Boolean Functions

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

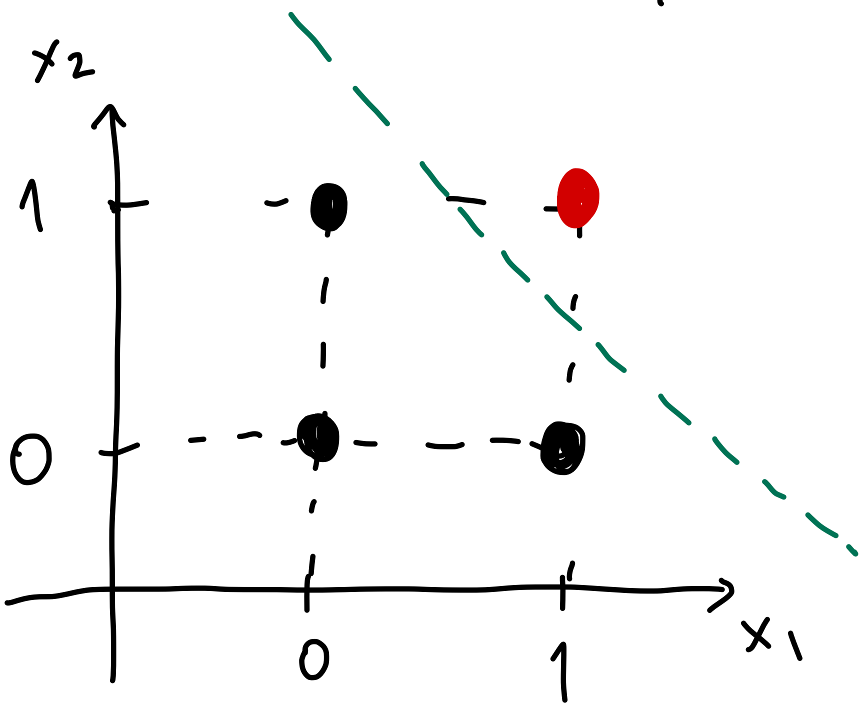
AND Function [x_1 AND x_2]

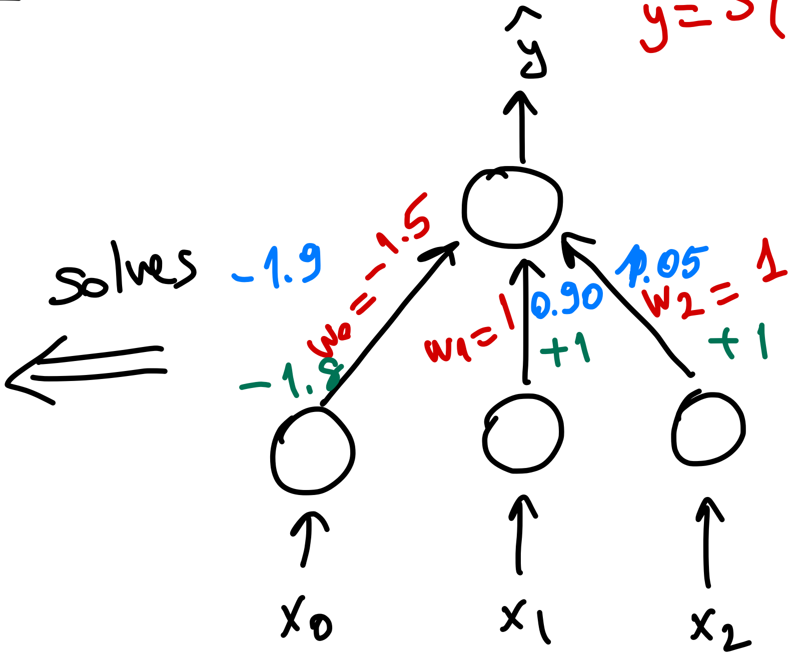
x_1	x_2	x_1 AND x_2
0	0	0
0	1	0
1	0	0
1	1	1

