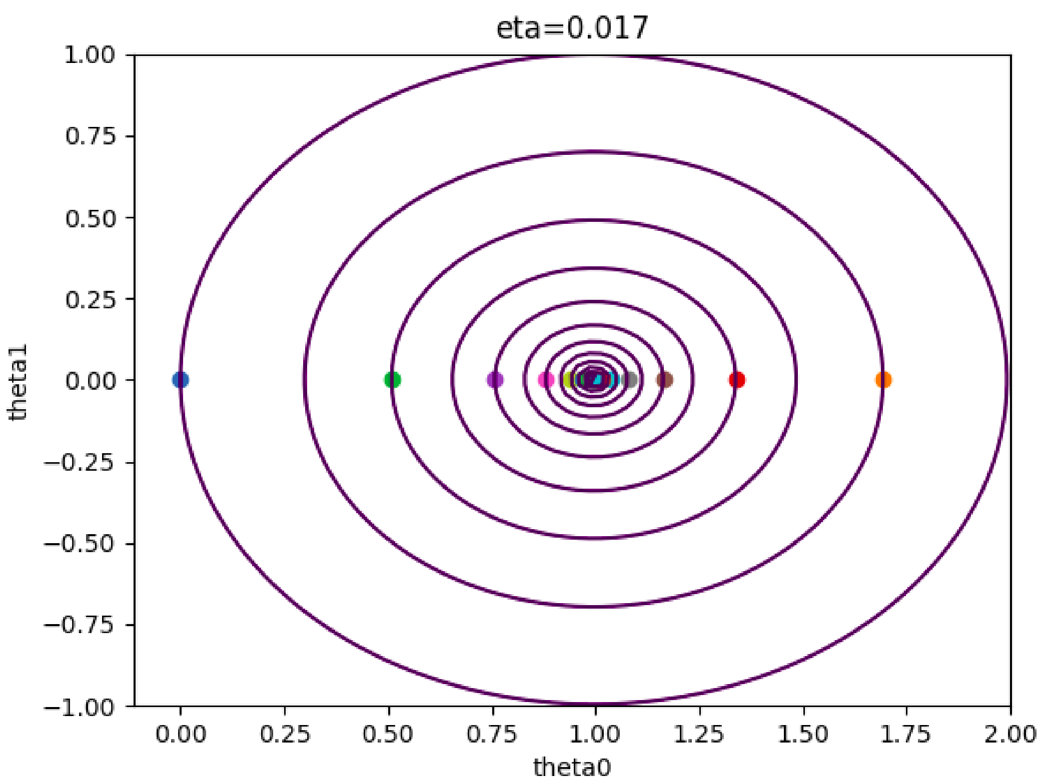
Part(d)
Plotted contour for cost function at each iteration.




eta=0.005
theta1
theta0
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


eta=0.009
theta1
theta0
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


eta=0.013
theta1
theta0
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


eta=0.017
theta1
theta0
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

Part(e)
The step size increases from η=0.001,0.005,0.009 and it take less iterations to converge.

Algorithm starts oscillating initially when η=0.013

For η=0.013 and 0.017 it oscillates initially but converges eventually.

