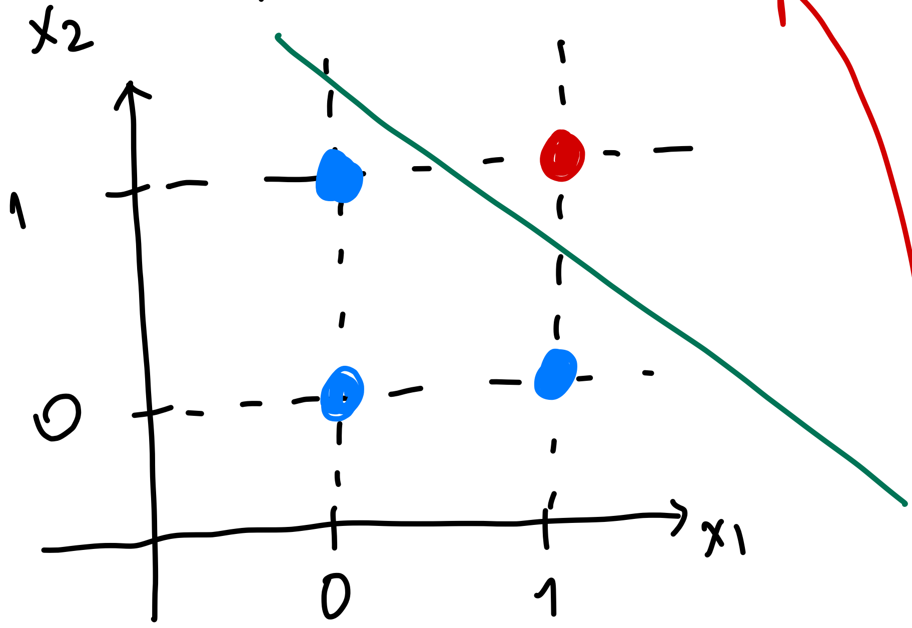
Boolean Functions

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

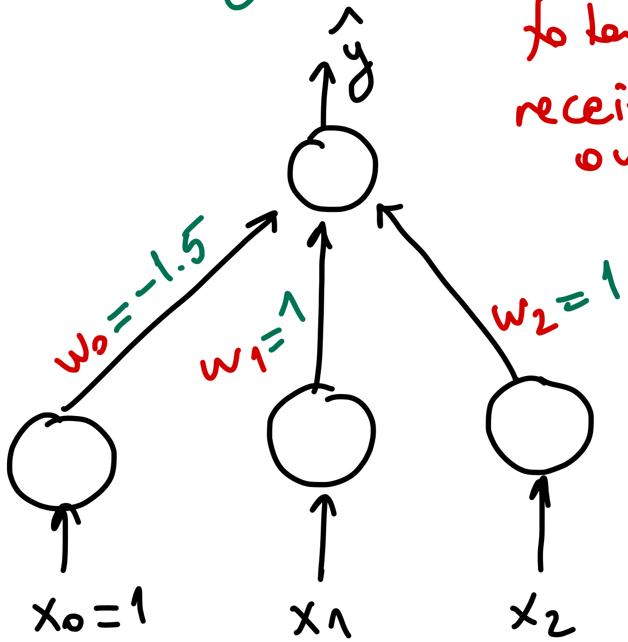
AND Function [x_1 AND x_2]

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



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

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



x_1	x_2	$\hat{y}$
0	0	$s(w_0 + w_1 . 0 + w_2 . 0) = 0$
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

