ML-ASSIGNMENT 1 SUBMISSION

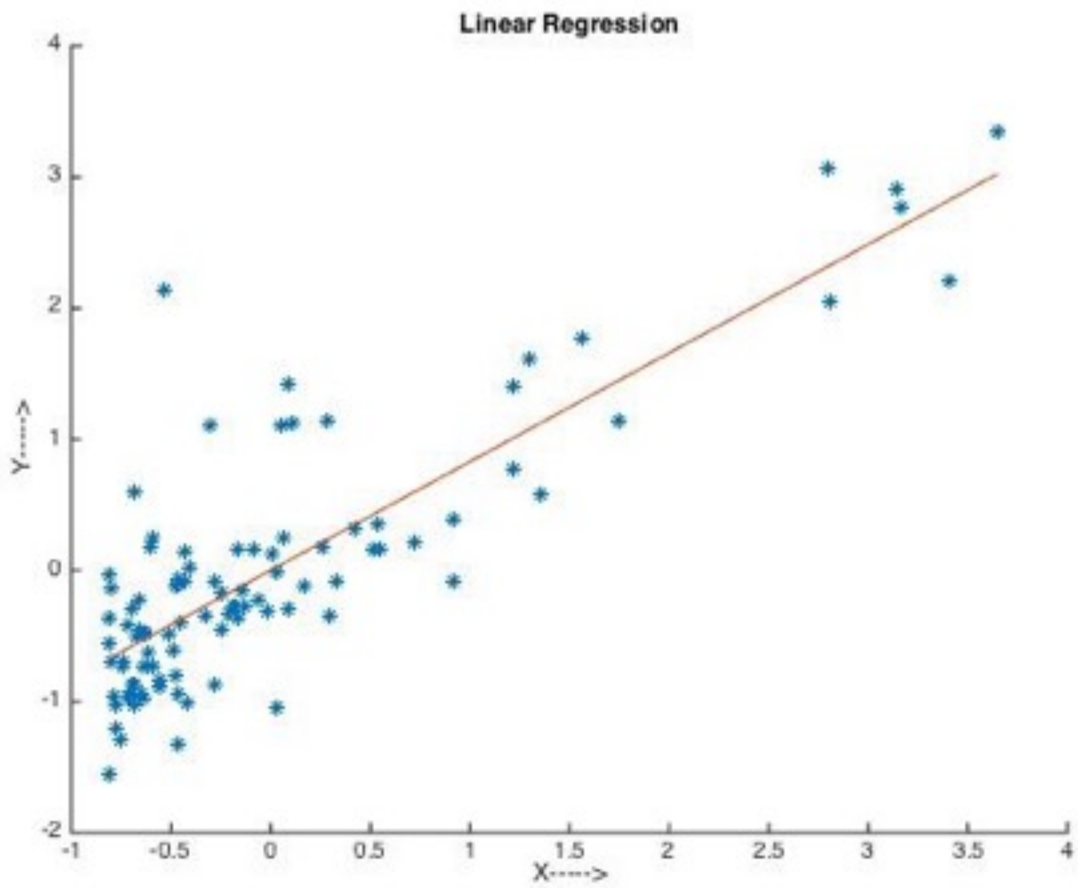
Q1 a.) The learning rate (eta)= 0.1

Stopping criteria = |J(Theta n)-J(Theta n-1)| < 0.00001

