① linear regression → $\hat{y} = X\theta \in \mathbb{R}$ linear combination of features

Regression model

② Classification →

→ Generative : NB $\quad P(y|x) = \dfrac{\overbrace{P(x|y)}^{\text{likelihood}}\ \overbrace{P(y)}^{\text{prior}}}{P(x)}$

→ Discriminative : logsitic Regression $\quad P(y|x) = ?$

LR: $P(y|x) = \underbrace{\dfrac{1}{1+e^{-(X\theta)}}}_{\text{sigmoid}}$

$X\theta = S = U = \Sigma$

$\log(\text{odds}) \Big]_{-\infty}^{+\infty} \in \mathbb{R}$

$\text{odds} = \dfrac{p}{1-p}\Big]_{0}^{\infty}$

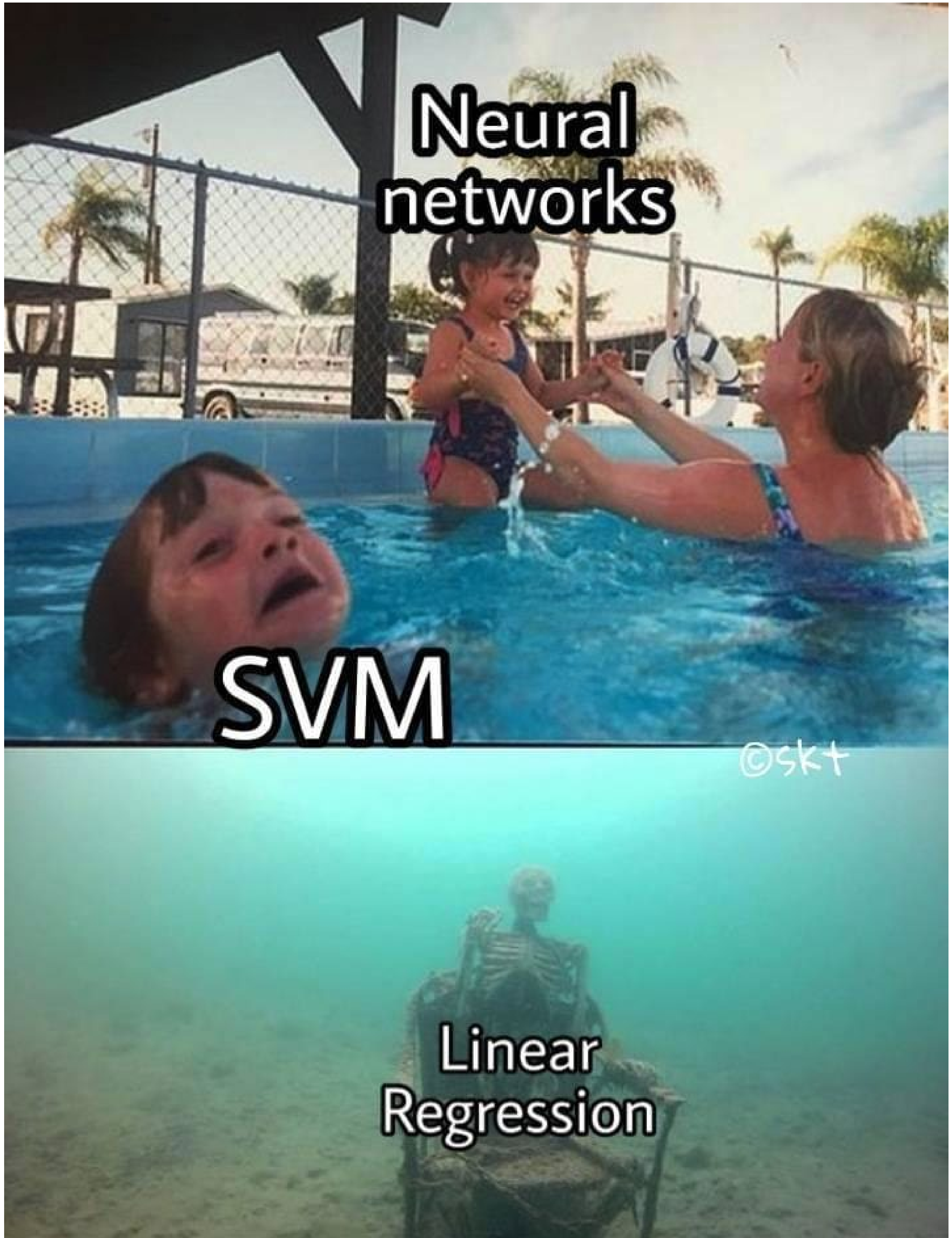
$\log\left(\dfrac{p}{1-p}\right) = X\theta$

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

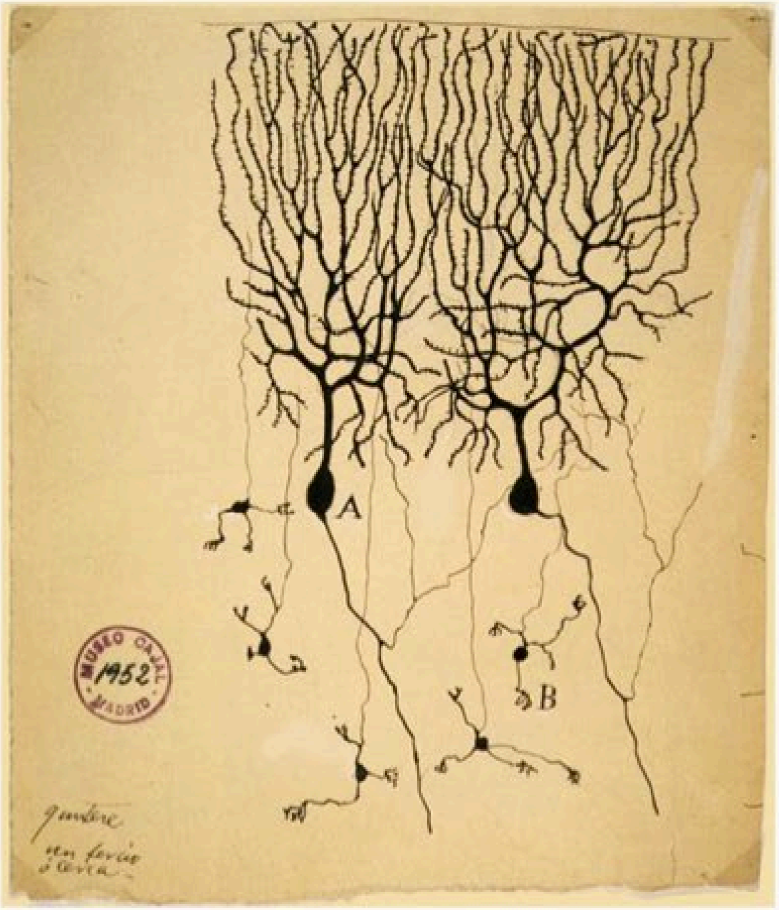
Neural Networks
Forward Pass and Back Propagation

