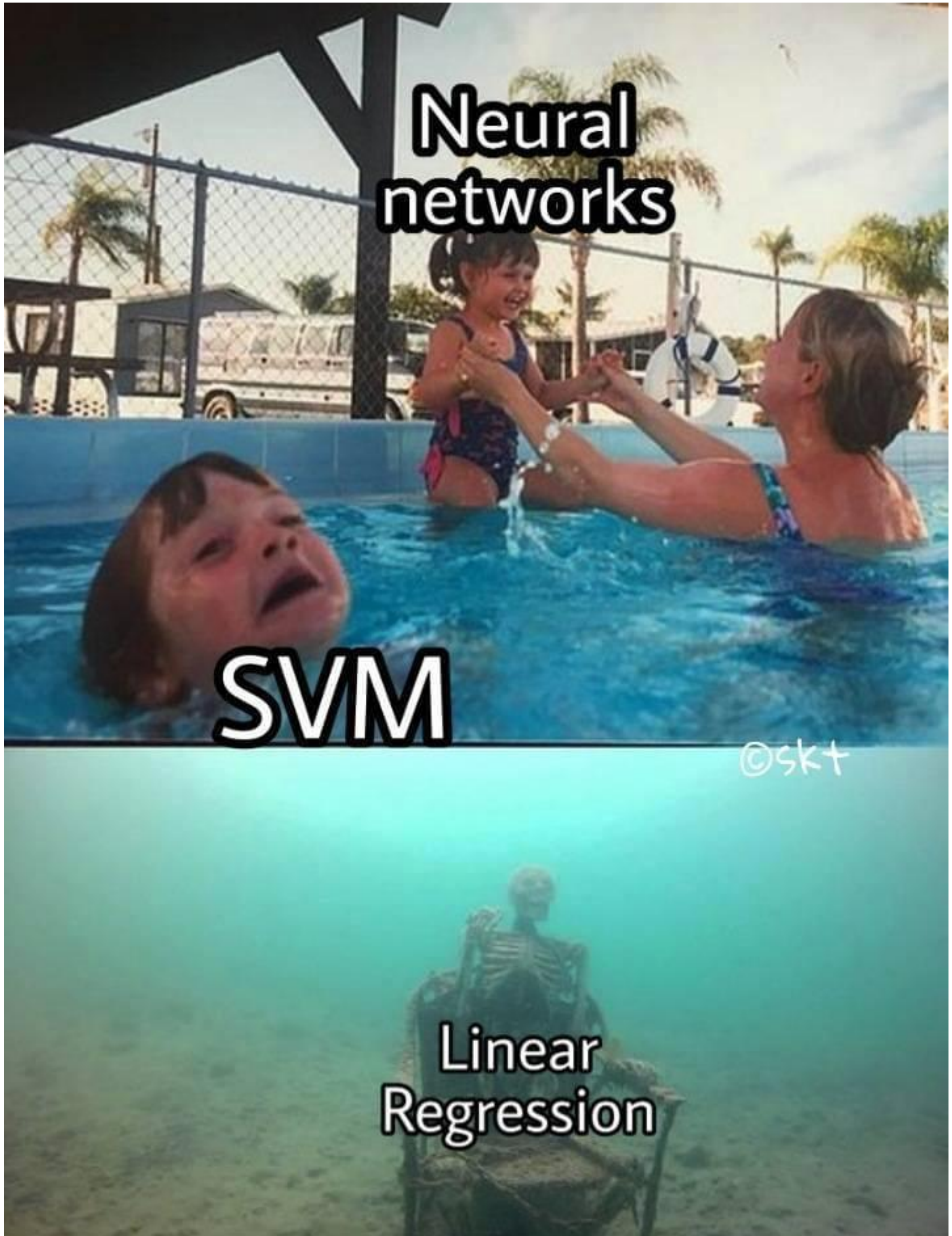
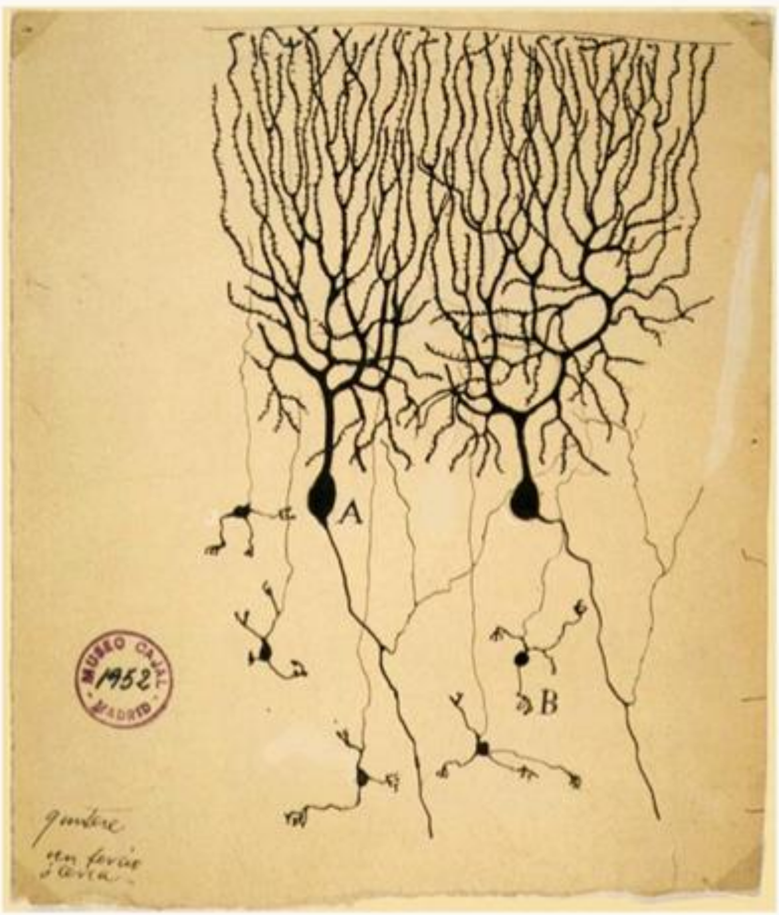
Neural Networks
Forward Pass and Back Propagation