Theta1= 0.8288

Theta0=..801* 10^-15

b.) Check Fig 1-b

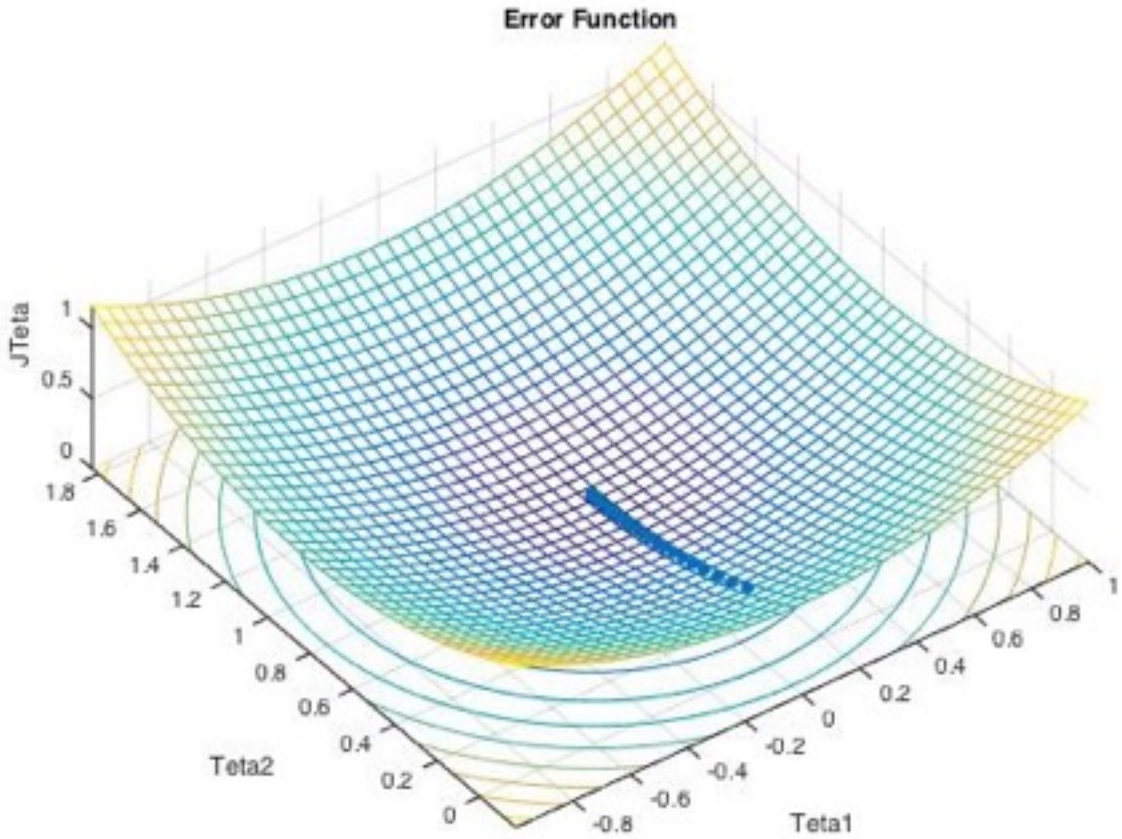


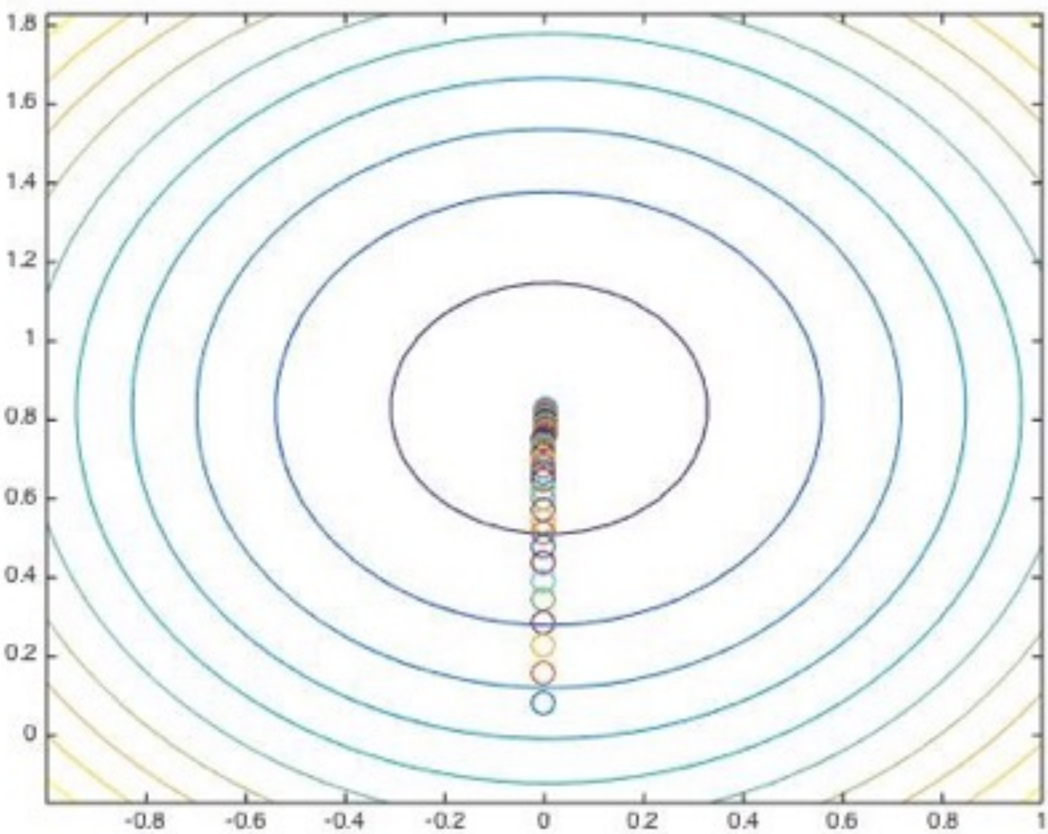
c.) Check Fig 1-c

d.) Check Fig 1-d

