$$\text{Error}(w \mid x_i, y_i) = \frac{1}{2}(y_i - \hat{y}_i)^2 = \frac{1}{2}(y_i - w^T . x_i)$$

$$\frac{\partial \text{Error}}{\partial w} = -(y_i - \hat{y}_i) . x_i \Rightarrow \Delta w = -\eta . \frac{\partial \text{Error}}{\partial w}$$

$$= \eta . (y_i - \hat{y}_i) . x_i$$

Perceptron for Binary classification

$$s(w^T . x_i) = \frac{1}{1 + \exp(-(w^T x_i))} = \hat{y}_i$$

hint: $\frac{\partial \log(\hat{y}_i)}{\partial w}$

$$\frac{\partial \log(f(w))}{\partial w} = \frac{1}{f(w)} . \frac{\partial f(w)}{\partial w}$$

$$\text{Error}(w \mid x_i, y_i) = -y_i \log(\boxed{\hat{y}_i}) - (1 - y_i) . \log(1 - \boxed{\hat{y}_i})$$

Exercise #7

$$\cancel{\frac{\partial \text{Error}}{\partial w}} = \boxed{\eta (y_i - \hat{y}_i) . x_i = \Delta w}$$

$$\hat{y}_{ic} = \frac{\exp(w_c^T x_i)}{\sum_{k=1}^{K} \exp(w_k^T x_i)}$$

$$\text{Error}(\{w_c\}_{c=1}^{K} \mid x_i, y_i) = -\sum_{c=1}^{K} y_{ic} . \log(\hat{y}_{ic})$$

Exercise #8

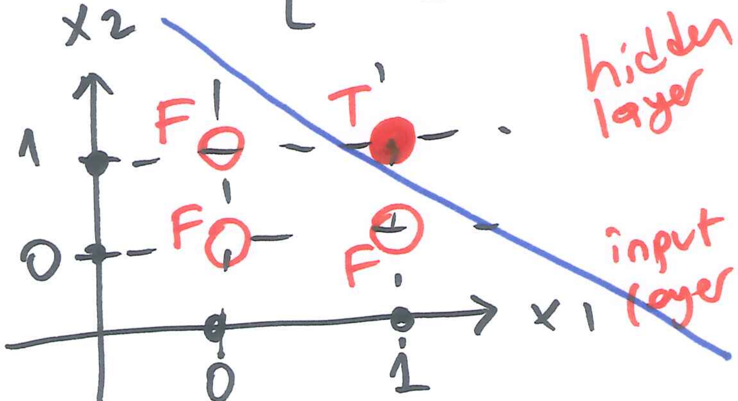
$$\boxed{\eta (y_{ic} - \hat{y}_{ic}) . x_i = \Delta w_c}$$

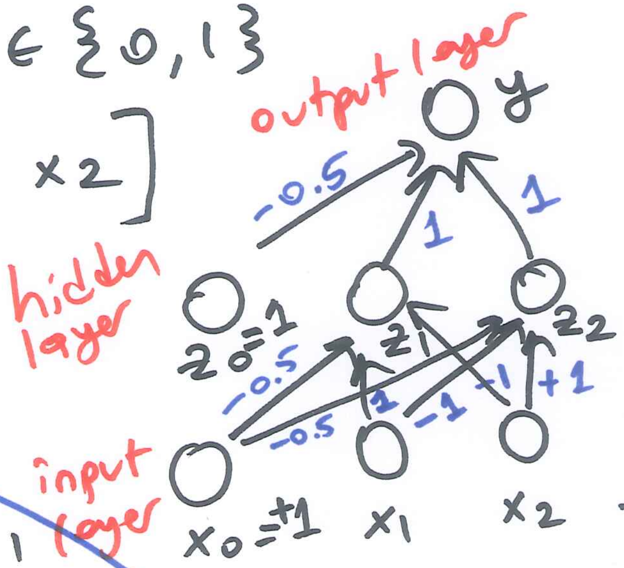
$$\underbrace{\text{Update}}_{\Delta} = \underbrace{\text{Learning Factor}}_{\eta} \times \Big[\underbrace{\text{True Output} - \text{Predicted Output}}_{[y-\hat{y}]}\Big] \underbrace{\times \text{Input}}_{x}$$

