COL774: Machine Learning | Date: February 12, 2019

Assignment 1 Report

Entry Number: 2016CS10363 | Name: Manish Tanwar

1 A

2 Locally Weighted Linear Regression:

(a) Linear Regression (unweighted):

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 0.327680 \\ 0.175316 \end{bmatrix}$$

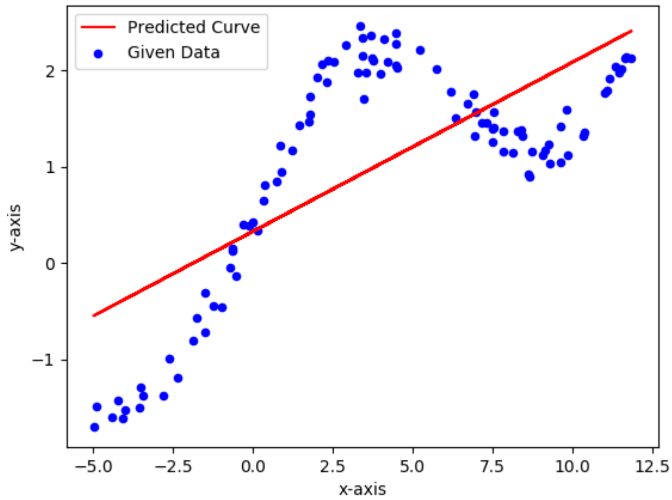

Figure 1: Linear Regressin Plot (Underfitting)

Linear Regression(unweighted) is not a good fit for the data as shown in the Figure 1(underfitting).

(b) Locally Weighted Linear Regression:

Weights : $w^{(i)} = \exp\left(-\frac{(x - x^{(i)})^2}{2\tau^2}\right)$ (where τ = Bandwidth Parameter)

Error Function: $J(\theta) = \frac{1}{2m}(X\theta - Y)^T W (X\theta - Y)$ (where $W = diag(w^{(i)})$)

Minima: $\nabla_\theta J(\theta) = 0 \quad \Rightarrow \quad \theta = (X^T W X)^{-1} X^T W Y$

Plots:

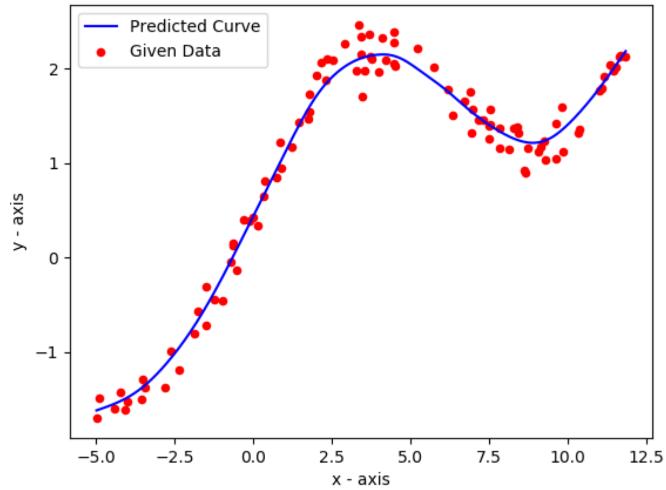

Figure 2: Locally Weighted Linear Regressin Plot ($\tau = 0.8$)

(c) Plots on Varying Bandwidth Parameter(τ)

