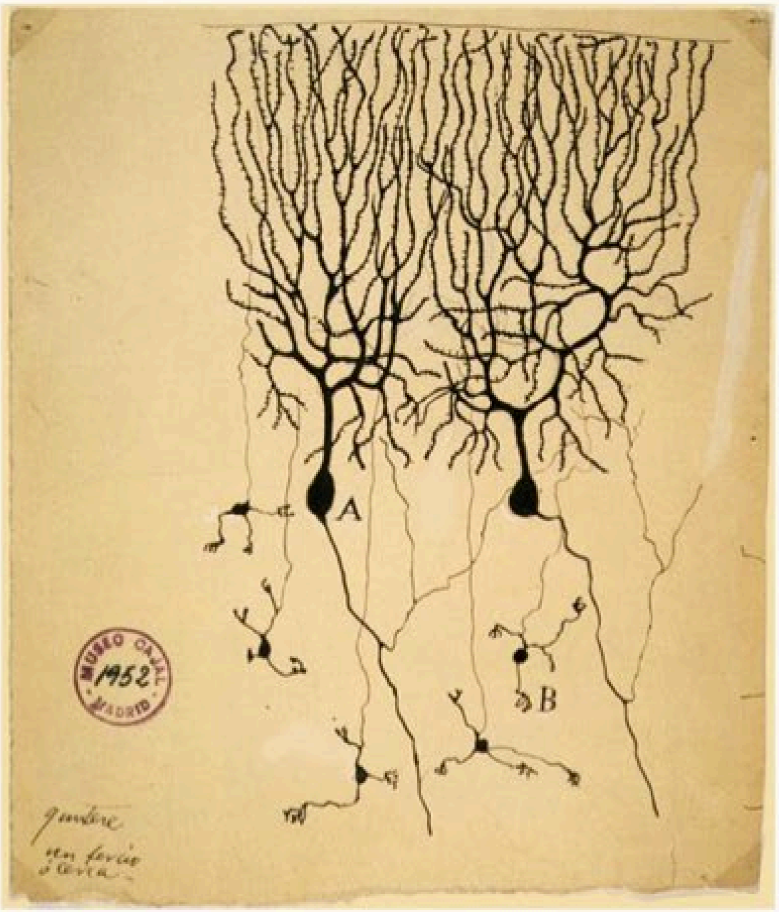
Neural Networks
Forward Pass and Back Propagation

Mahdi Roozbahani

Georgia Tech

Inspiration from Biological Neurons

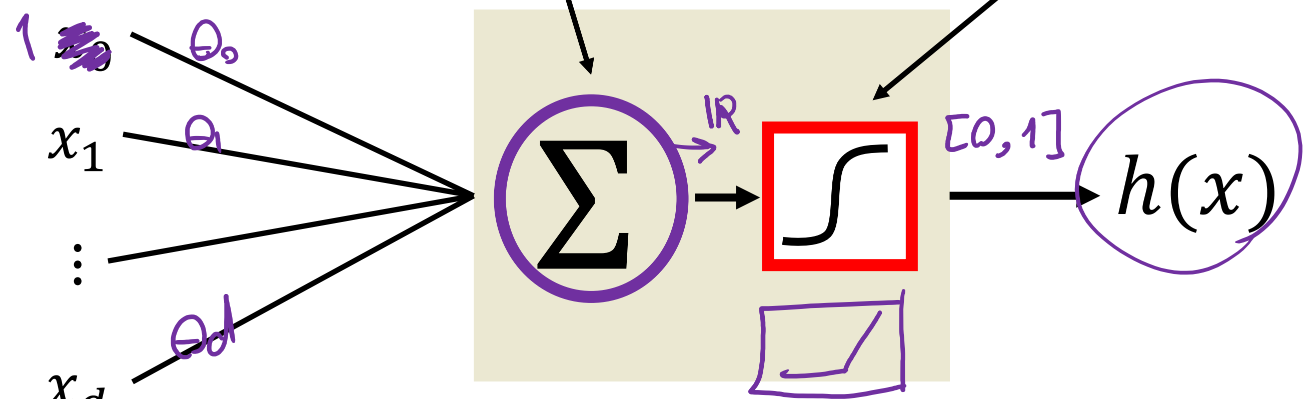


The first drawing of a brain cells by Santiago Ramón y Cajal in 1899

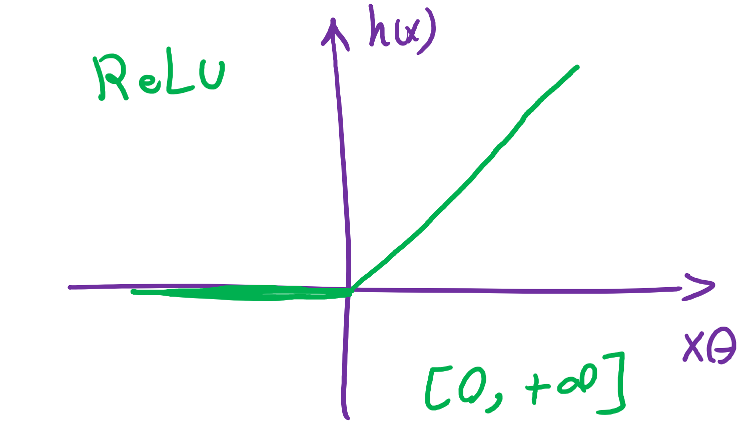
Neurons: core components of brain and the nervous system consisting of

