COL 774 - Assignment Writeup

Amartya Chaudhuri
2014MCS2117

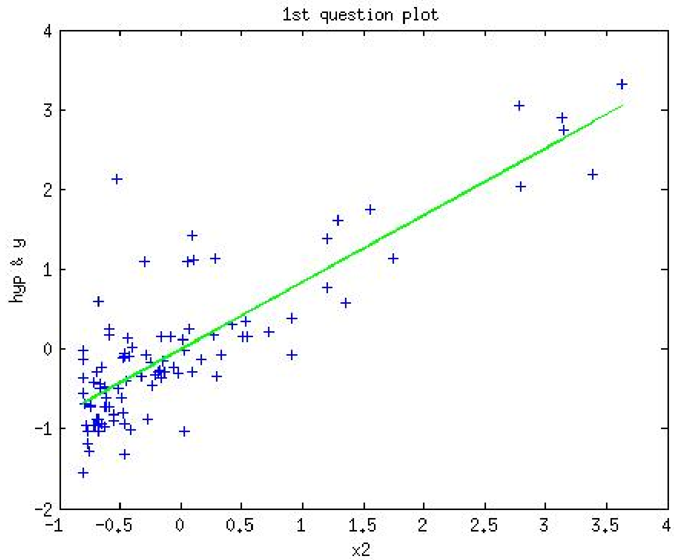
1. **Batch Gradient Descent**
   Following values were observed in assignment 1
   a. On keeping **learning rate** (η) = 1.5 , **convergence condition** as $|J(\theta) - J'(\theta)| < 0.0000005$ (two successive $J(\theta)$ values) upto a maximum of 5000 iterations, **parameter values** are coming to be [θ_0= 5.8393 , θ_1 = 4.6170] .

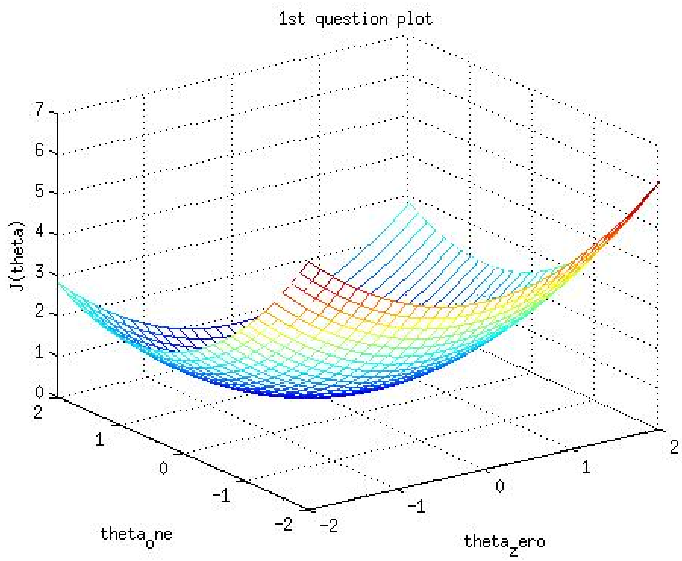
   e. changing learning parameters caused a change in no. of iterations were observed

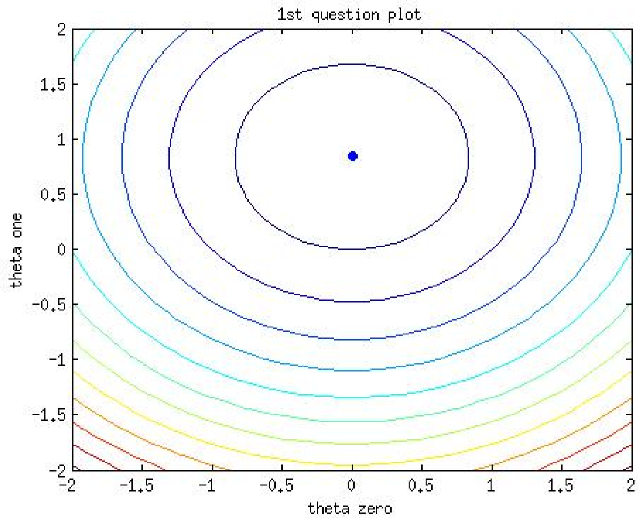
   η= 0.1 , no. of iterations = 80
   η= 0.5 , no. of iterations = 16
   η = 0.9 , no. of iterations = 6
   η = 1.3 , no of iterations = 10
   η = 2.0 , no of iterations = NA as gradient diverges and it never came to the optima.