Mahdi Roozbahani

Georgia Tech


Neural networks
SVM
©skt
Linear Regression

Inspiration from Biological Neurons

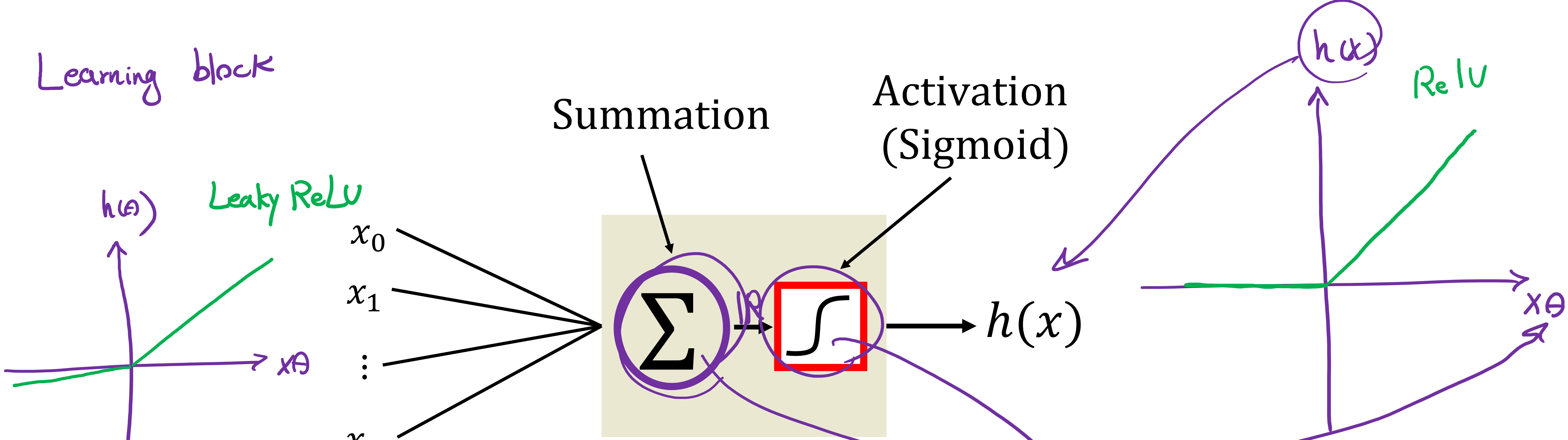


The first drawing of a brain cells by Santiago Ramón y Cajal in 1899

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

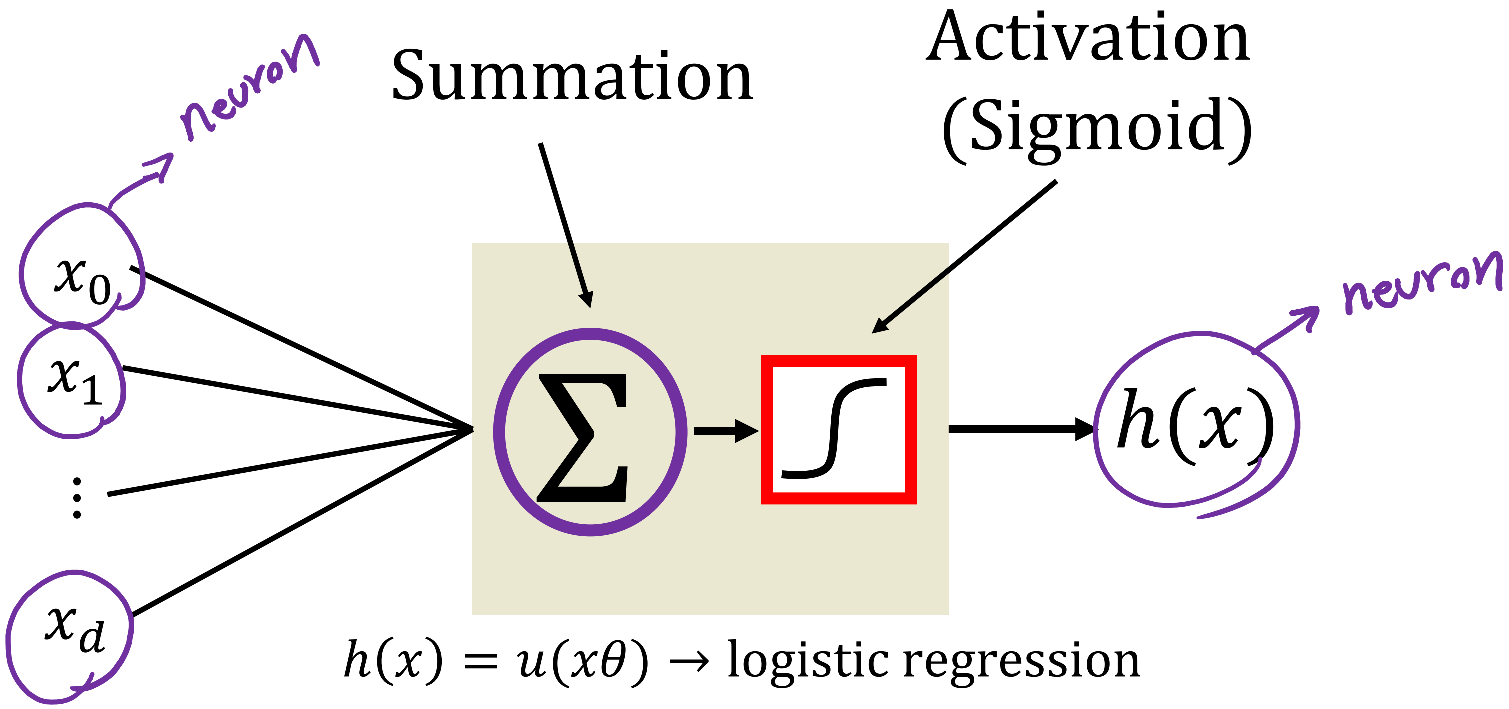
Learning block



$h(x) = u(x\theta) \rightarrow$ logistic regression