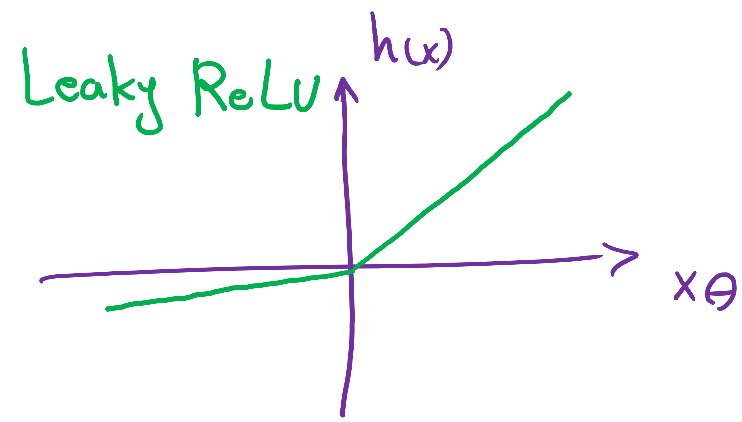
1. Dendrites that collect information from other neurons
2. An axon that generates outgoing spikes

$\Sigma = x\theta$



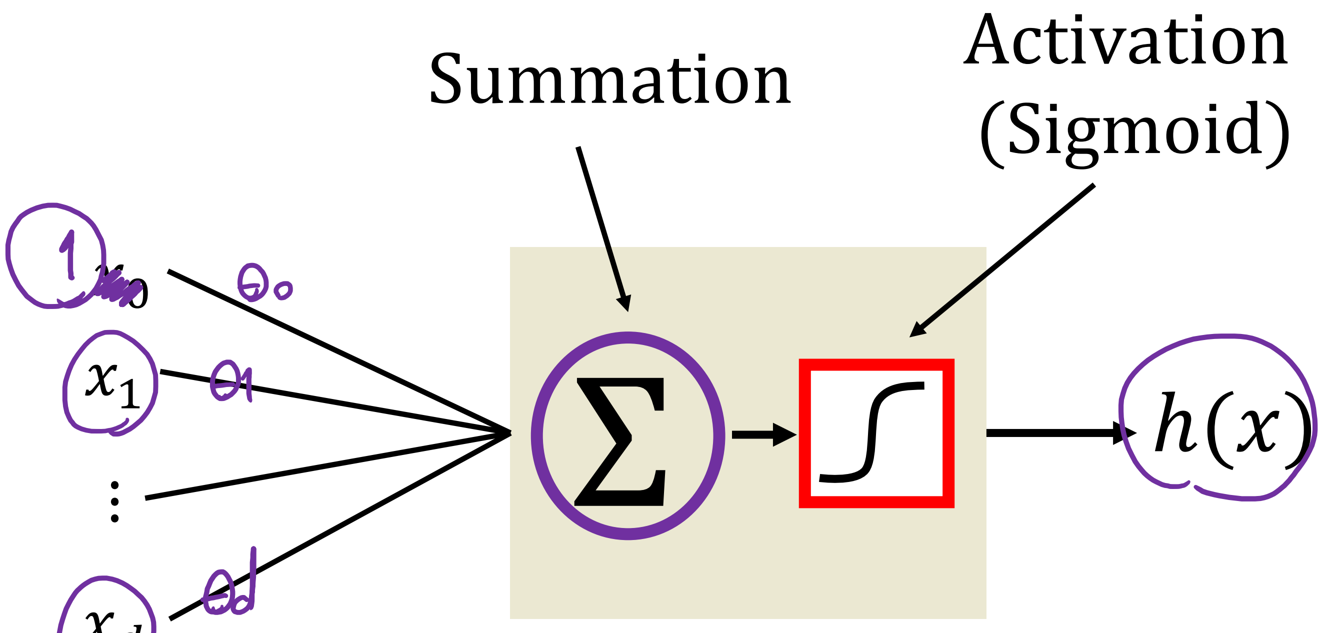
$h(x) = u(x\theta) \rightarrow$ logistic regression





