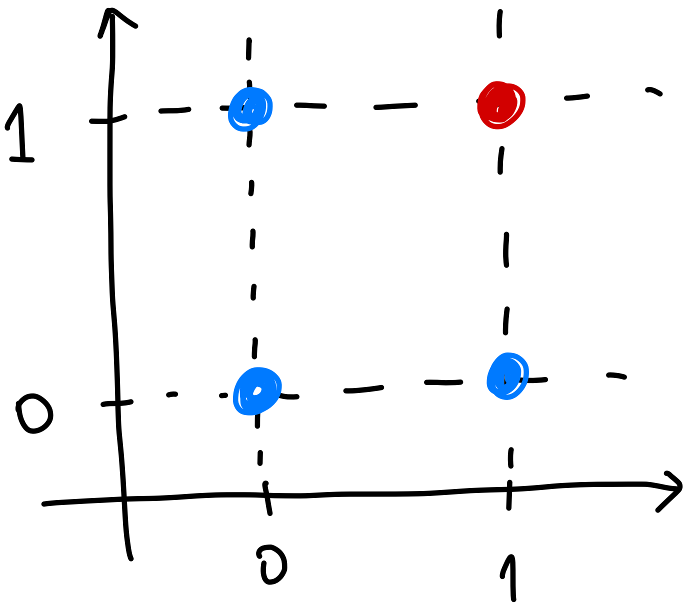
Boolean Functions

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

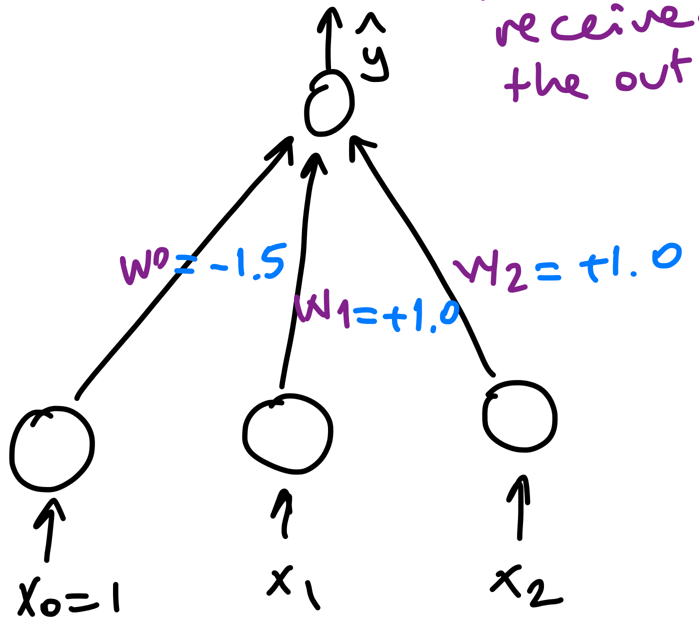
AND FUNCTION [x_1 AND x_2]

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



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

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



x_1	x_2	$\hat{y}$ ($w_0 = -1.5$, $w_1 = 1.0$, $w_2 = 1.0$)
0	0	$s(w_0 + w_1 . 0 + w_2 . 0) = 0$
0	1	$s(w_0 + w_1 . 0 + w_2 . 1) = 0$
1	0	$s(w_0 + w_1 . 1 + w_2 . 0) = 0$
1	1	$s(w_0 + w_1 . 1 + w_2 . 1) = 1$

XOR FUNCTION [x_1 XOR x_2]

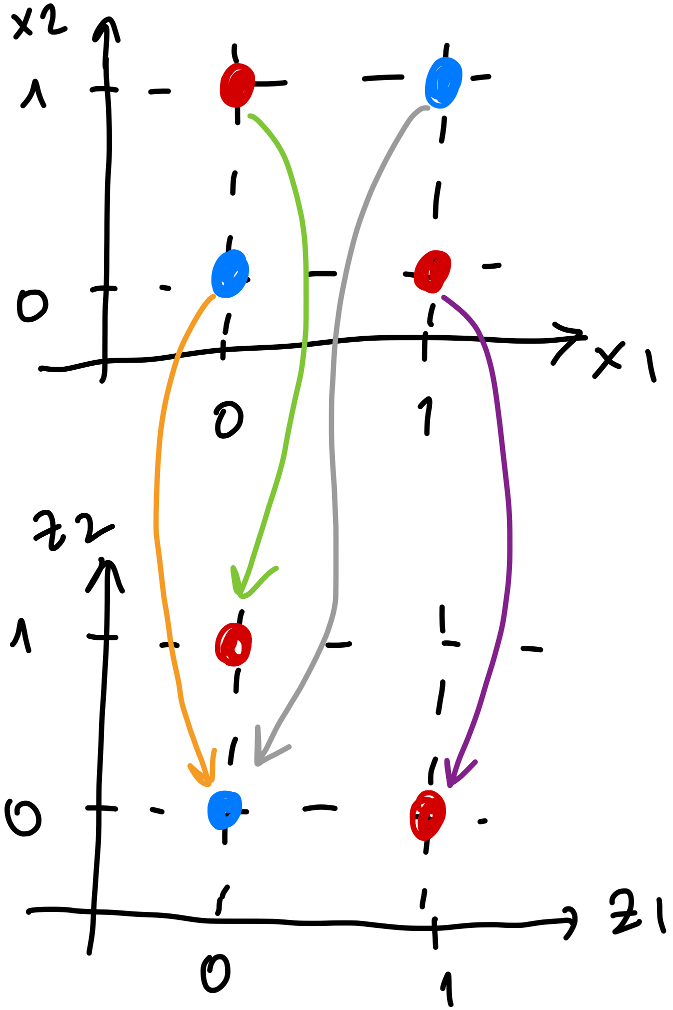
x_1	x_2	x_1 XOR x_2
0	0	0
0	1	1
1	0	1
1	1	0