$$output = activation(x\theta \ + b)$$

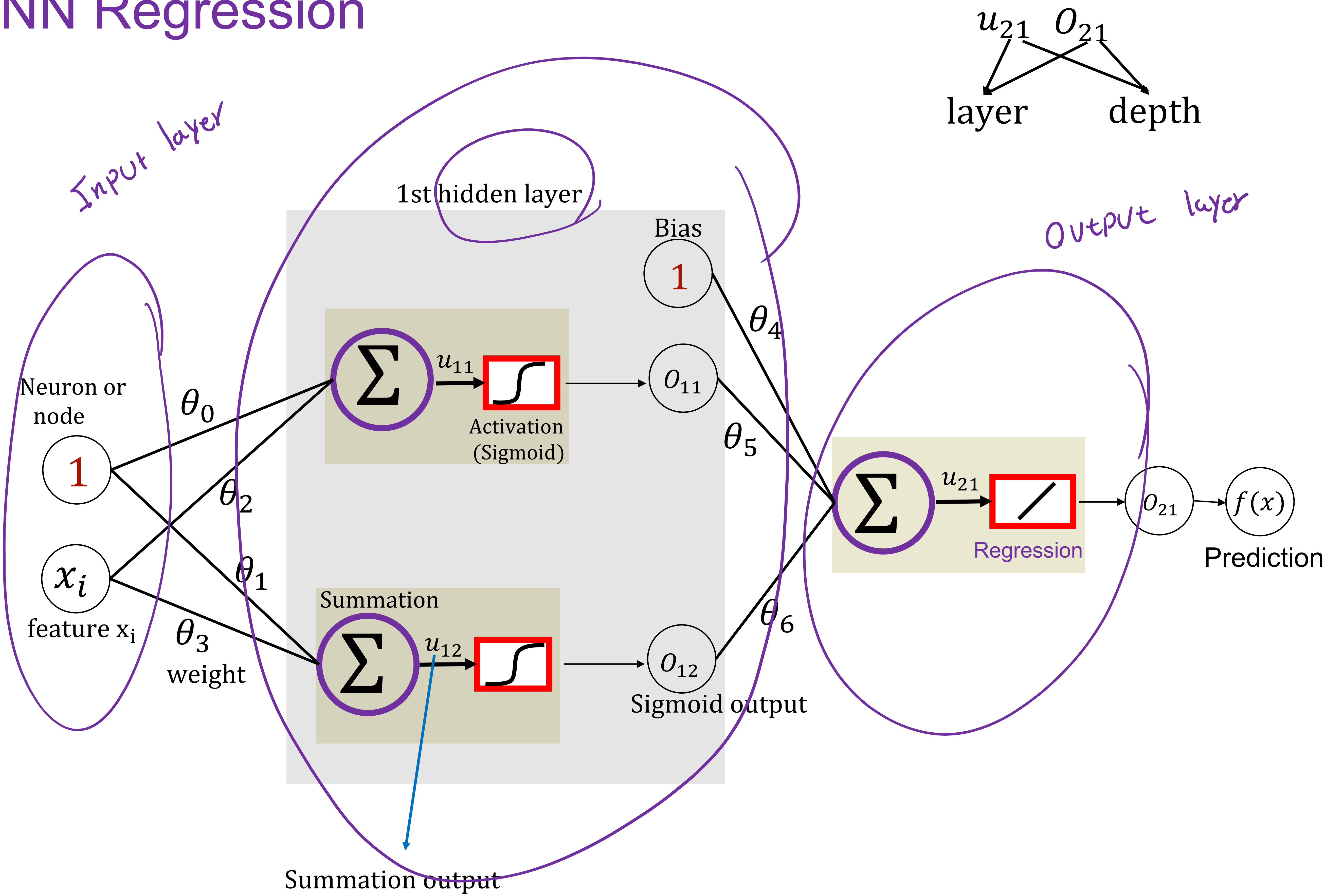
Name of the neuron	Activation function: $activation(z)$
Linear unit	z
Threshold/sign unit	$\operatorname{sgn}(z)$
Sigmoid unit	$\dfrac{1}{1+\exp(-z)}$
Rectified linear unit (ReLU)	$\max(0, z)$
Tanh unit	$\tanh(z)$



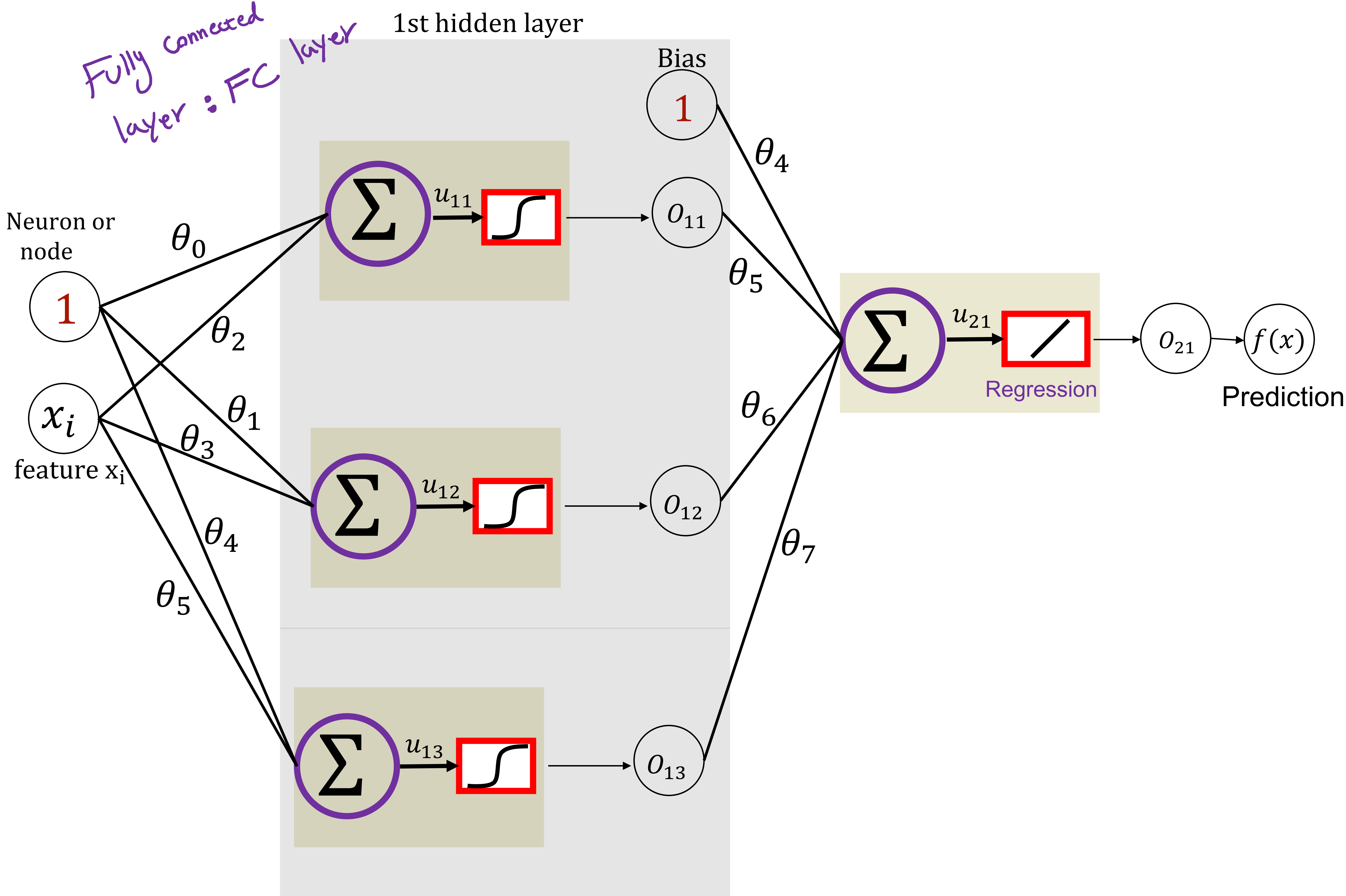
$h(x) = u(x\theta) \rightarrow$ logistic regression