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

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



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

u

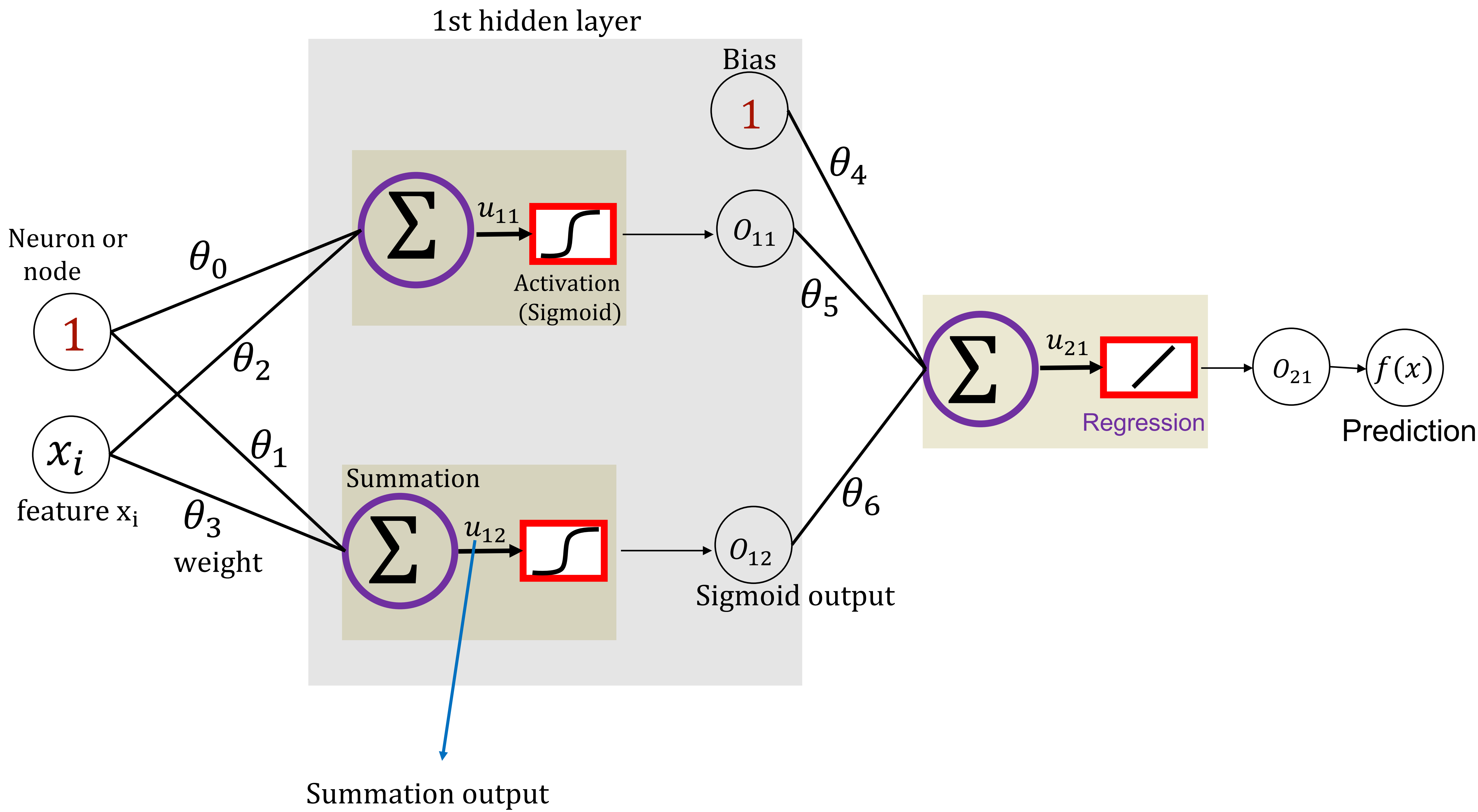
$$\Sigma = \theta_0 + \theta_1 x_1 + \cdots + \theta_d x_d = x\theta$$

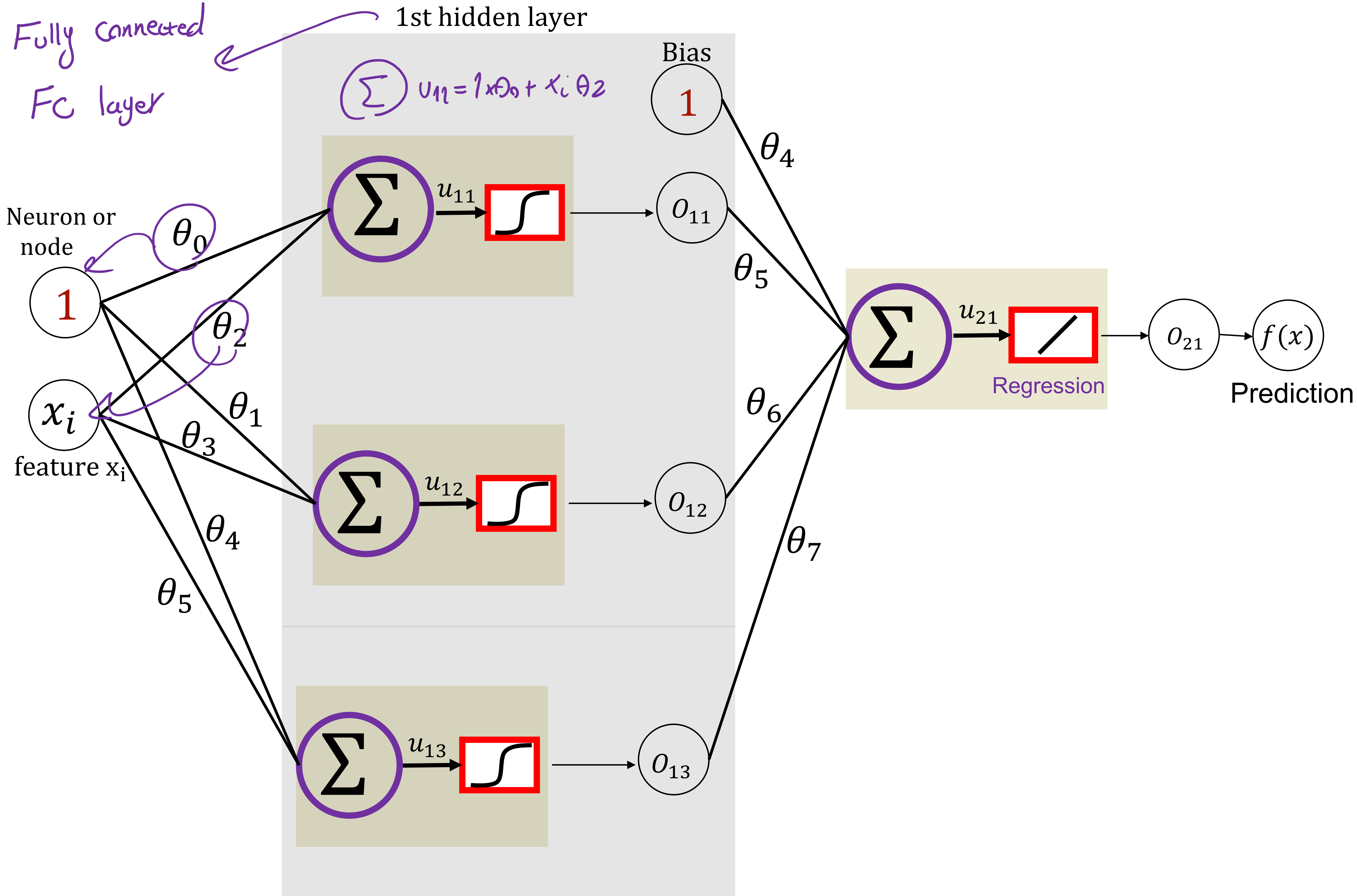
$$\Sigma \rightarrow \boxed{S} = \frac{1}{1 + \exp(-x\theta)} = \bigcirc$$