Mahdi Roozbahani

Georgia Tech


Neural
networks
SVM
©skt
Linear
Regression

Inspiration from Biological Neurons

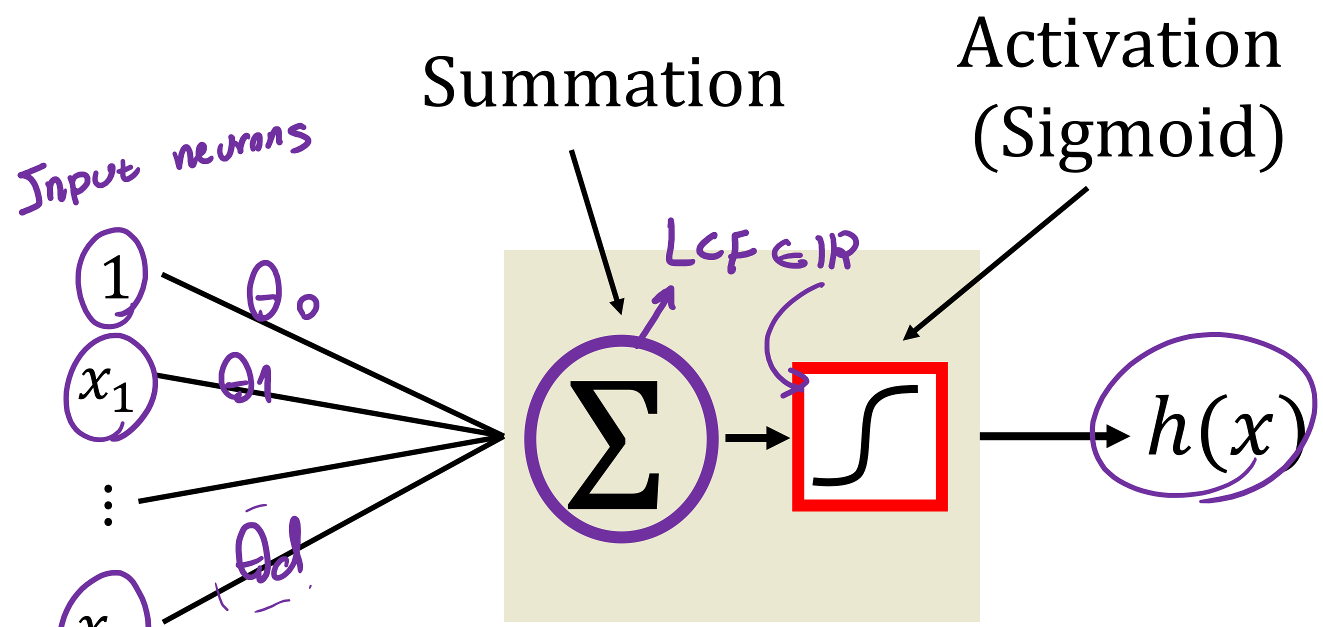


The first drawing of a brain cells by Santiago Ramón y Cajal in 1899