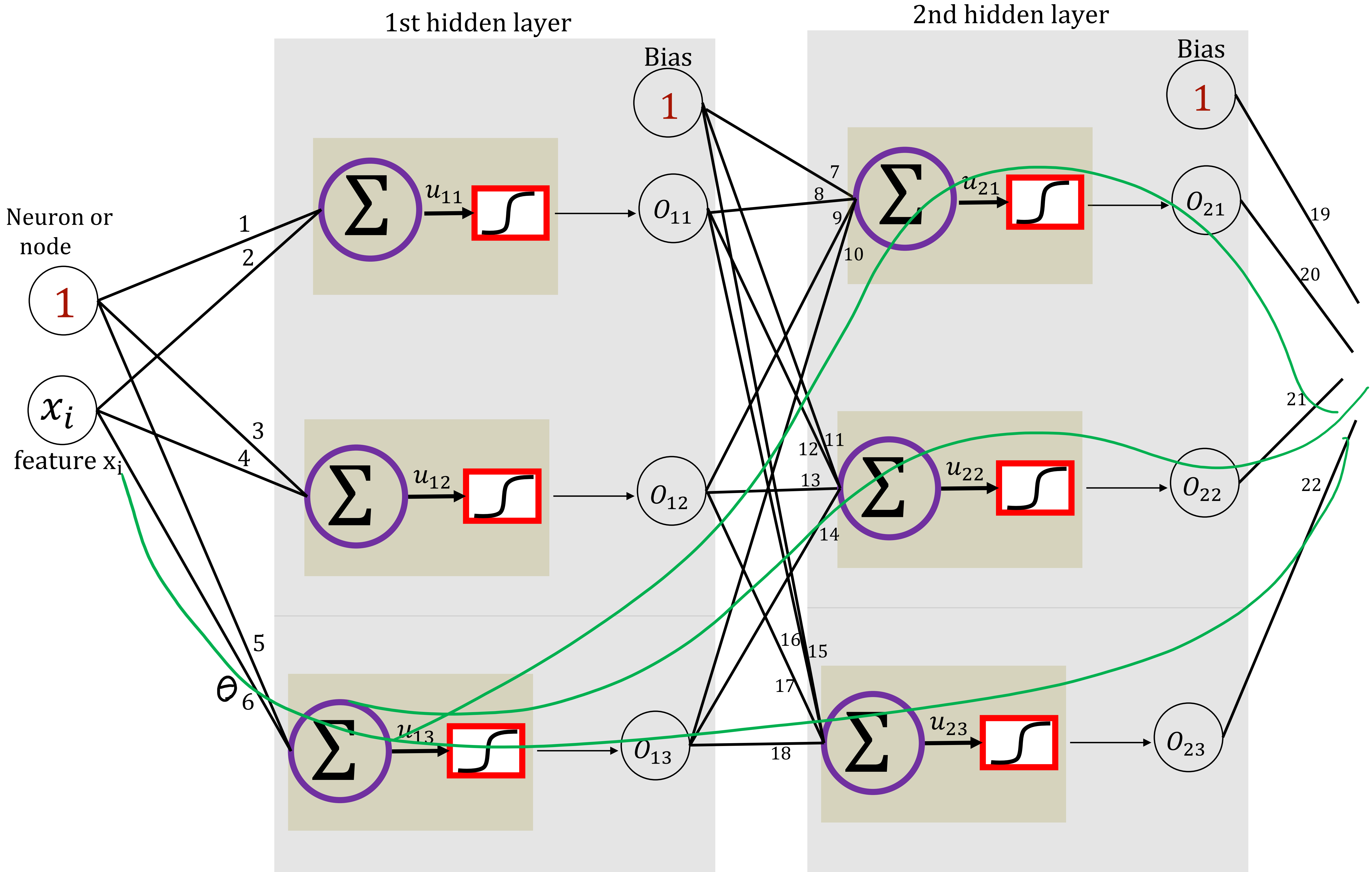
$\Sigma = X\theta$

NN Regression

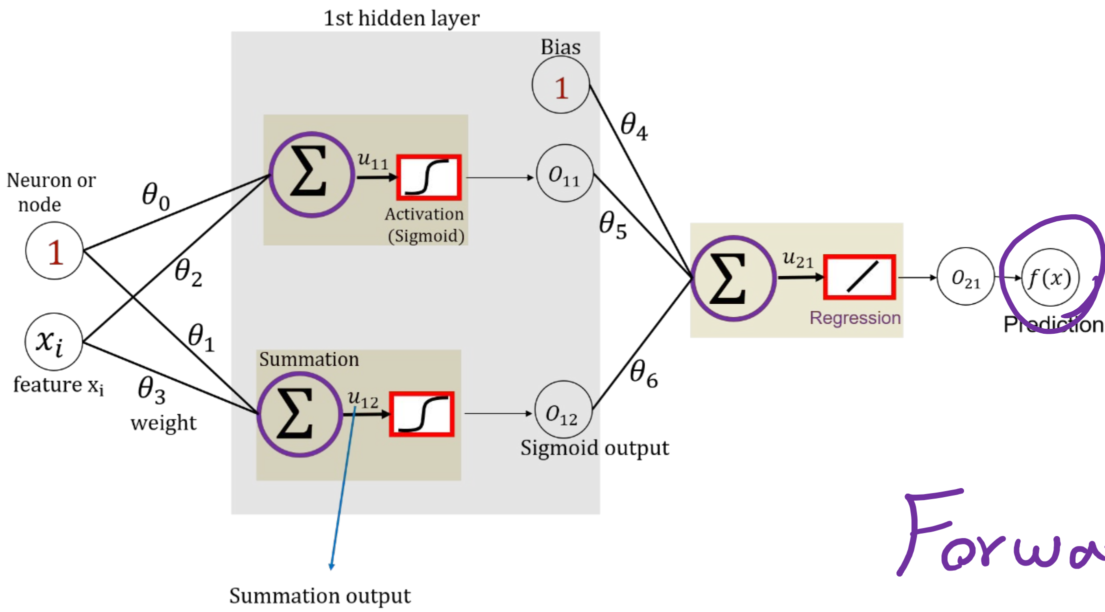


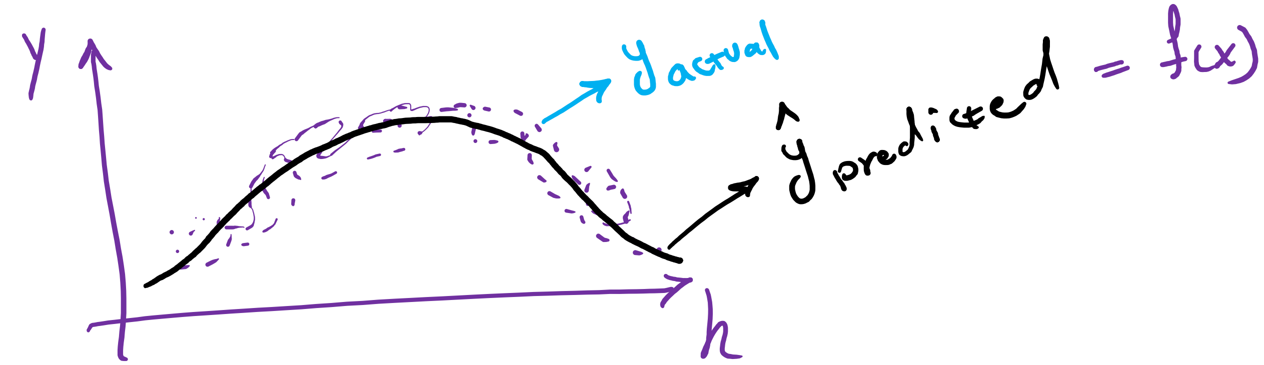
Fully connected layer : FC layer





1st hidden layer
2nd hidden layer
Bias
1
Bias
1
Neuron or node
1
x_i
feature x_i
u_{11}
u_{12}
u_{13}
o_{11}
o_{12}
o_{13}
u_{21}
u_{22}
u_{23}
o_{21}
o_{22}
o_{23}
θ
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