NN Regression

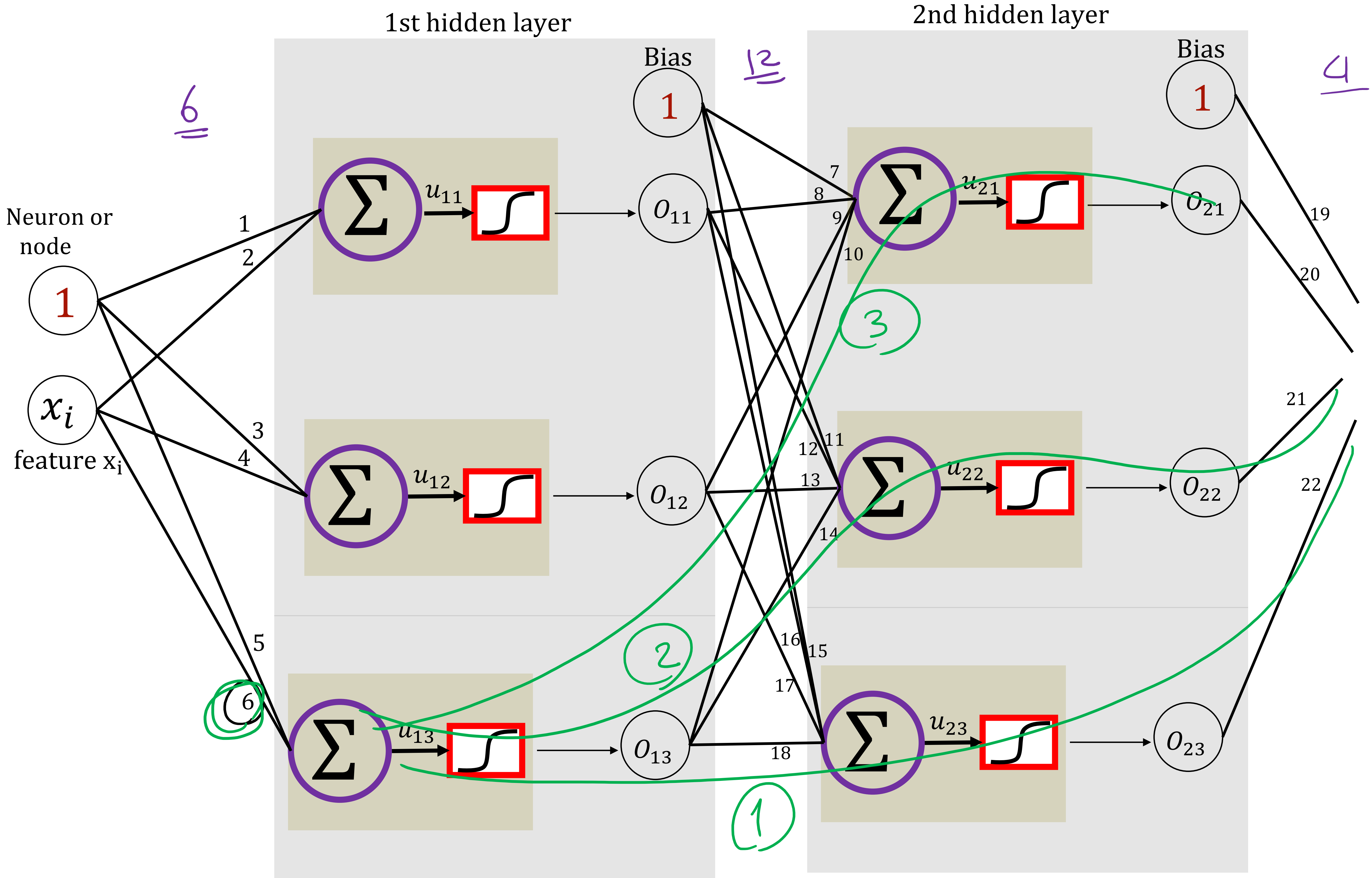
linear combination of features ← u_{21} O_{21} → output of activation function

