FYS-STK3155/4155, Lecture September 29

FYS-STK3155/4155, Sept 29

$$y_i = p(x_i) + \varepsilon_i$$

$$\int_{x \in D} p(x)\,dx = \underline{1}$$

$$p(x_i) \leq p(x_j) \quad \text{if} \quad x_i \leq x_j$$

cummulative probability

$$P(x) = \int_{-\infty}^{x} p(x')\,dx'$$

$$x \in [a, b] \qquad P(b) = \underline{1}$$

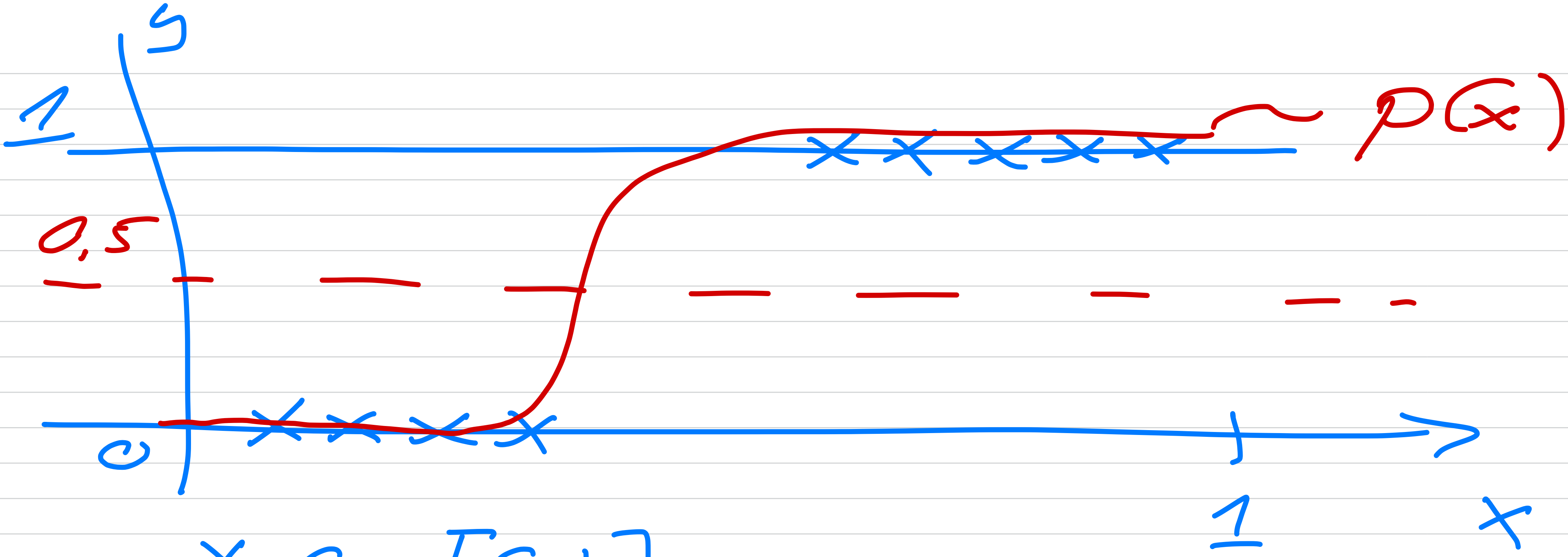
$$\mathbb{E}[x^n] = \int_{x \in D} x^n \, p(x) \, dx$$

Typical output of a binary problem

$$y_i = \{0, 1\}$$

Domain $D = \{(x_0, y_0), (x_1, y_1) \ldots (x_{n-1}, y_{n-1})\}$

all y-values are given by either 0 or 1



$x \in [0,1]$

$y = \{0,1\}$

$y = 0$ if $p(x) \leq 0.5$

$y = 1$ if $p(x) > 0.5$ $\quad (p \leq 1)$

Typical function p (sigmoid, logit etc)