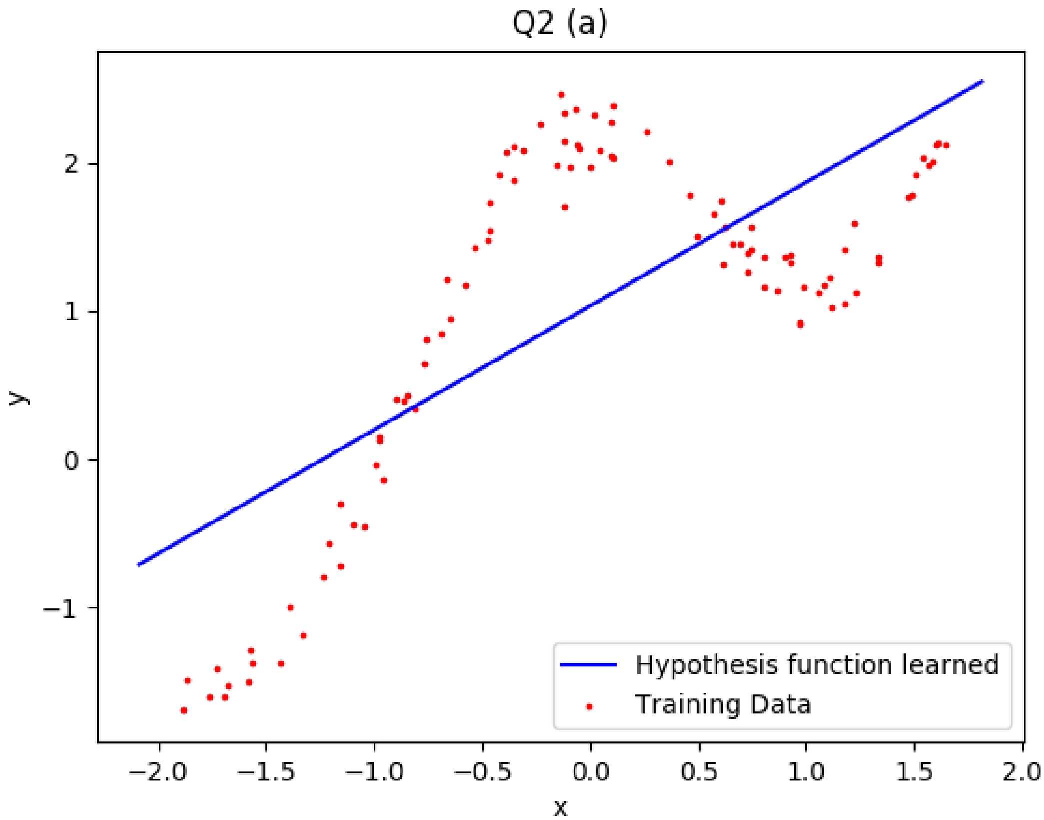
At η=0.021 and 0.025 it keeps diverging and does not converge.

Q2

Part(a)

Calculated θ using :

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



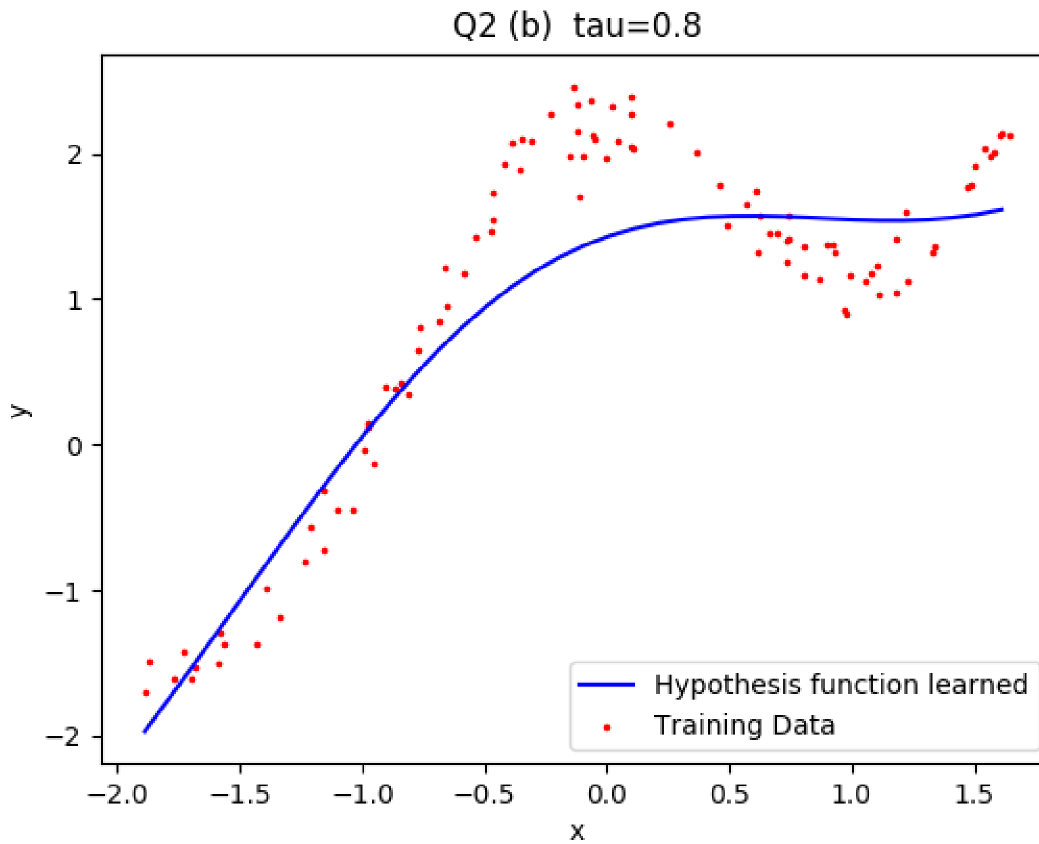
Part(b)