layer depth

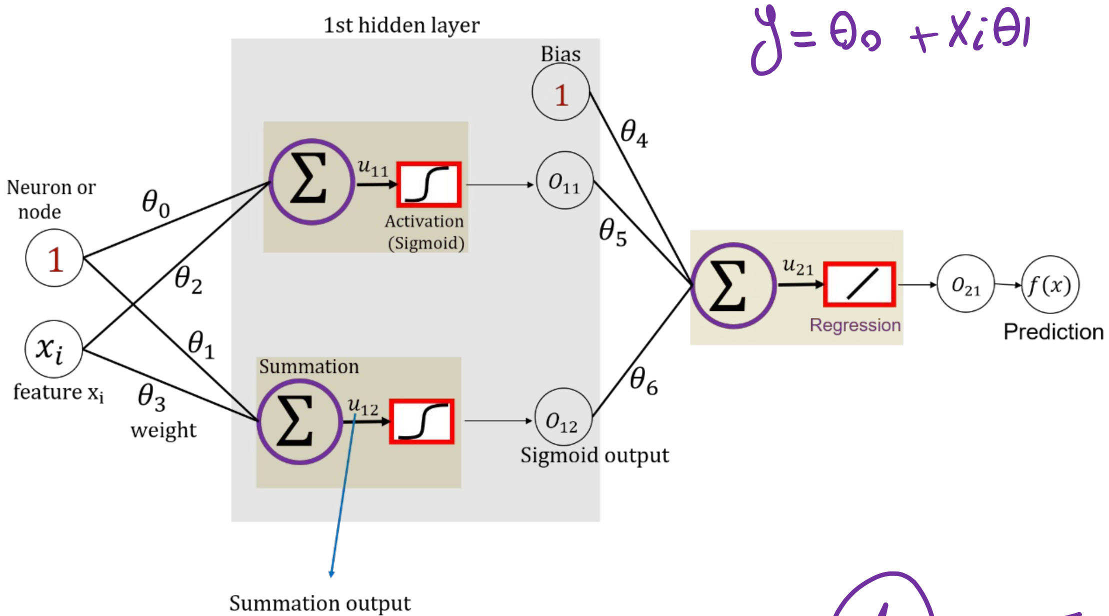




Fully connected
FC layer
1st hidden layer
Bias
$u_{11} = 1 \times \theta_0 + x_i \theta_2$
Neuron or node
θ_0
θ_2
θ_1
θ_3
θ_4
θ_5
feature x_i
u_{11}
u_{12}
u_{13}
o_{11}
o_{12}
o_{13}
θ_4
θ_5
θ_6
θ_7
u_{21}
Regression
o_{21}
$f(x)$
Prediction



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



$y = \theta_0 + x_i\theta_1$

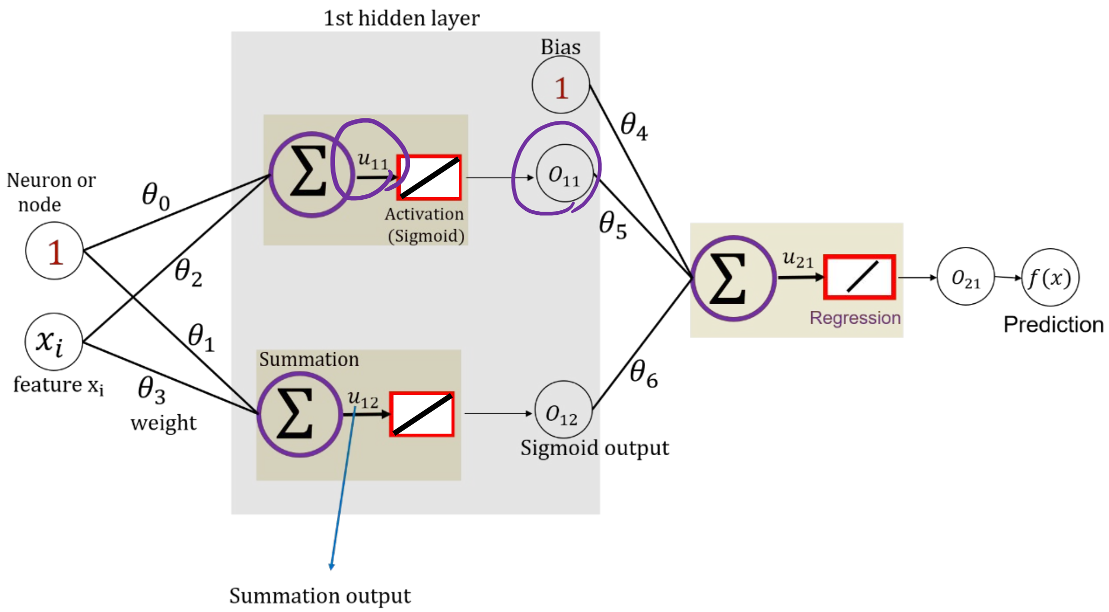
1. Initialize all Parameters (θs) ; Do not initialize them with zero values

