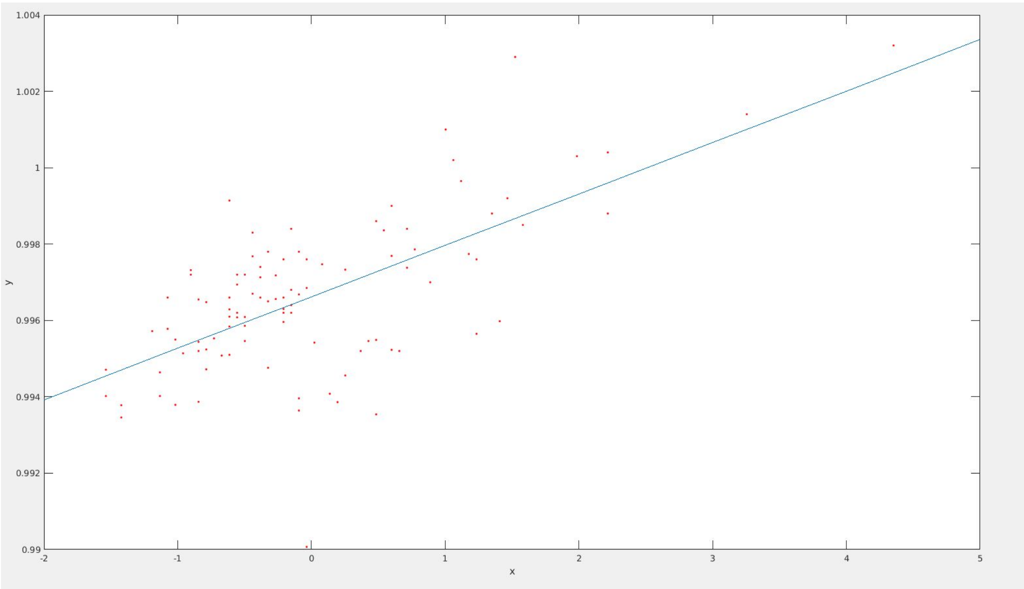
<u>Assignment-1</u>

1. **Linear Regression:**

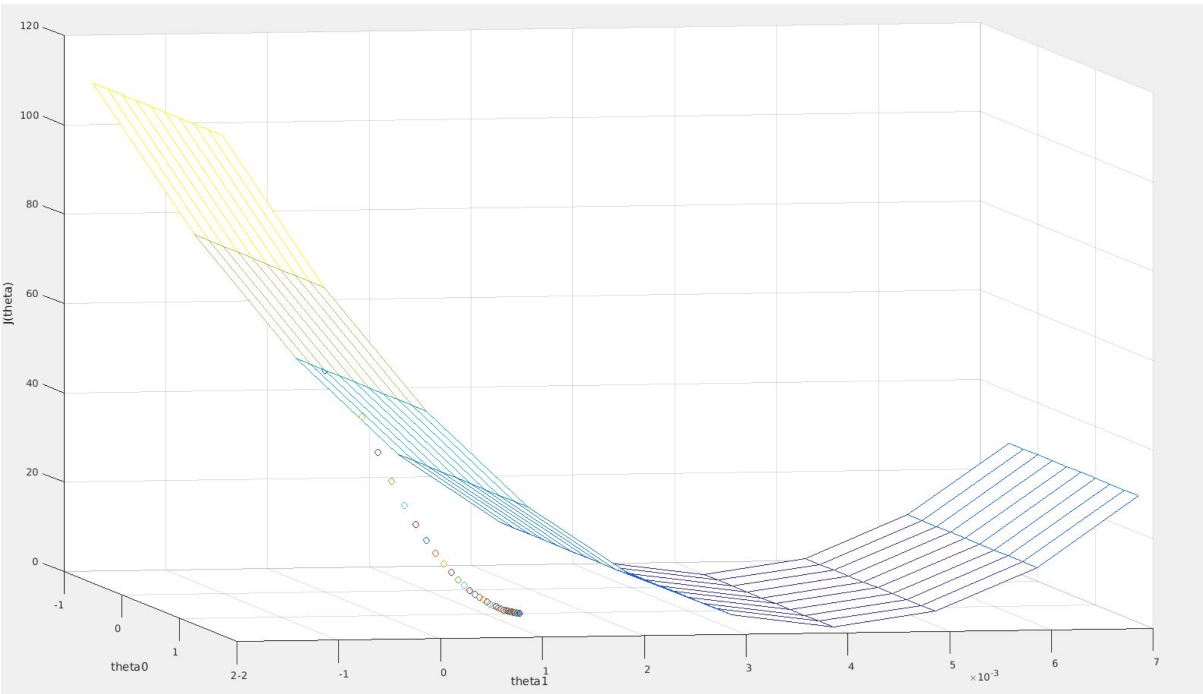
(a) The parameters are:

- Learning rate, $\eta = 0.001$
- Stopping Criteria : $||(\text{delta } \theta)|| < 0.00000001$ and $\text{delta}(J(\theta)) < 0.000000000001$
- Final $\theta = [0.9966, 0.0013]$

