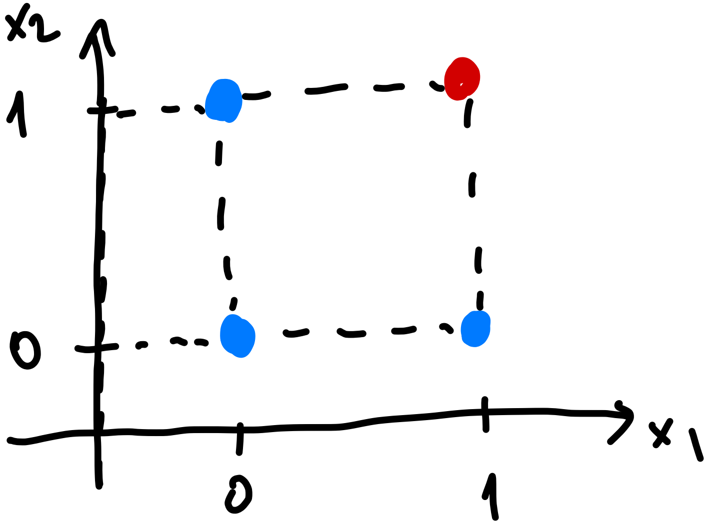
Boolean Functions

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

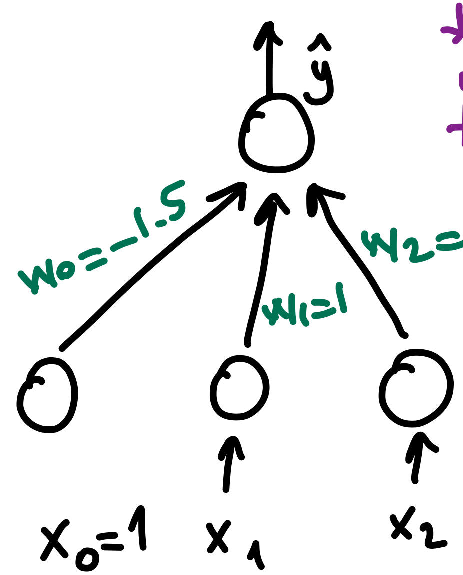
AND FUNCTION $[x_1 \text{ AND } x_2]$

x_1	x_2	x_1 AND x_2
0	0	0
0	1	0
1	0	0
1	1	1



$$s(a) = \begin{cases} 1 & \text{if } a > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{y} = s\big(\underbrace{w_0 + w_1 . x_1 + w_2 . x_2}_{\text{total signal received at the output layer.}}\big)$$