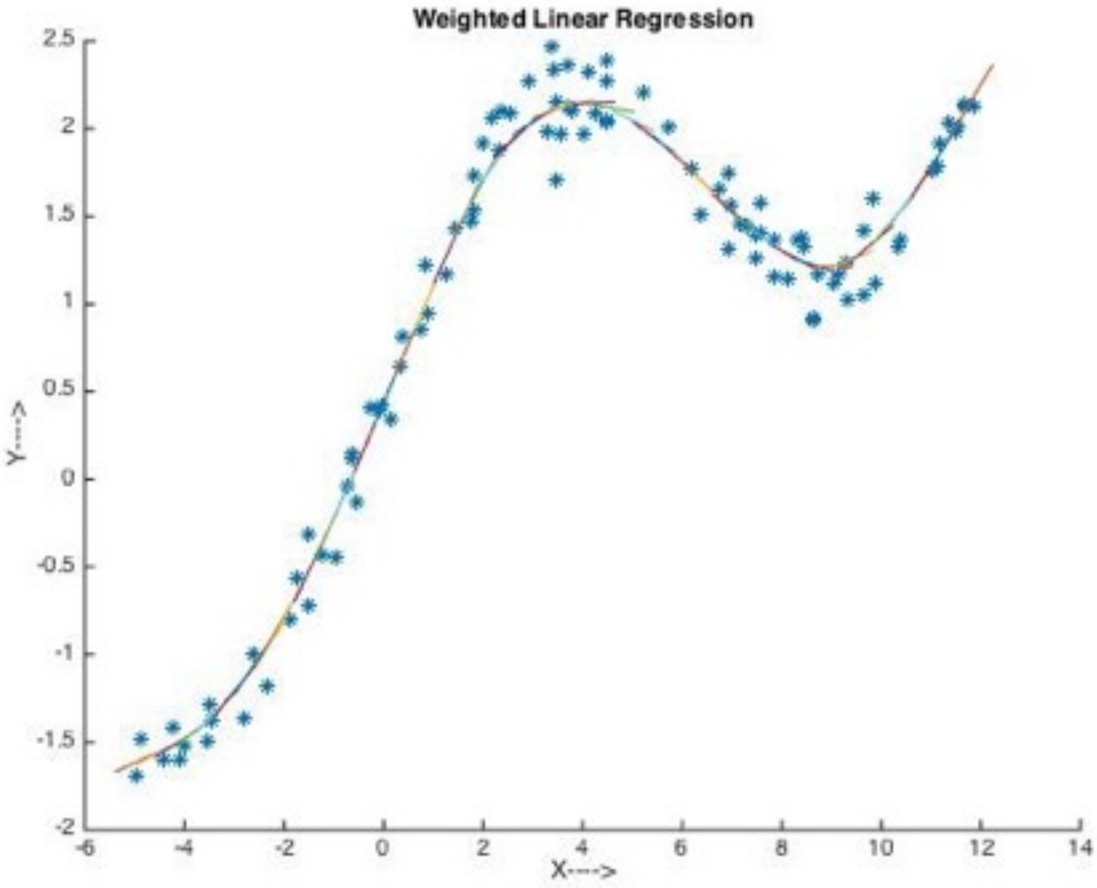
e.) The gradient converges at a rapid rate with increase in eta values until few values of eta like 0.1, 0.5, 0.9, 1.3.