Calculated θ for each query point x where $\theta^T x$ has to be calculated using:

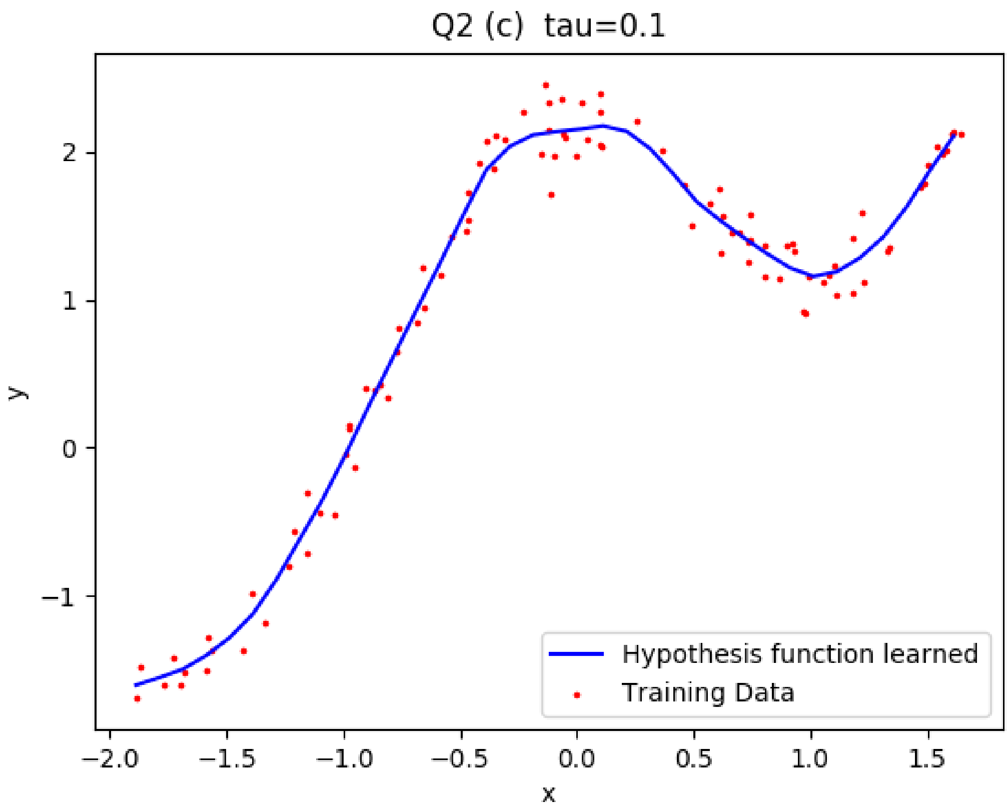
$$\theta = ((X^T W X)^{-1}) X^T W Y$$

Where W is a m*m diagonal matrix, W is calculated for each query point x