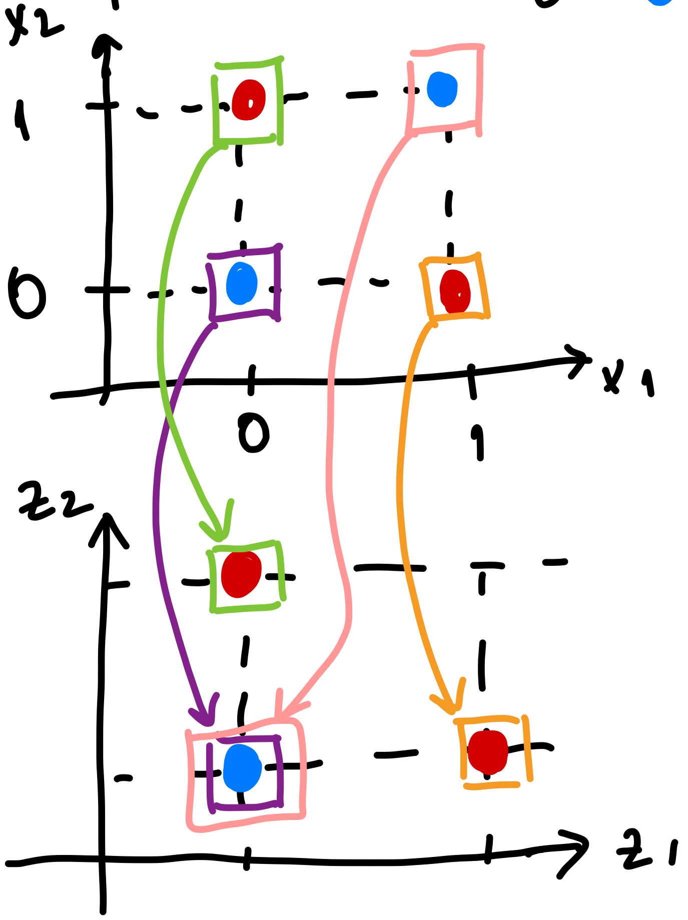


x_1	x_2	$\hat{y}$
0	0	$s(-1.5 + 1\times0 + 1\times0) = 0$
0	1	$s(-1.5 + 1\times0 + 1\times1) = 0$
1	0	$s(-1.5 + 1\times1 + 1\times0) = 0$
1	1	$s(-1.5 + 1\times1 + 1\times1) = 1$

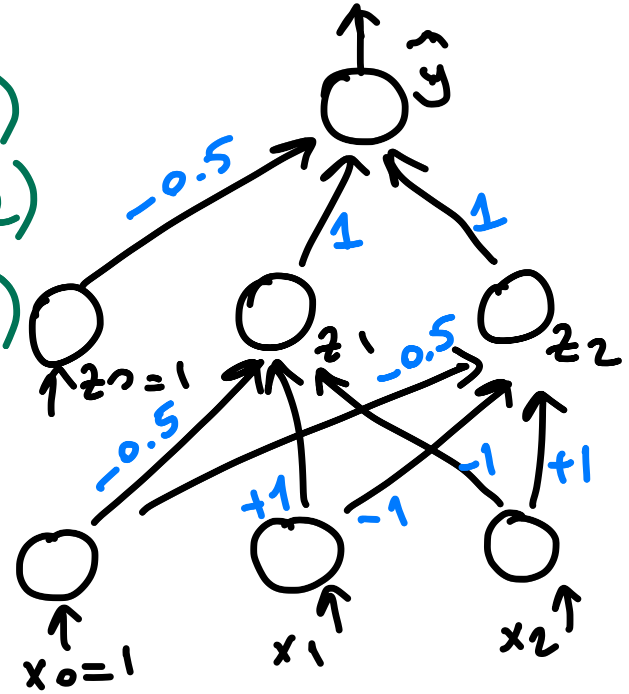
XOR FUNCTION [x_1 XOR x_2]

x_1	x_2	x_1 XOR x_2
0	0	0
0	1	1
1	0	1
1	1	0

$$w_0 + (w_1)\times(0) + (w_2)\times 0 \leq 0$$
$$w_0 + (w_1)\times(0) + (w_2)\times(1) > 0$$
$$w_0 + (w_1)\times(1) + (w_2)\times(0) > 0$$
$$w_0 + (w_1)\times(1) + (w_2)\times(1) \leq 0$$

