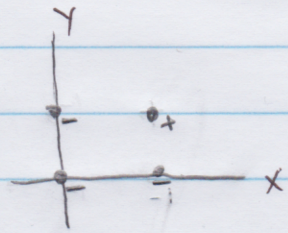
Henry Yang 204584728

PSET-2 : Math part

1. a)

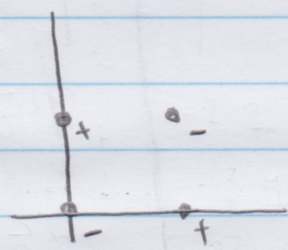


X	Y	AND
0	0	0
0	1	0
1	0	0
1	1	1

$$\boxed{① \ \vec{\theta} = \begin{bmatrix} -4 \\ 3 \\ 2 \end{bmatrix} \quad ② \begin{bmatrix} -7 \\ 5 \\ 3 \end{bmatrix}}$$

$$D = \begin{Bmatrix} (1,0,0), & -1 \\ (1,0,1), & -1 \\ (1,1,0), & -1 \\ (1,1,1), & 1 \end{Bmatrix}$$

b)



X	Y	XOR
0	0	0
0	1	1
1	0	1
1	1	0

$$D = \begin{Bmatrix} (1,0,0), & -1 \\ (1,0,1), & 1 \\ (1,1,0), & 1 \\ (1,1,1), & -1 \end{Bmatrix}$$

no valid perceptron as the data is not linearly separable.

gff(x)

2. a) $J(\vec{\theta}) = -\sum_{n=1}^{N} \left[y_n \log h_\theta(\vec{x}_n) + (1-y_n) \log (1 - h_{\vec{\theta}}(\vec{x}_n)) \right]$

$h_\theta(\vec{x}_n) = \sigma(\vec{\theta}^T \vec{x}_n)$

<u>note</u>: $\vec{\theta}^T \vec{x}_n = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_{D+1} x_{D+1}$

$$\frac{\partial h}{\partial \theta_j} = \sigma(\vec{\theta}^T \vec{x}_n)\left[1 - \sigma(\vec{\theta}^T \vec{x}_n)\right] x_j$$

$$\frac{\partial J}{\partial \theta_j} = -\sum_n \Big\{ y_n \frac{1}{\sigma(\vec{\theta}^T \vec{x}_n)} \left(\sigma(\vec{\theta}^T \vec{x}_n)\left[1-\sigma(\vec{\theta}^T \vec{x}_n)\right] x_j\right.$$
$$\left. + (1-y_n)(-1)\left(\frac{1}{1-\sigma(\vec{\theta}^T \vec{x}_n)}\left[\sigma(\vec{\theta}^T \vec{x}_n)\left[\sigma(\vec{\theta}^T \vec{x}_n)\right] x_j\right]\right)\right\}$$

$$= -\sum_n \left\{ y_n \left[1 - \sigma(\vec{\theta}^T \vec{x}_n)\right] x_j - (1-y_n)\left[\sigma(\vec{\theta}^T \vec{x}_n)\right] x_j \right\}$$

$$= \sum_n \left[\sigma(\vec{\theta}^T \vec{x}_n) - y_n\right] x_j$$

$$\boxed{\frac{\partial J}{\partial \theta_j} = \sum_{n=1}^{N} \left[h_{\vec{\theta}}(\vec{x}_n) - y_n\right] x_j}$$

b) $$\frac{\partial^2 J}{\partial \theta_j \partial \theta_K} = x_j \sum_n \sigma(\vec{\theta}^T \vec{x}_n)\left[1 - \sigma(\vec{\theta}^T \vec{x}_n)\right] x_K$$

$$= x_j x_K \sum_n h_{\vec{\theta}}(\vec{x}_n)\left[1 - h_{\vec{\theta}}(\vec{x}_n)\right]$$

$$\vec{x}_n x_n^T = \begin{bmatrix} x_1^2 & \dots x_1 x_K & \cdots x_1 x_n \\ \vdots & \ddots & \\ x_j x_1 & x_j x_K & \\ \vdots & & \ddots \; x_n^2 \\ x_n x_1 & & \end{bmatrix}$$