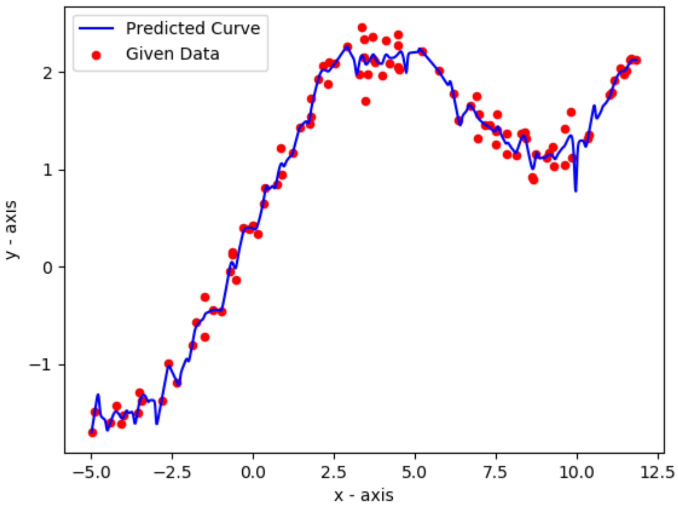

Figure 3: $\tau = 0.1$

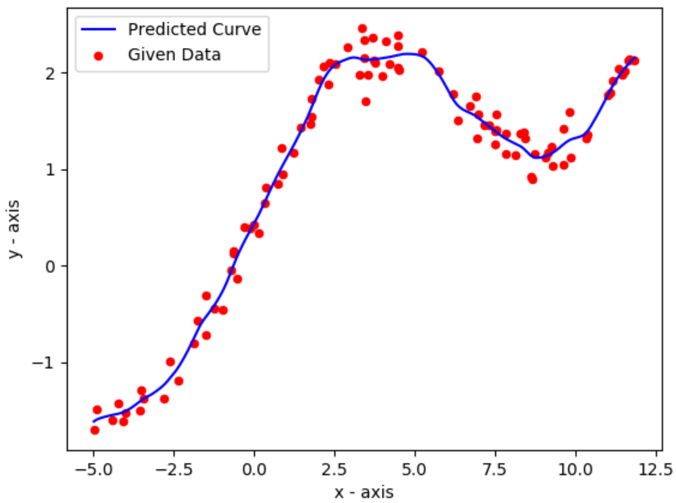

Figure 4: $\tau = 0.3$

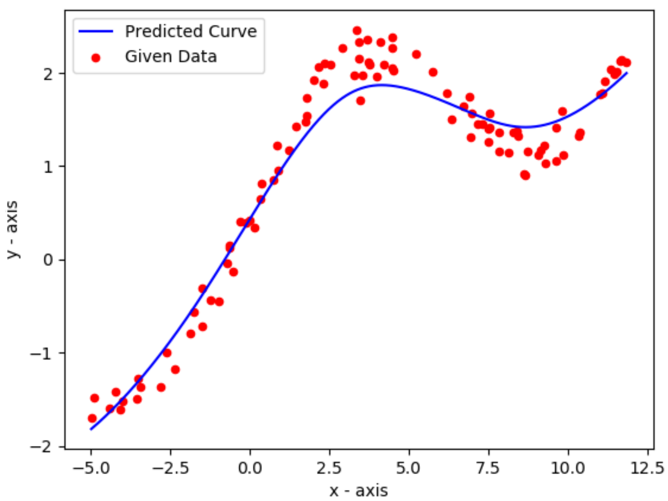

Figure 5: $\tau = 2$

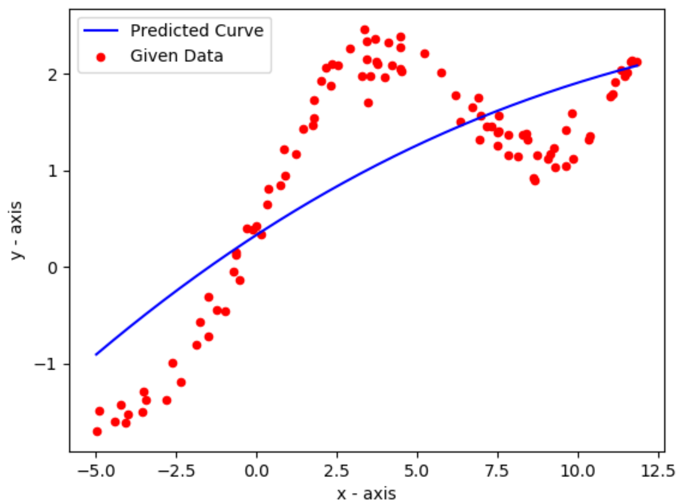

Figure 6: $\tau = 10$

Analysis:

- When Bandwidth Parameter(τ) is:
  - **Too Small:** the model **overfits** the training data.
  - **Too Large:** The model tries to **underfits** the training data.
- $\tau = 0.8$ works the best.