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

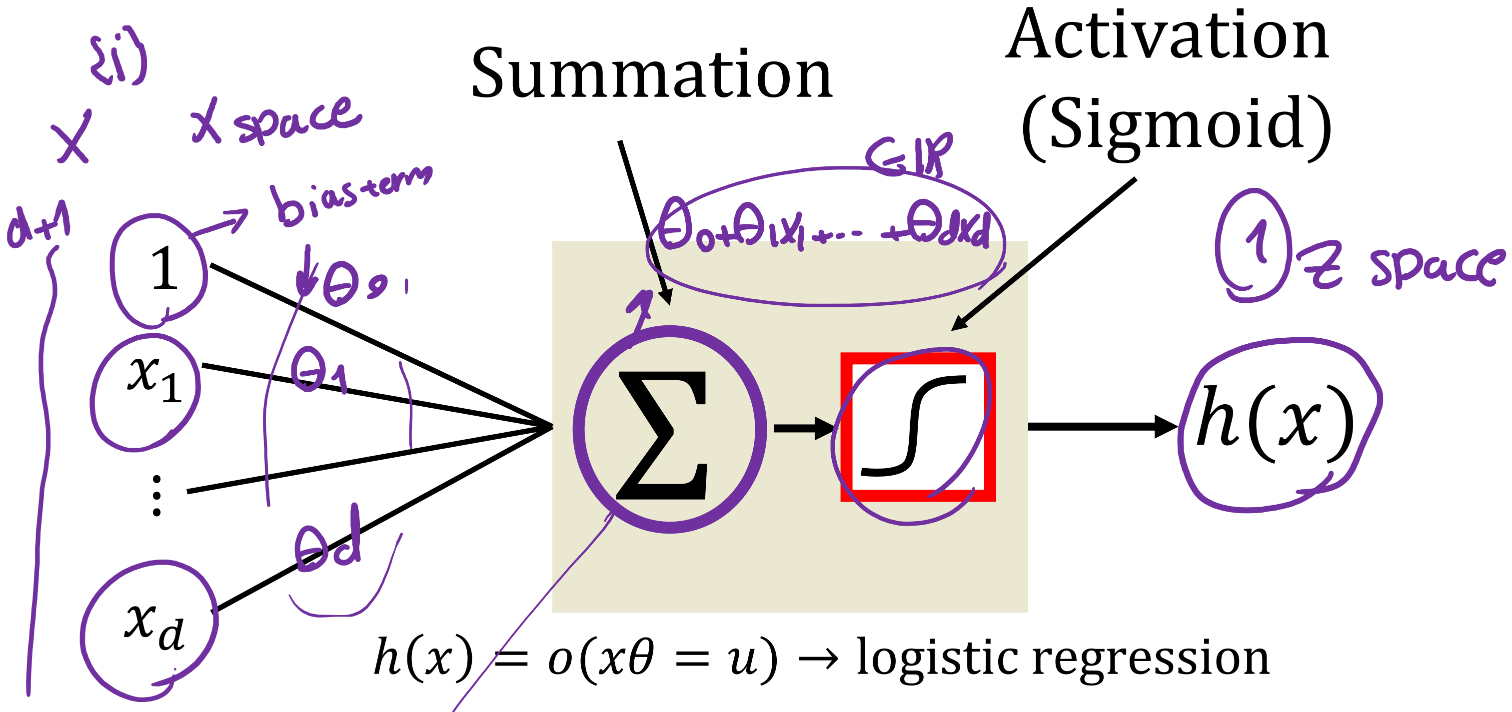
$X^{\{i\}}$



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

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

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