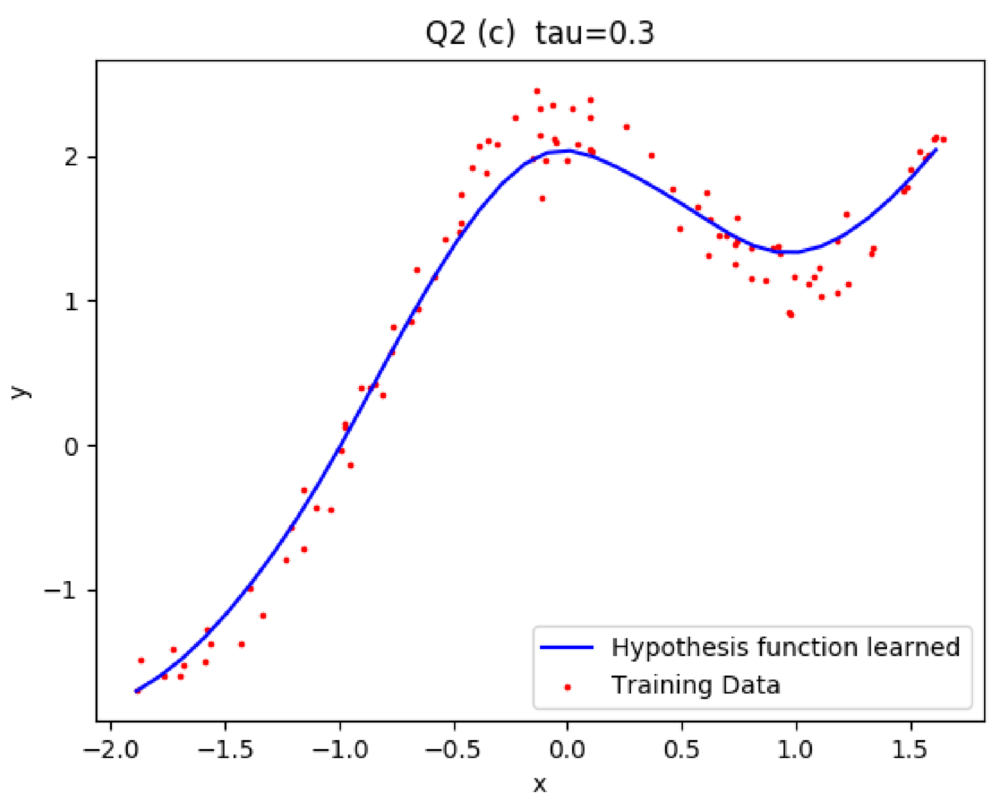
$$W_{ii}(x) = exp(-\frac{(x - x^{(i)})^2}{2\tau^2}) \qquad 0 \leq i < m$$



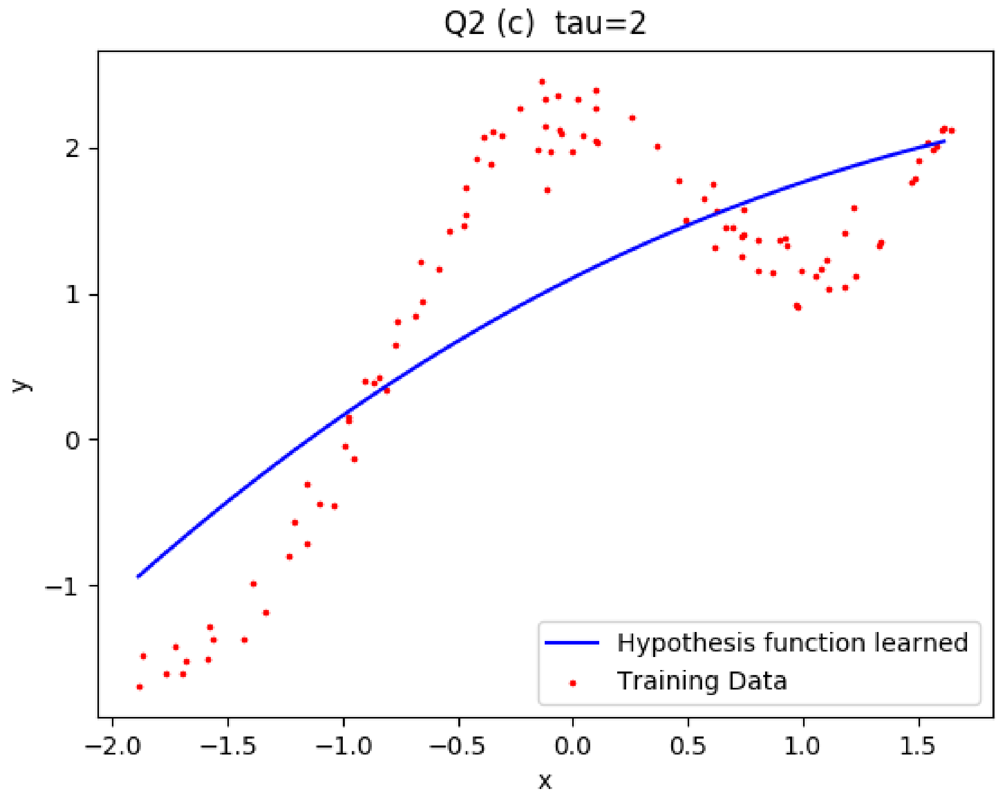
Part(c)
Tried tau={0.1,0.3,0.8,2,10} and plotted the hypothesis function learned.