3 Logistic Regression:

Log Likelihood: $$LL(\theta) = \sum_{i=1}^{m} y^{(i)} log(h_\theta(x^{(i)})) + (1 - y^{(i)}) log(1 - h_\theta(x^{(i)}))$$

$$\nabla_\theta LL(\theta) = X^T(Y - g(X\theta)) \qquad \text{(where } g(x) = \frac{1}{1 + \exp(-x)}\text{)}$$

Hessian Matrix: $$H = \nabla_\theta^2 LL(\theta) = -X^T D X$$

$$\text{(where } D = diag(\, g(x^{(i)T}\theta)(1 - g(x^{(i)T}\theta))\,)\text{)}$$

Newton's Method:

$$\theta^{(t+1)} = \theta^{(t)} - H^{-1} \nabla_\theta LL(\theta)\big|_{\theta_t}$$

Convergence Condition:

$$\left|\theta_j^{(t+1)} - \theta_j^{(t)}\right| < \epsilon \qquad \forall j \in 1, 2 ... n \qquad \text{(for a sufficiently small } \epsilon\text{)}$$

Resulting Parameters:

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 0.223295 \\ 1.962616 \\ -1.964861 \end{bmatrix}$$

Decision Boundary is the straight line boundary separating the region where $h_\theta(x) \geq 0.5$ (class $y = 1$) from where $h_\theta(x) \leq 0.5$ (class $y = 0$).

Plot:

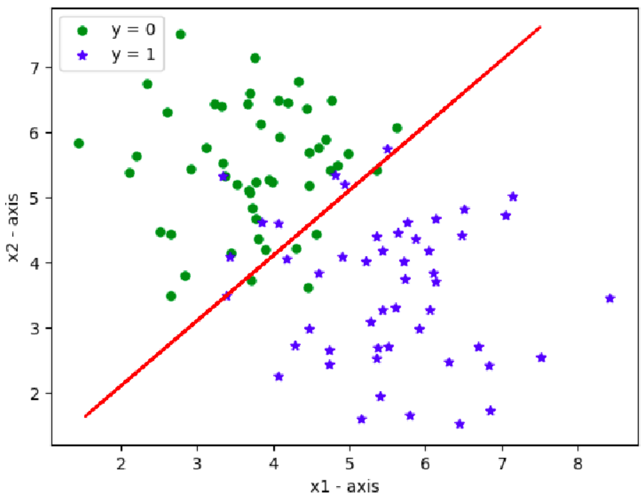

Figure 7: Logistic Regression : Given data and Linear Separator

4 Gaussian Discrmimant Analysis:

(a):

$$\phi = \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\{y^{(i)} = 1\} \qquad \mu_0 = \frac{\sum_{i=1}^{m} \mathbb{1}\{y^{(i)} = 0\}x^{(i)}}{\sum_{i=1}^{m} \mathbb{1}\{y^{(i)} = 0\}} \qquad \mu_1 = \frac{\sum_{i=1}^{m} \mathbb{1}\{y^{(i)} = 1\}x^{(i)}}{\sum_{i=1}^{m} \mathbb{1}\{y^{(i)} = 1\}}$$

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)} - \mu_{y^{(i)}})(x^{(i)} - \mu_{y^{(i)}})^T$$

which can be written as: $\Sigma = \frac{1}{m}W^T W$ (where $W = X - Y\mu_1^T - (1-Y)\mu_0^T$)

Resulting Parameters:

$$\phi = 0.5 \qquad \mu_0 = \begin{bmatrix} 98.38 \\ 429.66 \end{bmatrix} \qquad \mu_1 = \begin{bmatrix} 137.46 \\ 366.62 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 287.482 & -26.748 \\ -26.748 & 1123.25 \end{bmatrix}$$

(b) & (c):