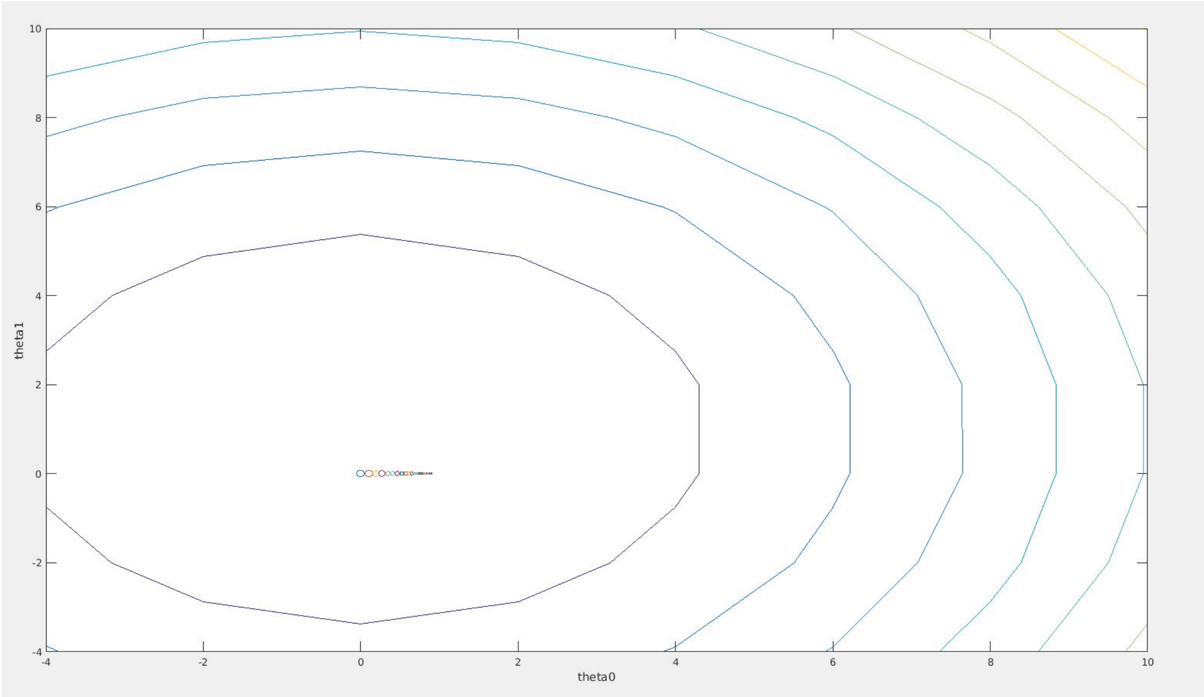
(b) The required plot is:
The line represents the hypothesis function obtained.



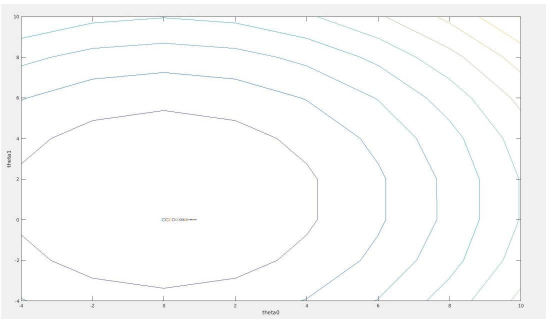
(c) The 3-D mesh is as follows:



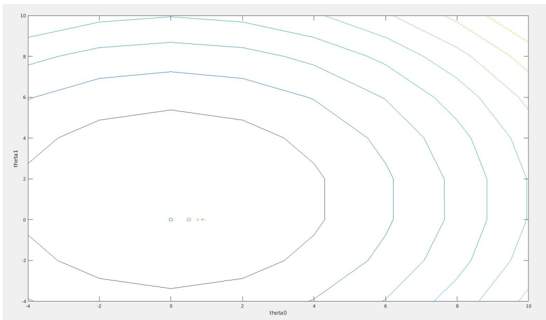
(d) The contours for above parameters are:



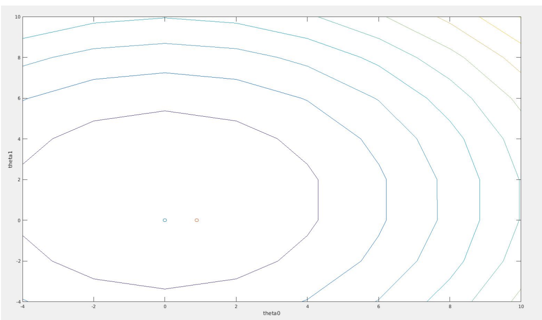
(e) The contours for varying η are:



