Practical 2.0

Neural Networks – feed forward (inference)

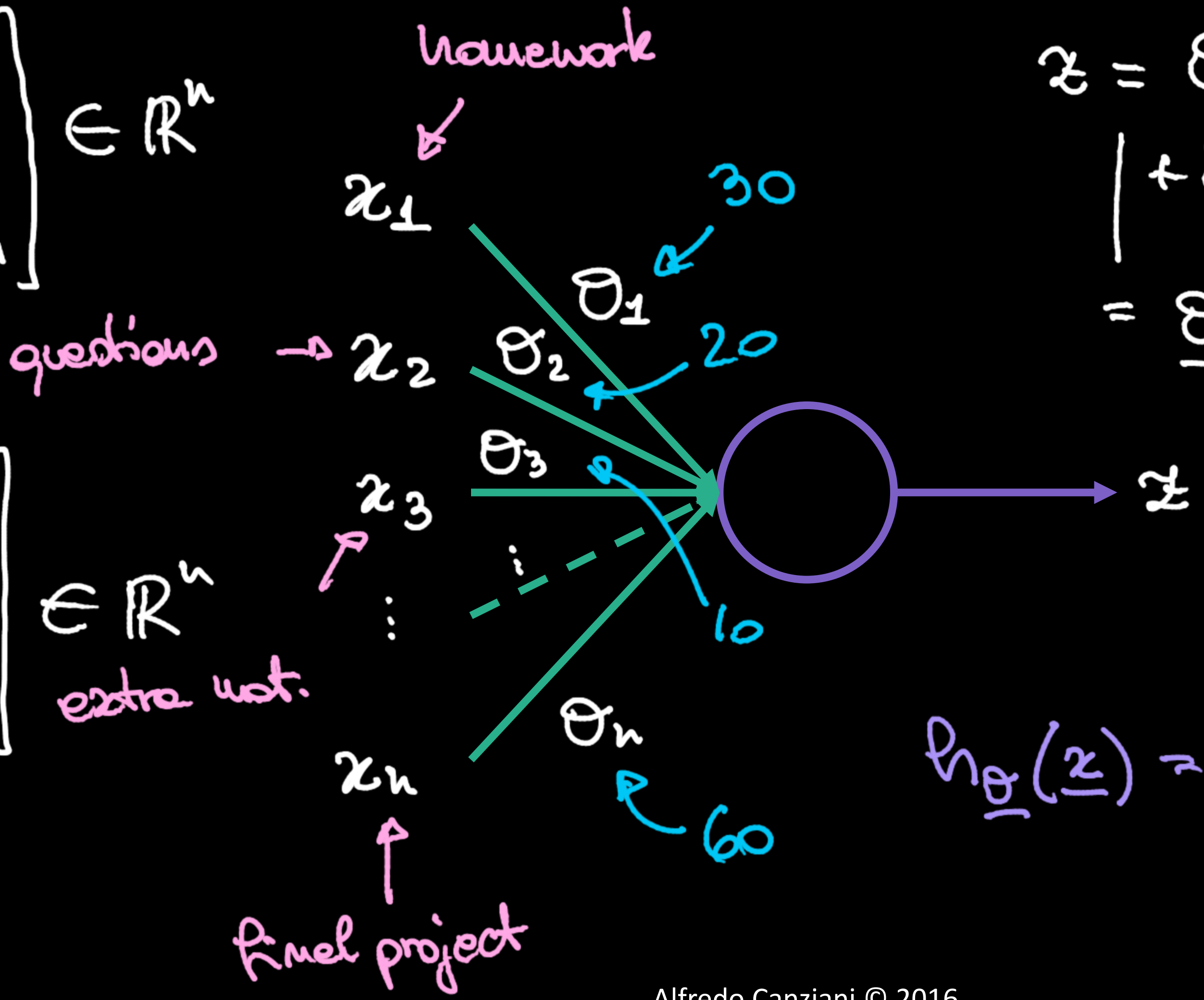
Overview

- Perceptron
  - Embedded threshold
  - Step activation function
- Logistic unit
- Neural network
  - Architecture
  - Equations

Perceptron (I) (1950s, Frank Rosenblatt)

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

$$\underline{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^n$$

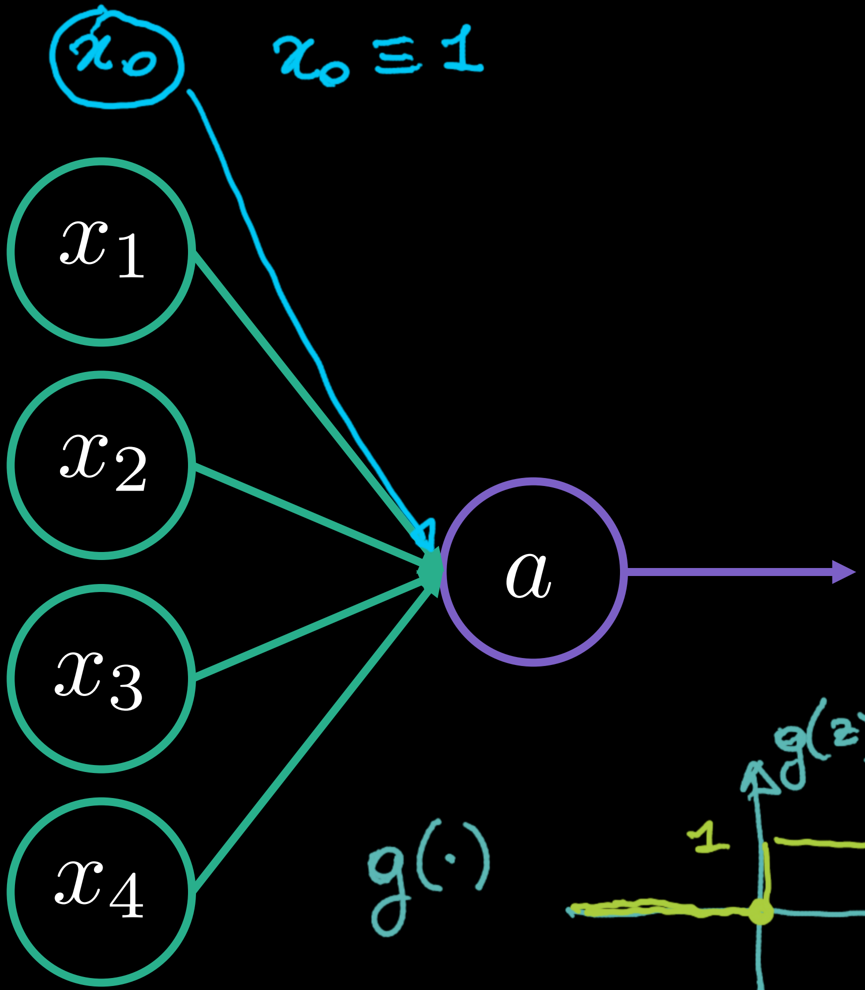


$$z = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n = \underline{\theta}^T \underline{x}$$

$$\boxed{T = 100 \text{ (bias)}}$$

$$h_{\underline{\theta}}(\underline{x}) = \begin{cases} 0, & z \le T \\ 1, & z > T \end{cases}$$

Perceptron (II)



$x_0 \equiv 1$

$$\underline{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n+1}$$

$$z := \underline{\theta}^T \underline{x} = \theta_0 \cdot 1 + \theta_1 x_1 + \dots + \theta_n x_n$$

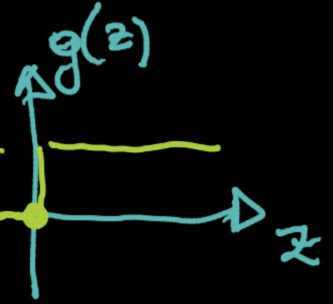
$$\underline{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n+1}$$

old $z - \underbrace{T}_{=\theta_0}$

$$h_{\underline{\theta}}(\underline{x}) = \begin{cases} 0, & z \leq 0 \\ 1, & z > 0 \end{cases}$$

$$= g(z) = g(\underline{\theta}^T \underline{x}) = a$$

$g(\cdot)$



$$\underline{\theta} = \begin{bmatrix} -100 \\ 30 \\ 20 \\ 10 \\ \dots \end{bmatrix} \qquad \underline{x} = \begin{bmatrix} 1 \\ [0,1] \\ [0,1] \\ [0,1] \end{bmatrix} \in \mathbb{R}^{n+1}$$

x_0

Logistic unit

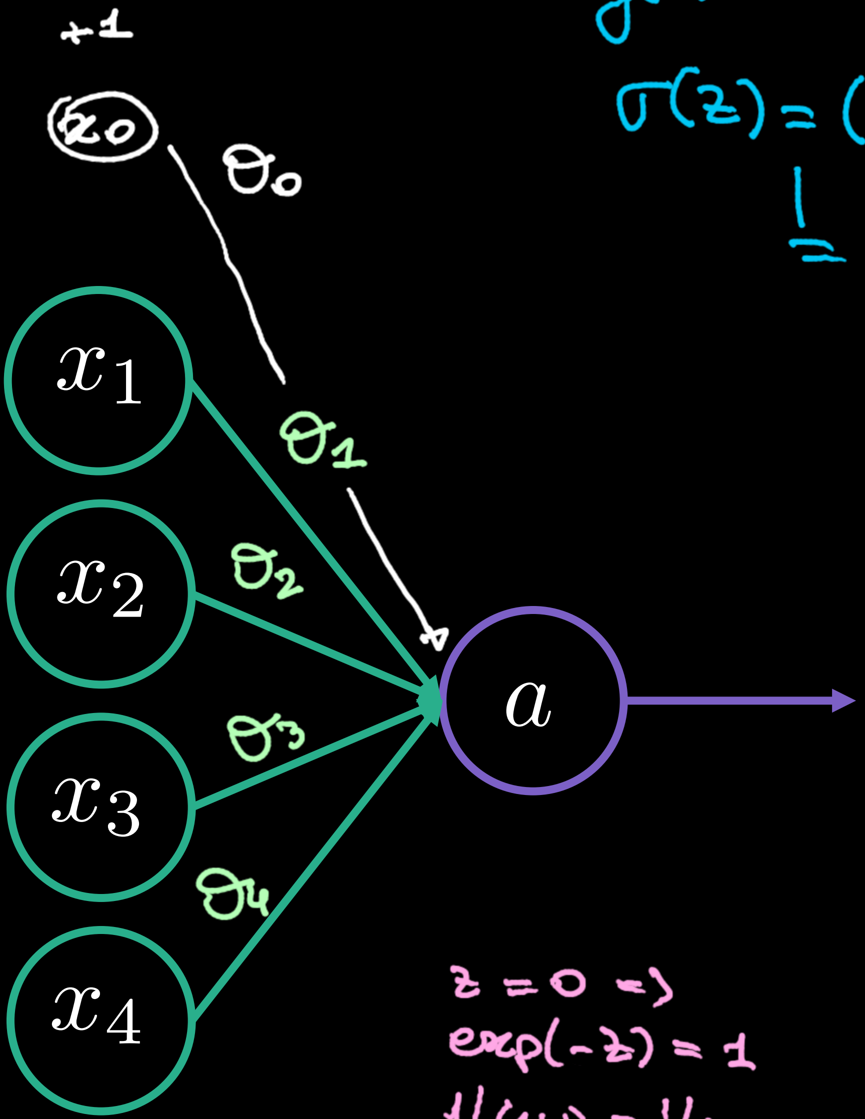
$\underline{x} \in \mathbb{R}^{n+1}$

$\underline{\theta} \in \mathbb{R}^{n+1}$

$z = \underline{\theta}^T \underline{x}$

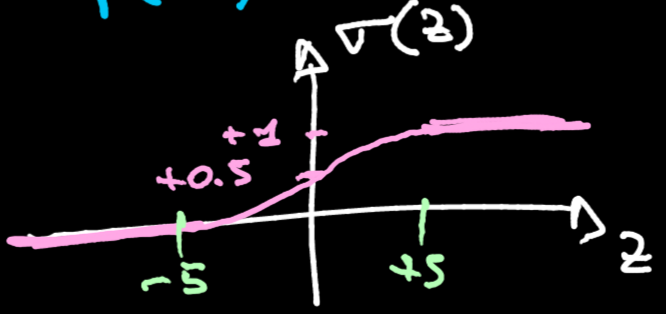
$h_{\underline{\theta}}(\underline{x}) = \sigma(z) \in [0,1]$

$h_{\underline{\theta}}(\underline{x}) \overset{?}{>} 0.5$



$g(\cdot) = \sigma(\cdot) \Rightarrow$

$$\sigma(z) = (1+\exp(-z))^{-1} = \frac{1}{1+\exp(-z)}$$



$z = 0 \Rightarrow$

$\exp(-z) = 1$

$1/(1+1) = 1/2$

$z \to -\infty$

$-z \to +\infty \Rightarrow \exp(-z) \to +\infty$

$1/(1+\exp(-z)) \to 0^+$

$z \to +\infty$

$-z \to -\infty \Rightarrow \exp(-z) \to 0^+$

$1/(1+\underbrace{\exp(-z)}_{\to 0^+}) \to 1$

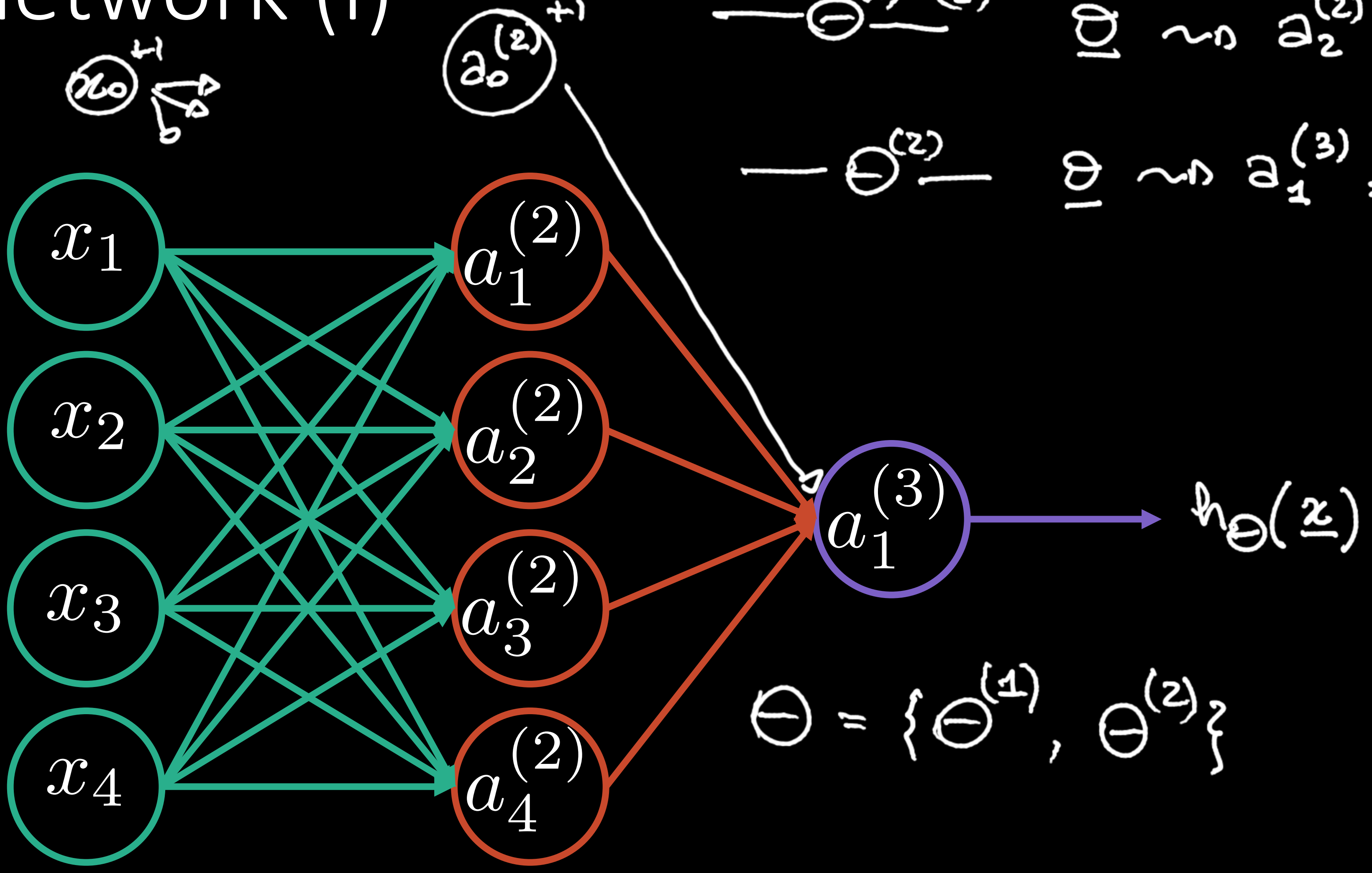
Neural network (I)

$s_2 = 4$

$\underline{\Theta}^{(1)} \rightsquigarrow a_1^{(2)}$

$\underline{\Theta}^{(1)}_{(2)} \rightsquigarrow a_2^{(2)}$

$\underline{\Theta}^{(2)} \rightsquigarrow a_1^{(3)} = h_\Theta(\underline{x})$



$\Theta = \{\Theta^{(1)}, \Theta^{(2)}\}$

$s_1 = 4$

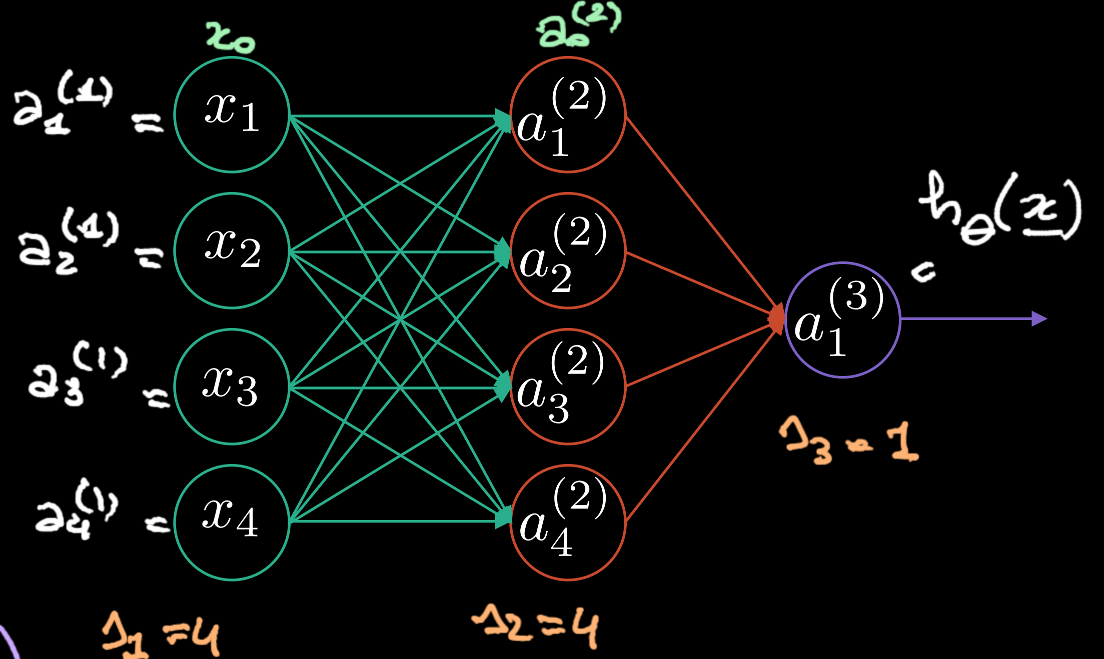
Input layer

Hidden layer

Output layer

$s_L = s_3 = 1$

Neural network (II)



$a_i^{(j)}$: i-th activation in layer j

$\Theta^{(j)}$: collection of mappings to go from layer j to j+1

s_j : # elements for layer j

$$a_1^{(2)} = \sigma\left(\Theta_{10}^{(1)} \overset{+1}{\overset{\parallel}{x_0}} + \Theta_{11}^{(1)} x_1 + \dots + \Theta_{14}^{(1)} x_4\right)$$

$$a_2^{(2)} = \sigma\left(\Theta_{20}^{(1)} x_0 + \Theta_{21}^{(1)} x_1 + \dots + \Theta_{24}^{(1)} x_4\right)$$

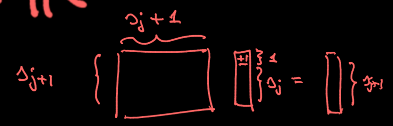
$$a_4^{(2)} = \sigma\left(\Theta_{40}^{(1)} x_0 + \Theta_{41}^{(1)} x_1 + \dots + \Theta_{44}^{(1)} x_4\right)$$

$$h_{\Theta}(\boldsymbol{x}) = a_1^{(3)} = \sigma\left(\Theta_{10}^{(2)} a_0^{(2)} + \Theta_{11}^{(2)} a_1^{(2)} + \dots + \Theta_{14}^{(2)} a_4^{(2)}\right)$$

$$\underline{a}^{(2)} = \begin{bmatrix} a_1^{(2)} \\ a_2^{(2)} \\ \vdots \\ a_4^{(2)} \end{bmatrix} \in \mathbb{R}^{s_2}$$

$$= \sigma\left(\Theta^{(1)} \underline{x}\right)$$

$$\Theta^{(j)} \in \mathbb{R}^{s_{j+1} \times (s_j + 1)}$$



$$\hat{\underline{a}}^{(2)} = \begin{bmatrix} +1 \\ \underline{a}^{(2)} \end{bmatrix} \Rightarrow h_\Theta(\underline{x}) = \sigma\left(\Theta^{(2)} \hat{\underline{a}}^{(2)}\right)$$