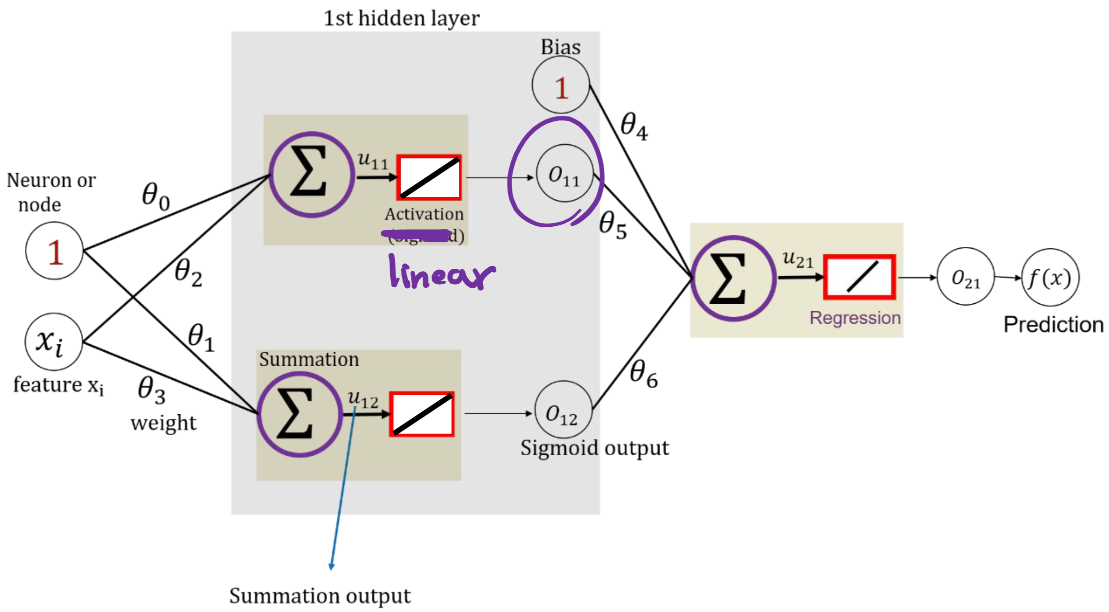




Forward Pass: It calculates all U_s and O_s

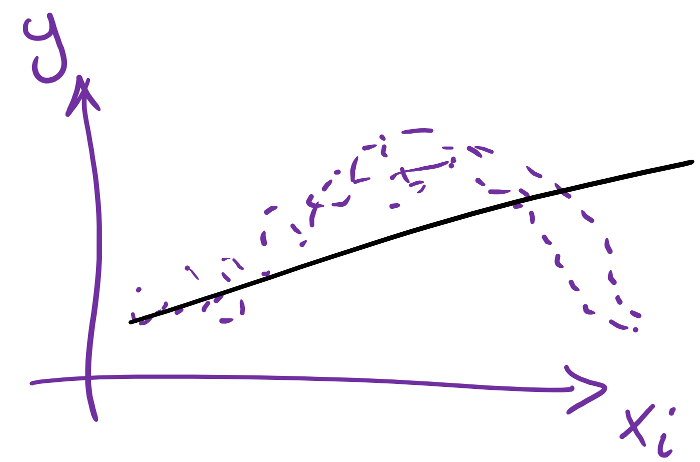
① Initialize parameters randomly

② we need to calculate U_s and ~~then O_s based~~ then on U_s

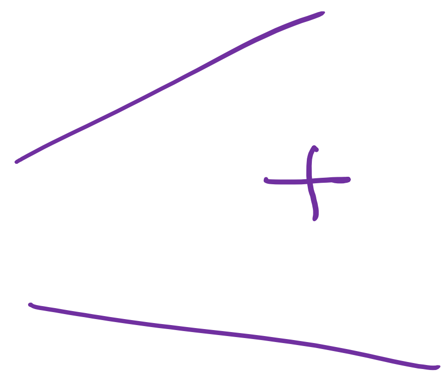


$$U_{11} = 1 \times \Theta_0 + \Theta_2 X_i = \Theta_0 + \Theta_2 X_i$$

$$U_{12} = \Theta_1 + \Theta_3 X_i$$



$$O_{11} = U_{11} = \Theta_0 + \Theta_2 X_i$$

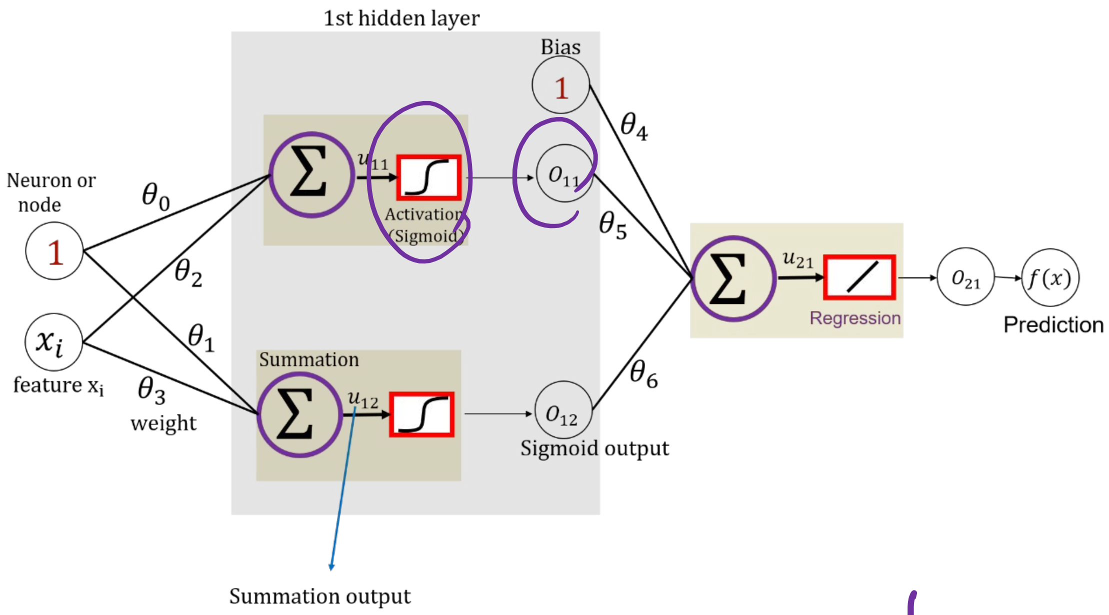


$$O_{12} = U_{12} = \Theta_1 + \Theta_3 X_i$$

$$U_{21} = \Theta_4 + \Theta_5 O_{11} + \Theta_6 O_{12} = \Theta_4 + \Theta_5 (\Theta_0 + \Theta_2 X_i) + \Theta_6 (\Theta_1 + \Theta_3 X_i)$$

$$f(x) = O_{21} = U_{21} = \underbrace{\Theta_4 + \Theta_5 \Theta_0 + \Theta_6 \Theta_1}_{\Theta_{new\,0}} + (\Theta_5 \Theta_2 + \Theta_6 \Theta_3) X_i$$