So we see a decrease in no. of iterations before convergence as we increase the η parameter but it increases again if increased beyond certain limit. The algorithm diverges and never reaches optimum when learning parameters > 2.

b. , c & d. Plots for 2d hypothesis , 3d mesh and contour for error function are given as follows


1st question plot
hyp & y
x2
4
3
2
1
0
-1
-2
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4


1st question plot
J(theta)
theta_one
theta_zero
7
6
5
4
3
2
1
0
2
1
0
-1
-2
-2
-1
0
1
2



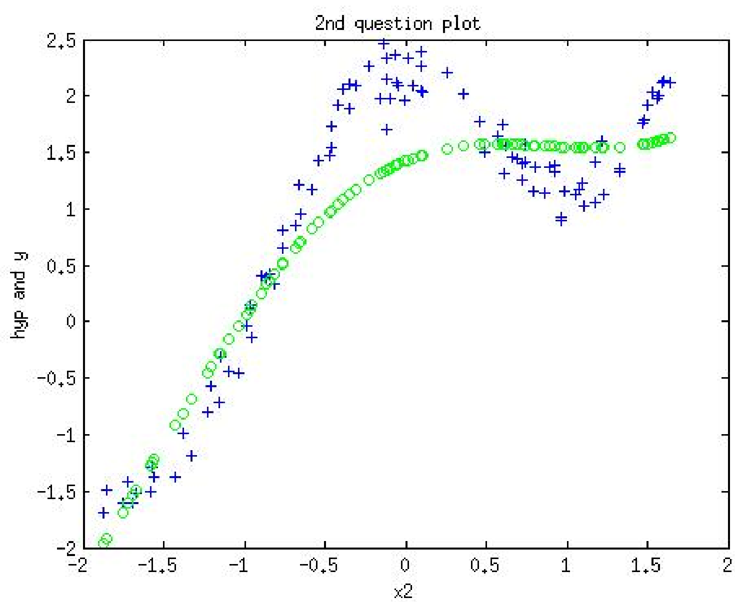
2. Locally weighted Logistic regression

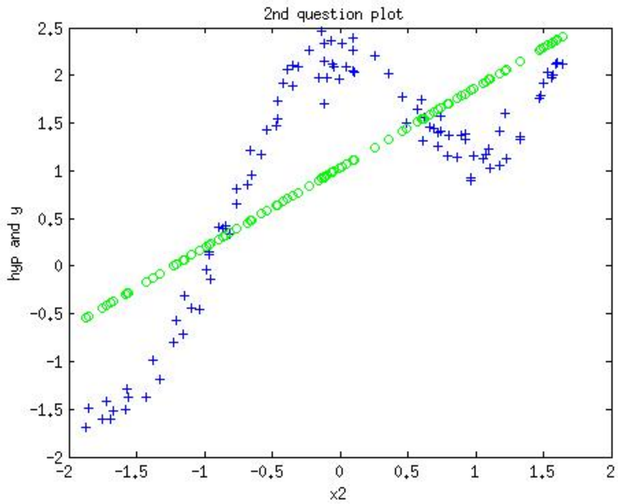
b &c. Fit is best when value of τ is 2 as the hypothesis is neither a gross overfit or underfit of the training data.

With small τ values , each training point is given weightage and the hypothesis is grossly overfitting the training data. As τ increases , the hypothesis curve become more biased and less overfitten . When increased beyond a level , the curve underfits the training data.

Plot of the hypothesis with data was as follows

τ value was taken to be 0.8

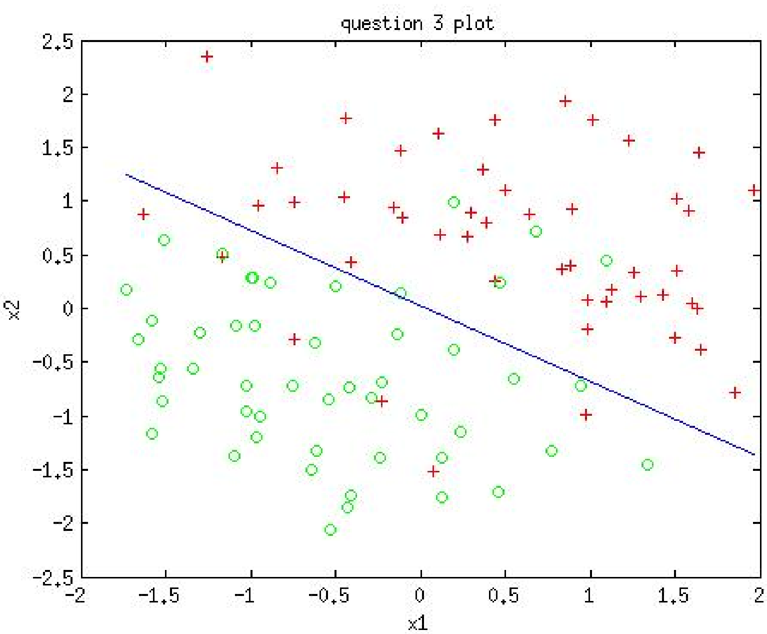




a. On fitting an unweighted linear hypothesis on the same data, we obtain the following curve (τ = 0.8)