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

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

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



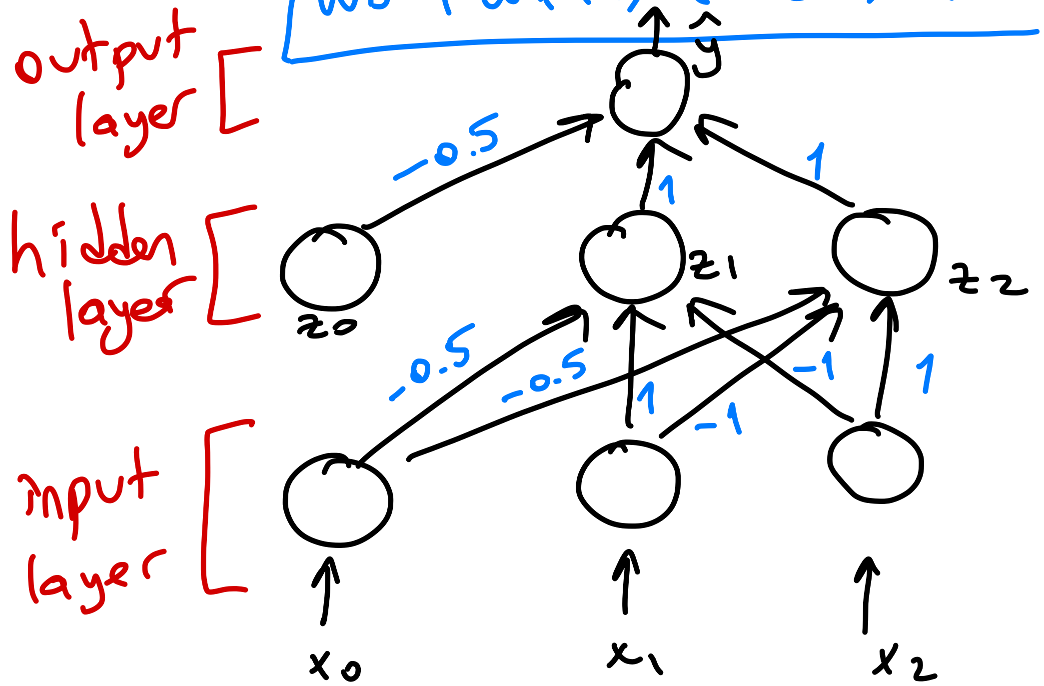


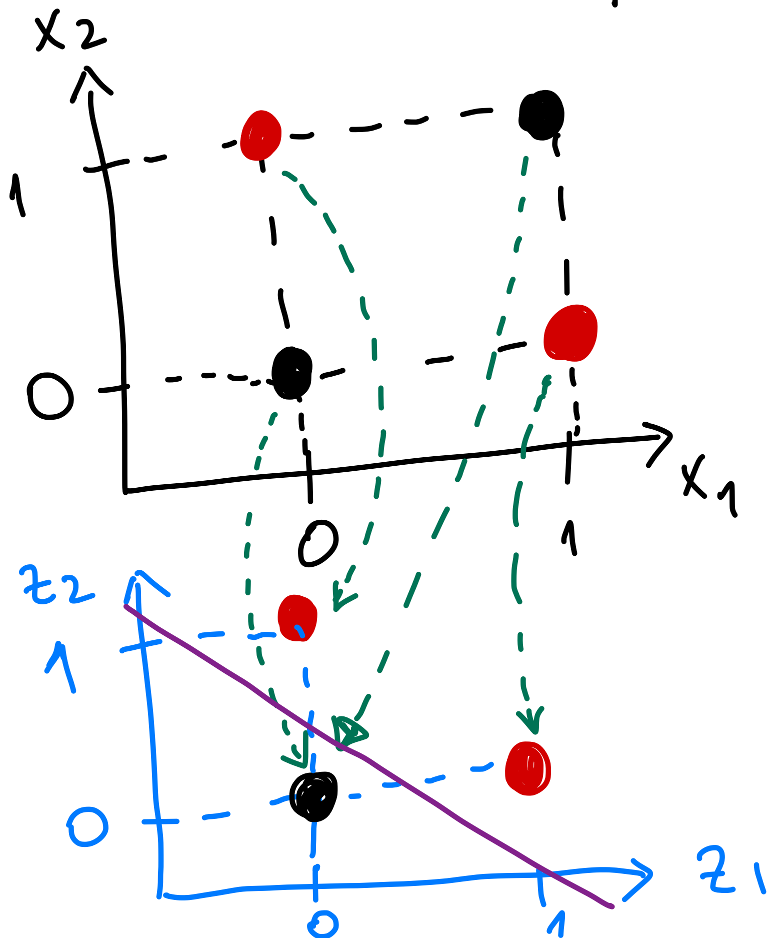
x_1	x_2	$\hat{y}$
0	0	0
0	1	0
1	0	0
1	1	1