$$= \Theta_{new\,0} + (\Theta_{1new}) X_i$$

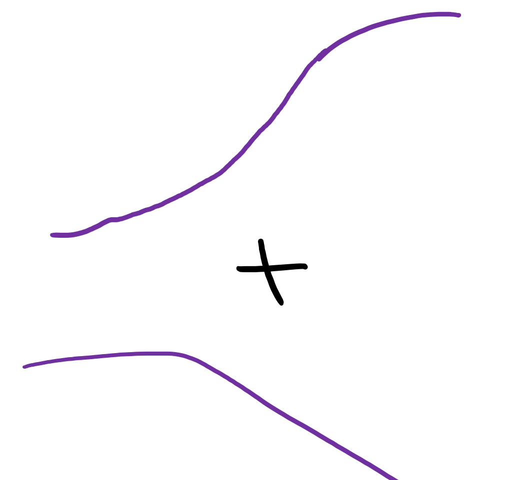


$$U_{11} = \theta_0 + \theta_2 x_i$$

$$U_{12} = \theta_1 + \theta_3 x_i$$

$$O = \frac{1}{1+e^{-U}}$$

$$O_{11} = \frac{1}{1+e^{-U_{11}}} = \frac{1}{1+e^{-(\theta_0+\theta_2 x_i)}}$$



$$O_{12} = \frac{1}{1+e^{-U_{12}}} = \frac{1}{1+e^{-(\theta_1+\theta_3 x_i)}}$$

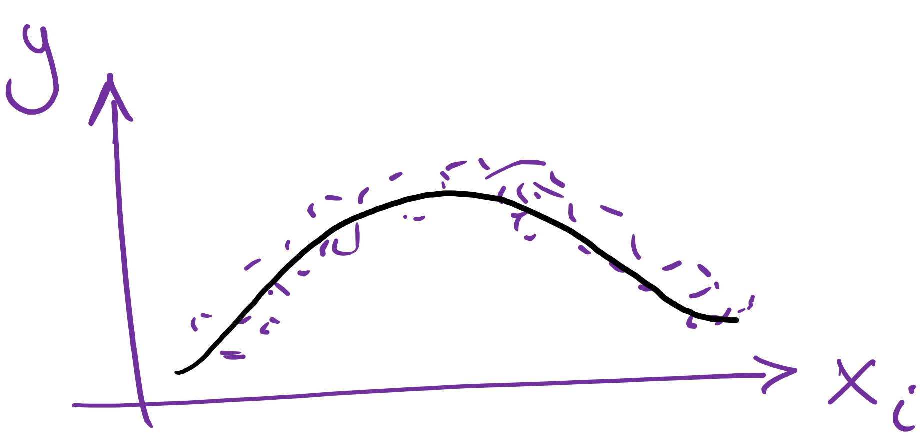
squash or stretch in y direction

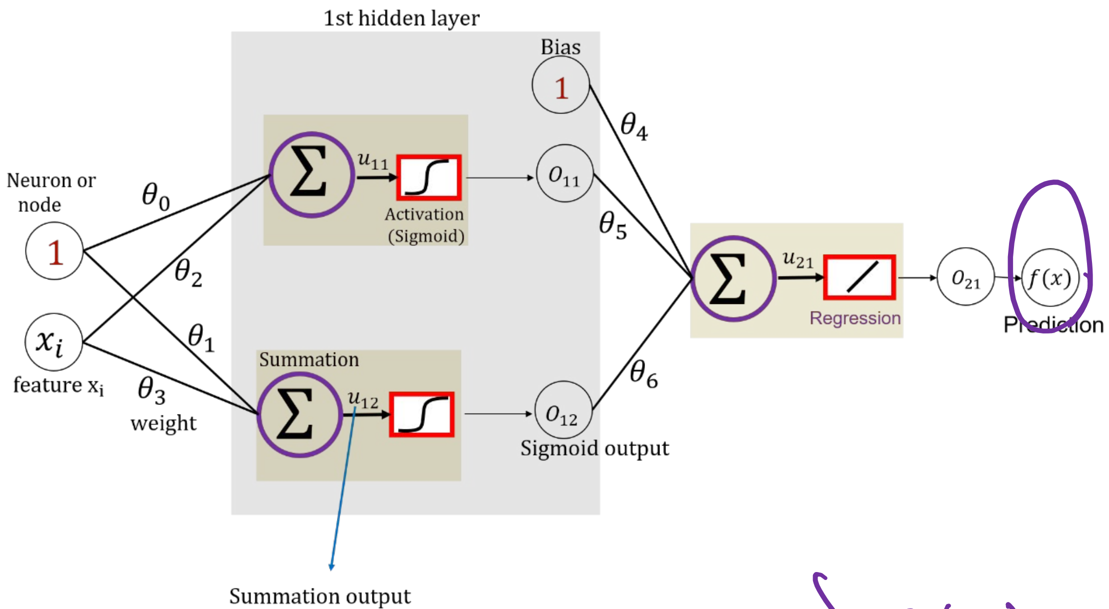
$$U_{21} = \theta_4 + \theta_5 O_{11} + \theta_6 O_{12} = O_{21} = f(x) = \theta_4 + \theta_5 \frac{1}{1+e^{-(\theta_0+\theta_2 x_i)}} + \theta_6 \frac{1}{1+e^{-(\theta_1+\theta_3 x_i)}}$$

translation in y direction

Translation in x direction

squash & stretch in x direction





Back Propagation

$f(x_i)$

$$E(\Theta) = \frac{1}{N}\sum_{i=1}^{N}(y_a - y_p)^2$$

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

$$\frac{\partial E(\Theta)}{\partial \Theta} = 0 = \nabla_{\Theta} E(\Theta) = \underline{-(y_a - f(x))}\,\frac{\partial f(x)}{\partial \Theta}$$

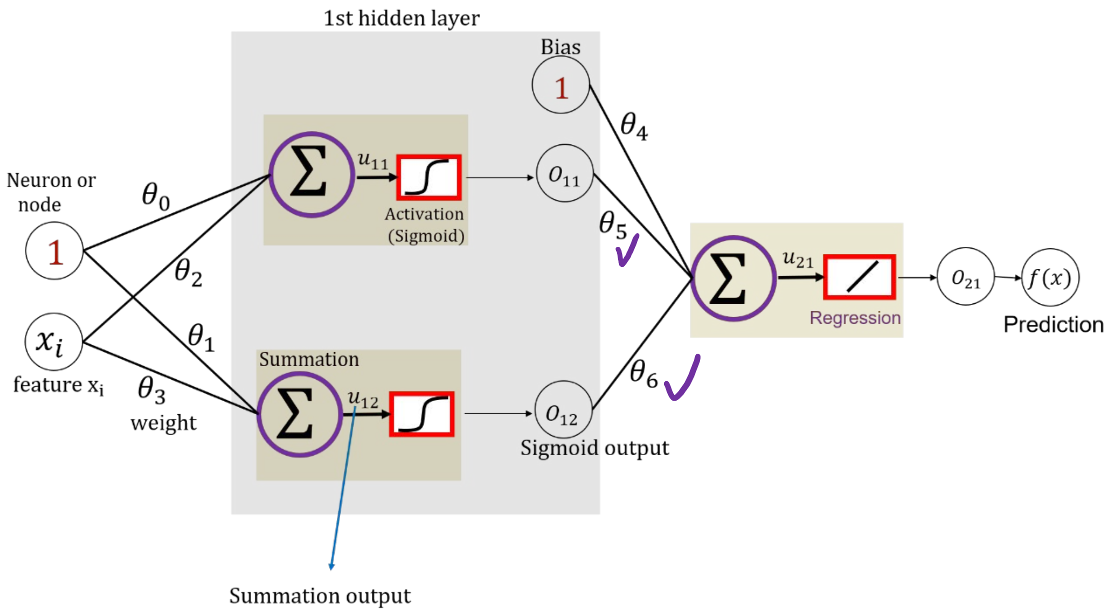
$$= \Delta \frac{\partial f(x)}{\partial \Theta}$$

$$\nabla_{\Theta} E(\Theta) = \Delta \frac{\partial f(x)}{\partial \Theta}$$

$$\Theta_0, \Theta_1, \Theta_2, \Theta_3, \Theta_4, \Theta_5, \Theta_6$$

$$\frac{\partial f(x)}{\partial \Theta_0}, \frac{\partial f(x)}{\partial \Theta_1}, \ldots\ldots, \frac{\partial f(x)}{\partial \Theta_6}$$