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

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

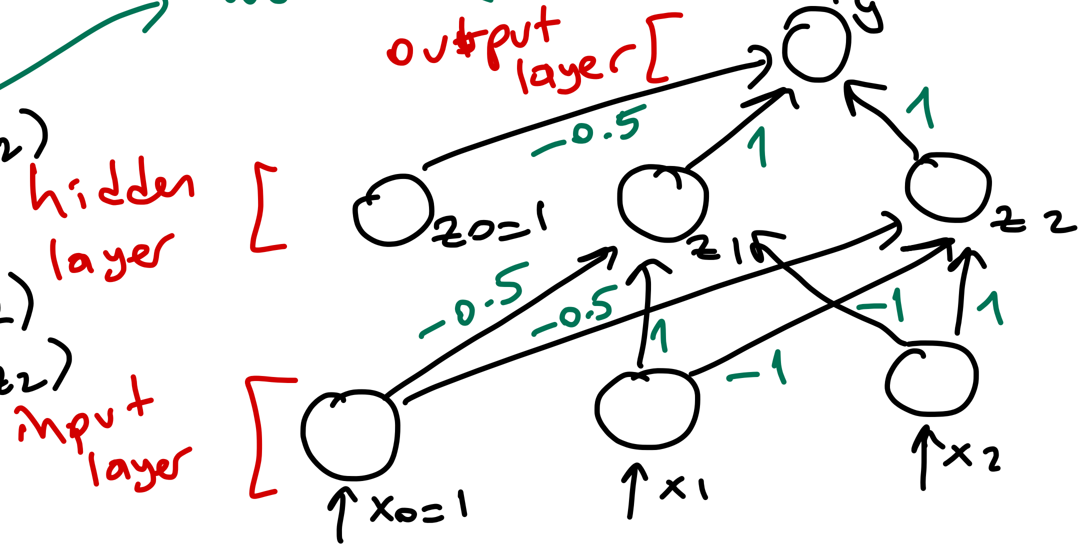
$w_0 + w_1.(1) + w_2(0) > 0$

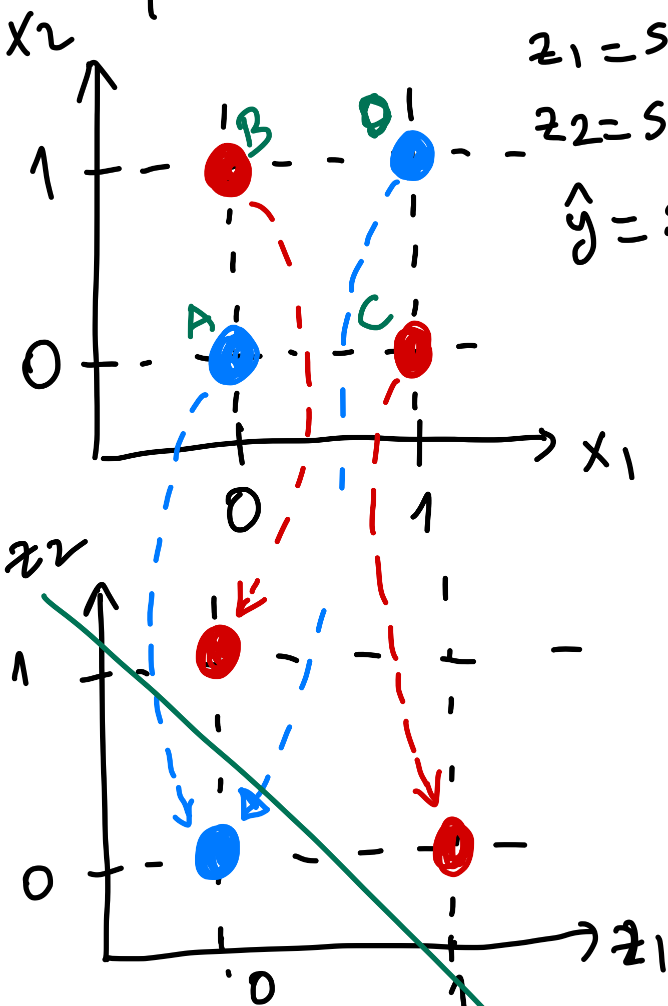
$w_0 + w_1(1) + w_2.(1) \leq 0$

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

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

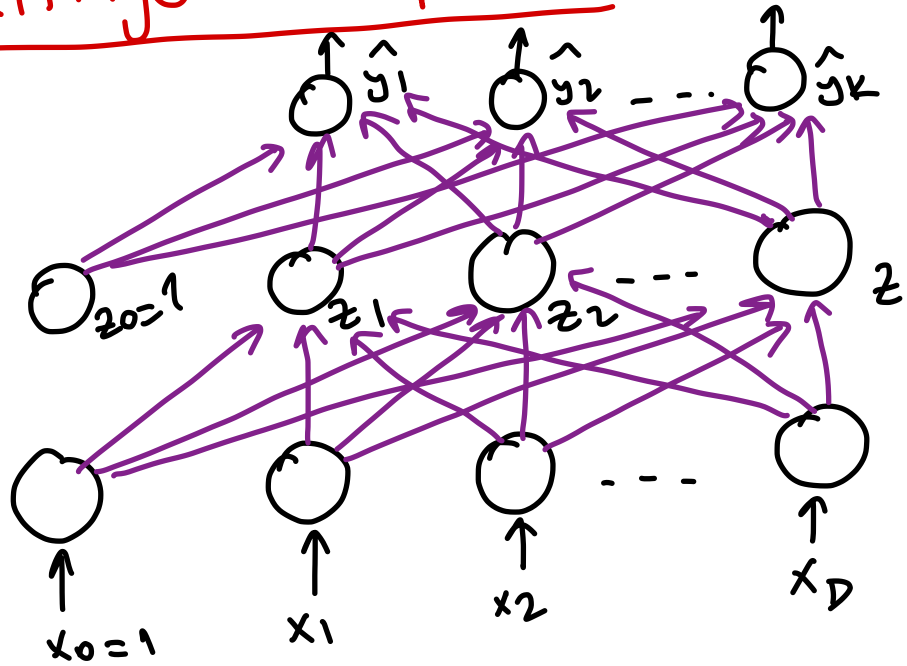
$\hat{y} = s(-0.5 + z_1 + z_2)$





	x_1	x_2	z_1	z_2	$\hat{y}$
A	0	0	0	0	0
B	0	1	0	1	1
C	1	0	1	0	1
D	1	1	0	0	0

Multilayer Perceptrons



] output layer } $(H+1).K \Rightarrow V$

] hidden layer } $(D+1)H \Rightarrow W$

] input layer.