Since $H_{j,k}$ represents $\frac{\partial^2 J}{\partial\theta_j \partial\theta_k}$, we see that $\vec{x}_n \vec{x}_n^T$ represents the constant part of the general form of $\frac{\partial^2 J}{\partial\theta_j \partial\theta_k}$, and that $\sum_n h_{\vec{\theta}}(\vec{x}_n)[1-h_{\vec{\theta}}(\vec{x}_n)]$ is the same as the summation portion of $\frac{\partial^2 J}{\partial\theta_j \partial\theta_k}$. Therefore, $H = \sum_{n=1}^{N} h_{\vec{\theta}}(\vec{x}_n)[1-h_\theta(x_n)]\vec{x}_n \vec{x}_n^T$.

c) $\vec{H} = \sum_{n=1}^{N} h_{\vec{\theta}}(\vec{x}_n)[1-h_{\vec{\theta}}(\vec{x}_n)]\, \vec{x}_n \vec{x}_n^T$

$$\vec{z}^T H \vec{z} = \vec{z}^T \underbrace{\left(\sum_{n=1}^{N} h_{\vec{\theta}}(\vec{x}_n)[1-h_{\vec{\theta}}(\vec{x}_n)]\right)}_{\text{scalar}} \vec{x}_n \vec{x}_n^T \vec{z}$$

$$= \left(\sum_{n=1}^{N} h_{\vec{\theta}}(\vec{x}_n)[1-h_{\vec{\theta}}(\vec{x}_n)]\right) \vec{z}^T \vec{x}_n \vec{x}_n^T \vec{z}$$

since $0 < h_{\vec{\theta}}(\vec{x}_n) < 1$, $\sum_{n=1}^{N} h_{\vec{\theta}}(\vec{x}_n)[1-h_{\vec{\theta}}(\vec{x}_n)] > 0$

$\vec{z}^T \vec{x}_n \vec{x}_n^T \vec{z} = (\vec{x}_n^T \vec{z})^2 > 0$ for all real vectors $\vec{z}$

Therefore, $\vec{z}^T \vec{H} \vec{z} \geq 0$ for all real vectors $\vec{z}$.

3. a) ① $\frac{\partial J}{\partial \theta_0} = 2 \sum_n w_n (\theta_0 + \theta_1 x_n - y_n)$

② $\frac{\partial J}{\partial \theta_1} = 2 \sum_n w_n (\theta_0 + \theta_1 x_n - y_n) x_n$

b) ① $2 \sum_n w_n (\theta_0 + \theta_1 x_n - y_n) = 0$

② $2 \sum_n w_n (\theta_0 + \theta_1 x_n - y_n) x_n = 0$

① $\sum_n w_n y_n = \theta_0 \sum_n w_n + \theta_1 \sum w_n x_n$

☆ let $W = \sum w_n$, $\bar{X}_w = \frac{\sum w_n x_n}{\sum w_n}$ (weighted avg. of x)

and $\bar{Y}_w = \frac{\sum w_n y_n}{\sum w_n}$ (weighted avg of y)

$$\boxed{\theta_0 = \bar{Y}_w - \theta_1 \bar{X}_w}$$

② In order to solve for θ_1, we have to substitute θ_1 for θ_0, and then take the partial derivative.

$$J = \sum_n w_n (\bar{Y}_w - \theta_1 \bar{X}_w + \theta_1 x_n - y_n)^2$$

$$\frac{\partial J}{\partial \theta_1} = 2 \sum_n w_n \left[(\bar{Y}_w - y_n) - \theta_1 (\bar{X}_w - x_n) \right] (\bar{X}_w - x_n) = 0$$

$$\boxed{\theta_1 = \frac{\sum_n w_n (\bar{Y}_w - y_n)(\bar{X}_w - x_n)}{\sum_n w_n (\bar{X}_w - x_n)^2}}$$