(i) Linear Boundary Equation:

$$2(\mu_1^T - \mu_0^T\Sigma^{-1})x + \mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1 - 2\log\left(\frac{1-\phi}{\phi}\right) = 0$$

(ii) Plot:

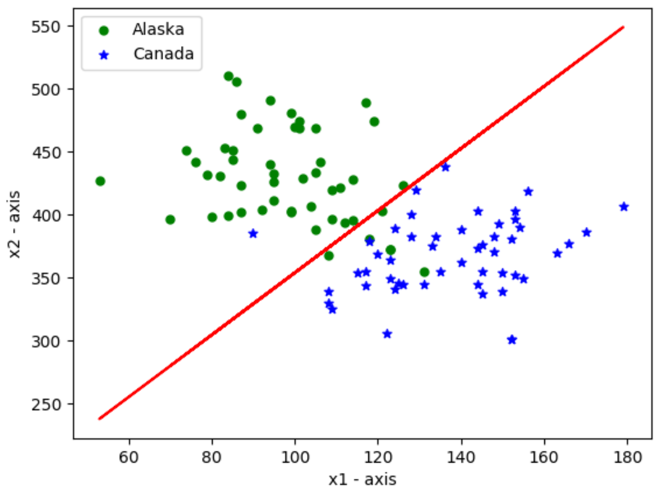

Figure 8: GDA : Given data and Linear Separator

(d):

$$\phi = \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\{y^{(i)} = 1\} \qquad \mu_0 = \frac{\sum_{i=1}^{m} \mathbb{1}\{y^{(i)} = 0\}x^{(i)}}{\sum_{i=1}^{m} \mathbb{1}\{y^{(i)} = 0\}} \qquad \mu_1 = \frac{\sum_{i=1}^{m} \mathbb{1}\{y^{(i)} = 1\}x^{(i)}}{\sum_{i=1}^{m} \mathbb{1}\{y^{(i)} = 1\}}$$

$$\Sigma_0 = (X - \mu_0^{ext})^T diag(y^{(i)})(X - \mu_0^{ext}) \quad \text{(where } \mu_0^{ext} \text{ is } m \times n \text{ matrix with each row} = \mu_0^T\text{)}$$

Resulting Parameters:

$$\phi = 0.5 \qquad \mu_0 = \begin{bmatrix} 98.38 \\ 429.66 \end{bmatrix} \qquad \mu_1 = \begin{bmatrix} 137.46 \\ 366.62 \end{bmatrix}$$

$$\Sigma_0 = \begin{bmatrix} 255.3956 & -184.3308 \\ -184.3308 & 1371.1044 \end{bmatrix} \qquad \Sigma_1 = \begin{bmatrix} 319.5684 & 130.8348 \\ 130.8348 & 875.3956 \end{bmatrix}$$

(e):

(i) Quadratic Boundary Equation:

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

(ii) Plot:

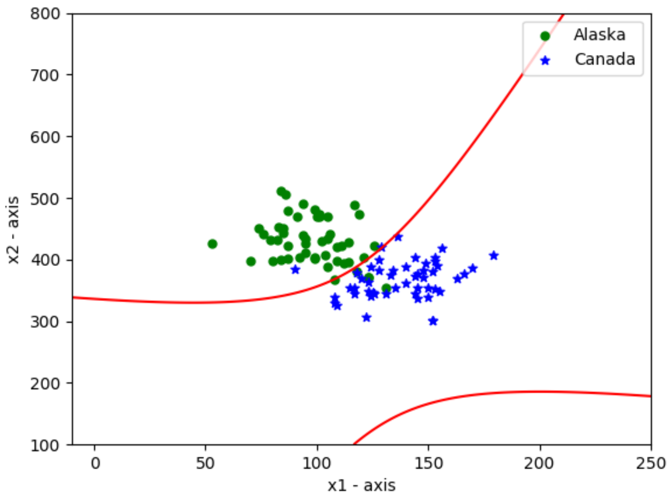

Figure 9: GDA : Given data and Quadratic Separator

(f) Analysis: