

Logistic Regression $\Rightarrow P(y|x) = \frac{P(y) P(x|y)}{P(x)} = \frac{1}{1 + \exp(-x\theta)} \quad [0, 1]$

\downarrow
 Posterior

$$\text{Odds} = \frac{p}{1-p} \in [0, +\infty) \quad x\theta \in \mathbb{R} \quad (-\infty, +\infty)$$

$$\frac{p}{1-p} \neq x\theta \Rightarrow \log\left(\frac{p}{1-p}\right) \in (-\infty, +\infty)$$

$$\log\left(\frac{p}{1-p}\right) = x\theta \Rightarrow \exp\left(\log\left(\frac{p}{1-p}\right)\right) = \exp(x\theta) \Rightarrow p = \frac{1}{1 + \exp(-x\theta)}$$

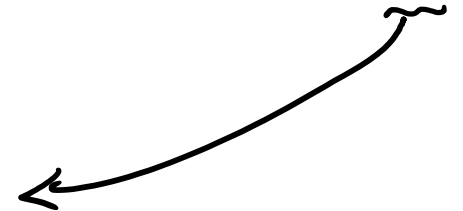
$$p(y|x) = \frac{1}{1 + \exp(-x\theta)}$$

① μ, Σ

→ Generative models: NB \Rightarrow
$$\underbrace{P(y|x)}_{\text{Posterior}} = \frac{\frac{N(x|M, \Sigma)}{P(x|y) P(y)}}{P(x)}$$

→ Discriminative models: Logistic regression
$$P(y|x) = \frac{1}{1 + \exp(-x\theta)}$$