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

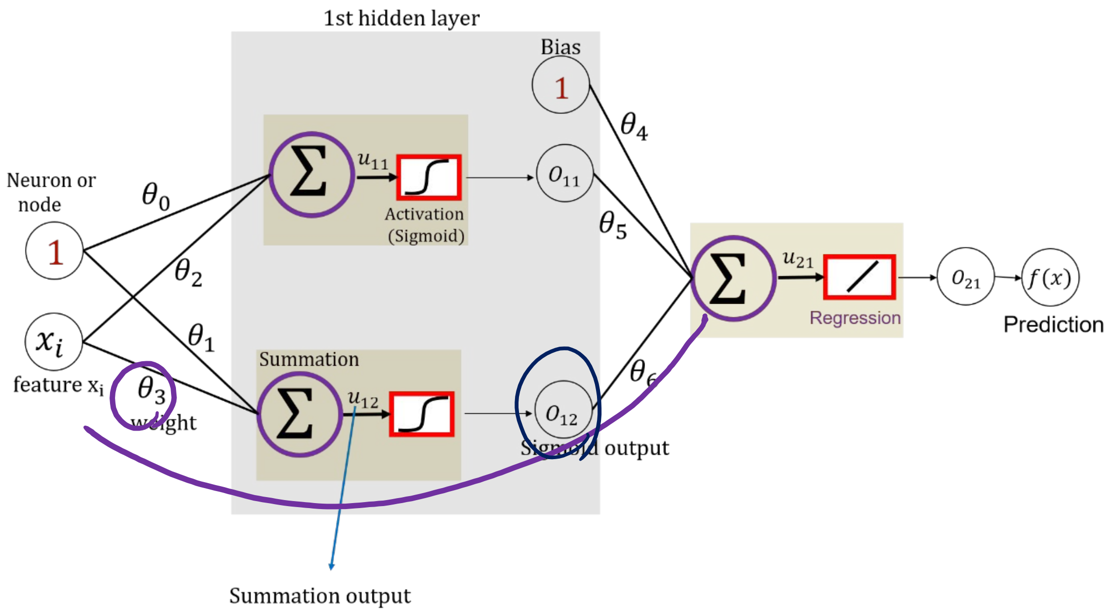


$$f(x) = O_{21} = U_{21} = \theta_4 + \theta_5 O_{11} + \theta_6 O_{12}$$

$$\frac{\partial f(x)}{\partial \theta_6} = O_{12} \qquad \theta_6^{\{t+1\}} \leftarrow \theta_6^{\{t\}} - \alpha \Delta O_{12}$$

$$\frac{\partial f(x)}{\partial \theta_5} = O_{11} \qquad \theta_5^{\{t+1\}} \leftarrow \theta_5^{\{t\}} - \alpha \Delta O_{11}$$

$$\frac{\partial f(x)}{\partial \theta_4} = 1 \qquad \theta_4^{\{t+1\}} \leftarrow \theta_4^{\{t\}} - \alpha \Delta$$



$$f(x) = \theta_4 + \theta_5 O_{11} + \theta_6 O_{12}$$

$$O_{12} = \frac{1}{1+e^{-U_{12}}}$$

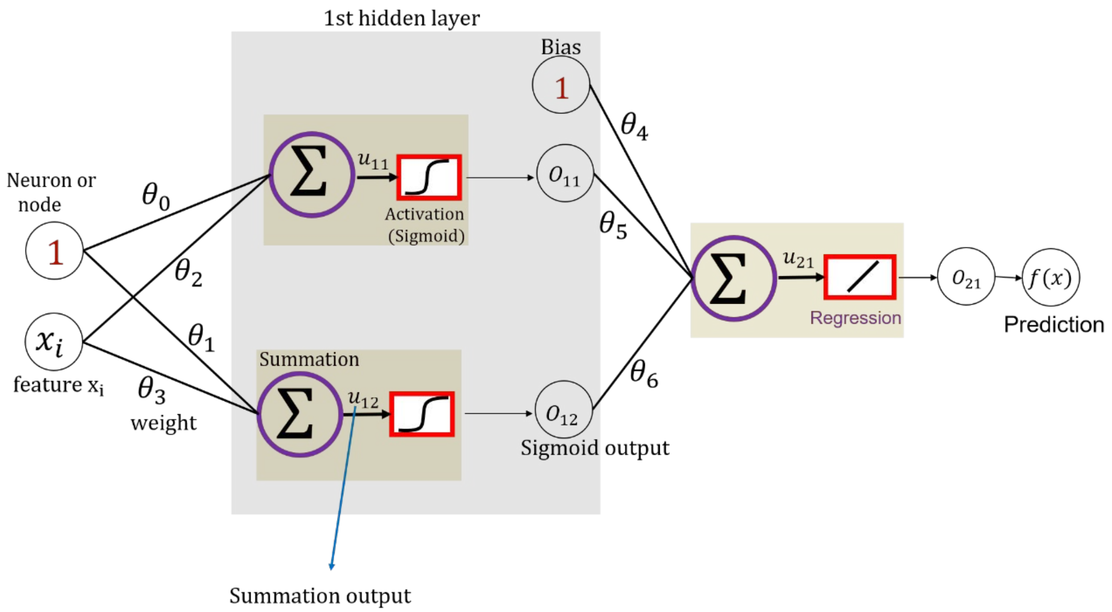
$$U_{12} = \theta_1 + \theta_3 X_i$$

$$\frac{\partial f(x)}{\partial \theta_3} = \frac{\partial f(x)}{\partial O_{12}} \frac{\partial O_{12}}{\partial U_{12}} \frac{\partial U_{12}}{\partial \theta_3} = \theta_6 \quad O_{12}[1-O_{12}] \quad X_i$$

$$O = \frac{1}{1+e^{-U}} = (1+e^{-U})^{-1} \qquad \frac{\partial O}{\partial U} = -1 \times -1 \times e^{-U} \times (1+e^{-U})^{-2} = \frac{e^{-U}}{(1+e^{-U})^2}$$

$$\frac{\partial O}{\partial U} = \frac{1+e^{-U}-1}{(1+e^{-U})^2} = \frac{1}{1+e^{-U}}\left[\frac{1+e^{-U}}{1+e^{-U}} - \frac{1}{1+e^{-U}}\right] = \frac{1}{1+e^{-U}}\left[1 - \frac{1}{1+e^{-U}}\right]$$

$$\frac{\partial O}{\partial U} = O[1-O]$$



$$\theta_3^{\{t+1\}} \Leftarrow \theta_3^{\{t\}} - \alpha \Delta \theta_6 O_{12} [1 - O_{12}] x_i$$

Vanishing gradients

Exploding gradients