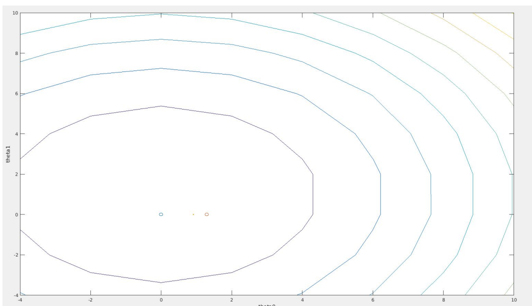
η= 0.001



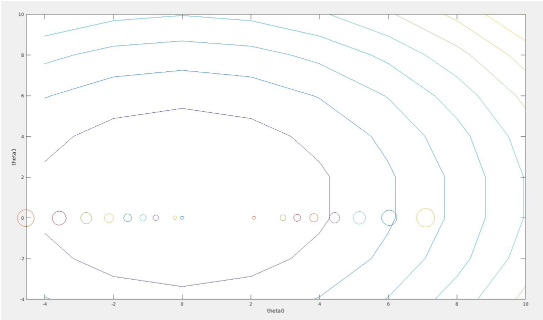
η= 0.005



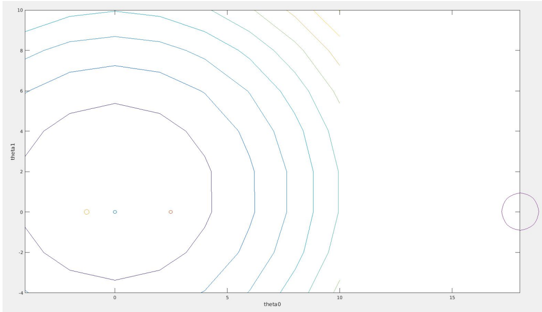
η= 0.009



η= 0.013



η= 0.021

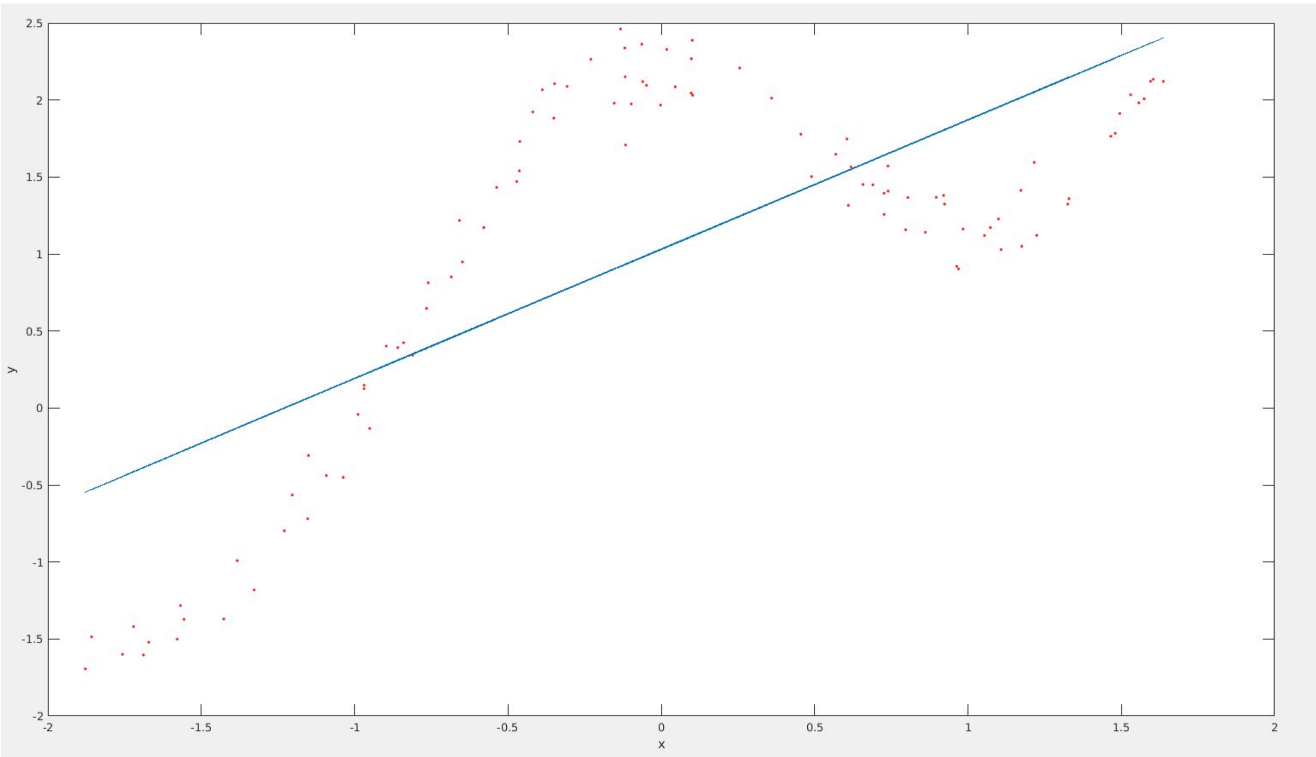


η= 0.025

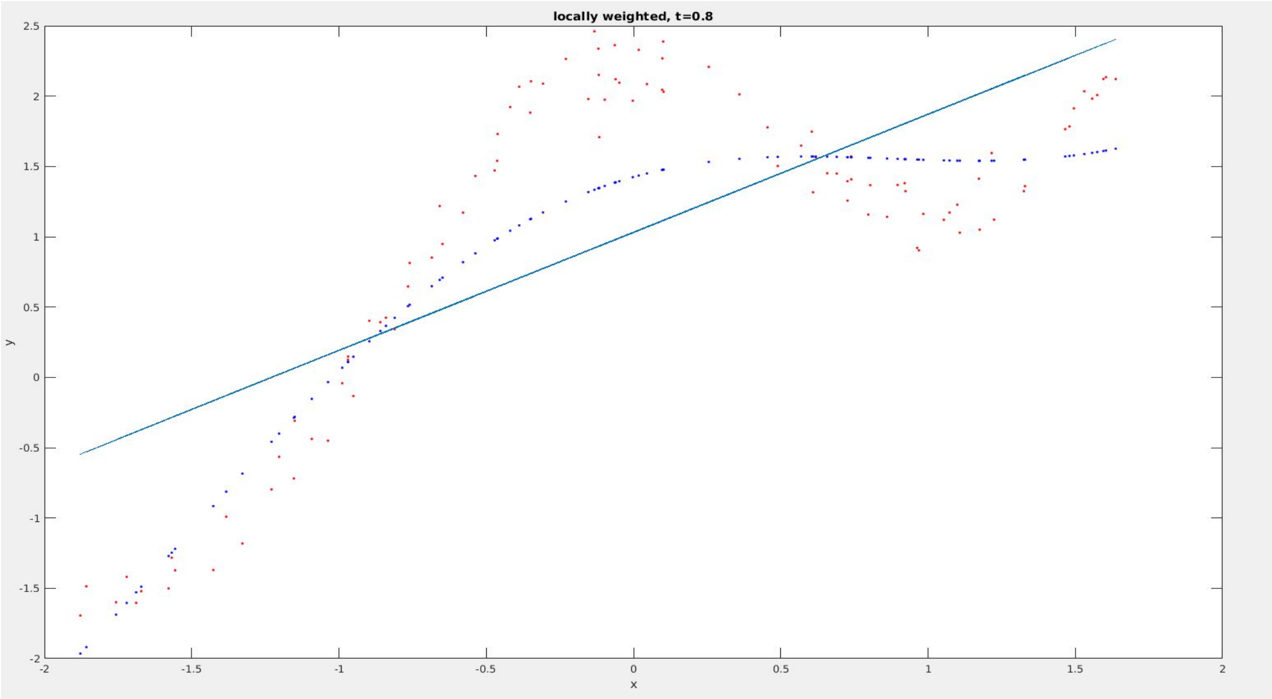
We observe that for that when η values are small, then gradient descent iterates for a longer time and hence more points are plotted. The algorithm finally converges at a theta close to the