$\#$ of parameters $\Rightarrow$ $\begin{cases} \text{perceptron} \Rightarrow (D+1)K \\ \text{multilayer perceptron} \Rightarrow (D+1).H + (H+1).K \end{cases}$

hidden nodes $\Rightarrow$ $z_h = s_1\left(\overbrace{w_h^T}^{1\times(D+1)} . \overbrace{x}^{(D+1)\times 1}\right)$

↳ activation function at the hidden layer

output nodes $\Rightarrow$ $\hat{y}_k = s_2\left(\overbrace{v_k^T}^{1\times(H+1)} . \overbrace{z}^{(H+1)\times 1}\right)$

↳ activation function at the output layer

Multiclass Classification

$S_1 \Rightarrow$ sigmoid $\qquad S_2 \Rightarrow$ softmax

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

$\{(x_i, y_i)\}_{i=1}^N$

$x_i \in \mathbb{R}^D \quad y_i \in \{1, 2, .., K\}$

y_{ic} / one-hot-encoding.

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

$W \in \mathbb{R}^{H \times (D+1)}$

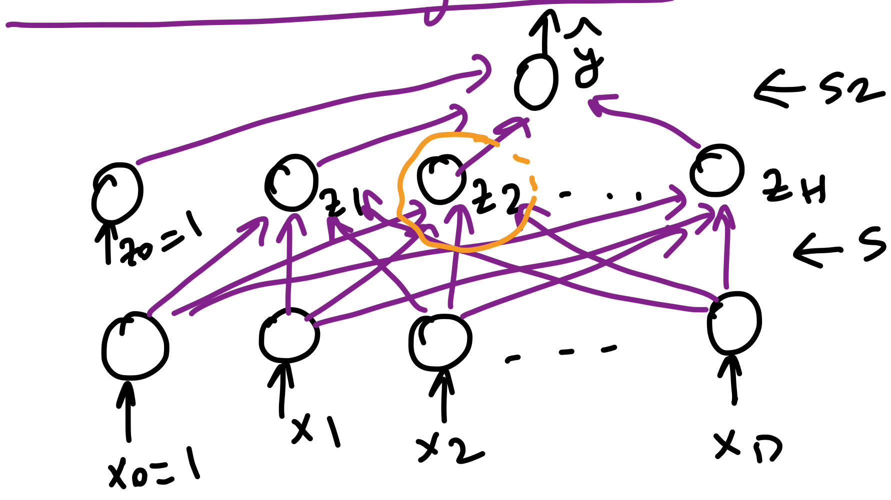
$$\frac{\partial \text{Error}_i}{\partial w_{hd}} = \boxed{\frac{\partial \text{Error}_i}{\partial \hat{y}_{ic}}} \frac{\partial \hat{y}_{ic}}{\partial z_{ih}} \frac{\partial z_{ih}}{\partial w_{hd}}$$

$V \in \mathbb{R}^{K(H+1)}$

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

Hint: $\frac{\partial f(z)}{\partial x} = \frac{\partial f(z)}{\partial z} \frac{\partial z}{\partial x}$ [Chain Rule]

Nonlinear Regression



$s_1 \Rightarrow$ sigmoid

$\leftarrow \hat{y}_i \in \mathbb{R}^1$

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

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

$s_2 \Rightarrow$ linear

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

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

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

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

$$\frac{\partial \text{Error}_i}{\partial v_h} = \frac{1}{2} \cdot (y_i - v^T \cdot z_i)^{②} = \frac{1}{2}\left(y_i - \left[\sum_{k=1}^{H} v_k \cdot z_{ik} + v_0\right]\right)^2$$

$$= \frac{1}{2} \cdot 2 \cdot (y_i - v^T \cdot z_i) \cdot (-z_{ih})$$

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

$v_1 \cdot z_{i1} + \dots + v_h \cdot z_{ih} + \dots + v_H \cdot z_{iH}$

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

$$\hat{y}_i = v^T \cdot z_i = \sum_{k=1}^{H} v_k \cdot z_{ik} + v_0$$

$$w_h^T \cdot x_i = \sum_{d=1}^{D} w_{hd} \cdot x_{id} + x_{i0}$$

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

$$\boxed{\begin{aligned} \Delta v_h &= \eta \cdot (y_i-\hat{y}_i)\cdot z_{ih} \\ \Delta w_{hd} &= \eta \cdot (y_i-\hat{y}_i)\cdot v_h \cdot z_{ih}\cdot(1-z_{ih})\cdot x_{id} \end{aligned}}$$

Binary classification

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

$$\hat{y}_i = \text{sigmoid}(v^T \cdot z_i)$$

$$v^T \cdot z_i = \sum_{k=1}^{H} v_k \cdot z_{ik} + v_0$$

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

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

$$\frac{\partial \text{Error}_i}{\partial v_h} = \frac{\partial \text{Error}_i}{\partial \hat{y}_i} \boxed{\frac{\partial \hat{y}_i}{\partial v_h}}$$

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

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

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

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

Exercise

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