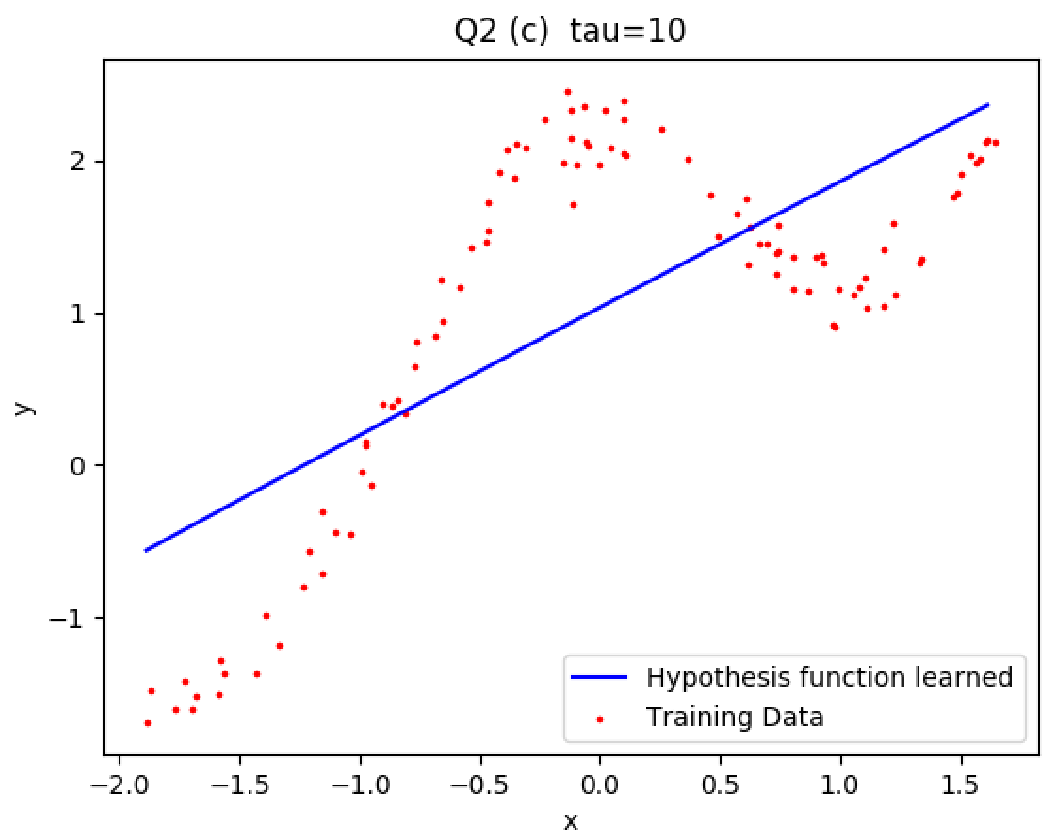


Q2 (c) tau=0.1
Hypothesis function learned
Training Data
x
y



Q2 (c) tau=0.3
Hypothesis function learned
Training Data
x
y


Q2 (c) tau=2
Hypothesis function learned
Training Data
x
y
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
−1
0
1
2


Q2 (c) tau=10
Hypothesis function learned
Training Data
x
y
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
−1
0
1
2

Out of tau={0.1,0.3,0.8,2,10}, I think tau=0.3 is best, because there is no overfitting and no underfitting.

When tau is too small, tau=0.1 then overfitting happens.
When tau is too large, tau=10 then underfitting happens and the function learned is almost a straight line.
For tau=2,10 the curve does not represent the data at all.