linear combination of
features

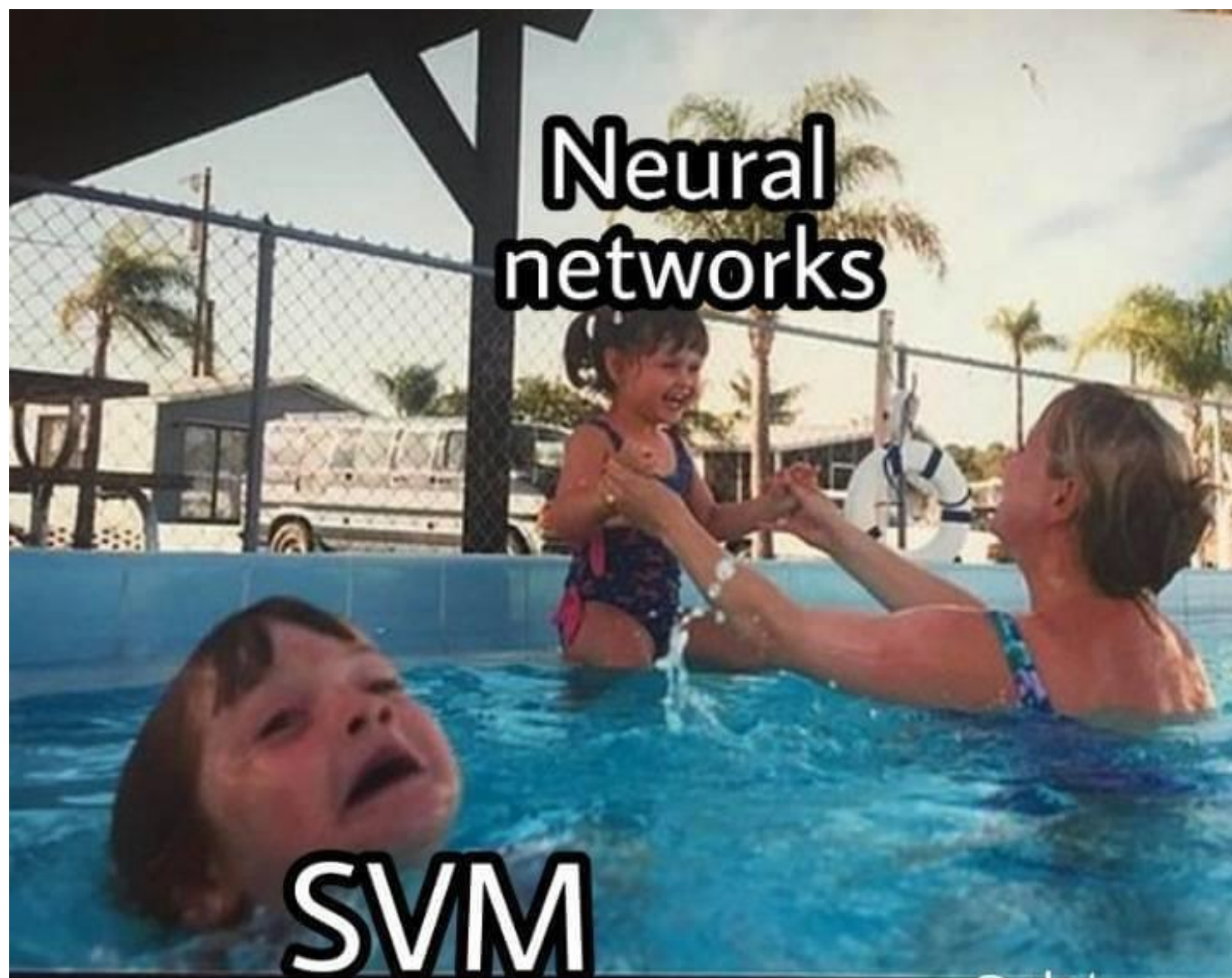


Neural Networks

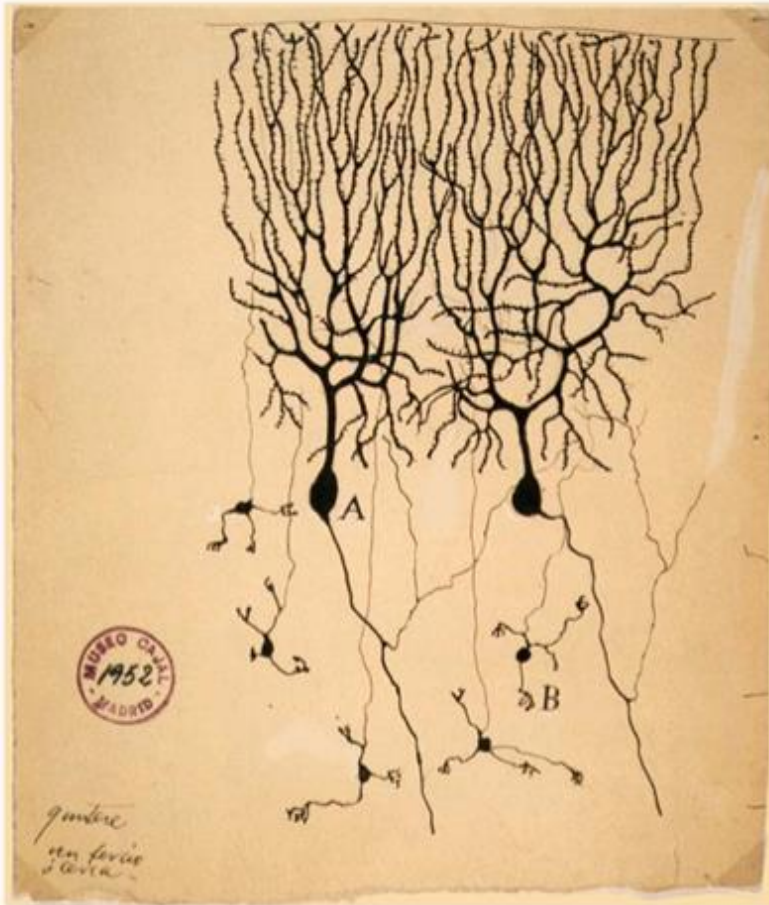
Forward Pass and Back Propagation

Mahdi Roozbahani

Georgia Tech



Inspiration from Biological Neurons



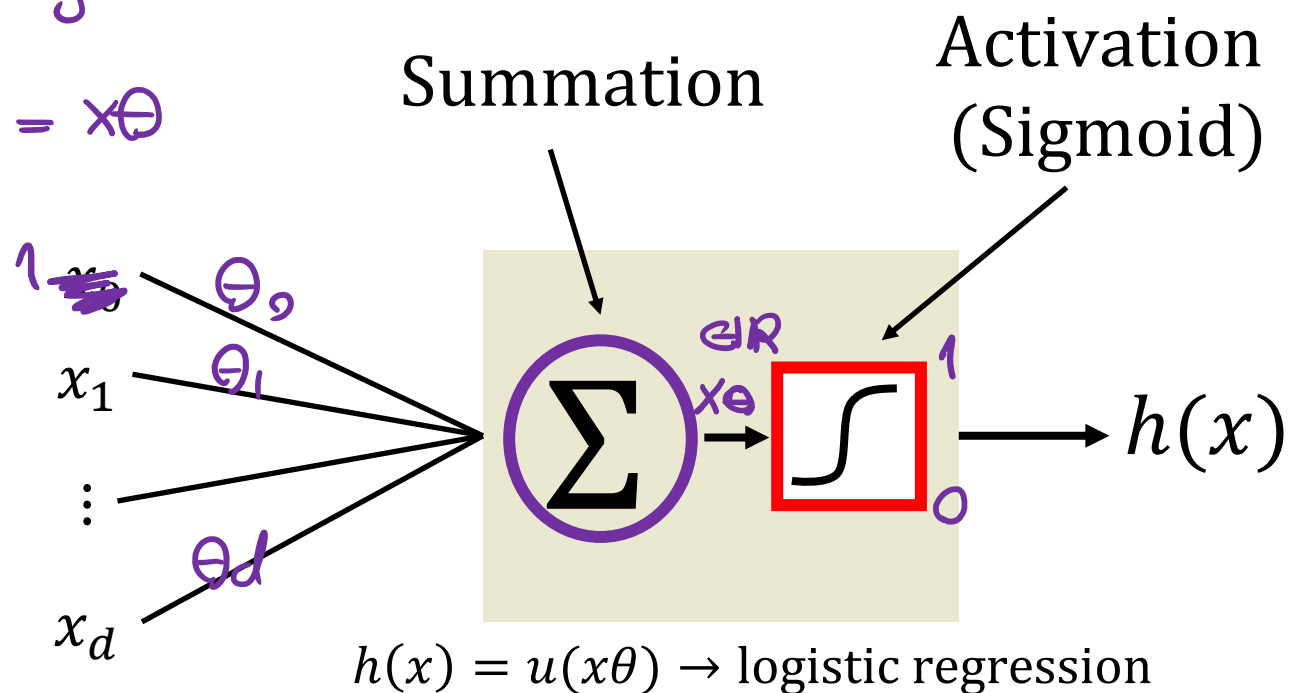
The first drawing of a brain cells by Santiago Ramón y Cajal in 1899