For eta= 2.1, 2.5 there is no convergence.





Q2 a.) Fig 2-a

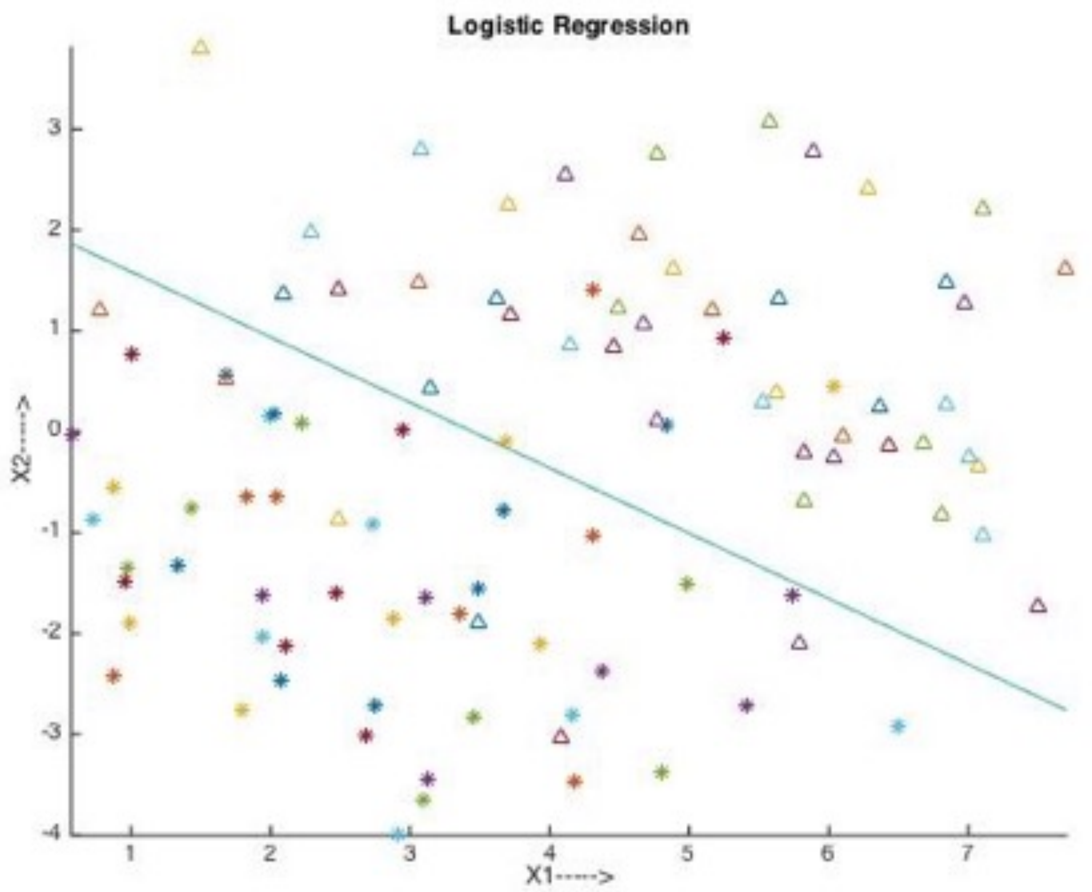


b.) -

c.) At small values of tou the line passes through almost all the points making the line a very bad predictor.

At bigger values of tou the weighted linear regression will behave similar to simple gradient descent model.
So, we must take an optimum value for tou.

Q3 a.) Theta=[-2.620511597178009; 0.760371535897073; 1.171946741565785] b.) Fig 3-b