$$w_0 + \cancel{w_1.(0)} + \cancel{w_2(0)} \leq 0$$

$$w_0 + w_1.(0) + w_2(1) > 0$$

$$w_0 + w_1.(1) + w_2(0) > 0$$

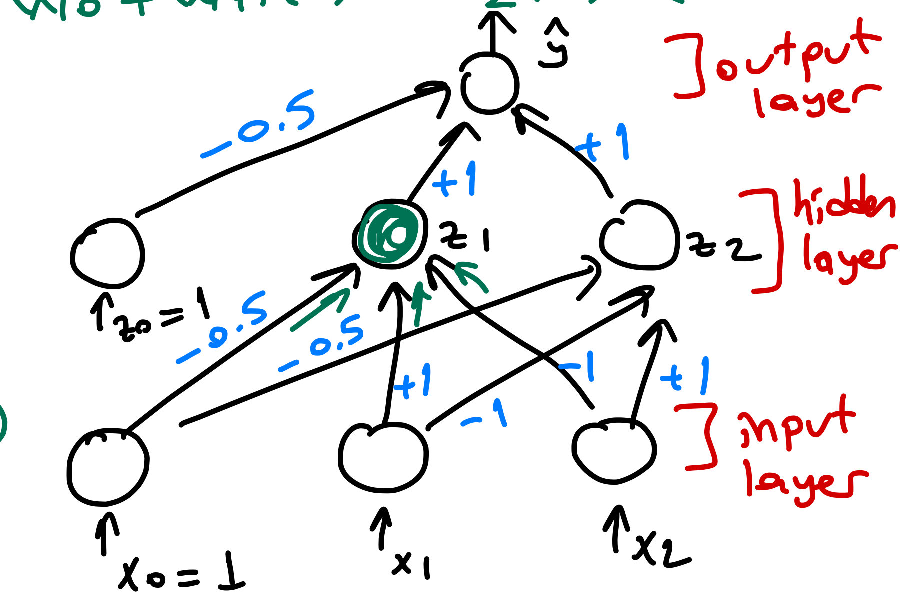
$$w_0 + w_1.(1) + w_2(1) \leq 0$$



$$z_1 = s(-0.5 + x_1 - x_2)$$

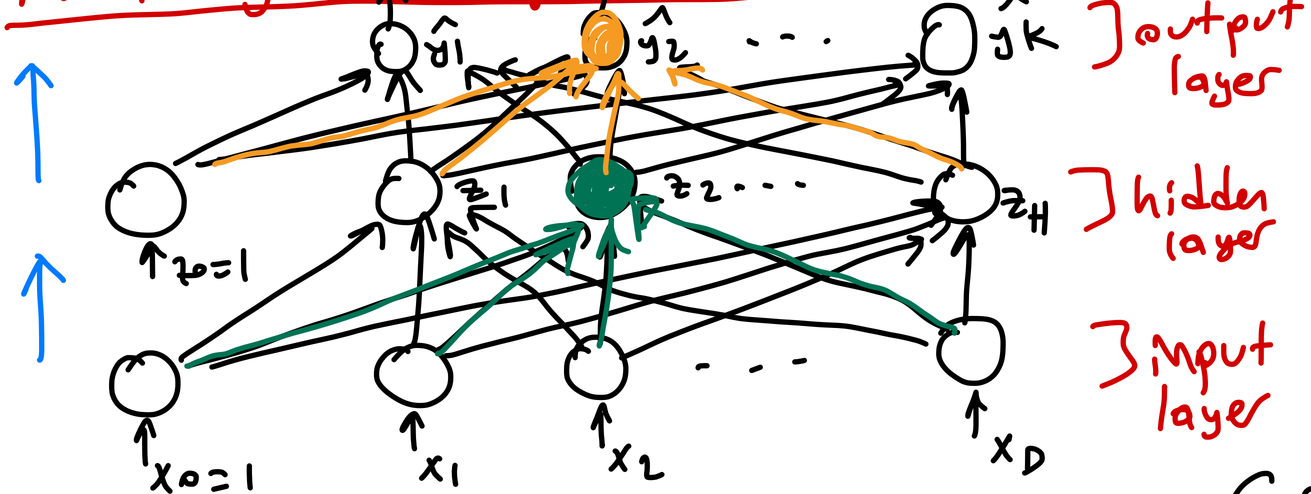
$$z_2 = s(-0.5 - x_1 + x_2)$$

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



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



$\}\ (H+1) \times K \Rightarrow V$

$\}\ (D+1) \times H \Rightarrow W$

$z = f(x)$

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

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

\# of model parameters $\Rightarrow$
- perceptron $\Rightarrow (D+1) \times K$
- multilayer perceptron $\Rightarrow (D+1) \times H + (H+1) \times K$

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

↳ activation func. at the hidden layer

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

↳ activation func. at the output layer

Multiclass classification

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

$s_1 \Rightarrow$ sigmoid

$\boxed{s_2 \Rightarrow \text{softmax}}$

$$z_h = \text{sigmoid}(w_h^T . x)$$

$$\hat{y}_c = \text{softmax}(v_c^T . z)$$

constant

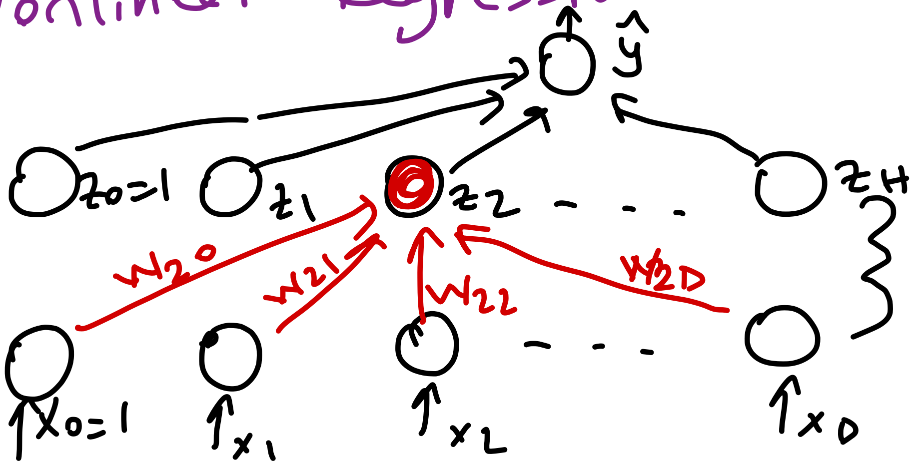
$$\text{Error}_i = - \sum_{c=1}^{K} y_{ic} \log(\hat{y}_{ic})$$

$$\Rightarrow \hat{y}_{ic} = \text{softmax}\left[v_c^T . \begin{bmatrix} \text{sigmoid}(w_1^T . x_i) \\ \vdots \\ \text{sigmoid}(w_H^T . x_i) \end{bmatrix}\right]$$

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