Q3

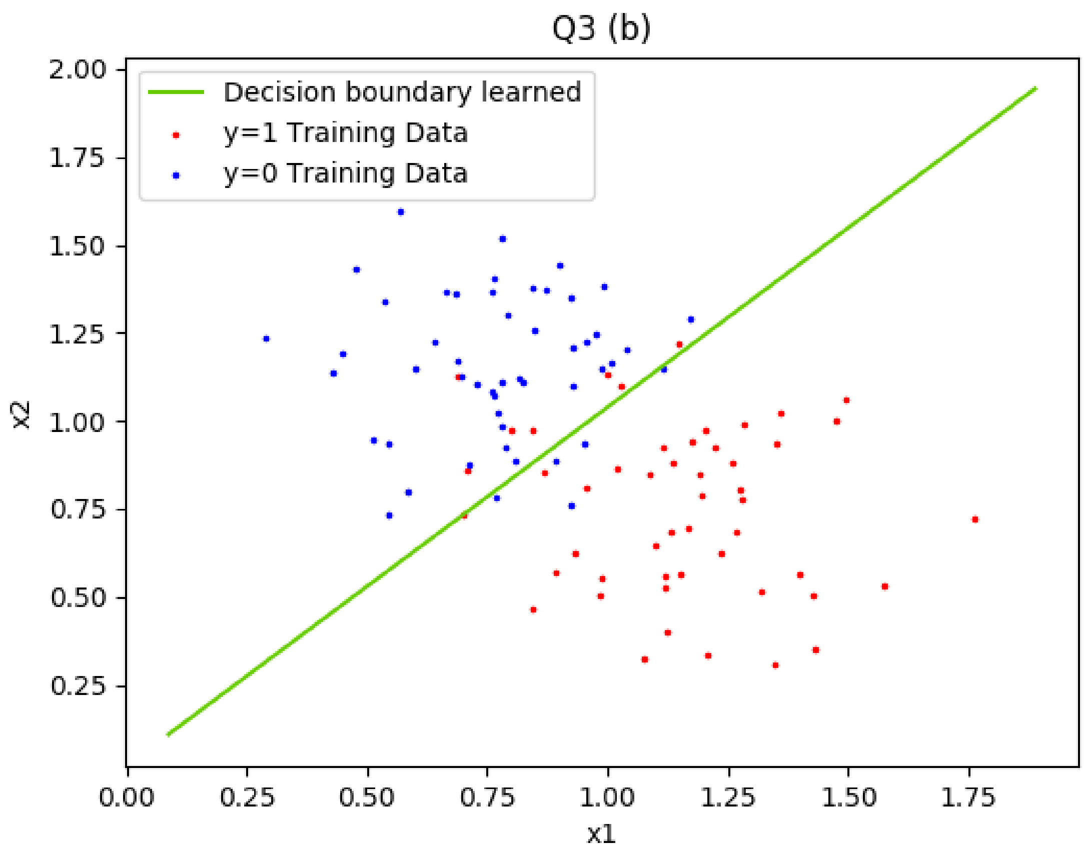
Part(a)
Implement Newton's method for optimizing L(θ).
Used Newton's method to find roots of L'(θ)

Convergence condition used: If L2 norm of difference between theta of 2 consecutive iterations is less than 0.000001 then stop

θ obtained is:
θ = [0.18002598 9.27432171 -9.10218919]

Part(b)

Plotted decision boundary learned with the training data.



Q4
Alaska is 0 and Canada is 1

Part (a)

Calculated μ0, μ1, ∑ using the formulas told in class.

μ0=[0.80651536 1.07283018]

μ1=[1.14318082 0.91362299]

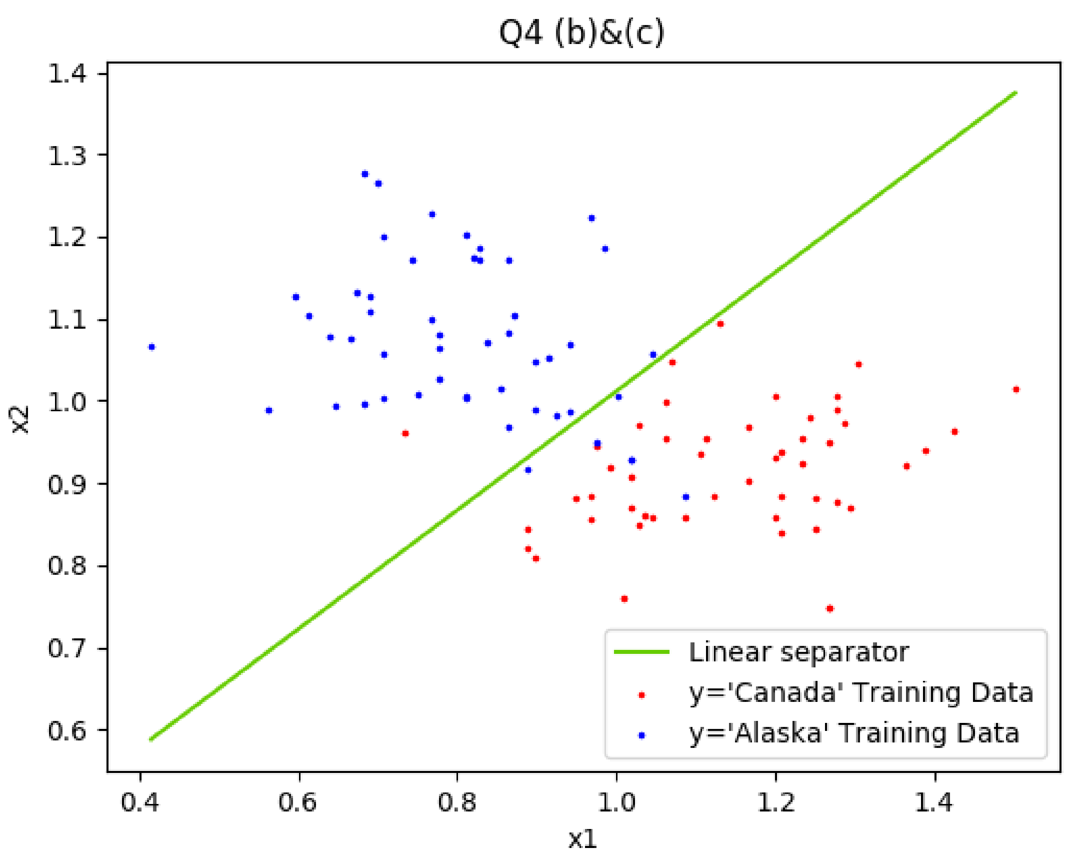
∑=[[0.0213353 -0.00058194]
 [-0.00058194 0.00716423]]