Booleen Functions

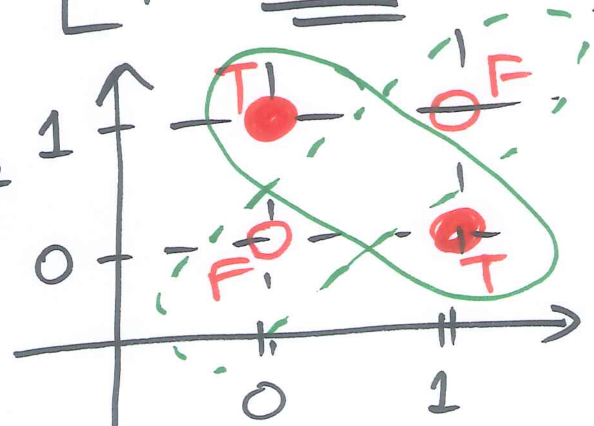
$x_1 \in \{0,1\}$ $x_2 \in \{0,1\}$

$[x_1 \ \underline{\text{AND}} \ x_2]$



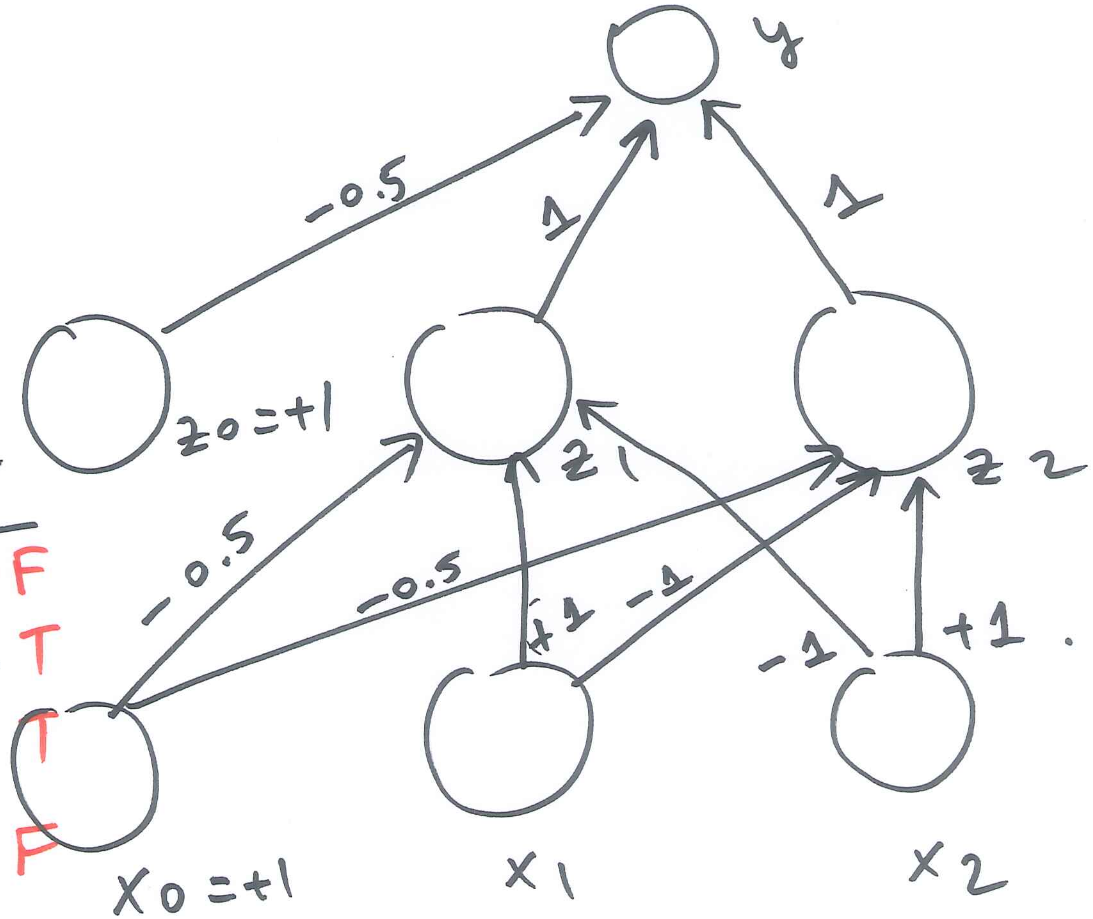


$[x_1 \ \underline{\text{XOR}} \ x_2]$



x_1	x_2	z_1	z_2	y
0	0	0	0	0 ✓
0	1	0		
1	0			
1	1			

x_1	x_2	z_1	z_2	y	
0	0	0	0	0	F
0	1	0	1	1	T
1	0	1	0	1	T
1	1	0	0	0	F



$x_1 = 0 \quad x_2 = 0$

$z_1 = -0.5 + 1.0 - 1.1 = -1.5$

$z_2 = -0.5 - 1.0 + 1.1 = 0.5$

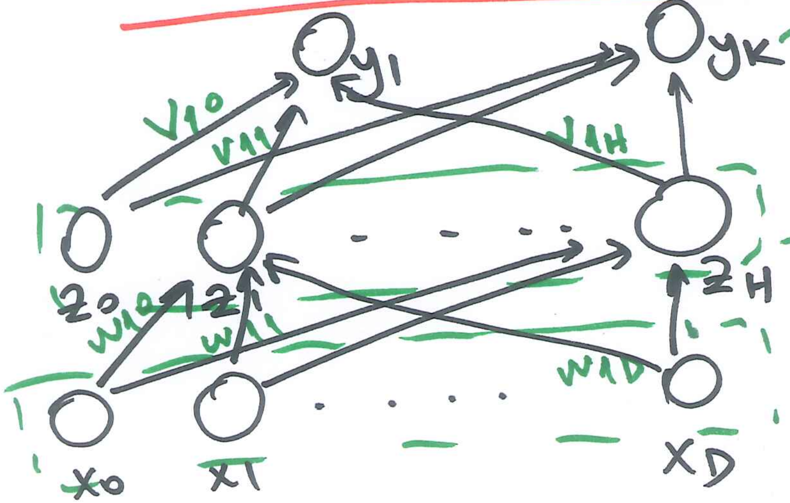