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

Nonlinear Regression



$s_1 \Rightarrow$ sigmoid $\quad \boxed{s_2 \Rightarrow \text{linear}}$.

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

$$z_{ih} = \text{sigmoid}\Big(\overbrace{\underbrace{w_h^T}_{w_2^T} . x_i}^{\sum_{d=1}^{D} w_{hd}.x_{id}}\Big)$$

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

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

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

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

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

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

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

Binary Classification

$s_1 \Rightarrow$ sigmoid

$$\boxed{s_2 \Rightarrow \text{sigmoid}}$$

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

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

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

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

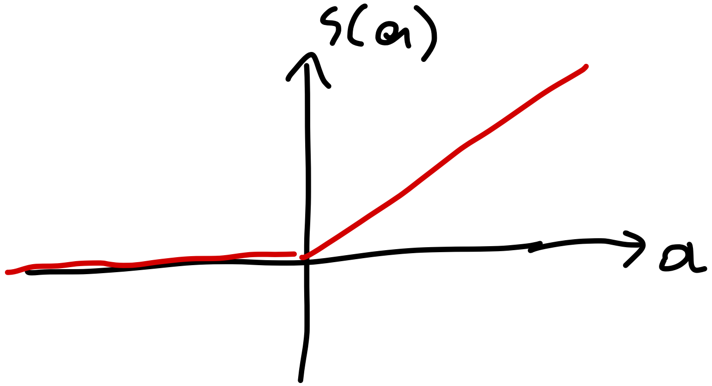
$$= -\left[y_i(1-\hat{y}_i) + (1-y_i)(-\hat{y}_i) \right] . z_{ih}$$

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

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

Exercise

$$\frac{\partial Error_i}{\partial w_{hd}} = -(y_i - \hat{y}_i) . v_h . z_{ih} . (1 - z_{ih}) . x_{id}$$



ReLU Rectified Linear Unit

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

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