2. Forward Pass → Calculate Us and θs
   Left to Right

3. Backpropagation to optimize θs

4. Check for convergence

Forward Pass

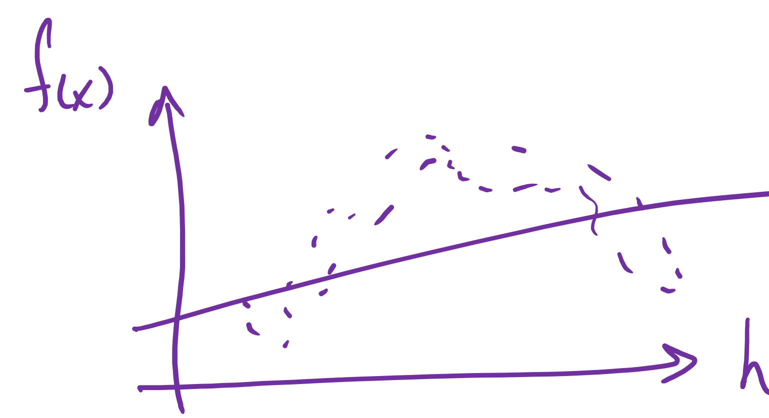


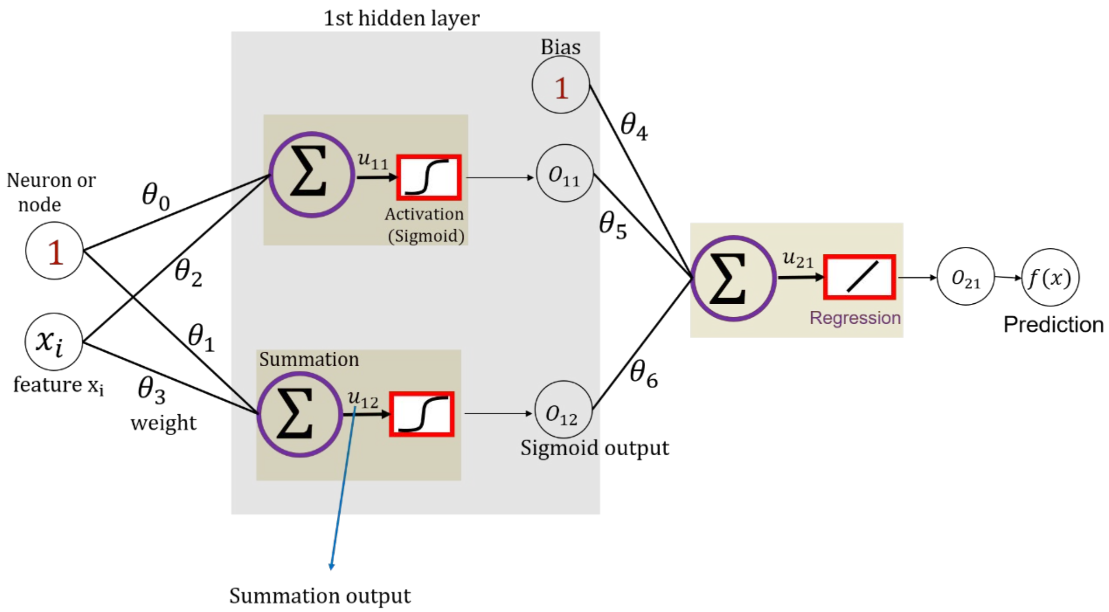
$$U_{11} = \theta_0 + \theta_2 X_i \qquad O_{11} = U_{11} = \theta_0 + \theta_2 X_i$$

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

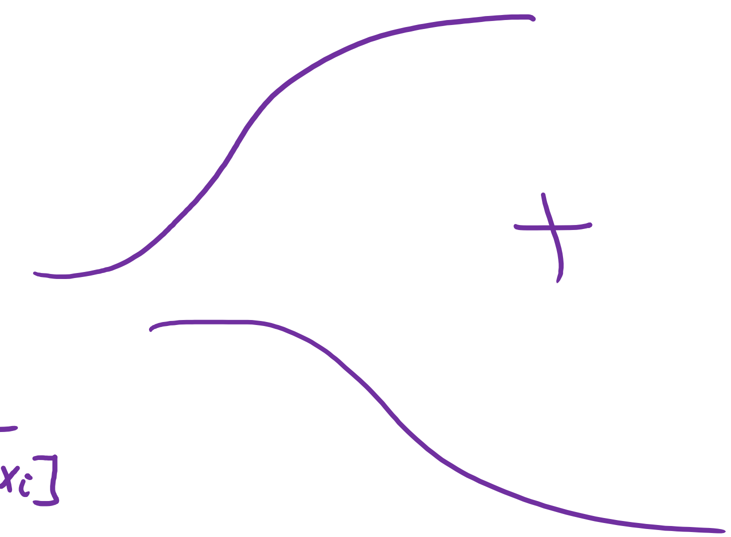
$$U_{21} = \theta_4 + \theta_5 O_{11} + \theta_6 O_{12} = \theta_4 + \theta_5 (\theta_0 + \theta_2 X_i) + \theta_6 (\theta_1 + \theta_3 X_i)$$

$$U_{21} = O_{21} = f(x)$$





$$u_{11} = \theta_0 + \theta_2 x_i \Rightarrow o_{11} = \frac{1}{1+\exp(-u_{11})} = \frac{1}{1+\exp(-\theta_0-\theta_2 x_i)}$$



$$u_{12} = \theta_1 + \theta_3 x_i \Rightarrow o_{12} = \frac{1}{1+\exp(-u_{12})} = \frac{1}{1+\exp^{-[\theta_1+\theta_3 x_i]}}$$

$\theta_5 \rightarrow$ Squash it in y direction

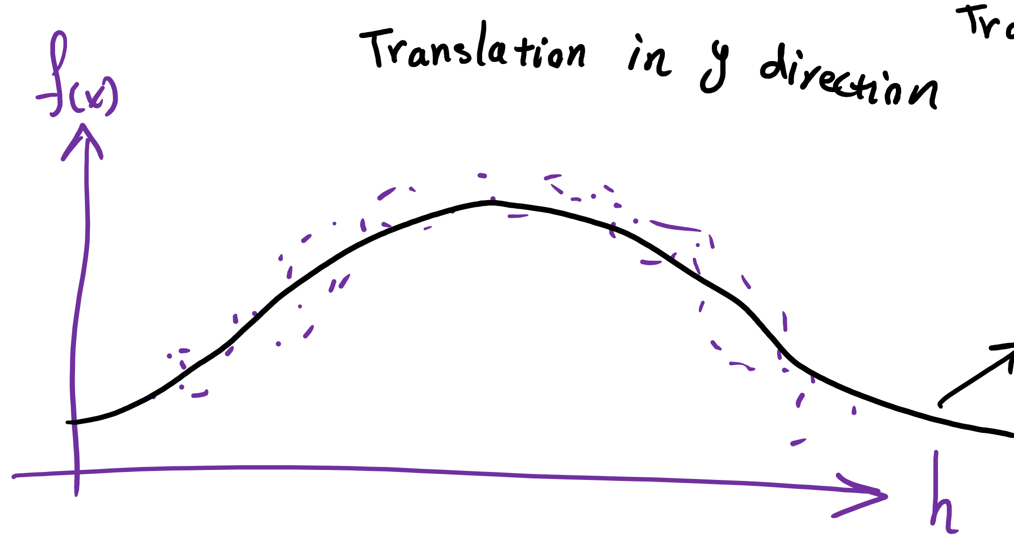
$$u_{21} = \theta_4 + \theta_5 o_{11} + \theta_6 o_{12} = \theta_4 + \frac{\theta_5}{1+\exp^{-[\theta_0+\theta_2 x_i]}} + \frac{\theta_6}{1+\exp^{-[\theta_1+\theta_3 x_i]}}$$