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

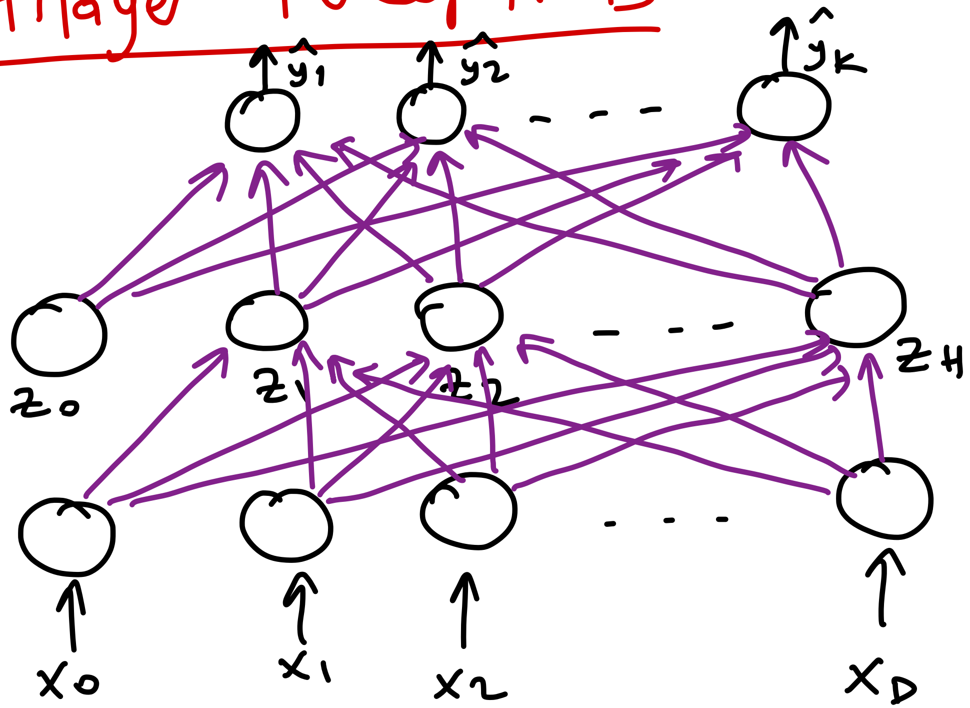
$$w_0 + w_1(0) + w_2(0) \le 0$$
$$w_0 + w_1(0) + w_2(1) > 0$$
$$w_0 + w_1(1) + w_2(0) > 0$$
$$w_0 + w_1(1) + w_2(1) \le 0$$





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
(H+1) K weights

] hidden layer
(D+1) H weights

] input layer

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

sigmoid

$$\text{hidden nodes} \Rightarrow z_h = s_1\Big(\underbrace{w_h^T}_{1\times(D+1)} \cdot \underbrace{x}_{(D+1)\times 1}\Big)$$

softmax

$$\text{output nodes} \Rightarrow \hat{y}_k = s_2\Big[\underbrace{v_k^T}_{1\times(H+1)} \cdot \underbrace{z}_{(H+1)\times 1}\Big]$$