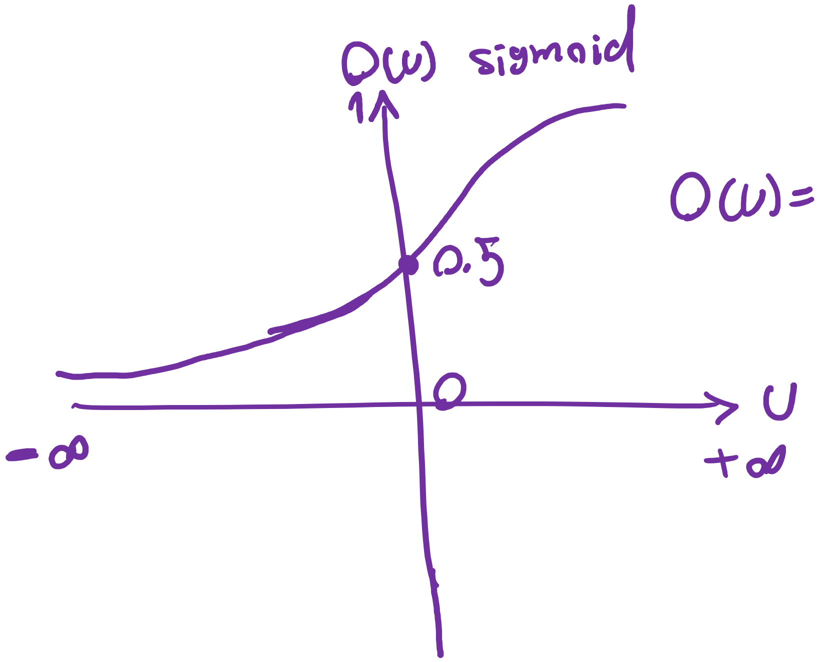


$h(x) = o(x\theta = u) \rightarrow$ logistic regression

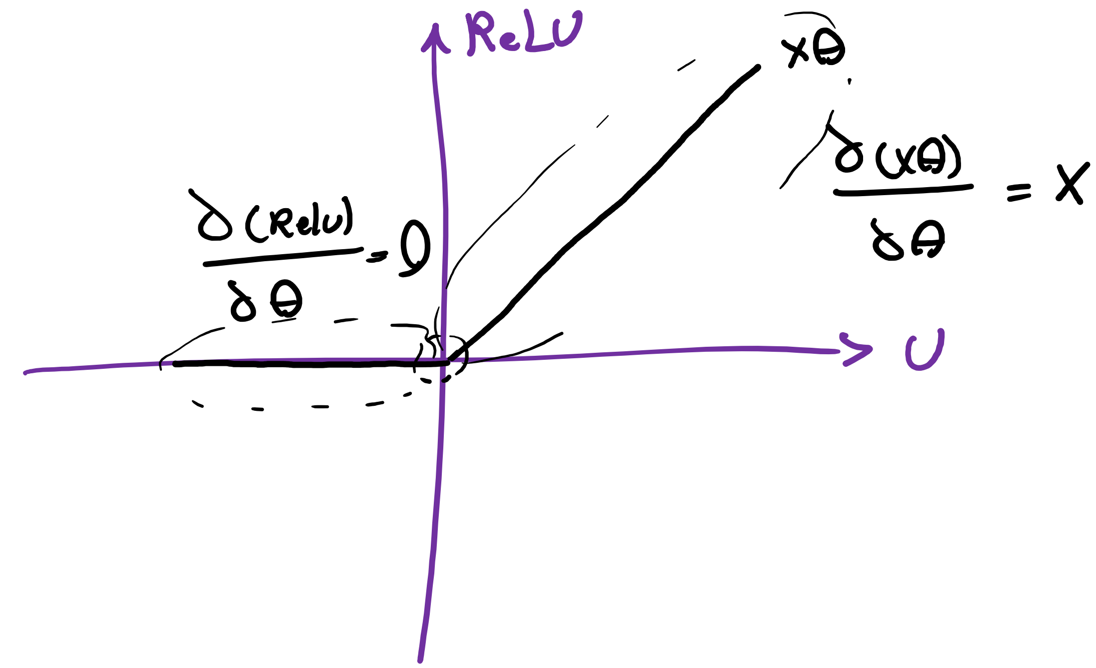
$$\sum = LCF = \theta_0 + \theta_1 x_1 + \dots + \theta_d x_d = S = U$$

