Lecture 24: Deep neural networks

Professor Ilias Bilionis

Mathematical description of dense deep neural networks

PREDICTIVE
SCIENCE LABORATORY

Mathematics of neural networks

$$f: \mathbb{R}^d \longrightarrow \mathbb{R}^q, \quad y = \underline{f}(x;\theta)$$

parameters

$L=1$:

$$z = \underbrace{W^{(0)}x}_{\text{linear}} + \underbrace{b^{(0)}}_{\text{bias}} \;;\; W^{(0)} \in \mathbb{R}^{n\times d}, \quad b^{(0)} \in \mathbb{R}^n$$

affine trans.

# of neurons

$$x^{(1)} = h(z) = (h(z_1), \dots, h(z_n))$$

activation function

$$y = W^{(1)}x^{(1)} + b^{(1)}, \quad W^{(1)} \in \mathbb{R}^{q\times n}, \quad b^{(1)} \in \mathbb{R}^q$$

$$y = W^{(1)}h(W^{(0)}x + b^{(0)}) + b^{(1)}$$

$$\theta = (\underbrace{(W^{(0)}, b^{(0)})}_{\text{internal}}, \underbrace{(W^{(1)}, b^{(1)})}_{\text{output}})$$

PREDICTIVE SCIENCE LABORATORY

10

L layers ; $\underbrace{n^{(0)} = d}_{\text{input layer}}, \underbrace{n^{(1)}, n^{(2)}, \dots, n^{(L)}}_{L \text{ hidden (latent) layers}}, \underbrace{n^{(L+1)} = q}_{\text{output layer}}$

$$x^{(0)} = x$$

$$z^{(0)} = W^{(0)}x^{(0)} + b^{(0)},$$

$$x^{(1)} = h(z^{(0)})$$

$$\vdots$$

$$i = 0, \dots, L-1 \begin{cases} z^{(i)} = W^{(i)}x^{(i)} + b^{(i)} \\ x^{(i+1)} = h(z^{(i)}) \end{cases} ; \; W^{(i)} \in \mathbb{R}^{n^{(i)} \times n^{(i-1)}}, \quad b^{(i)} \in \mathbb{R}^{n^{(i)}}$$

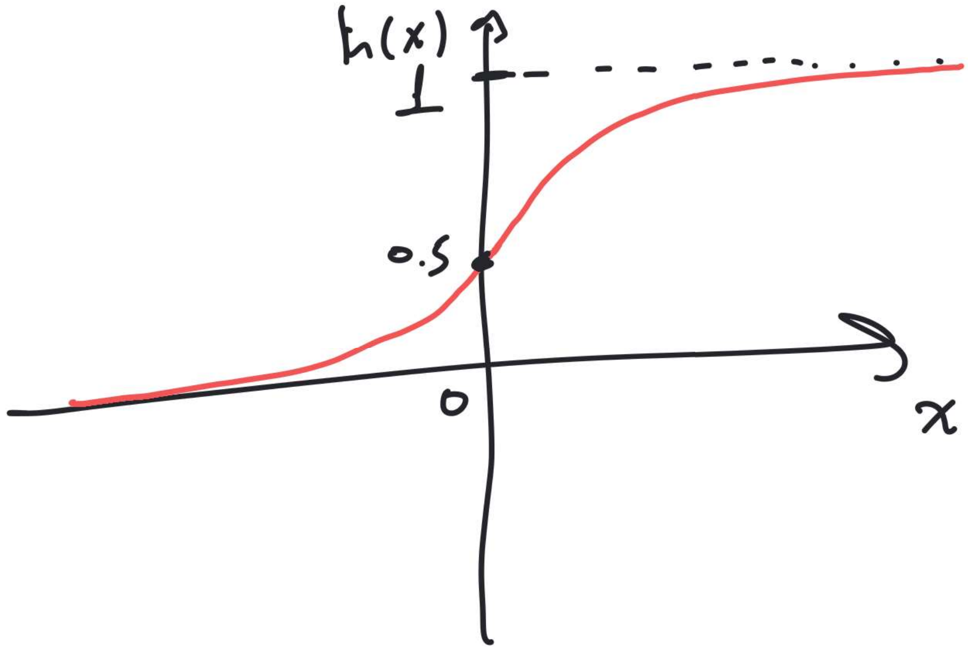
output: $y = W^{(L)}x^{(L)} + b^{(L)}$

$$\theta = \{W^{(i)}, b^{(i)}\}_{i=0}^{L}$$

Not the only possible architecture for a DNN, just the simplest

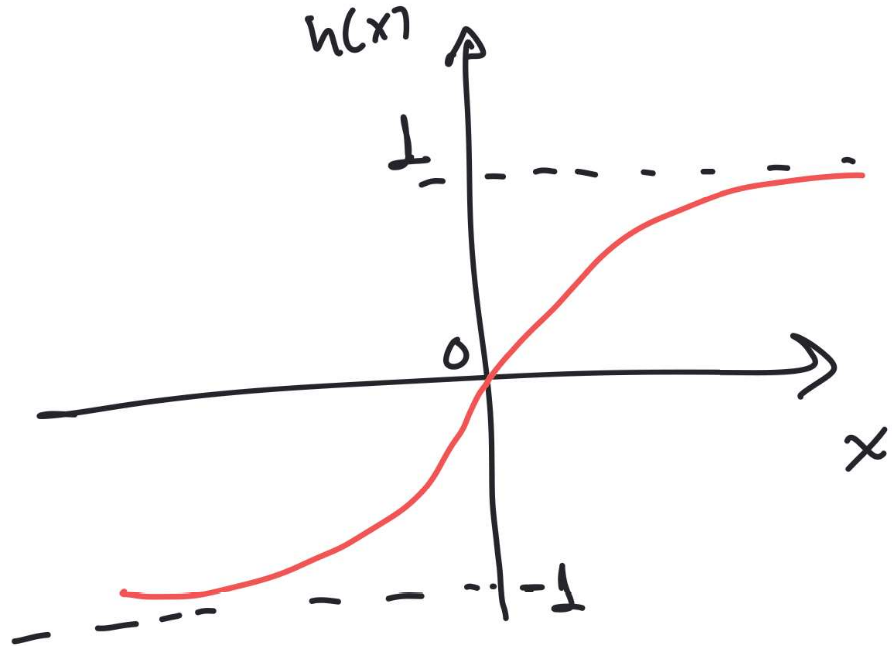
The sigmoid activation function

$$h(x) = \frac{1}{1+e^{-x}} \in (0,1)$$



The TanH activation function

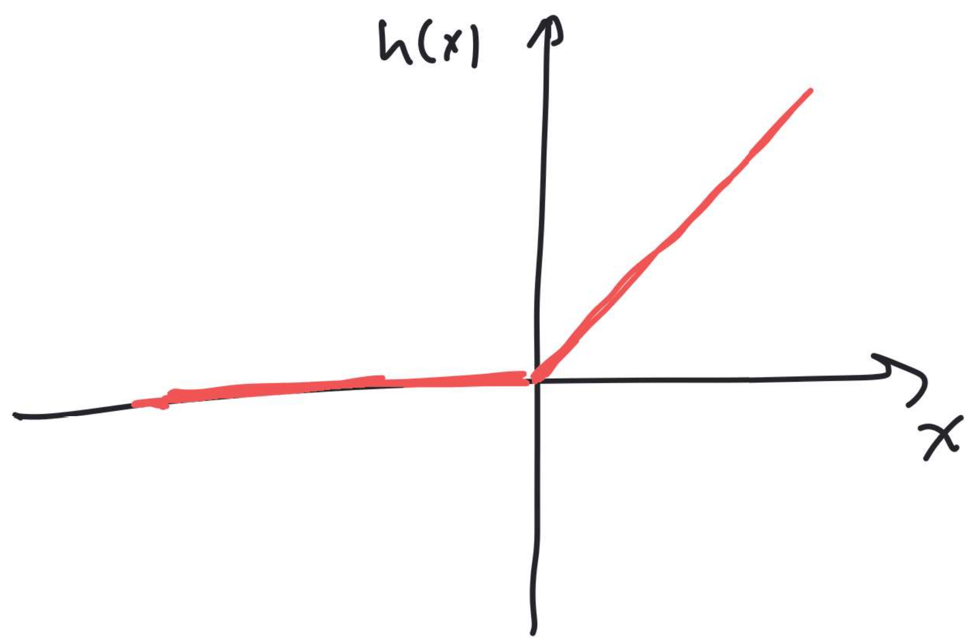
$$h(x) = \tanh x = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \in (-1, 1)$$



The rectified linear unit

ReLU

$$h(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$$



PREDICTIVE
SCIENCE LABORATORY