Translation in y direction

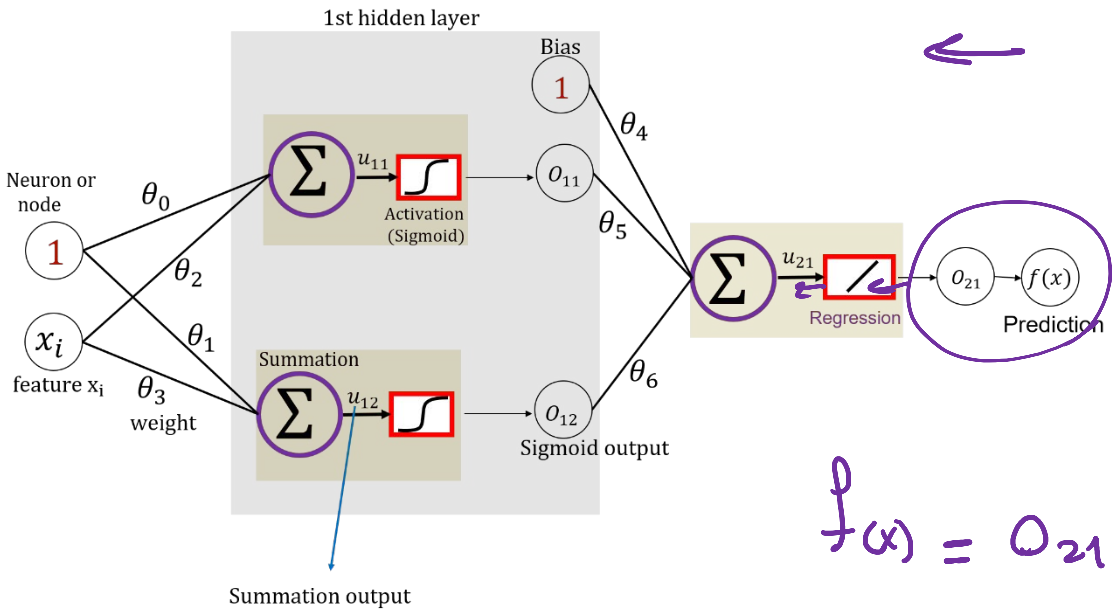
Translation in x direction

Squash it in x direction

$$u_{21} = o_{21} = f(x)$$



y predicted



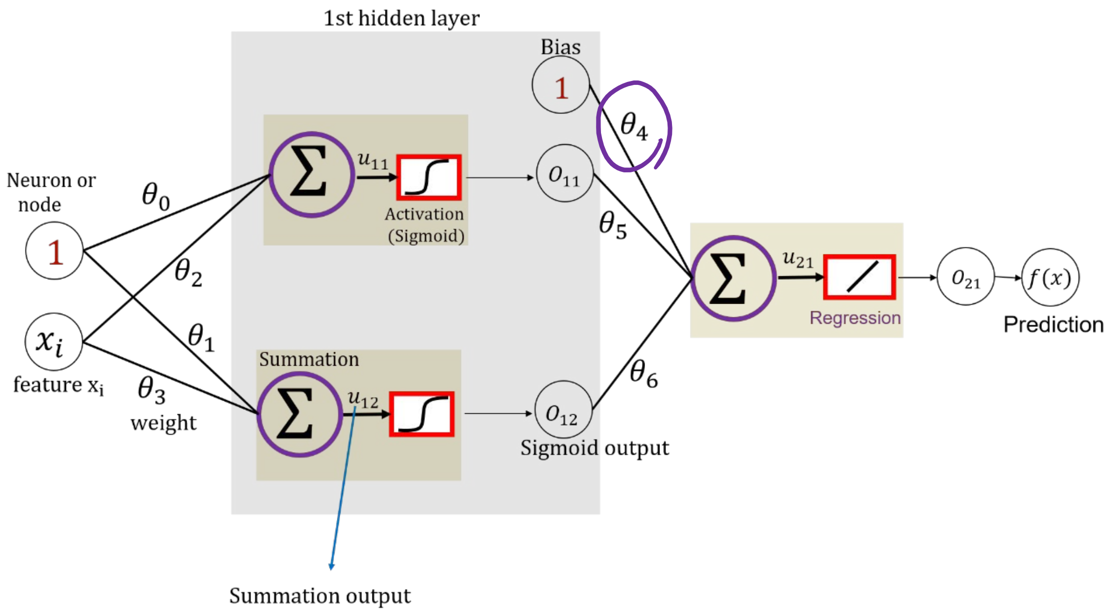
$$E(\Theta) = \frac{1}{N} \sum_{i=1}^{N} (y_a - f(x))^2$$