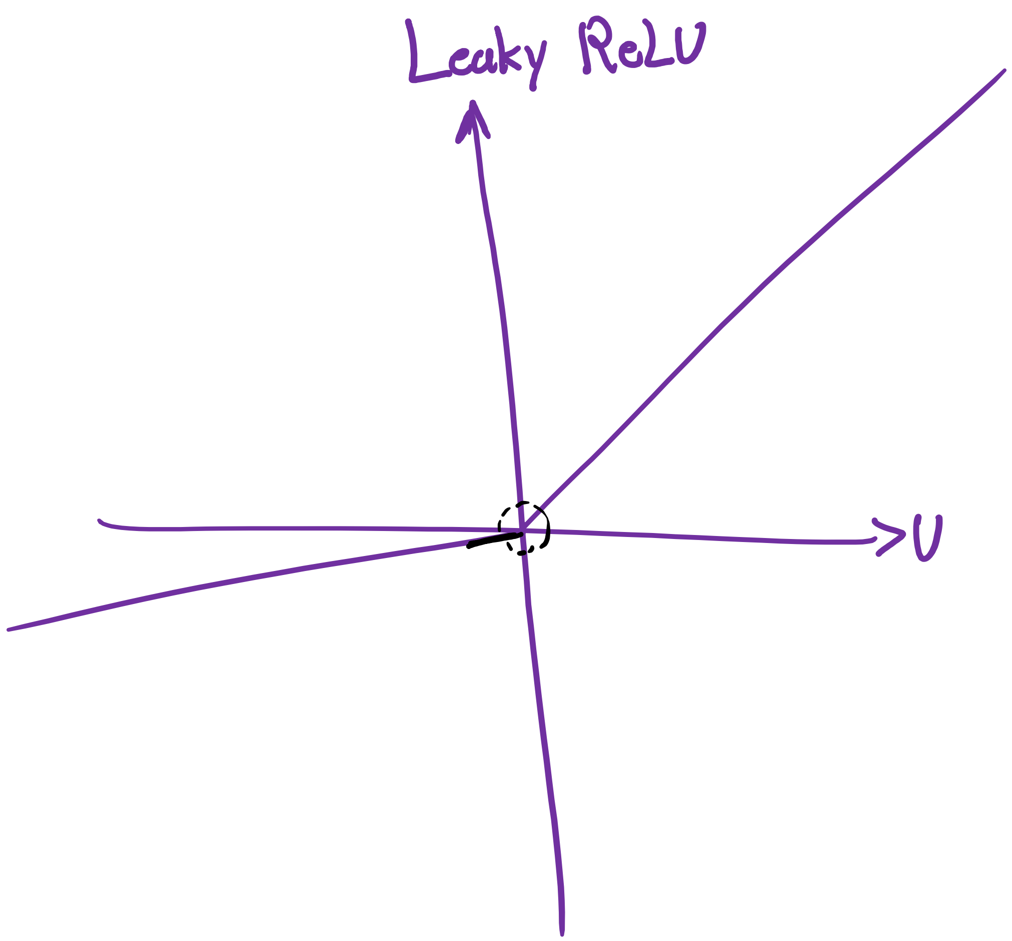
output

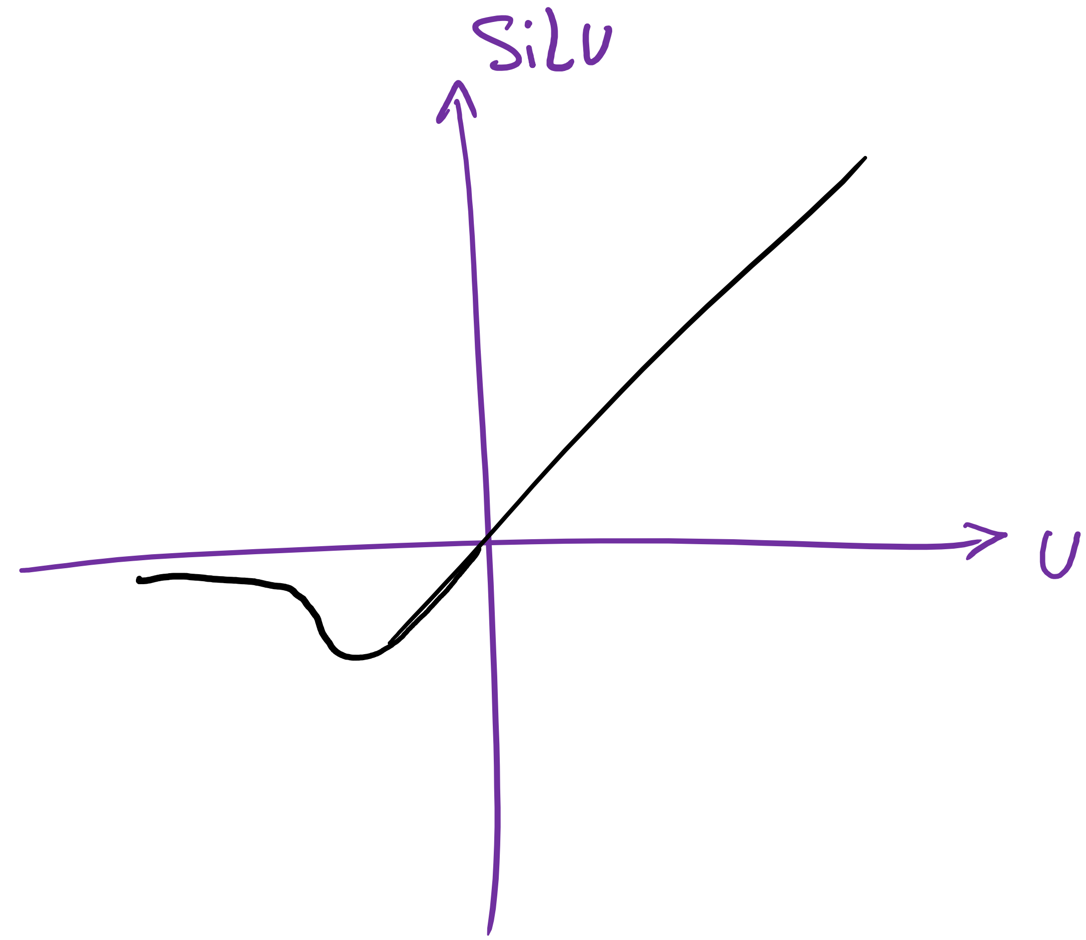
$$\underbrace{o(U)}_{\text{sigmoid}} = O = \frac{1}{1+\exp(-U)} = \frac{1}{1+\exp(-(\theta_0+\theta_1x_1+\dots+\theta_dx_d))}$$



