CS188, Winter 2017
Problem Set 2:
Due 2/9/2017

Shannon Phu

1 Problem 1

(a) w1 = 1, w2 = 1, b = -1

(b) XOR's feature values are not linearly separable so it cannot be represented by a perceptron.

2 Problem 2

(a)

$$h_\theta = \frac{1}{1+e^{-\theta x_n}} \tag{1}$$

$$log(h_\theta(x_n)) = -log(1+e^{-\theta x_n}) \tag{2}$$

$$log(1-h_\theta(x_n)) = -\theta x_n - log(1+e^{-\theta x_n}) \tag{3}$$

$$J(\theta) = \sum_{n=1}^{N} y_n log(1+e^{-\theta x_n}) + (1-y_n)(\theta x_n + log(1+e^{-\theta x_n})) \tag{4}$$

$$= \sum_{n=1}^{N} -y_n\theta x_n + \theta x_n + log(1+e^{-\theta x_n}) \tag{5}$$

$$= \sum_{n=1}^{N} -y_n\theta x_n + log(1+e^{\theta x_n}) \tag{6}$$

$$\frac{\partial J(\theta)}{\partial\theta} = \sum_{n=1}^{N} -x_n y_n + \frac{x_n e^{\theta x_n}}{e^{\theta x_n}+1} \tag{7}$$

$$\frac{\partial J(\theta)}{\partial\theta} = \sum_{n=1}^{N} -x_n y_n + x_n h_\theta(x_n) \tag{8}$$

(b)

$$\frac{\partial^2 J(\theta)}{\partial\theta_j\theta_k} = \sum_{n=1}^{N} x_{n_j} h'_\theta(x_n) \tag{9}$$

because

$$h'_\theta(x_n) = h_\theta(x_n)(1-h_\theta(x_n)), \tag{10}$$

$$\frac{\partial^2 J(\theta)}{\partial\theta_j\theta_k} = \sum_{n=1}^{N} x_{n_j} h_\theta(x_n)(1-h_\theta(x_n))x_n^T \tag{11}$$

(c) Because H is written as

$$H = \sum_{n=1}^{N} x_{n_j} h_\theta(x_n)(1-h_\theta(x_n))x_n^T, \tag{12}$$

and we know that the function h, representing the sigmoid function, must be between 0 and 1 hence making

$$h_\theta(x_n)(1-h_\theta(x_n)) \tag{13}$$

positive. For J to be convex, H must be positive semi-definite. Therefore we can exclude Equation 13 from Equation 12, only paying attention to x_n.

$$z^T x_n x_n^T z = (x_n^T z)^T (x_n^T z) = (x_n^T z) \cdot (x_n^T z) \tag{14}$$

The dot product is positive which means that the Hessian matrix is positive semi-definite thus J is convex.

3 Problem 3

(a)

$$\frac{\partial J(\theta)}{\partial \theta_0} = 2\sum_{n=1}^{N} w_n(\theta_0 + \theta_1 x_{n,1} - y_n) \tag{15}$$

$$\frac{\partial J(\theta)}{\partial \theta_1} = 2\sum_{n=1}^{N} w_n(\theta_0 + \theta_1 x_{n,1} - y_n)x_{n,1} \tag{16}$$

(b) A system of equations was made from part (a) as follows:

$$\sum_{n=1}^{N} w_n y_n = \theta_0 \sum_{n=1}^{N} w_n + \theta_1 \sum_{n=1}^{N} w_n x_{n,1} \tag{17}$$

$$\sum_{n=1}^{N} w_n x_{n,1} y_n = \theta_0 \sum_{n=1}^{N} w_n x_{n,1} + \theta_1 \sum_{n=1}^{N} w_n x_{n,1}^2 \tag{18}$$

Solving for theta 0 and 1:

$$\theta_0 = \frac{\sum_{n=1}^{N} w_n y_n - \theta_1 \sum_{n=1}^{N} w_n x_{n,1}}{\sum_{n=1}^{N} w_n} \tag{19}$$

Plugging it into Equation 18, I can solve for theta 1:

$$\theta_1 = \frac{\sum_{n=1}^{N} w_n \sum_{n=1}^{N} w_n x_{n,1} y_n - \sum_{n=1}^{N} w_n y_n \sum_{n=1}^{N} w_n x_n}{\sum_{n=1}^{N} w_n \sum_{n=1}^{N} w_n x_{n,1}^2 - (\sum_{n=1}^{N} w_n x_{n,1})^2} \tag{20}$$

Plugging Equation 20 back into Equation 19, we can solve for theta 0 as well:

$$\theta_0 = \frac{\sum_{n=1}^{N} w_n y_n}{\sum_{n=1}^{N} w_n} - \frac{\sum_{n=1}^{N} w_n x_{n,1}}{\sum_{n=1}^{N} w_n} \cdot \theta_1 \tag{21}$$