$$p(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}$$

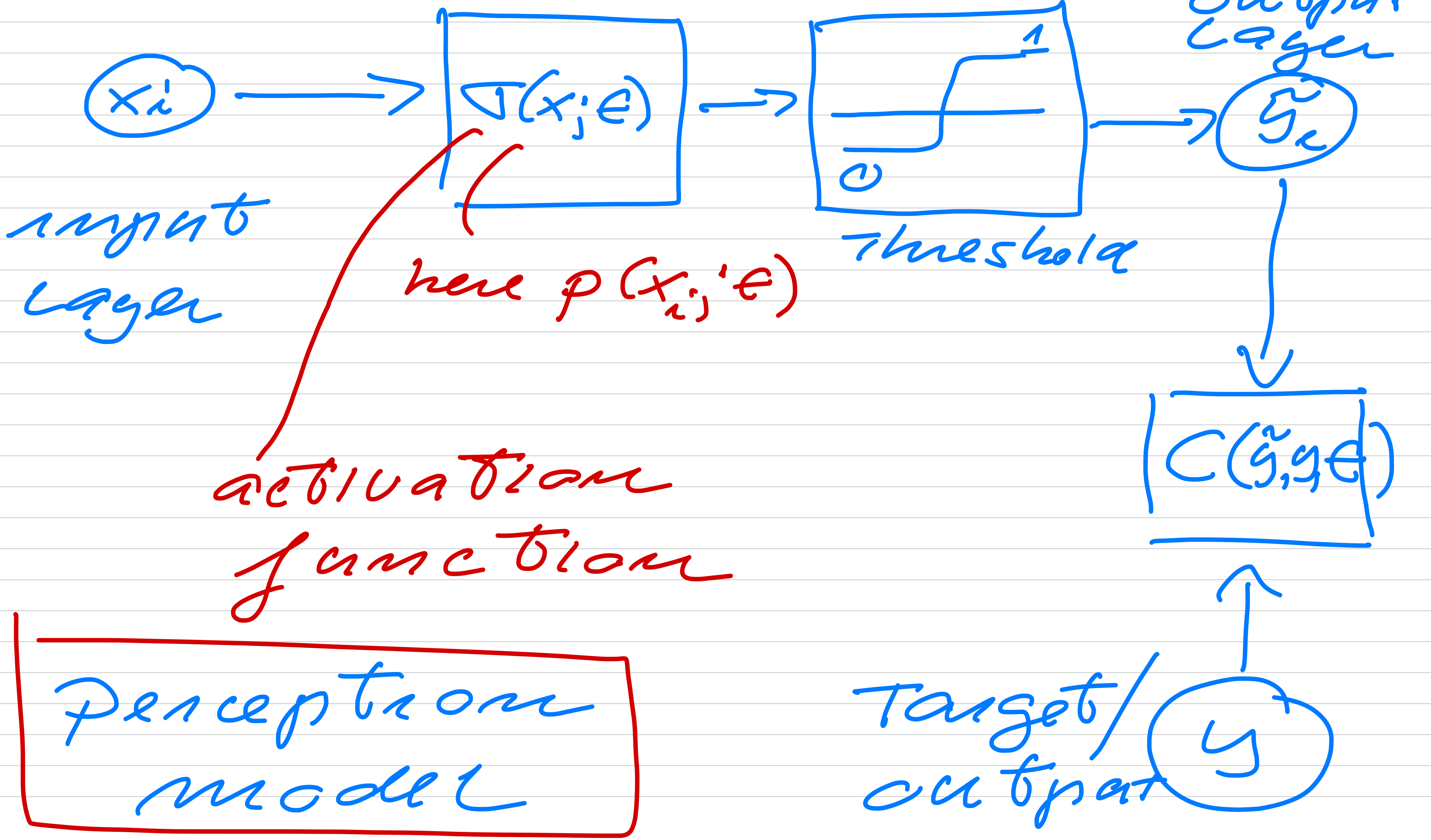
$$y_i = \underbrace{p(x_i)}_{p_i} + \varepsilon_i$$

$y_i = 1$ then $p_i = \frac{1}{1+e^{-x_i}}$

$y_i = 0$ then $p_i = 1 - \frac{1}{1+e^{-x_i}}$

$$\sum_{i=0}^{1} p_i = \underbrace{p_0}_{y_i=0} + \underbrace{p_1}_{y_i=1}$$

Graphical representation



Simplified:

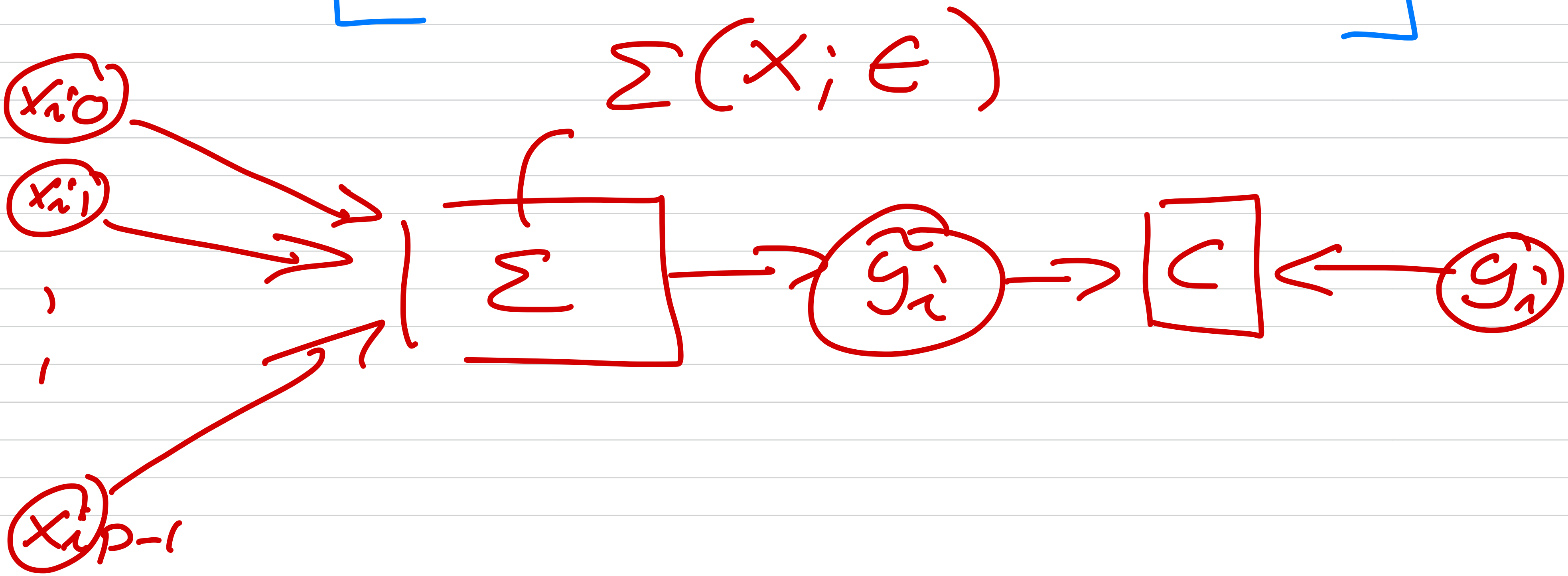
$$x_i \longrightarrow \boxed{\sigma(x_i;\theta)} \longrightarrow \hat{y}_i \longrightarrow \boxed{C(\hat{y}_i; y_i; \theta)} \longleftarrow y_i$$

input layer

output layer

$$\underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \; C(\hat{y}, y, \theta)$$

$$X = \begin{bmatrix} x_{00} & x_{01} & \cdots & x_{0\,p-1} \\ x_{10} & x_{11} & & \\ \vdots & & \ddots & \\ x_{i0} & x_{i1} & \cdots\ddots & x_{i\,p-1} \\ \vdots & & \ddots & \\ x_{m-1\,0} & & & x_{m-1\,p-1} \end{bmatrix}$$



simplest case

$$P(x_i) = \frac{1}{1+e^{-(\theta_0+\theta_1 x_i)}}$$

$$\left(\frac{1}{1+e^{-(\theta_0+\theta_1 x_i^1+\theta_2 x_i^1 + \dots \theta_{p-1} x_i^f)}} \right)$$