minima. But as we the values of η are increased, the error function converges faster, and in less number of iterations. At η= 0.013, the points oscillate but then converge to the minima. As we reach η = 0.021 and η = 0.025, the value of J(theta) overshoots and the minima is skipped and thus the algorithm fails to converge.

2. **Locally Weighted Linear Regression:**

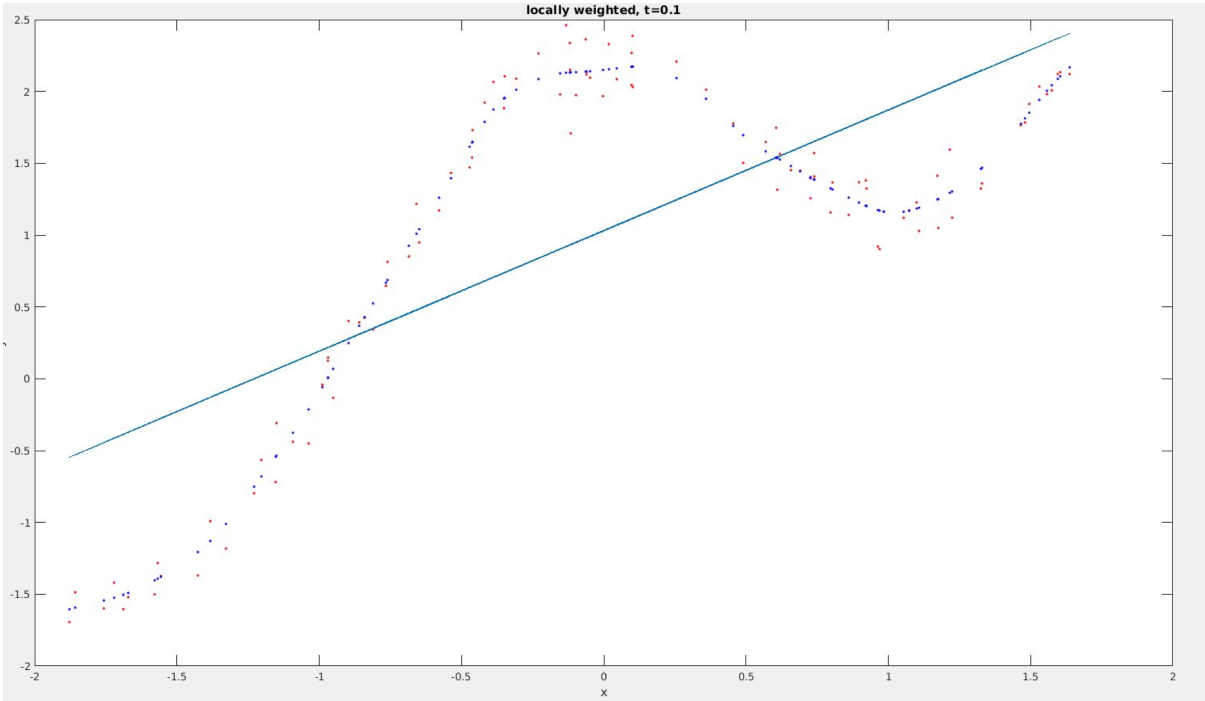
(a) The graph for unweighted linear regression is:



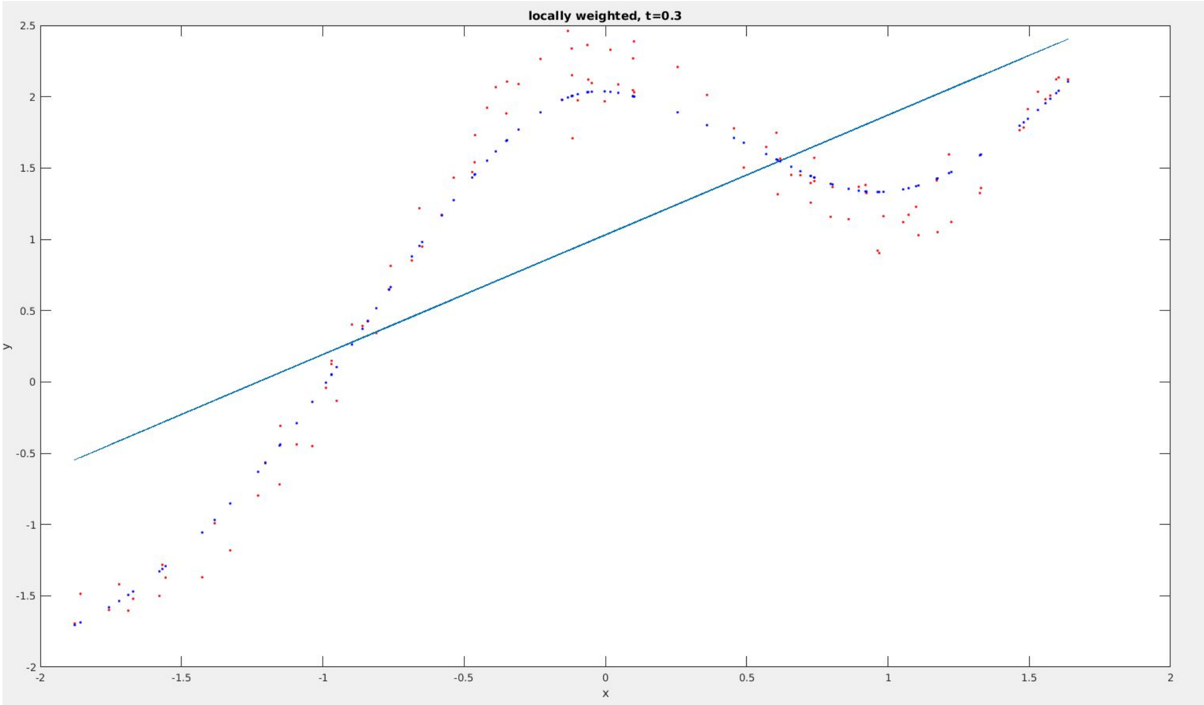
(b) The graph for weighted linear regression is (with τ=0.8):



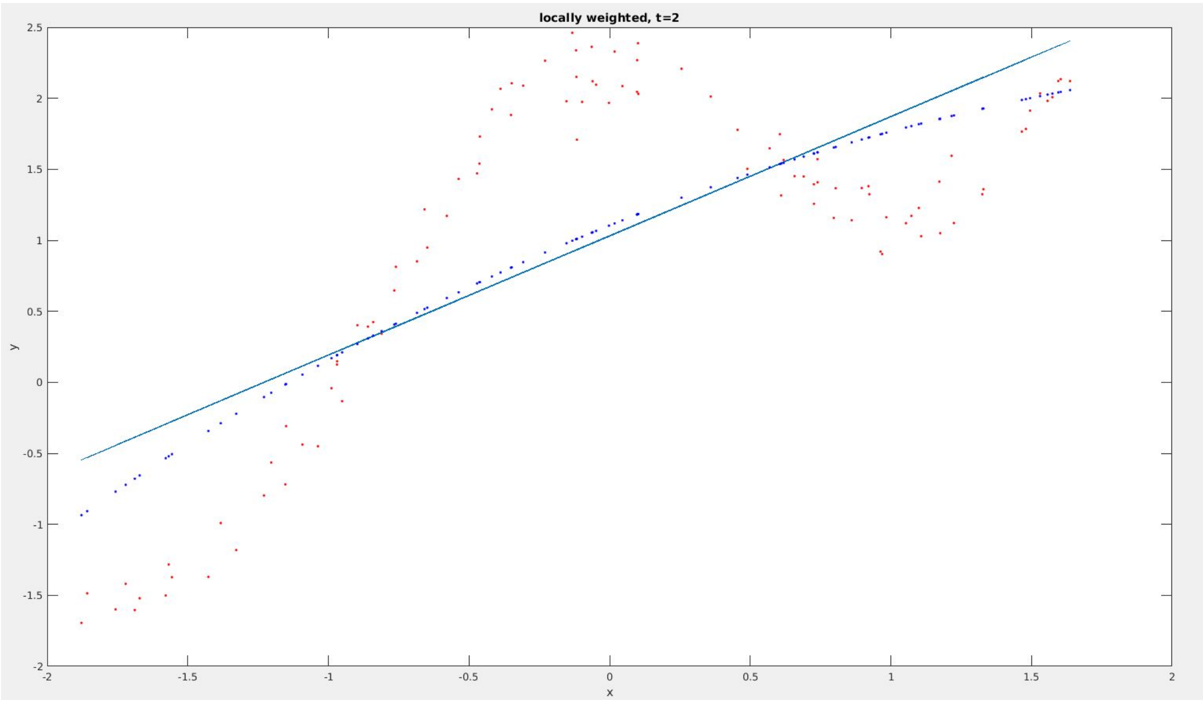
(c) The graph for weighted linear regression is for variable τ:



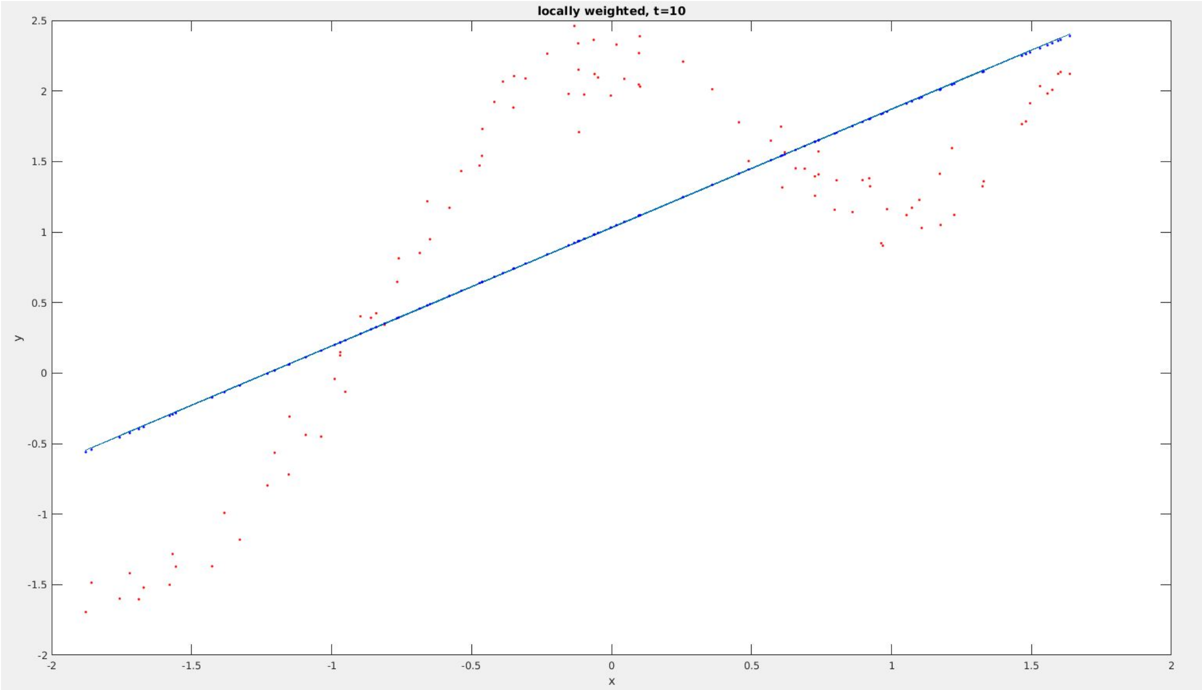
For τ=0.1



For τ=0.3



For τ=2



For τ=10

We observe that unweighted linear regression underfits the data. When we use weighted linear regression with τ=0.8, it fits the data pretty well. As we decrease the value of τ from 0.8 to 0.3 to 0.1, the curve starts to fit the data too much and at τ=0.1 it clearly overfits the data. When value of τ is increased beyond 1, then the curve starts to reach the line obtained by unweighted linear regression as the weights tend to become 1 (since weight is inversely proportional to τ^2), thus underfitting the data.
We observe that τ=0.3 is the optimal value among the values given.