Part (b)&(c)

When ∑0=∑1 the equation of separator is:

$$X^T\Sigma^{-1}(\mu_0-\mu_1)+(\mu_0-\mu_1)^T\Sigma^{-1}X+(\mu_1^T\Sigma^{-1}\mu_1-\mu_0^T\Sigma^{-1}\mu_0)=2log_e(\frac{\phi}{1-\phi})$$

This is the equation of a straight line, plotted the above line.



Part(d)

$\mu 0$ and $\mu 1$ are same as before.

$\mu 0$=[0.5 0.40325768 0.53641509]
$\mu 1$=[0.5 0.57159041 0.45681149]

Calculated $\sum 0$ and $\sum 1$ using the formula given in assignment pdf.

$\sum 0$=
[[0.01895402 -0.00401041]
[-0.00401041 0.00874507]]

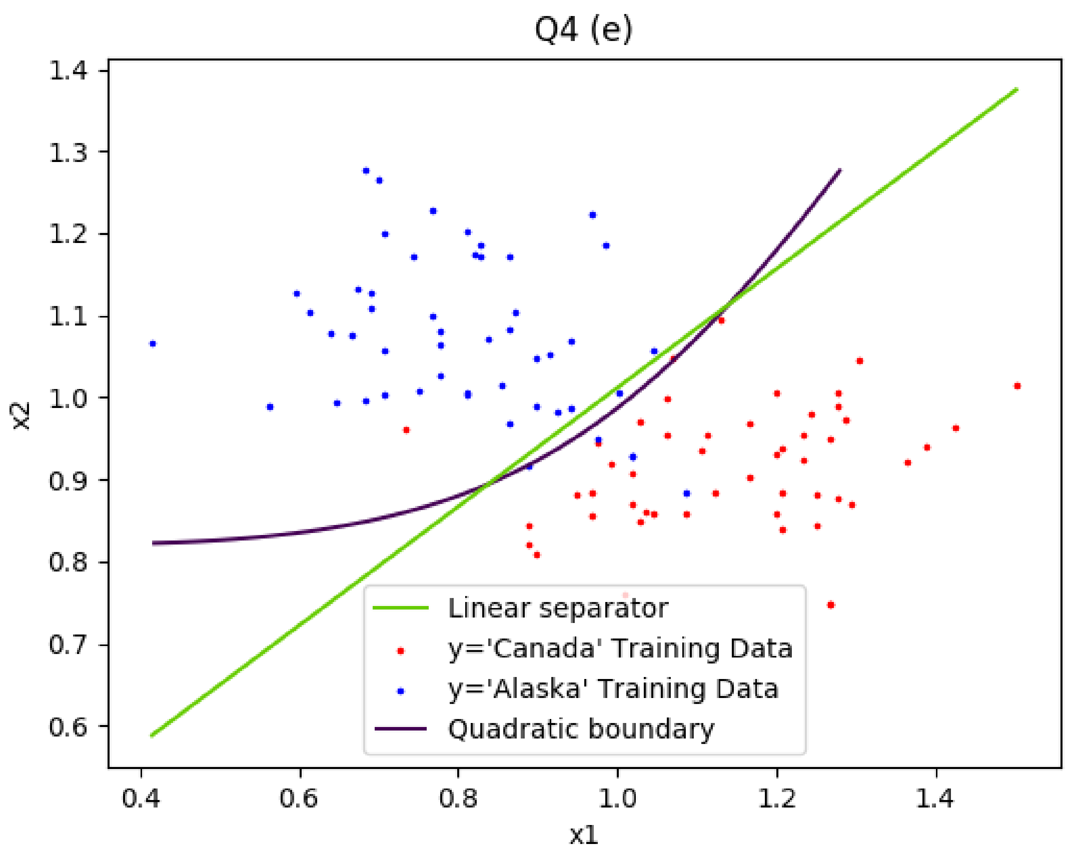
$\sum 1$=
[[0.02371657 0.00284652]
[0.00284652 0.00558338]]