$$C(\theta) = P(D|\theta) = \prod_{i=0}^{n-1} \underbrace{P(x_i|\theta)}$$

$$p_i^{\,y_i} (1-p_i)^{1-y_i}$$

$$y_i = \{0, 1\}$$

$$p_i = \frac{1}{1+e^{-x_i}}$$

$$\underset{\theta \in \mathbb{R}^p}{\operatorname{argmax}} \; C(\theta)$$

$$\underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} - \log P(D|\theta)$$

$$\Rightarrow - \log P(D|\theta)$$

$$= - \sum_{i=0}^{n-1} \left\{ y_i \log p_i + (1-y_i) \log (1-p_i) \right\}$$

$$\frac{\partial \varepsilon}{\partial \theta} = - X^T (y - p)$$

$$p, y \in \mathbb{R}^n \qquad X \in \mathbb{R}^{n \times p}$$

$$\frac{\partial c}{\partial \theta} = g(\theta) \in \mathbb{R}^p$$

θ is included in $p(x_i; \theta)$

$$\frac{\partial^2 c}{\partial \theta^2} = X^T W X = H(\theta)$$

depends on θ

$w_{ii} = p_i(1-p_i)$

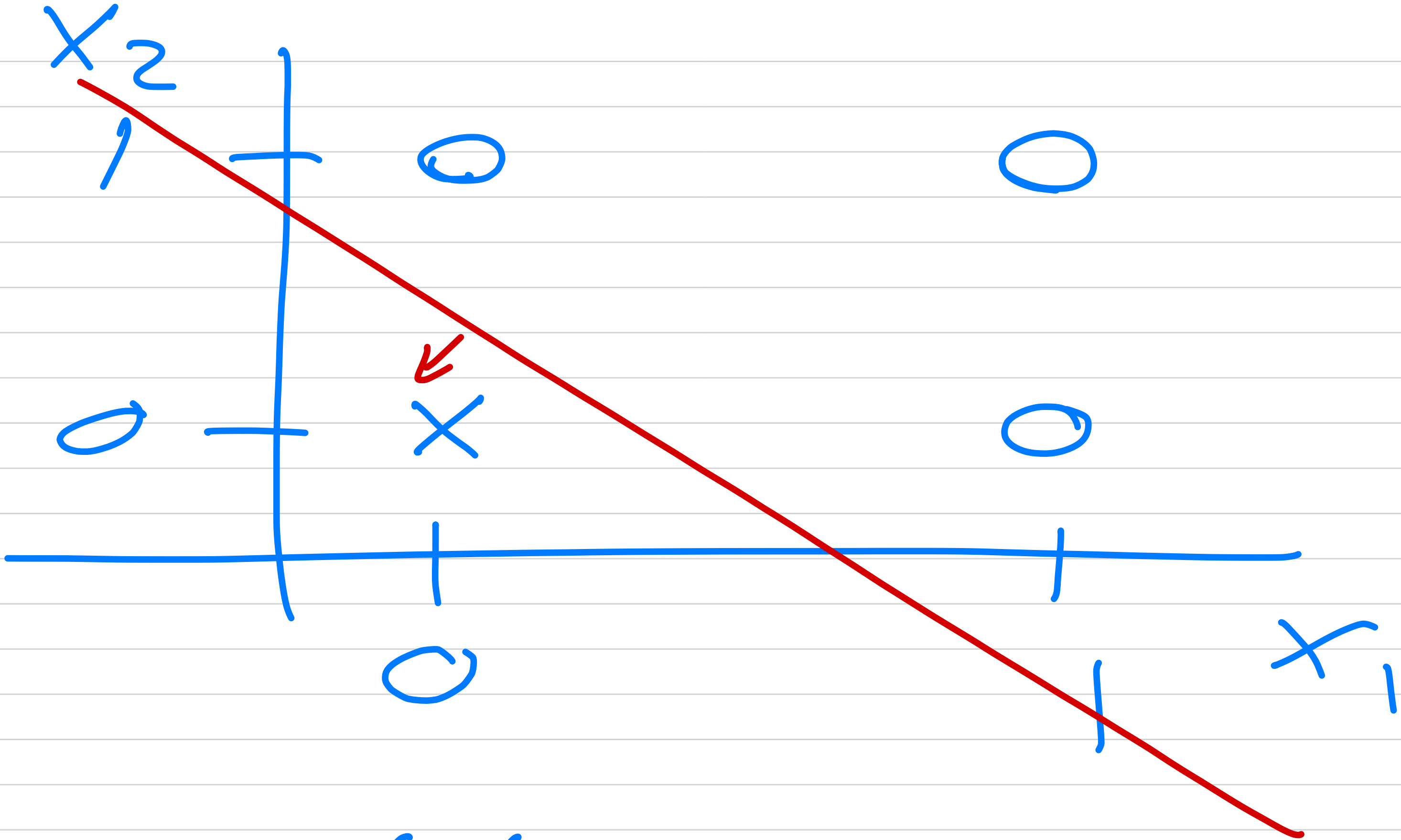
$w_{ij} = 0$ if $i \neq j$

$$\theta^{(n+1)} = \theta^{(n)} - \underbrace{\left(H(\theta^{(n)})\right)^{-1}}_{\eta} g(\theta^{(n)})$$

$$= \theta^{(n)} - \eta\, g(\theta^{(n)})$$

Example OR gate

inputs		output
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1



x = 0

o = 1

model

$$\tilde{y}_i = x_1(i)\, w_1 + x_2(i)\, w_2 + b$$

$$X = \left\{ [0,0]^T, [0,1]^T, [1,0]^T, [1,1]^T \right\}$$

$$y = 0 = [0\ 0] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + b$$

$$y = 1 = [0\ 1] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + b$$

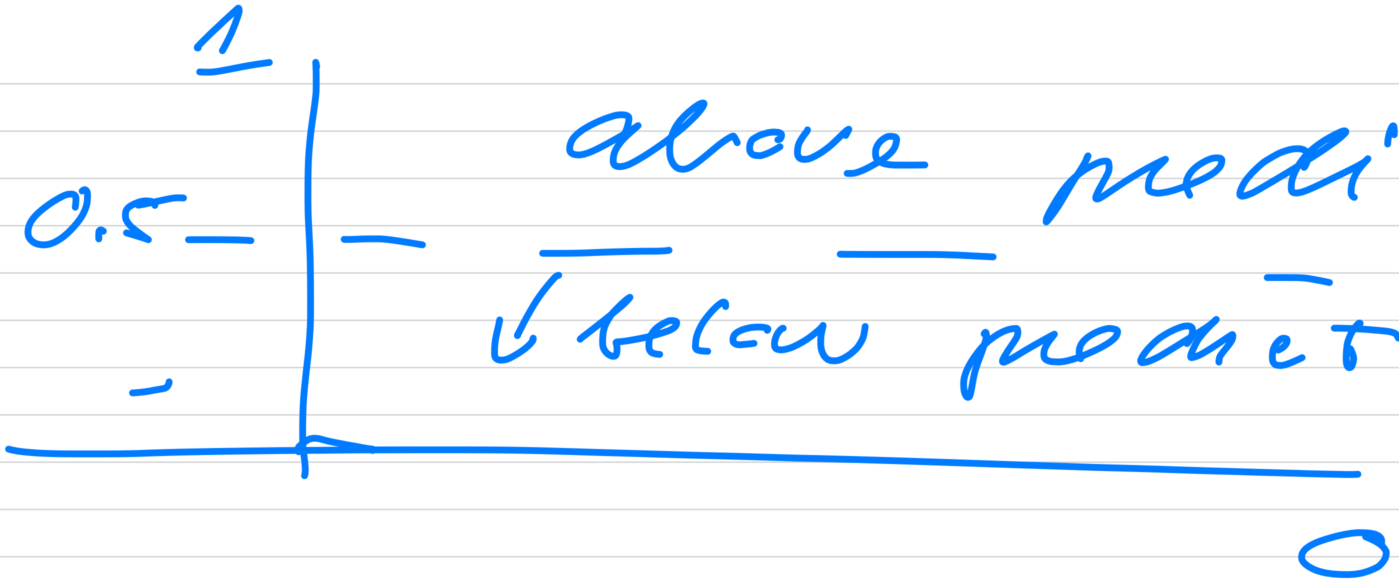
$$\Theta = \{ b, w_1, w_2 \}$$

$$
X = \begin{array}{c}
\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \\
\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ 1 & x_1 & x_2 \end{array}
\end{array}
\qquad (X^T X)^{-1} X^T y \longrightarrow \underline{\underline{OLS}}
$$