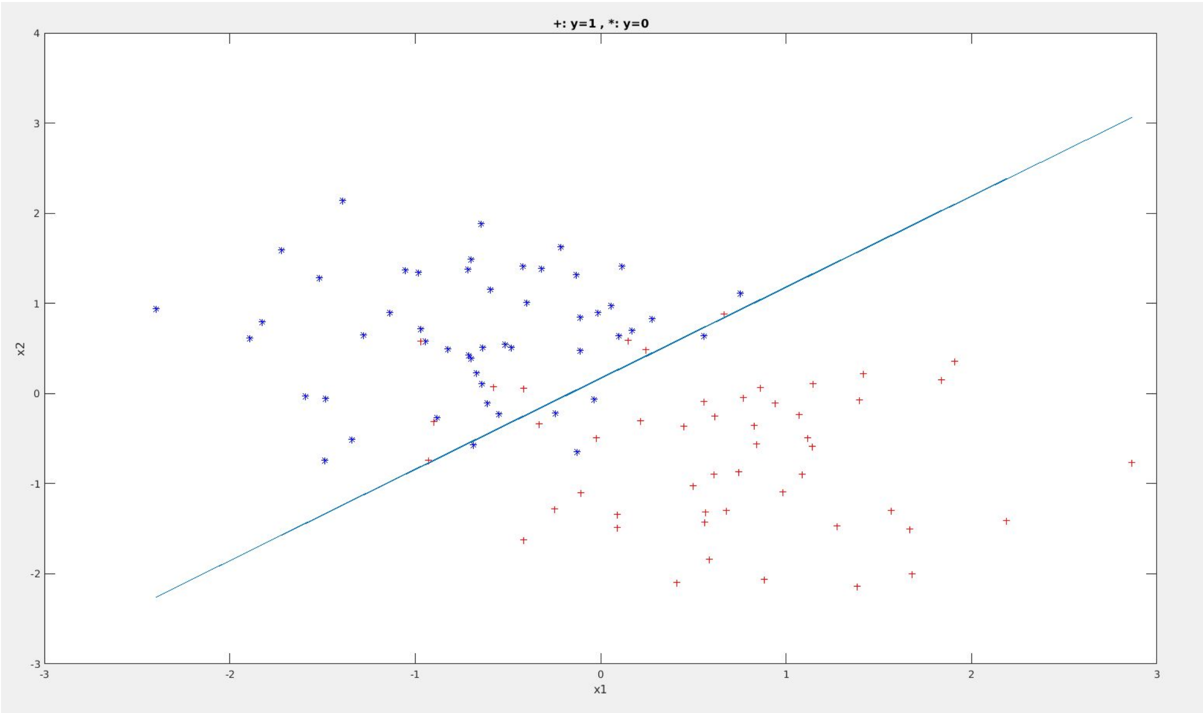
3. **Logistic Regression:**

(a) The parameters are:

- Stopping Criteria : ||(delta θ)||(infinity) {i.e. infinity norm of delta θ} < 0.000001
- Final θ = [-0.1179, -0.7060, 0.6983]

In each iteration, the values of θ obtained are normalised otherwise large differences in the three values of theta result in an almost singular Hessian matrix and thus it's inverse is not computed correctly.

(b) The graph for decision boundary is:

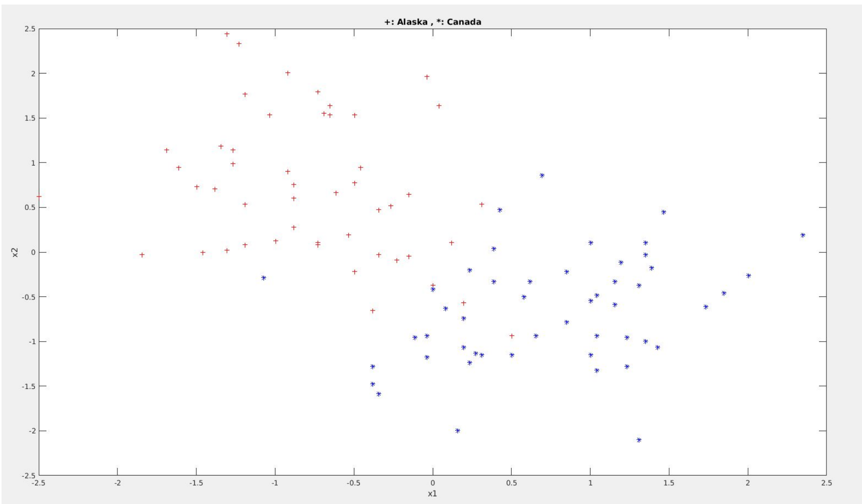


4. Gaussian Discriminant Analysis:

(a) The values obtained are:

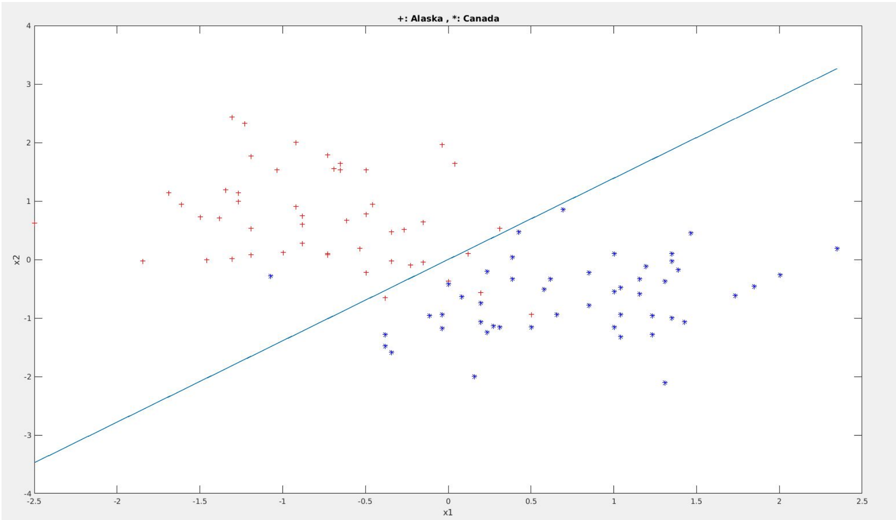
- $\varphi = 0.5$
- $\mu 0 = (-0.7515 \quad 0.6817)$
- $\mu 1 = (0.7515 \quad -0.6817)$
- $\Sigma = (0.4252 \quad -0.0222$
  $-0.0222 \quad 0.5253)$

(b) The points for Alaska and Canada are:



(c) Decision boundary for same Σ:

$$(\mu_0 - \mu_1)^T *(\Sigma^{-1})* x + \tfrac{1}{2}(\mu_1^T*(\Sigma^{-1})*(\mu_1) - (\mu_0)^T*(\Sigma^{-1})*(\mu_0)) + \log(1 - \varphi/\varphi) = 0$$

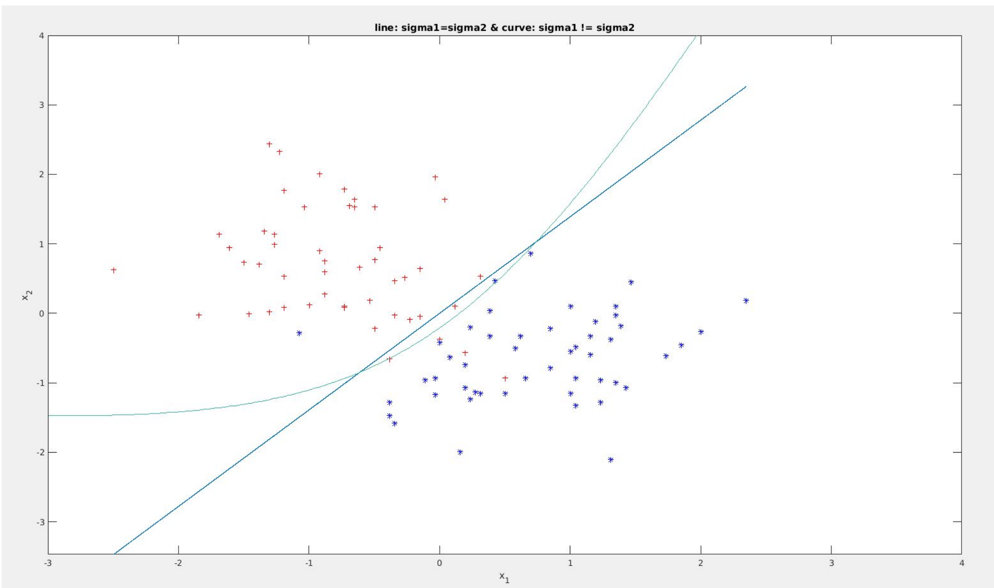


(d) The required values obtained are:

- $\mu0 = (-0.7515 \quad 0.6817)$
- $\mu1 = (0.7515 \quad -0.6817)$
- $\Sigma0 = \begin{pmatrix} 0.3778 & -0.1533 \\ -0.1533 & 0.6413 \end{pmatrix}$
- $\Sigma1 = \begin{pmatrix} 0.4727 & 0.1088 \\ 0.1088 & 0.4094 \end{pmatrix}$
- $\Phi = 0.5$

(e) Decision boundary for different Σ:

$$x^T*\tfrac{1}{2}(\Sigma_0^{-1} - \Sigma_1^{-1})*x + (\mu_1^T\Sigma_1^{-1} - \mu_0^T\Sigma_0^{-1})*x - \tfrac{1}{2}*(\mu_1\Sigma_1^{-1}\mu_1 - \mu_0\Sigma_0^{-1}\mu_0) + \log(\varphi*(||\Sigma_0||^{0.5})/(1-\varphi)*(||\Sigma_1||^{0.5})) = 0$$



(f) Observations:

- We observe that for same covariance matrix, the $x^T * \frac{1}{2}((\Sigma_0^{-1}) - (\Sigma_1^{-1})) * x$ term gets cancelled (as $\Sigma_0=\Sigma_1=\Sigma$). Thus for same Σ, we get a linear graph whereas for different Σ, we get a hyperbolic curve.
- One curve of the hyperbola obtained pass through the points and serves as a better decision boundary than the line because the variance of Alaska points is more than that of Canada which is incorporated into the quadratic curve.