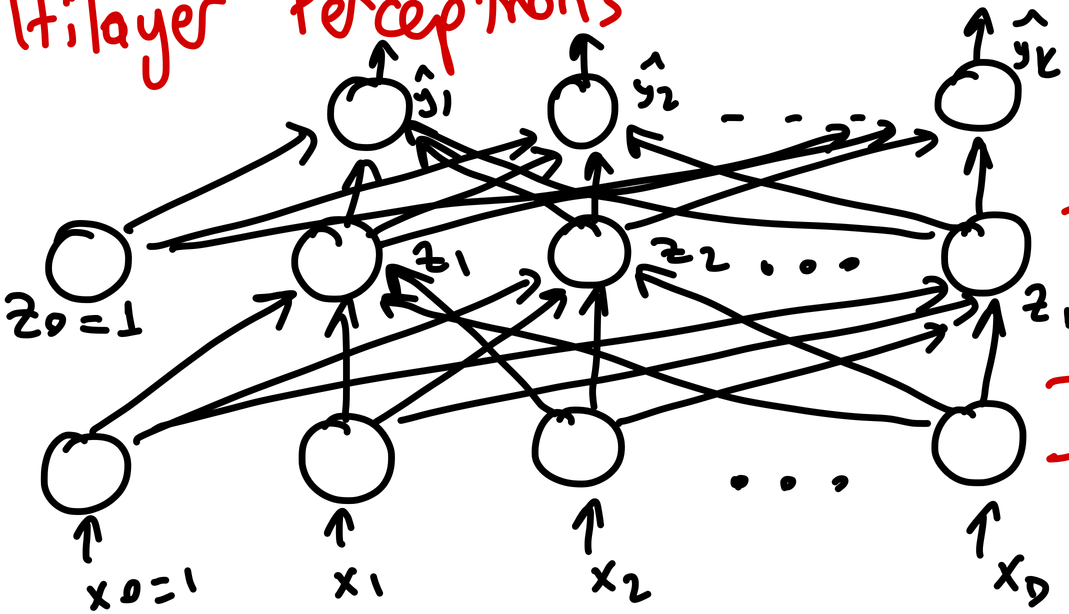


$$z_1 = s(-0.5 + x_1 - x_2)$$
$$z_2 = s(-0.5 - x_1 + x_2)$$
$$\hat{y} = s(-0.5 + z_1 + z_2)$$



x_1	x_2	z_1	z_2	$\hat{y}$
0	0	0	0	0
0	1	0	1	1
1	0	1	0	1
1	1	0	0	0

Multilayer Perceptrons



] output layer

] hidden layer

] input layer.

$\underbrace{(H+1) \times K}_{V}$

$\underbrace{(D+1) \times H}_{W}$

$$\# \text{ of model parameters} = \begin{cases} \text{perceptron} & (D+1) \times K \\ \text{multilayer perceptron} & (D+1) \times H + (H+1) \times K \end{cases}$$

$z = f(x)$

$\hat{y} = g(z)$

$\hat{y} = g(f(x))$

hidden nodes $\Rightarrow z_h = s_1(w_h^T \cdot x)$

$\hookrightarrow$ act. function at the hidden layer

output nodes $\Rightarrow \hat{y}_k = s_2(v_k^T \cdot z)$

$\hookrightarrow$ act. function at the output layer

Multiclass Classification

$\mathcal{X} = \{(x_i, y_i)\}$ $\quad x_i \in \mathbb{R}^D$ $\quad y_i \in \{1, 2, \ldots K\}$

$s_1 \Rightarrow$ sigmoid $\qquad s_2 \Rightarrow$ softmax

$$z_h = \text{sigmoid}(w_h^T . x) \qquad \hat{y}_c = \text{softmax}(v_c^T . z)$$

$$\text{Error}_i = -\sum_{c=1}^{K} y_{ic} . \log(\hat{y}_{ic}) \Rightarrow \hat{y}_{ic} = \text{softmax}\left[v_c^T . \left[\quad \right] \right]$$

$$\frac{\partial \text{Error}_i}{\partial w_{hd}} = \boxed{\frac{\partial \text{Error}_i}{\partial \hat{y}_{ic}}} \cdot \frac{\partial \hat{y}_{ic}}{\partial z_{ih}} \cdot \frac{\partial z_{ih}}{\partial w_{hd}}$$