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

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

constant

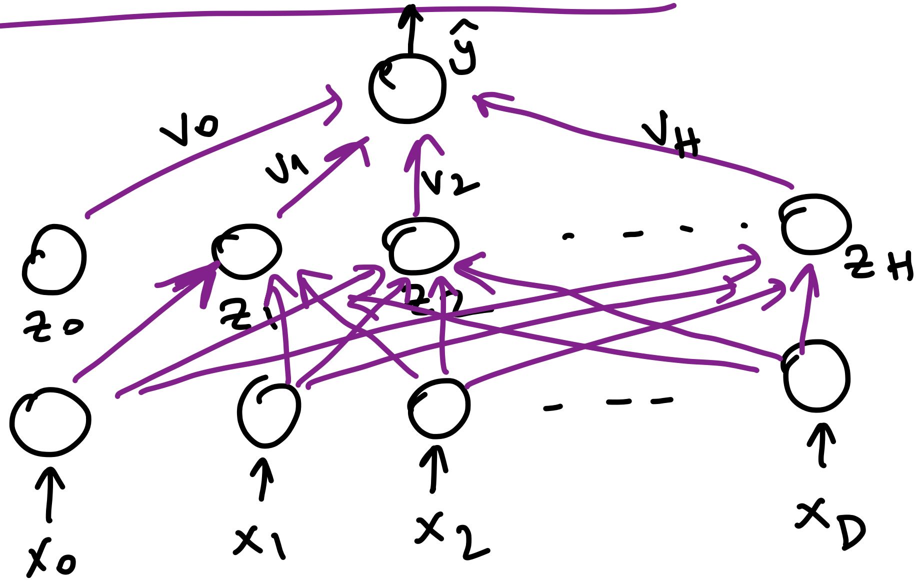
$$\frac{\partial \mathrm{Error}_i}{\partial w_{hd}} = \frac{\partial \mathrm{Error}_i}{\partial \hat{y}_{ic}} \frac{\partial \hat{y}_{ic}}{\partial z_h} \frac{\partial z_h}{\partial w_{hd}}$$

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

Hint: $$\frac{\partial f(z)}{\partial x} = \frac{\partial f(z)}{\partial z} \frac{\partial z}{\partial x}$$

$$f(z) \rightarrow g(x)$$

Nonlinear Regression



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

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

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

$$\hat{y}_i = \underbrace{V^T}_{1\times(H+1)} . \underbrace{z_i}_{(H+1)\times 1}$$

$$z_{ih} = \text{sigmoid}\left(\underbrace{w_h^T}_{1\times(D+1)} . \underbrace{x_i}_{(D+1)\times 1}\right)$$

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

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

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

v_h

$$= \frac{1}{2} . 2 (y_i - V^T . z_i)(-z_{ih}) = -(y_i - \hat{y}_i) z_{ih}$$

$$v_1 . z_{i1} + v_2 . z_{i2} + \dots + v_H . z_{iH} + v_0$$

$$\frac{\partial \left[\sum_{i=1}^{N} a_i . x_i\right]}{\partial x_2} = a_2$$

$a_1 . x_1 + a_2 . x_2 + \dots + a_N . x_N$

$$\frac{\partial Error_i}{\partial w_{hd}} = \frac{\partial Error_i}{\partial \hat{y}_i} \, \frac{\partial \hat{y}_i}{\partial z_{ih}} \cdot \frac{\partial z_{ih}}{\partial w_{hd}}$$

$$-\frac{1}{\cancel{2}}\cancel{2}(y_i-\hat{y}_i) \qquad v_h \qquad z_{ih}(1-z_{ih}) \cdot x_{id}$$

$$z_{ih} = \text{sigmoid}\left(w_h^T \cdot x_i\right) = \text{sigmoid}\left[\sum_{e=1}^{D} w_{he} \cdot x_{ie} + w_{h0}\right]$$

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

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

Binary Classification

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

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

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

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

$$\hat{y}_i \cdot (1-\hat{y}_i) \cdot z_{ih}$$

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

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

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

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

Exercise#7

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