3. Newton's method

theta values observed -
[θ_0 = -0.0472, θ_1 = 1.4675 , θ_2 = 2.0764]



Plot of the Newton's method is given above.

4. Gaussian discriminant analysis (GDA)

Following observations were made in assignment 4

a. Covariance (Σ)

1.5548	-1.0468
-1.0468	1.4547

mean of alaska (μ_0)

0.7515	-0.6817

mean of canada (μ_1)

-0.7515	0.6817

4. c Equation of the linear boundary was coming to be

$$x^T\Sigma^{-1}\mu_1 = \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1$$

where x , Σ , μ_1 are given in the question

4.d

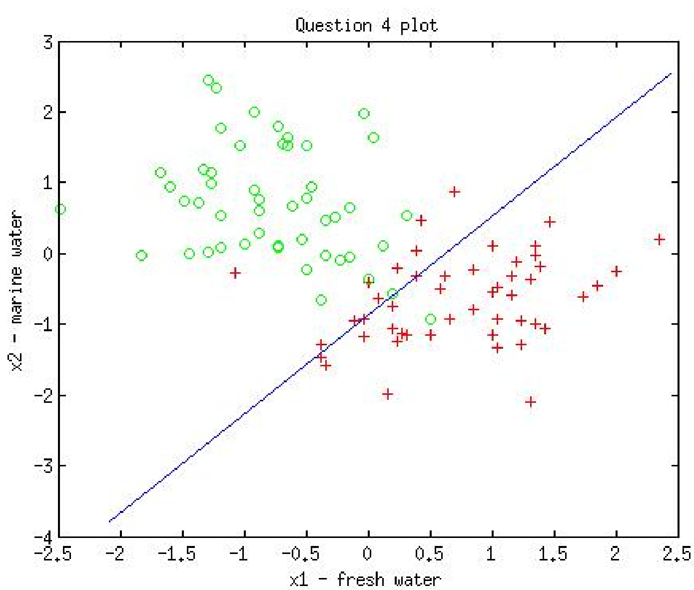
Covariance of alaska (Σ_0)

0.4727	0.1088
0.1088	0.4094

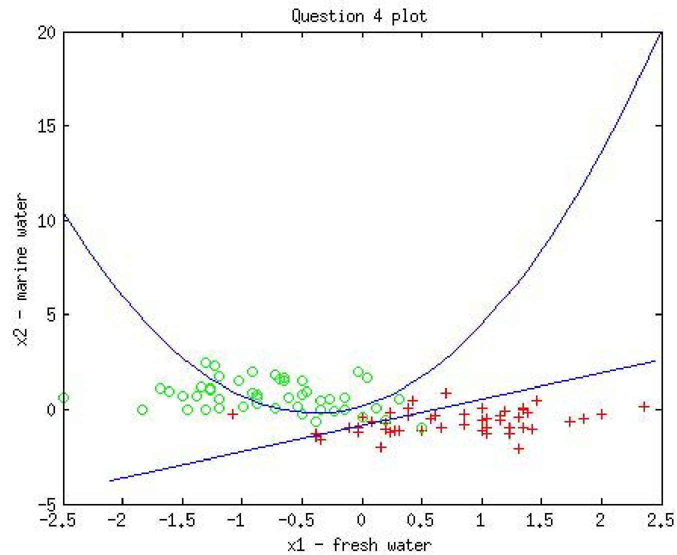
Covariance of canada (Σ_1)

0.3778	-0.1533
-0.1533	0.6413

4. b) Plot of linear boundary



4.e) Plot of quadratic boundary



Equation of the quadratic boundary

$$(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)-(x-\mu_1)^T\Sigma_2^{-1}(x-\mu_1)+log\frac{\Sigma_2}{\Sigma_1}=0$$

4.f) From the above diagram, the data is more linearly separable here and linear discriminant has a better fit . The quadratic separator is however more general with more bias and effective when the covariance of two classes of data differ.