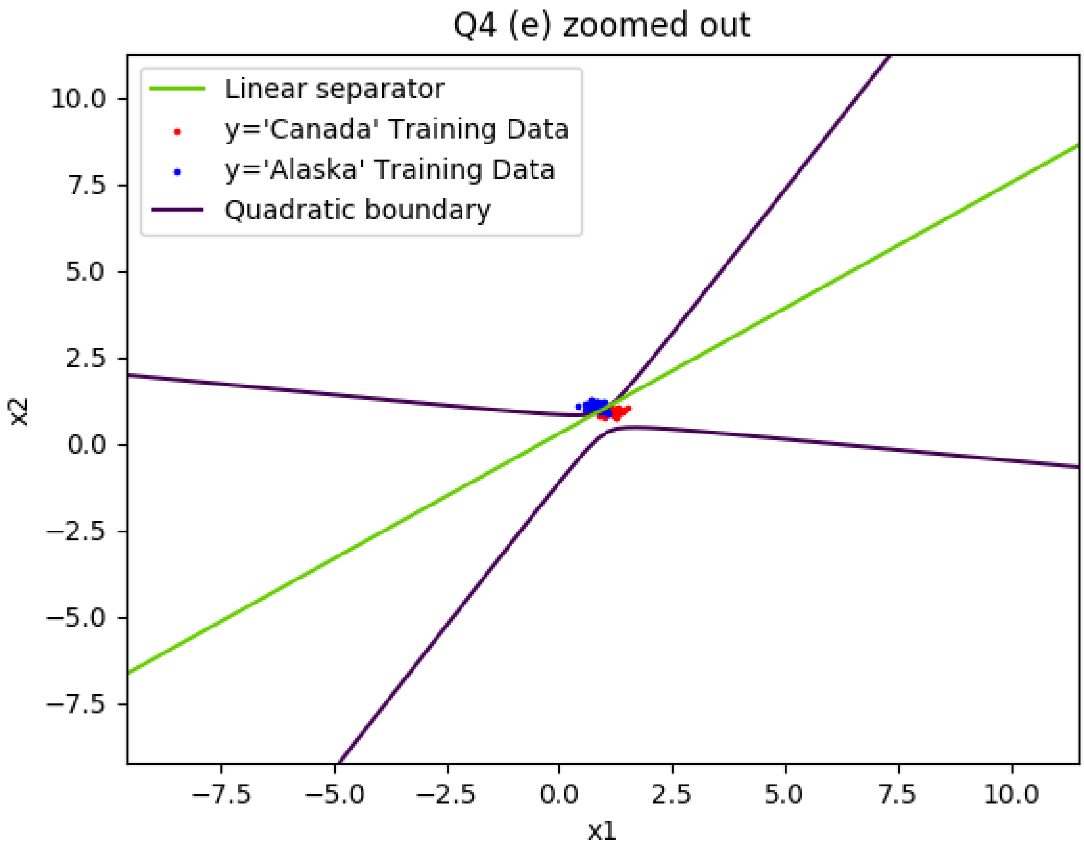
Part(e)

When $\sum 0 \neq \sum 1$ the equation of separator is:

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

Plotted quadratic separator as a contour of above equation at Z=0





Part(f)

When $\sum 0 = \sum 1$ the equation of separator is a straight line because the quadratic terms cancel.
When $\sum 0 \neq \sum 1$ the equation of separator is quadratic in x1 and x2.

The quadratic separator is actually a hyperbola, and it intersects linear separator at two points.

The linear separator classifies few 'Alaska' points as 'Canada' , but quadratic separator bends and correctly classifies those points as 'Alaska'.