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

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

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

$$\frac{\partial f(x)}{\partial \Theta_4}, \frac{\partial f(x)}{\partial \Theta_5}, \frac{\partial f(x)}{\partial \Theta_6}, \frac{\partial f(x)}{\Theta_3}, \frac{\partial f(x)}{\Theta_1}, \frac{\partial f(x)}{\partial \Theta_2}, \frac{\partial f(x)}{\partial \Theta_0}$$



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

SGD

$$\theta_4^{\{t+1\}} \leftarrow \theta_4^{\{t\}} - \alpha \nabla_\theta E(\theta)$$

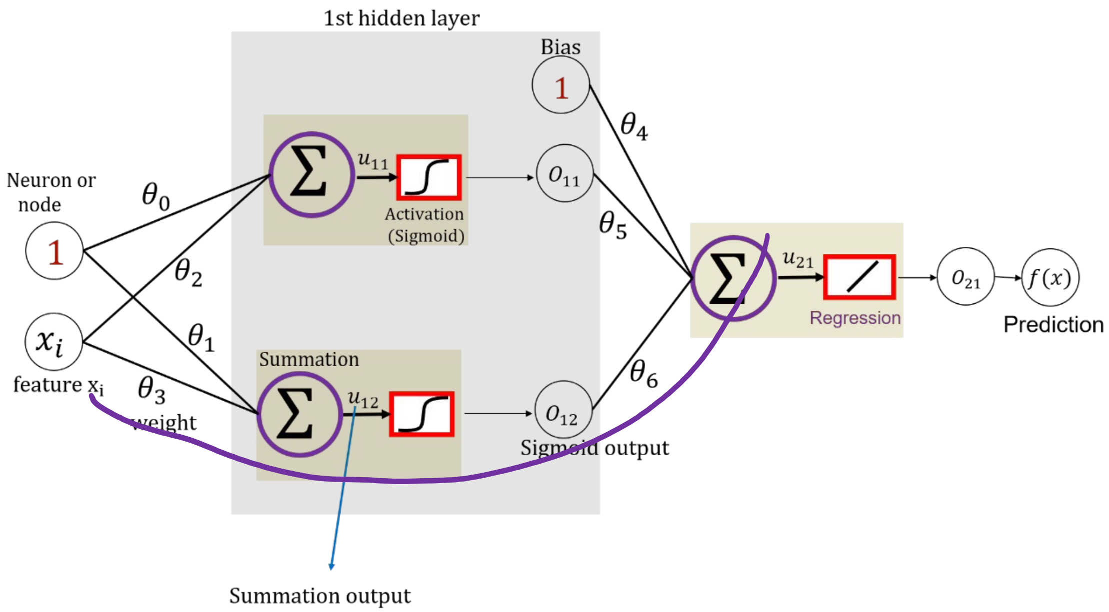
$$\theta_4^{\{t+1\}} \leftarrow \theta_4^{\{t\}} - \alpha \Delta \frac{\partial f(x)}{\partial \theta_4}$$

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

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

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

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



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

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

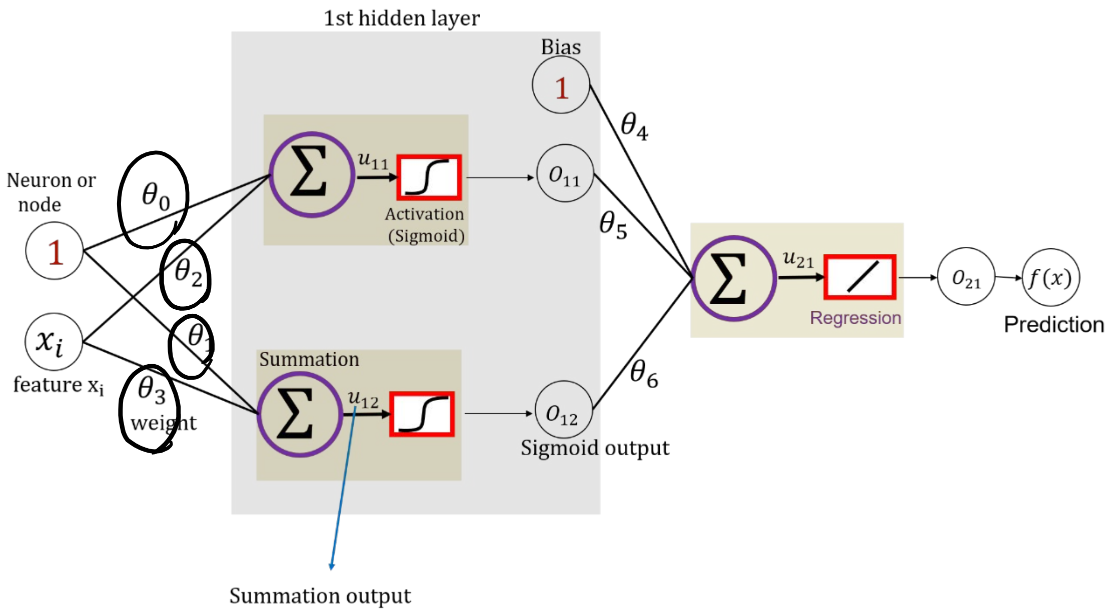
$$\frac{\partial f(x)}{\partial \theta_3} = \frac{\partial f(x)}{\partial O_{12}} \; \frac{\partial O_{12}}{\partial U_{12}} \; \frac{\partial U_{12}}{\partial \theta_3} = \theta_6 \; O_{12}[1 - O_{12}] \; X_i$$

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

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

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

$$= O[1 - O]$$



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

Vanishing gradient issue