Neurons: core components of brain and the nervous system consisting of

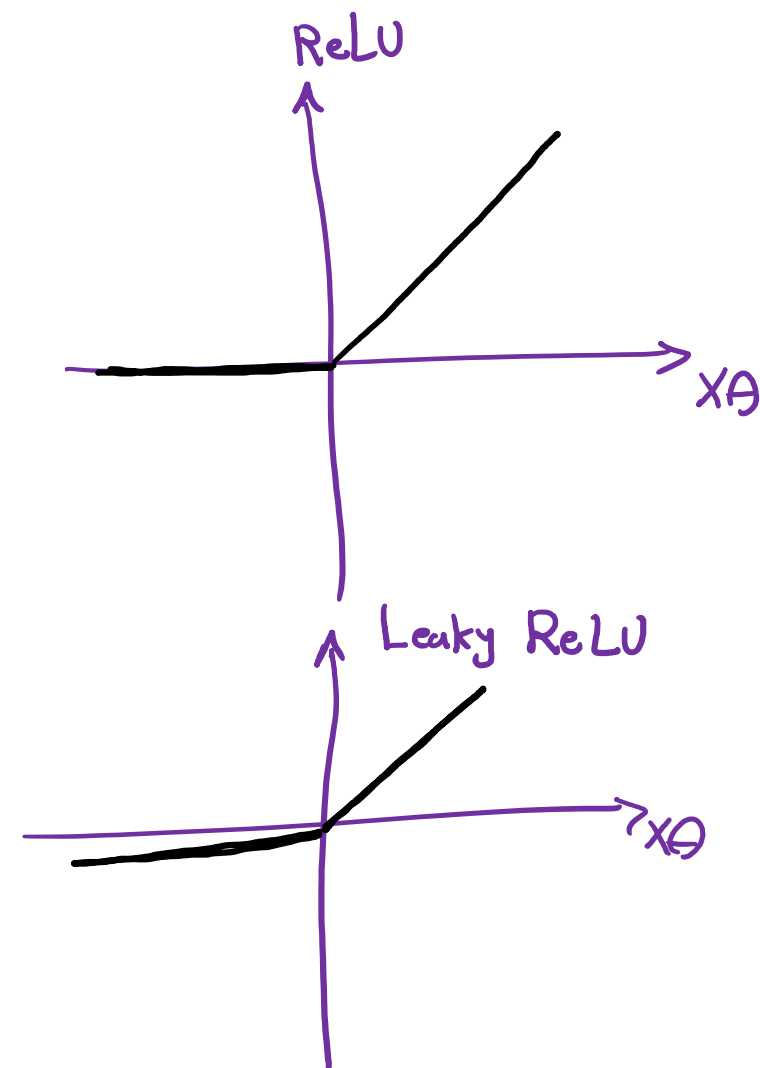
1. Dendrites that collect information from other neurons
2. An axon that generates outgoing spikes

A learning block

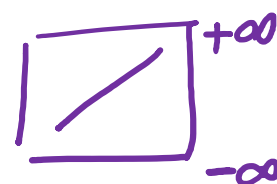
$$\textcircled{\Sigma} = \theta_0 + \theta_1 x_1 + \dots = x\theta$$

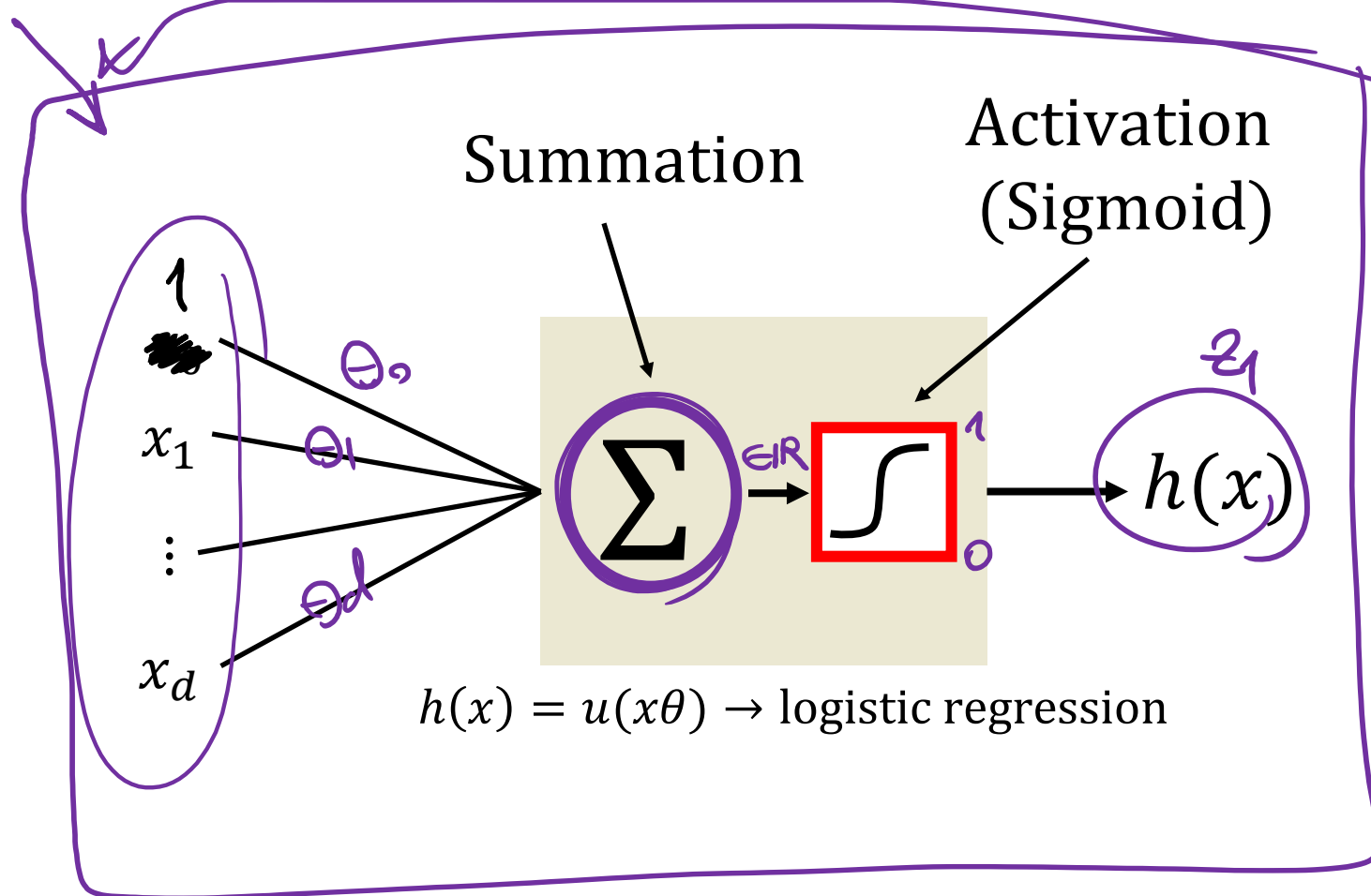


$$\text{output} = \text{activation}(x\theta + b)$$



Name of the neuron		Activation function: $\text{activation}(z)$	
Linear unit	linear	z	
Threshold/sign unit	non-linear	$\text{sgn}(z)$	
Sigmoid unit	non-linear	$\frac{1}{1 + \exp(-z)}$	
Rectified linear unit (ReLU)	non-linear	$\max(0, z)$	
Tanh unit	non-linear	$\tanh(z)$	





a learning block

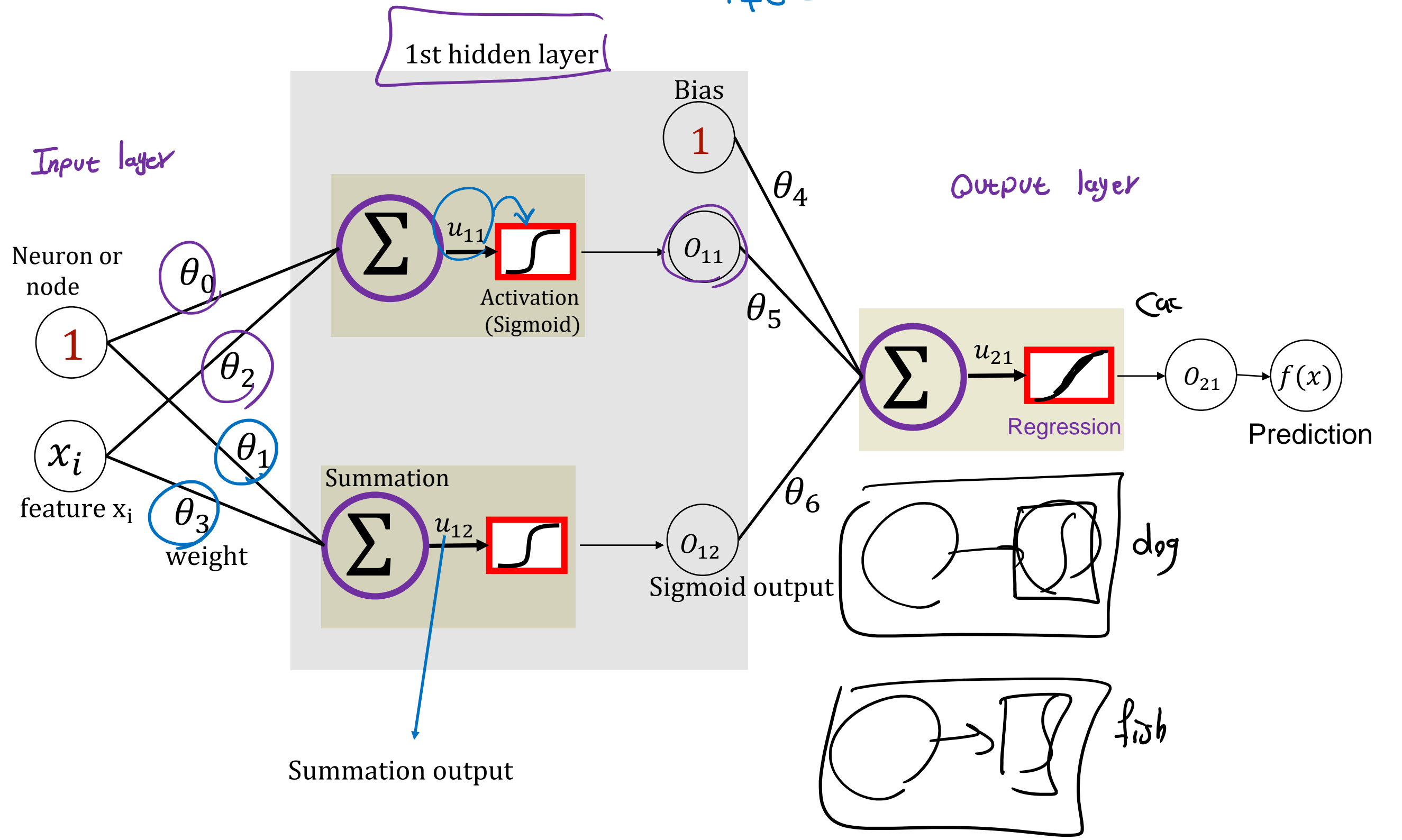
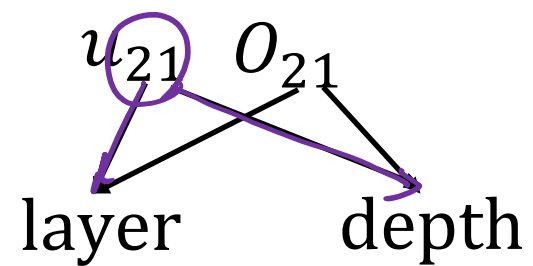
$$X = \begin{bmatrix} x_1 & \dots & x_d \\ \vdots & & \vdots \\ x_n & \dots & x_d \end{bmatrix} \quad n \times d$$

$$\Sigma = X\theta \in \mathbb{R}$$

NN Regression

$$\Sigma = u_{11} = X\theta$$

$$\sigma = \frac{1}{1 + e^{-u_{11}}}$$



1st hidden layer

FC layer

Fully connected layer

Neuron or
node

1

x_i

feature x_i

θ_0

θ_2

θ_1

θ_3

θ_4

θ_5

Σ

u_{11}



Bias
1

o_{11}

θ_4

θ_5

Σ

u_{12}



o_{12}

θ_6

θ_7

Σ

u_{13}



o_{13}

Σ

u_{21}



Regression

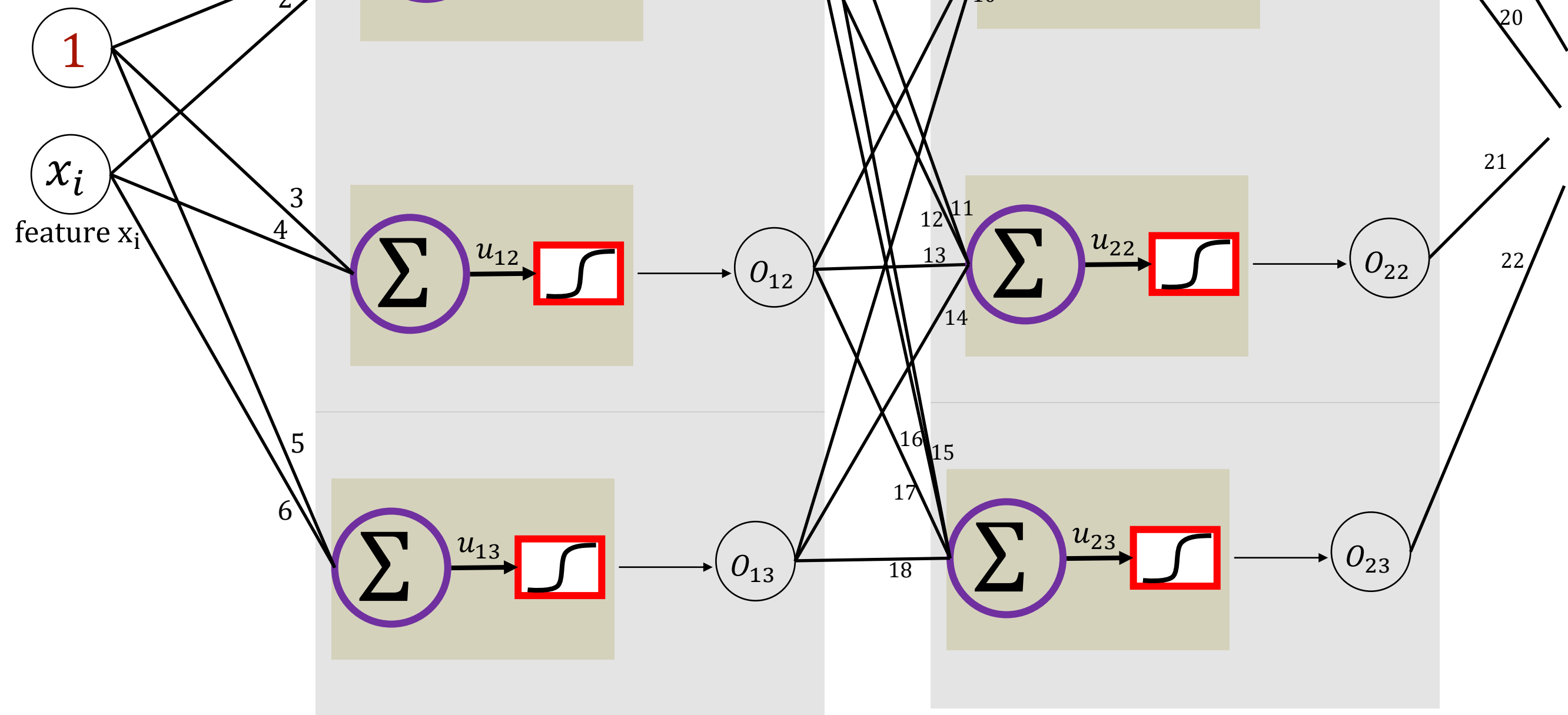
o_{21}

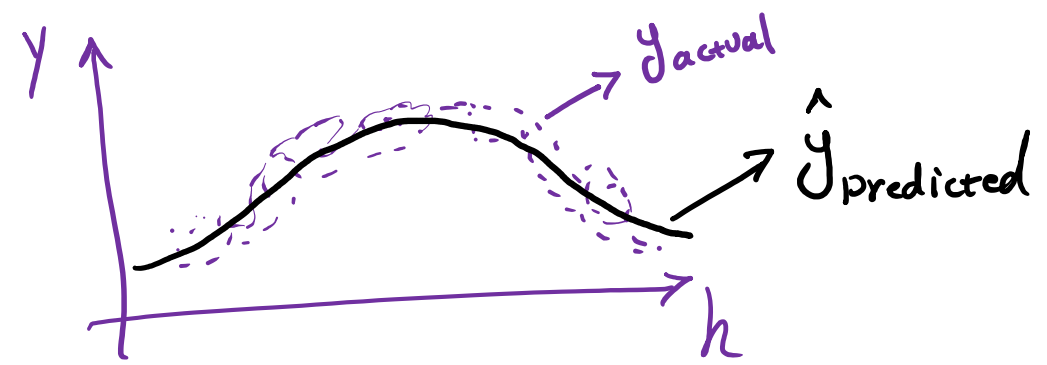
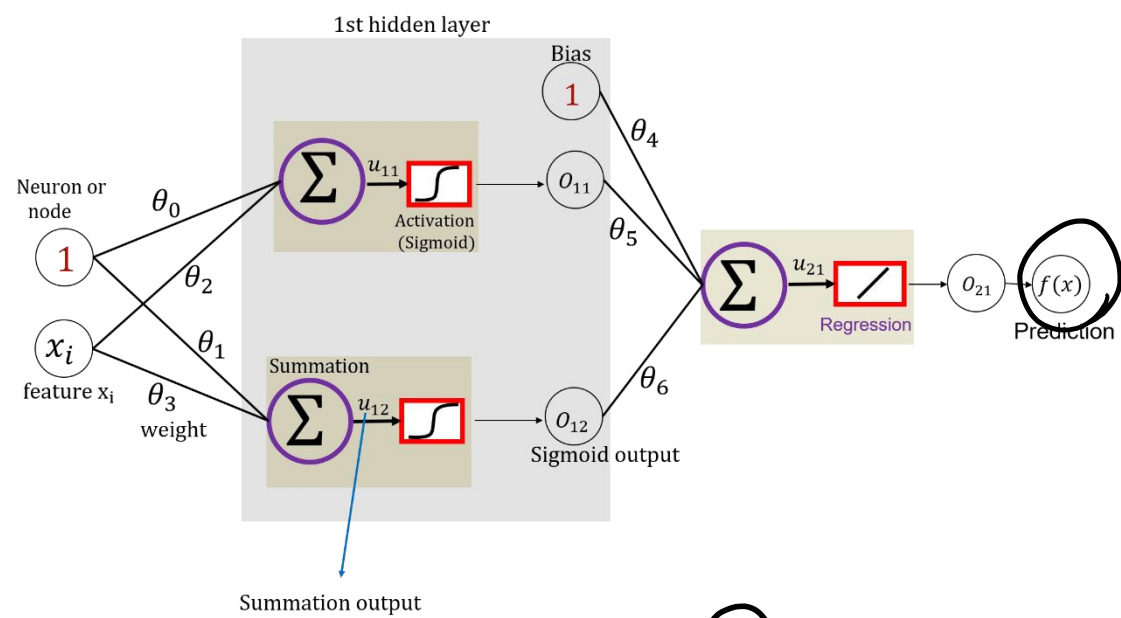
$f(x)$

Prediction

1st hidden layer

2nd hidden layer

Neuron or
node

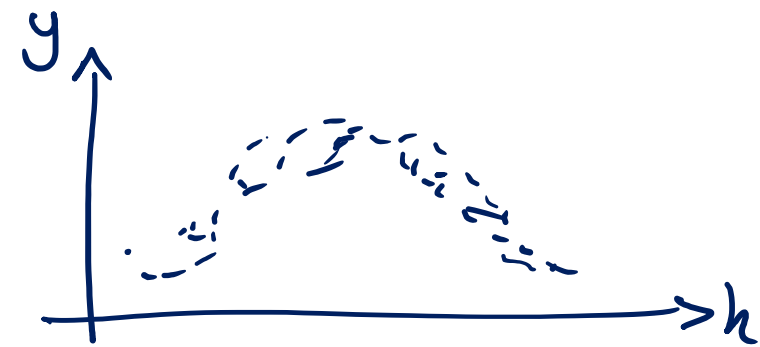
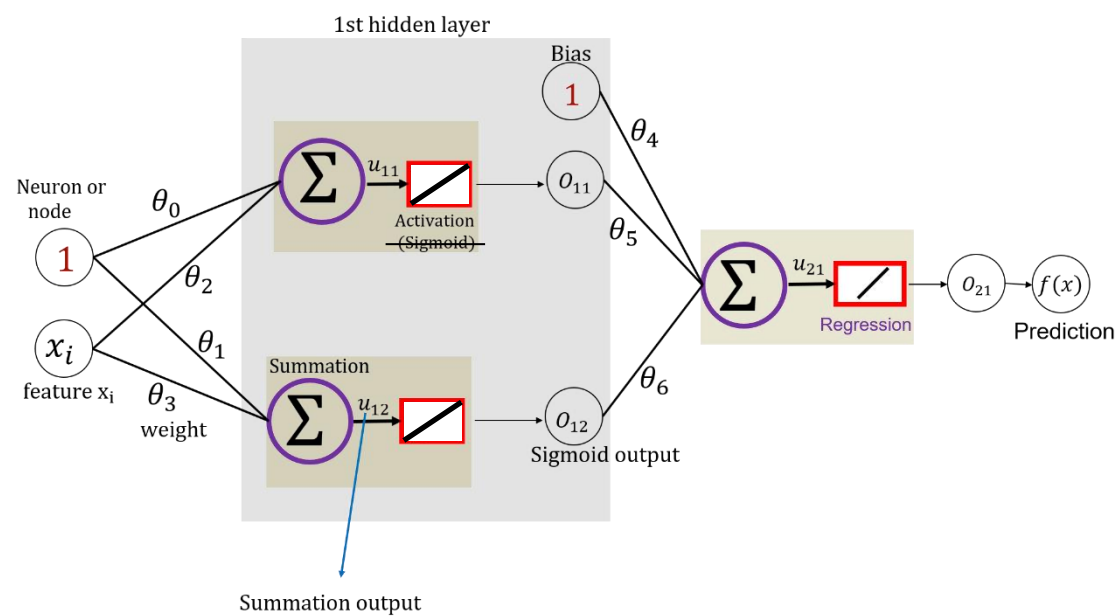


① Initialize all the parameters $\Theta_s \in \{\theta_0, \dots, \theta_6\}$
 \hookrightarrow Do not use zero

② Forward Pass $\longrightarrow \begin{cases} \rightarrow U_s \\ \rightarrow Q_s \end{cases}$

③ Back-propagation \longrightarrow Optimize parameters or Θ_s

④ Check for convergence



$$u_{11} = \theta_0 + \theta_2 x_i \Rightarrow o_{11} = u_{11} = \theta_0 + \theta_2 x_i$$

$$u_{12} = \theta_1 + \theta_3 x_i \Rightarrow o_{12} = u_{12} = \theta_1 + \theta_3 x_i$$

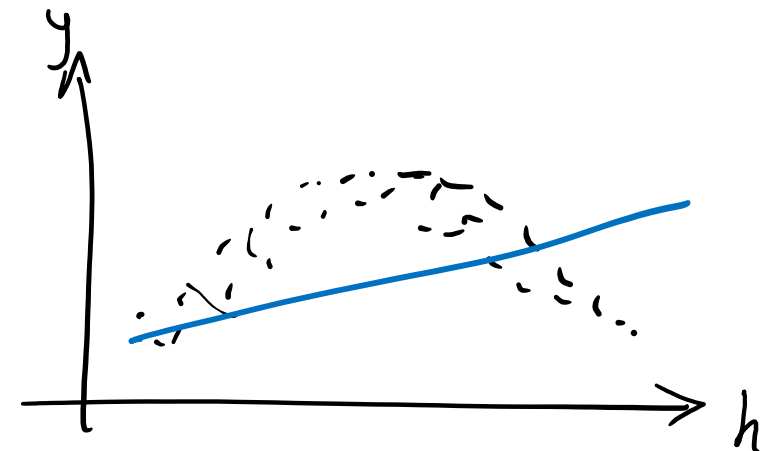


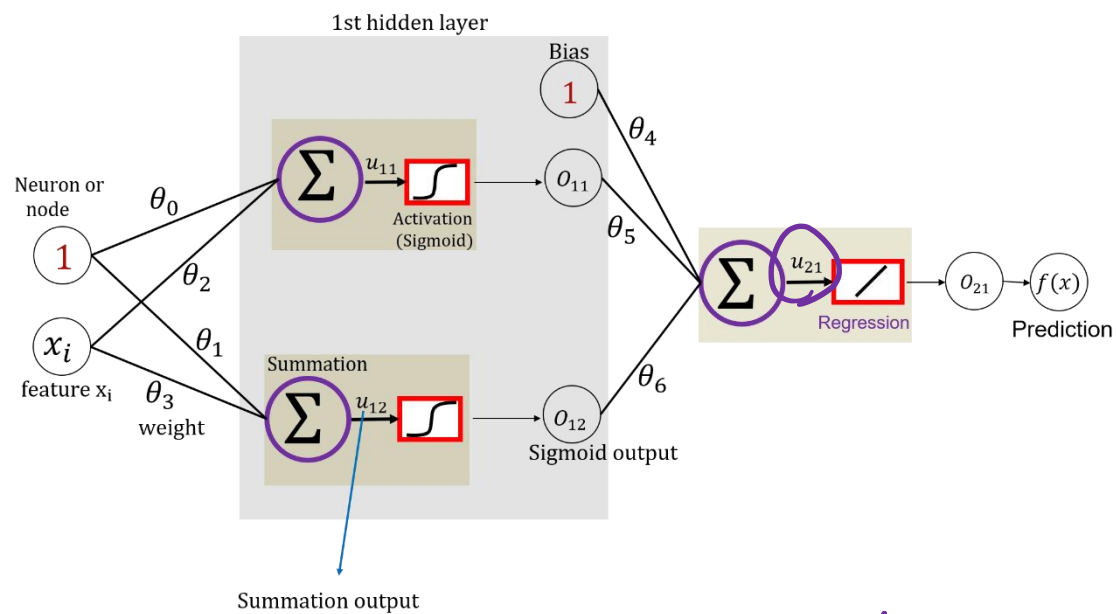
$$u_{21} = \theta_4 + \theta_5 o_{11} + \theta_6 o_{12} = o_{21} = f(x)$$

$$f(x) = \theta_4 + \theta_5 \theta_0 + \theta_5 \theta_2 x_i + \theta_6 \theta_1 + \theta_6 \theta_3 x_i$$

$$f(x) = \underbrace{\theta_4 + \theta_5 \theta_0 + \theta_6 \theta_1}_{\theta_0} + \underbrace{(\theta_5 \theta_2 + \theta_6 \theta_3)}_{\theta_1} x_i$$

$$f(x) = \theta_0 + \theta_1 x_i$$

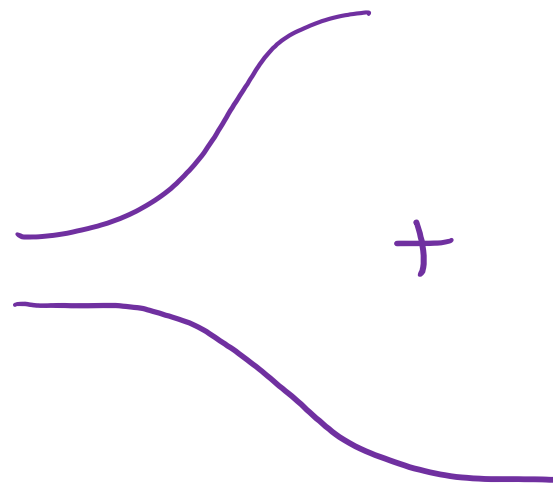




$$\Theta * \frac{1}{1+e^{-u}}$$

$$u_{11} = \Theta_0 + \Theta_2 x_i \Rightarrow O_{11} = \frac{1}{1+e^{-u_{11}}} = \frac{1}{1+e^{-(\Theta_0 + \Theta_2 x_i)}}$$

$$u_{12} = \Theta_1 + \Theta_3 x_i \Rightarrow O_{12} = \frac{1}{1+e^{-u_{12}}} = \frac{1}{1+e^{-(\Theta_1 + \Theta_3 x_i)}}$$



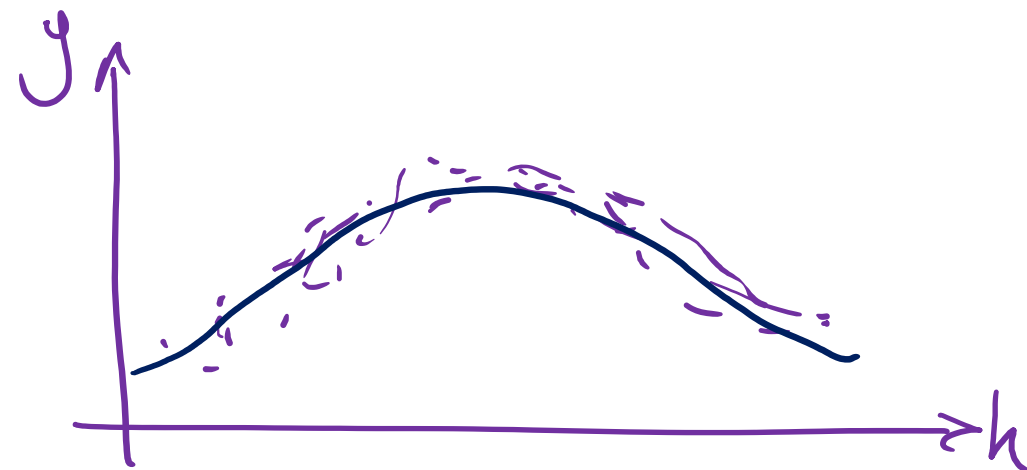
$$u_{21} = \Theta_4 + \Theta_5 O_{11} + \Theta_6 O_{12} \Rightarrow O_{21} = u_{21} = f(x)$$

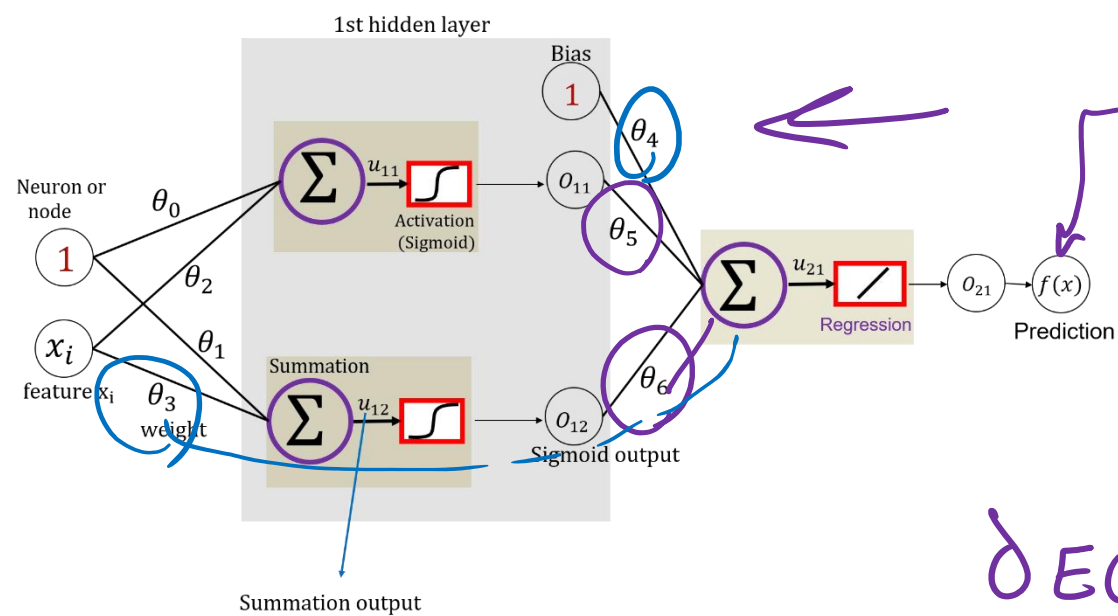
$$f(x) = \Theta_4 + \frac{\Theta_5}{1+e^{-(\Theta_0 + \Theta_2 x_i)}} + \frac{\Theta_6}{1+e^{-(\Theta_1 + \Theta_3 x_i)}}$$

Translation in
y direction

Translation in
x direction

Squash or stretch
in x direction





Back Propagation

$$GD \Rightarrow E(\Theta) = L(\Theta) = \frac{1}{2N} \sum_{i=1}^N (y_a - \hat{y}_p)^2$$

$$SGD \Rightarrow E(\Theta) = L(\Theta) = \frac{1}{2} (y_a - f(x))^2$$

$$\frac{\partial E(\Theta)}{\partial \Theta} = \underbrace{-(y_a - f(x))}_{\Delta} \frac{\partial f(x)}{\partial \Theta} = \Delta \frac{\partial f(x)}{\partial \Theta}$$

$$\Theta_0, \Theta_1, \dots, \Theta_6$$

$$f(x) = O_2 = u_2 = \Theta_4 + \Theta_5 O_{11} + \Theta_6 O_{12}$$

$$\Theta_6^{\{t+1\}} \leftarrow \Theta_6^{\{t\}} - \alpha \frac{\partial E(\Theta)}{\partial \Theta_6} = \Theta_6^{\{t\}} - \alpha \Delta \frac{\partial f(x)}{\partial \Theta_6} = \Theta_6^{\{t\}} - \alpha \Delta O_{12}$$

$$\Theta_5^{\{t+1\}} \leftarrow \Theta_5^{\{t\}} - \alpha \frac{\partial E(\Theta)}{\partial \Theta_5} = \Theta_5^{\{t\}} - \alpha \Delta \frac{\partial f(x)}{\partial \Theta_5} = \Theta_5^{\{t\}} - \alpha \Delta O_{11}$$

$$\Theta_4^{\{t+1\}} \leftarrow \Theta_4^{\{t\}} - \alpha \frac{\partial E(\Theta)}{\partial \Theta_4} = \Theta_4^{\{t\}} - \alpha \Delta \frac{\partial f(x)}{\partial \Theta_4} = \Theta_4^{\{t\}} - \alpha \Delta \times 1$$

