2.1) Multilayer perceptron:

Assumed training dataset $\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$ where $x^{(i)} \in R^n$ & $y^{(i)} \in \{0, 1\}$

Error function = $E(\{w_i\}, v)$

$$\text{MSE} = \frac{1}{2} \sum_{i=1}^{m} (y^{(i)} - \hat{y}^{(i)})^2$$

According to Backpropagation algorithm, we update $(i-1)^{th}$ layer parameter after updating the parameter of i^{th} layer.

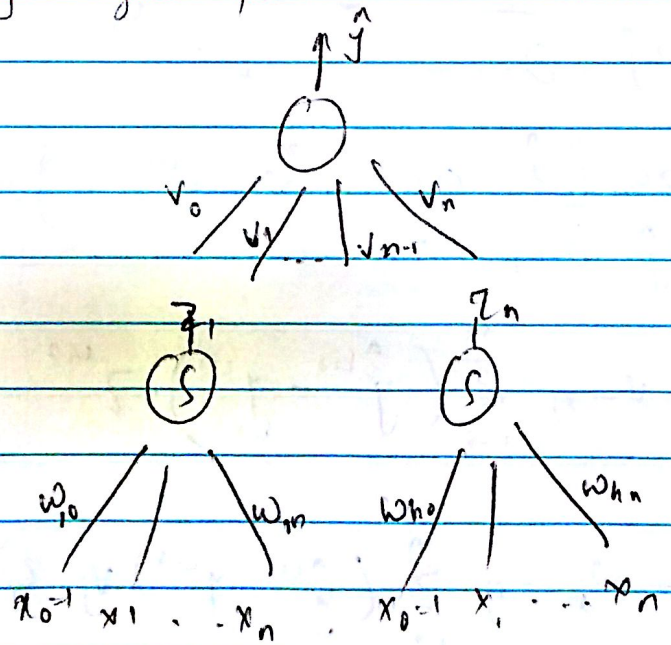
From assumption,

x → Original input

w - parameter of input layer

v - parameter of hidden

$\hat{y}$ - final prediction.



Minimizing the error:

$$V \leftarrow V - \eta \frac{\partial E}{\partial V}$$

Chain rule - $\frac{\partial E}{\partial V} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V}$

$$\therefore V \leftarrow V - \eta \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial V}$$

$$w_j = w_j - \eta \frac{\partial E}{\partial w_j}$$

Chain Rule, $\frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z_j} \frac{\partial z_j}{\partial w_j}$

$$\therefore w_j = w_j - \eta \cdot \frac{\partial E}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z_j} \frac{\partial z_j}{\partial w_j}$$

$$z_j = h(w_j^T x)$$

$$\frac{\partial z_j}{\partial w_j} = z_j (1 - z_j) x$$

$$\hat{y} = V^T z$$

$$\frac{\partial \hat{y}}{\partial v} = z$$

1. $$\frac{\partial E}{\partial v} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial v} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot z^{(i)}$$

$$\frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z_j} \frac{\partial z_j}{\partial w_j} = \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) v_j z_j^{(i)} (1 - z_j^{(i)}) x^{(i)}$$

$$\therefore V \leftarrow V - \eta \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \cdot z^{(i)}$$

4 $$w_j \leftarrow w_j - \eta \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) v_j z_j^{(i)} (1 - z_j^{(i)}) x^{(i)}$$

Update equations are same as that of calculated using maximum likelihood function in class.