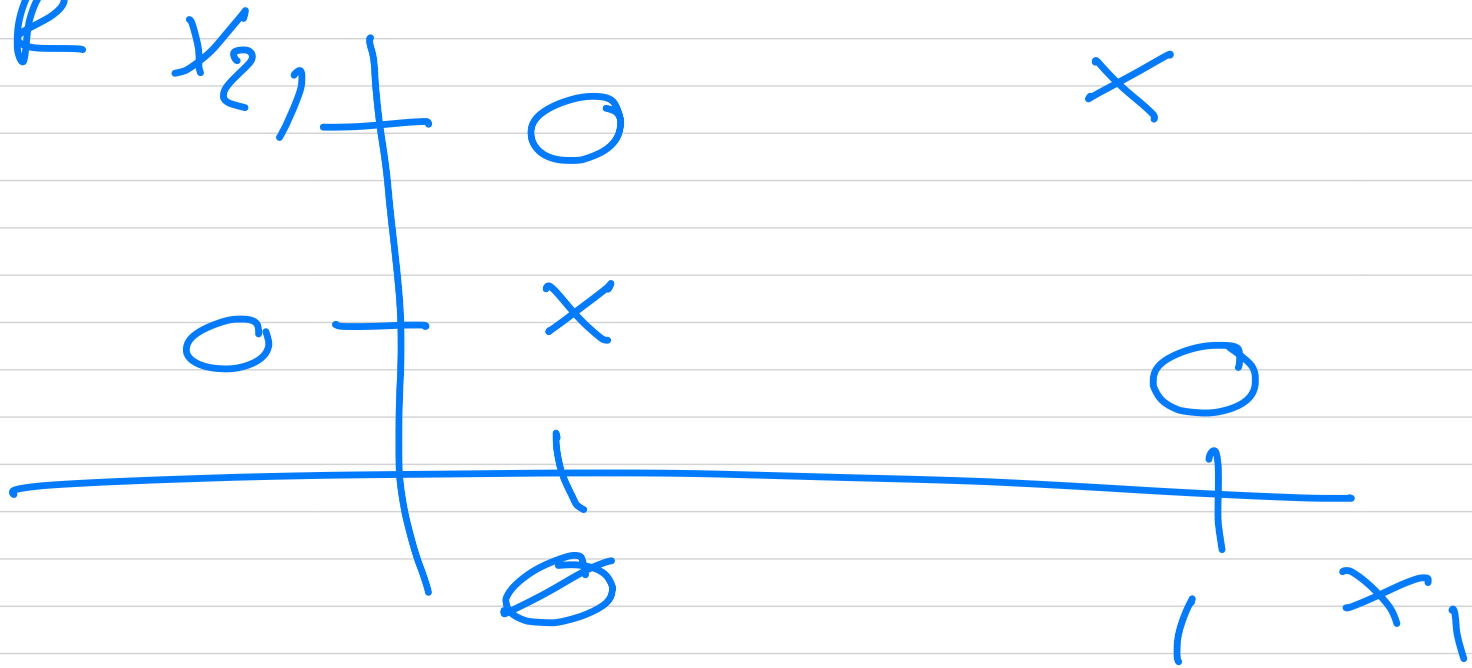
$$
\theta = \left[1/4 ,\ 1/2 ,\ 1/2 \right]
$$

$$
\hat{y} = X \cdot \theta = \left[1/4 ,\ 3/4 ,\ 3/4 ,\ 5/4 \right]
$$

$$
y = \left[0 \quad 1 \quad 1 \quad 1 \right]
$$



XOR



AND gate

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1