$y = -0.5 + 1.0 + 1.1 = +0.5$

$x_1 = 1 \quad x_2 = 1$

$z_1 = -0.5 + 1.1 - 1.1 = -0.5$

$z_2 = -0.5 - 1.1 + 1.1 = -0.5$

MULTILAYER PERCEPTRON



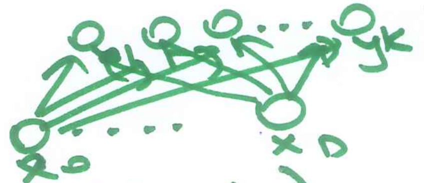
output layer (2nd layer)

$K(H+1)$ weights

$H(D+1)$ weights

hidden layer (1st layer)

input layer.



$K(D+1)$

$$z_h = \text{sigmoid}(w_h^T . x) = \frac{1}{1+\exp(-w_h^T . x)}$$

$$y_c = v_c^T \cdot z$$

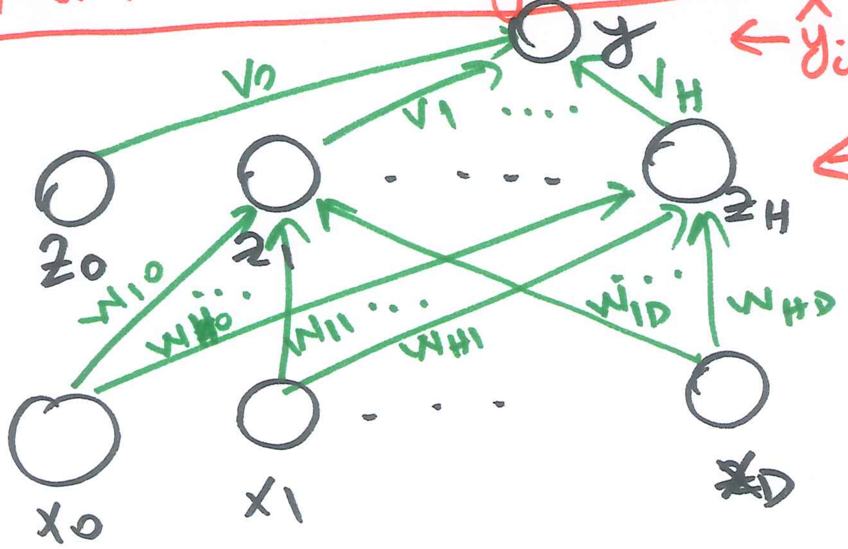
$z_1 = w_1^T . x$ $\quad z_2 = w_2^T . x \; \ldots\ldots$