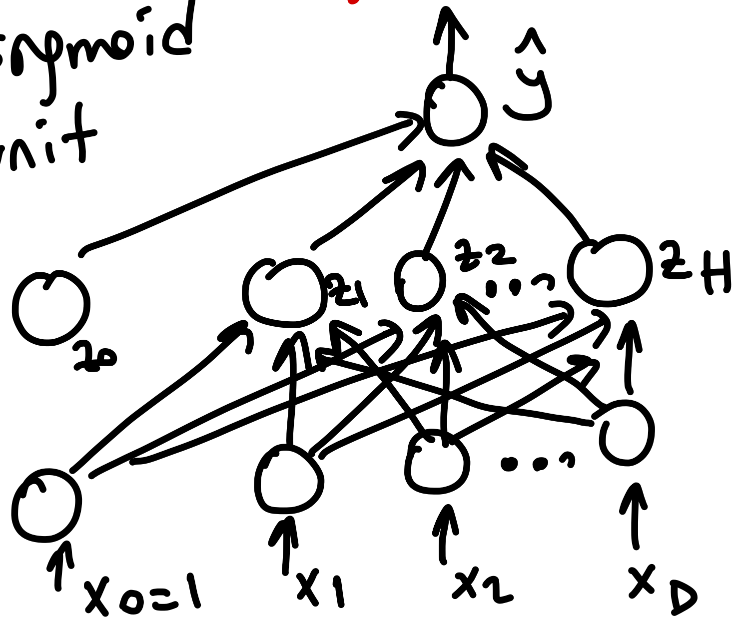
$$\frac{\partial \text{Error}_i}{\partial v_{ch}} = \boxed{\frac{\partial \text{Error}_i}{\partial \hat{y}_{ic}}} \cdot \frac{\partial \hat{y}_{ic}}{\partial v_{ch}}$$

Nonlinear Regression

$$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N} \qquad x_i \in \mathbb{R}^D \quad y_i \in \mathbb{R}$$

$s_1 \Rightarrow$ sigmoid

$s_2 \Rightarrow$ unit



$$\boxed{\hat{y}_i = v^T . z_i}$$

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

$$\boxed{\hat{y}_i = \sum_{k=1}^{H} v_k . z_{ik} + v_0 . z_{i0}}$$

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

$$\frac{\partial \text{Error}_i}{\partial v_h} = \frac{\partial \left[\frac{1}{2}(y_i - \hat{y}_i)^2\right]}{\partial v_h} = \frac{1}{2} . 2 . (y_i - v^T . z_i) . (-z_{ih})$$

$$= -(y_i - \hat{y}_i) . z_{ih}$$

$$\frac{\partial \text{Error}_i}{\partial w_{hd}} = \underbrace{\frac{\partial \text{Error}_i}{\partial \hat{y}_i}}_{-\frac{1}{2} . 2 . (y_i - \hat{y}_i)} . \underbrace{\frac{\partial \hat{y}_i}{\partial z_{ih}}}_{v_h} . \underbrace{\frac{\partial z_{ih}}{\partial w_{hd}}}_{z_{ih}(1 - z_{ih}) . x_{id}}$$

$$\Delta v_h = \eta\,(y_i - \hat{y}_i)\,.\,z_{ih}$$

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

Binary Classification

$s_1 \Rightarrow$ sigmoid $\quad s_2 \Rightarrow$ sigmoid

$$\hat{y}_i = \text{sigmoid}\,(\underbrace{v^T.z_i}_{\sum_{k=1}^{H} v_k.z_{ik} + v_0})$$

$$z_{ih} = \text{sigmoid}\,(\underbrace{w_h^T.x_i}_{\sum_{d=1} w_{hd}.x_{id} + w_{h0}})$$

$$\text{Error}_i = -\left[y_i.\log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right]$$

$$\frac{\partial \text{Error}_i}{\partial v_h} = \underbrace{\frac{\partial \text{Error}_i}{\partial \hat{y}_i}}_{-\left[y_i \frac{1}{\hat{y}_i} + (1-y_i)\frac{(-1)}{(1-\hat{y}_i)}\right]} \cdot \underbrace{\frac{\partial \hat{y}_i}{\partial v_h}}_{\hat{y}_i(1-\hat{y}_i)z_{ih}}$$

$$= -\left[y_i \cdot (1-\hat{y}_i) + (1-y_i)(-\hat{y}_i) \right] \cdot z_{ih}$$

$$= -\left[y_i - \cancel{y_i \hat{y}_i} - \hat{y}_i + \cancel{y_i \hat{y}_i} \right] \cdot z_{ih}$$

$$= -\left[y_i - \hat{y}_i \right] \cdot z_{ih}$$

Exercise: show that

$$\frac{\partial \, Error_i}{\partial w_{hd}} = -(y_i - \hat{y}_i) \cdot v_h \cdot z_{ih} \left[1 - z_{ih} \right] \cdot x_{id}$$