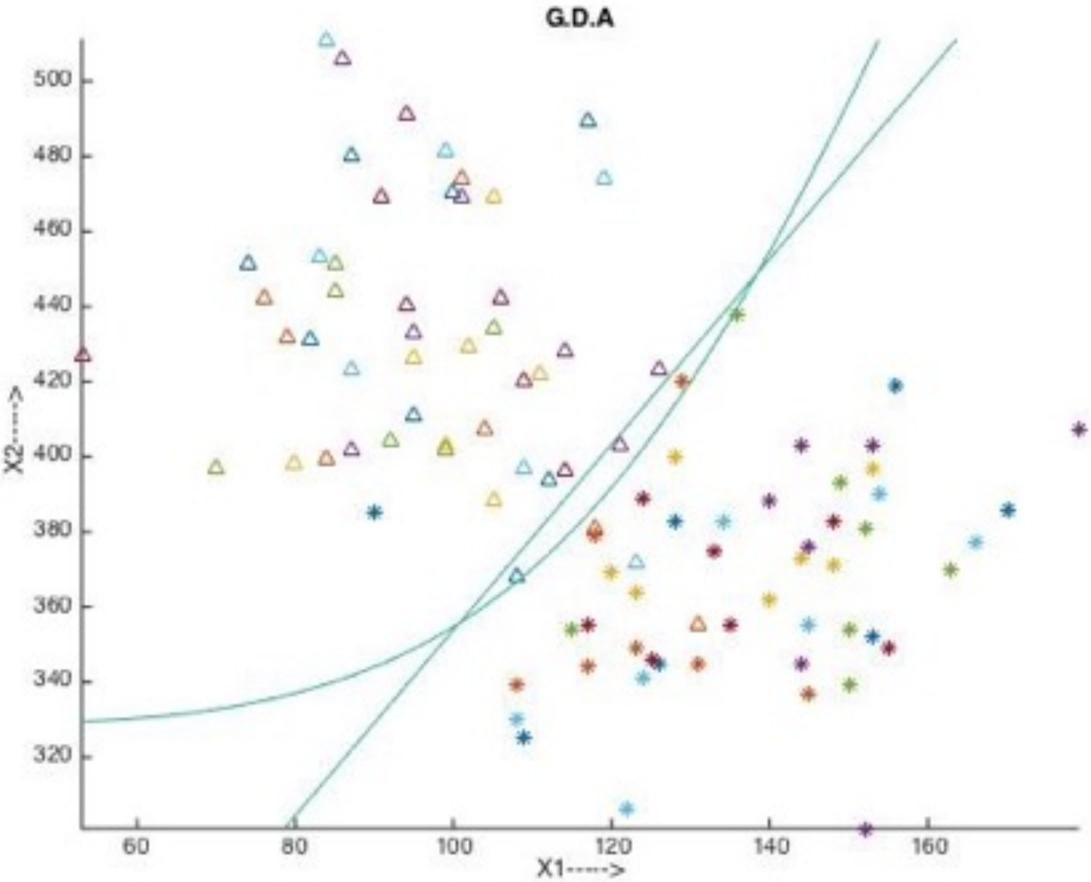


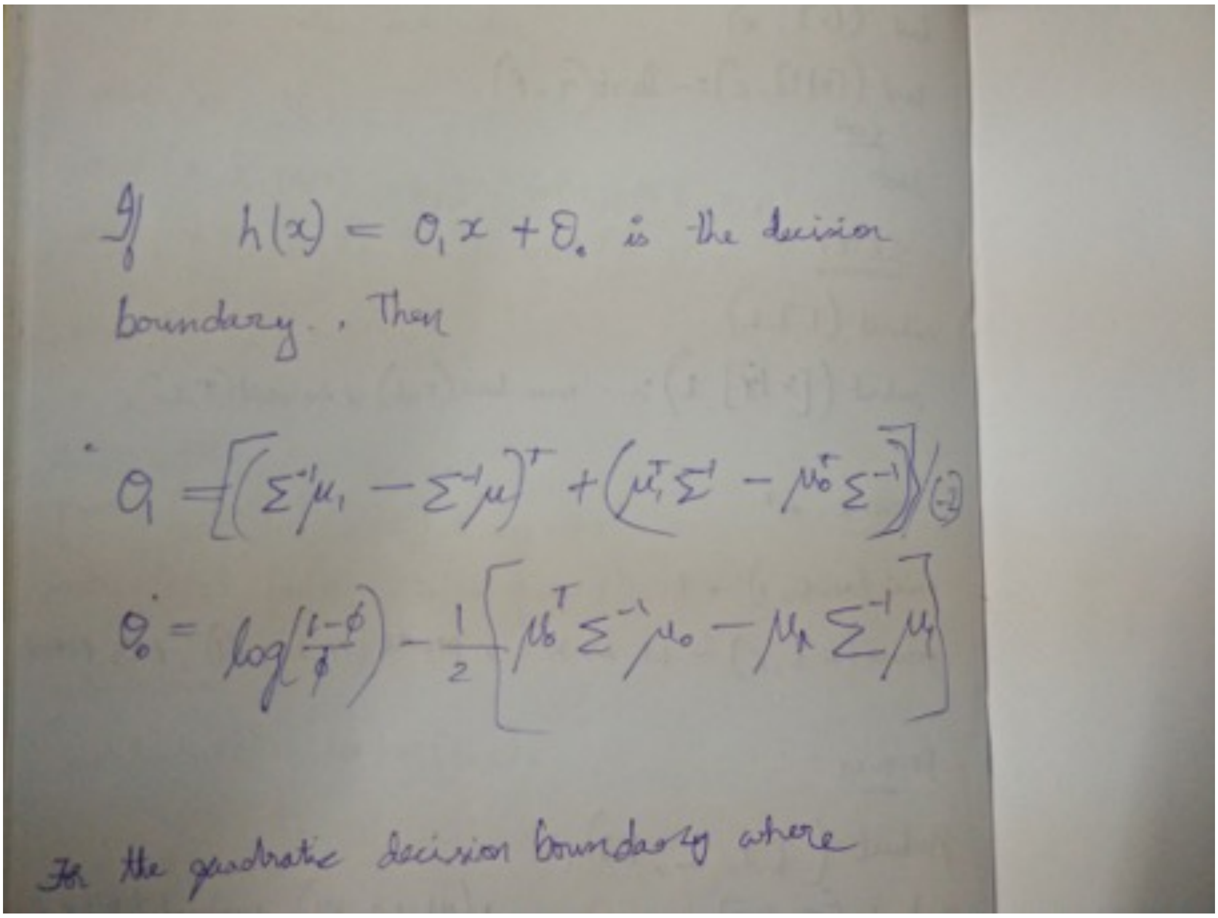
Q4 a.) Mu0 = [137.4600; 366.6200]
Mu1= [98.3800; 429.6600]
Cov = [287.4820, -26.7480; -26.7480, 0.001233]

b.) Fig 4-b

c.) Fig 4-b

d.) Mu0 = [137.4600; 366.6200]
Mu1= [98.3800; 429.6600]
Cov0=[319.5684, 130.8348; 130.8348, 875.3956]
Cov1=[255.3956, -184.3308; -184.3308, 0.0013711]





If $h(x) = \theta_1 x + \theta_0$ is the decision boundary. , Then

$$\theta_1 = \left[\left(\Sigma^{-1}\mu_1 - \Sigma^{-1}\mu\right)^T + \left(\mu_1^T \Sigma^{-1} - \mu_0^T \Sigma^{-1}\right)\right] / (-2)$$

$$\theta_0 = \log\left(\frac{1-\phi}{\phi}\right) - \frac{1}{2}\left[\mu_0^T \Sigma^{-1} \mu_0 - \mu_1 \Sigma^{-1} \mu_1\right]$$

For the quadratic decision boundary where e.) The linear decision boundary can be described as shown above and the quadratic decision boundary is shown below.

$$\theta_0 \log\left(\frac{1-\phi}{\phi}\right) - \frac{1}{2}\left[\mu_0 \Sigma \mu_0 - \mu_1 \Sigma \mu_1\right]$$

For the quadratic decision boundary where $\Sigma_0 \neq \Sigma_1$

$$x^T(\Sigma_0^{-1} - \Sigma_1^{-1})x + (\mu_1^T\Sigma_1^{-1} - \mu_0^T\Sigma_0^{-1})x + x^T(\Sigma_1^{-1}\mu_1 - \Sigma_0^{-1}\mu_0) + (\mu_0^T\Sigma_0^{-1}\mu_0 - \mu_1\Sigma_1^{-1}\mu_1) = 0.$$

f.) The logistic regression for classification can give a linear equation that can separate two known classes of objects, whereas the gaussian discriminant analysis is a generative algorithm which can separate objects into different classes.

In many cases, the quadratic or any higher degree polynomial can better classify the data. In the obtained Fig 4-b we can clearly notice the two decision boundaries and which one is the best fit.