$y_1 = v_1^T \cdot [w_1^T . x \quad w_2^T . x \; \ldots\ldots]^T.$

$y_1 = \underbrace{\boxed{v_{11} . w_1^T}}_{\tilde{w}_1^T} . x + \underbrace{\boxed{v_{12} . w_2^T}}_{\tilde{w}_2^T} . x + \ldots\ldots$

$$\frac{\partial \text{Error}}{\partial w_{hd}} = \frac{\partial \text{Error}}{\partial \hat{y_c}} \cdot \frac{\partial \hat{y_c}}{\partial z_h} \frac{\partial z_h}{\partial w_{hd}}$$

Nonlinear Regression



$\leftarrow \hat{y}_i \in \mathbb{R}$

$\leftarrow z_i \in \mathbb{R}^H$

$\leftarrow x_i \in \mathbb{R}^D$

$$\hat{y}_i = v^T \cdot z_i = \sum_{h=1}^{H} v_h \cdot z_{ih} + v_0$$

$$z_{ih} = \text{sigmoid}(w_h^T x_i)$$

$$W = ?$$

$$v = ?$$

$$\text{Error}(W, v \mid \mathcal{X}) = \frac{1}{2} \sum_{i=1}^{N} (y_i - \underline{\underline{\hat{y}_i}})^2 = \frac{1}{2} \sum_{i=1}^{N} \text{Error}_i$$

$$= \frac{1}{2} \sum_{i=1}^{N} \left(y_i - \left[\sum_{k=1}^{H} v_k \cdot z_{ik} + v_0 \right] \right)^2$$

$$\frac{\partial \text{Error}}{\partial v_h} = \cancel{\frac{1}{2}} \cdot \cancel{2} \cdot \sum_{i=1}^{N} \left(y_i - \left[\sum_{k=1}^{H} v_k \cdot z_{ik} + v_0 \right] \right) \cdot -z_{ih}$$

$$\boxed{= - \sum_{i=1}^{N} (y_i - \hat{y}_i) \cdot z_{ih}}$$

$$\frac{\partial \left(\sum_{i=1}^{N} x_i \cdot a_i \right)}{\partial x_3} = a_3$$

$$\Delta v_h = \eta \sum_{i=1}^{N} (y_i - \hat{y}_i) \cdot z_{ih}$$

$$\frac{\partial Error}{\partial w_{hd}} = \boxed{\sum_{i=1}^{N}} \underbrace{\frac{\partial Error_i}{\partial \hat{y}_i}}_{= -(y_i - \hat{y}_i)} \cdot \underbrace{\frac{\partial \hat{y}_i}{\partial z_{ih}}}_{v_h} \cdot \underbrace{\frac{\partial z_{ih}}{\partial w_{hd}}}_{z_{ih}.(1-z_{ih})\, x_{id}}$$

$$Error_i = \frac{1}{2}(y_i - \hat{y}_i)^2$$

$$\hat{y}_i = \sum_{k=1}^{H} v_k . z_{ik} + v_0$$

$$z_{ih} = sigmoid(\underbrace{w_h^T . x_i}_{\sum_{d=1}^{D} w_{hd} . x_{id}})$$

$$\boxed{\frac{\partial Error}{\partial w_{hd}} = \sum_{i=1}^{N} (y_i - \hat{y}_i) . v_h . z_{ih}(1 - z_{ih}) . x_{id}}$$

$$\Delta w_{hd} = \eta \sum_{i=1}^{N} (y_i - \hat{y}_i) . v_h . z_{ih}(1 - z_{ih}) . x_{id}$$

Binary Classification

$$\hat{y}_i = sigmoid\left(\sum_{h=1}^{H} v_h . z_{ih} + v_0\right)$$

$$Error(w, v \mid \mathcal{X}) = -\sum_{i=1}^{N} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right]$$

$$\Delta w_{hd} = \eta . \sum_{i=1}^{N} (y_i - \hat{y}_i) . v_h . z_{ih} . (1 - z_{ih}) . x_{id}$$

$$\Delta v_h = \eta . \sum_{i=1}^{N} (y_i - \hat{y}_i) . z_{ih} .$$