$$\sigma(u) = \frac{1}{1+\exp(-u)} = \frac{1}{2}$$

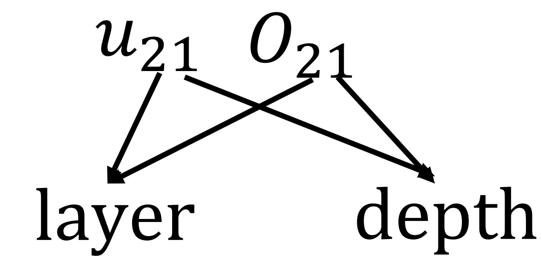


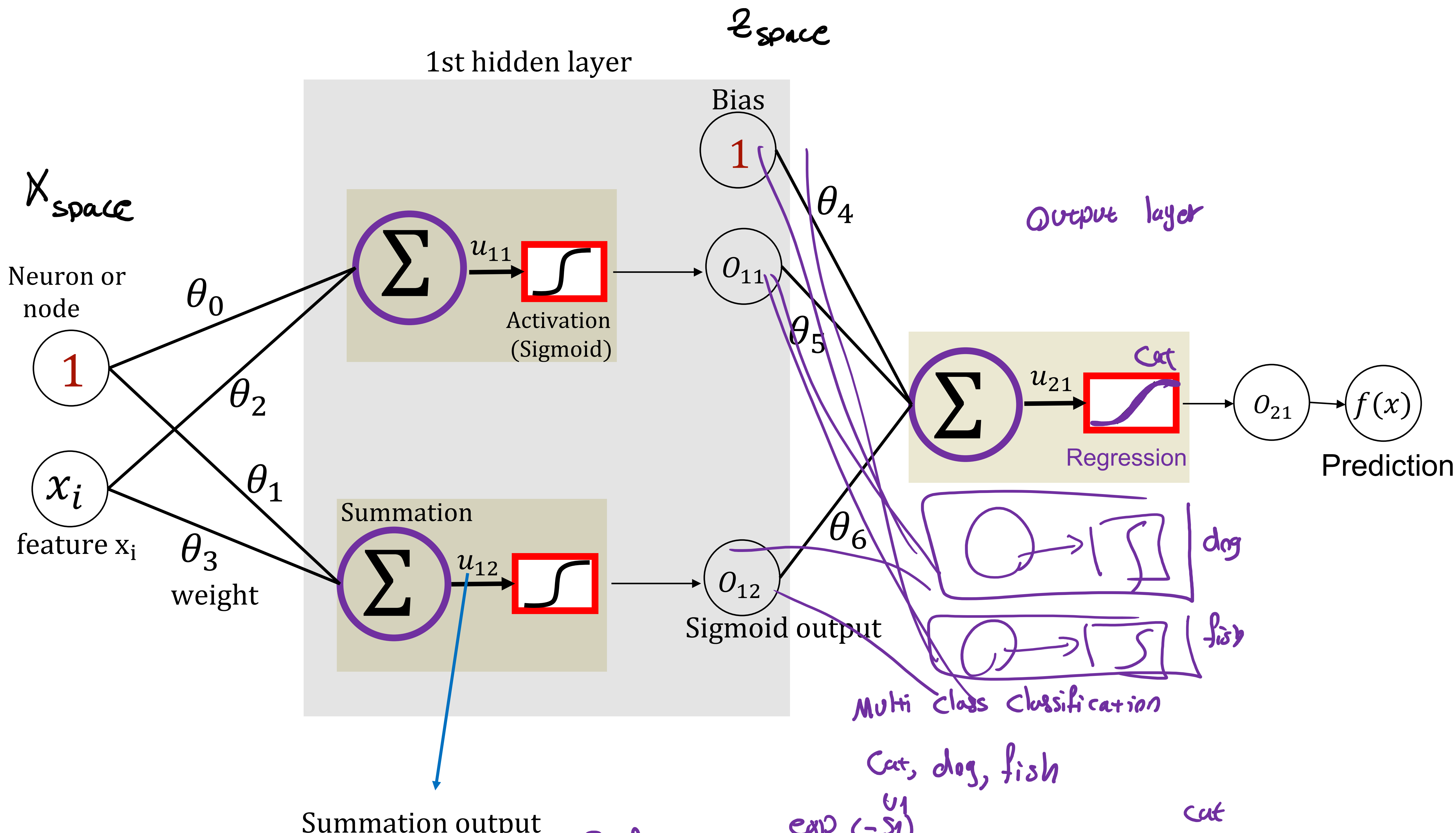




NN Regression

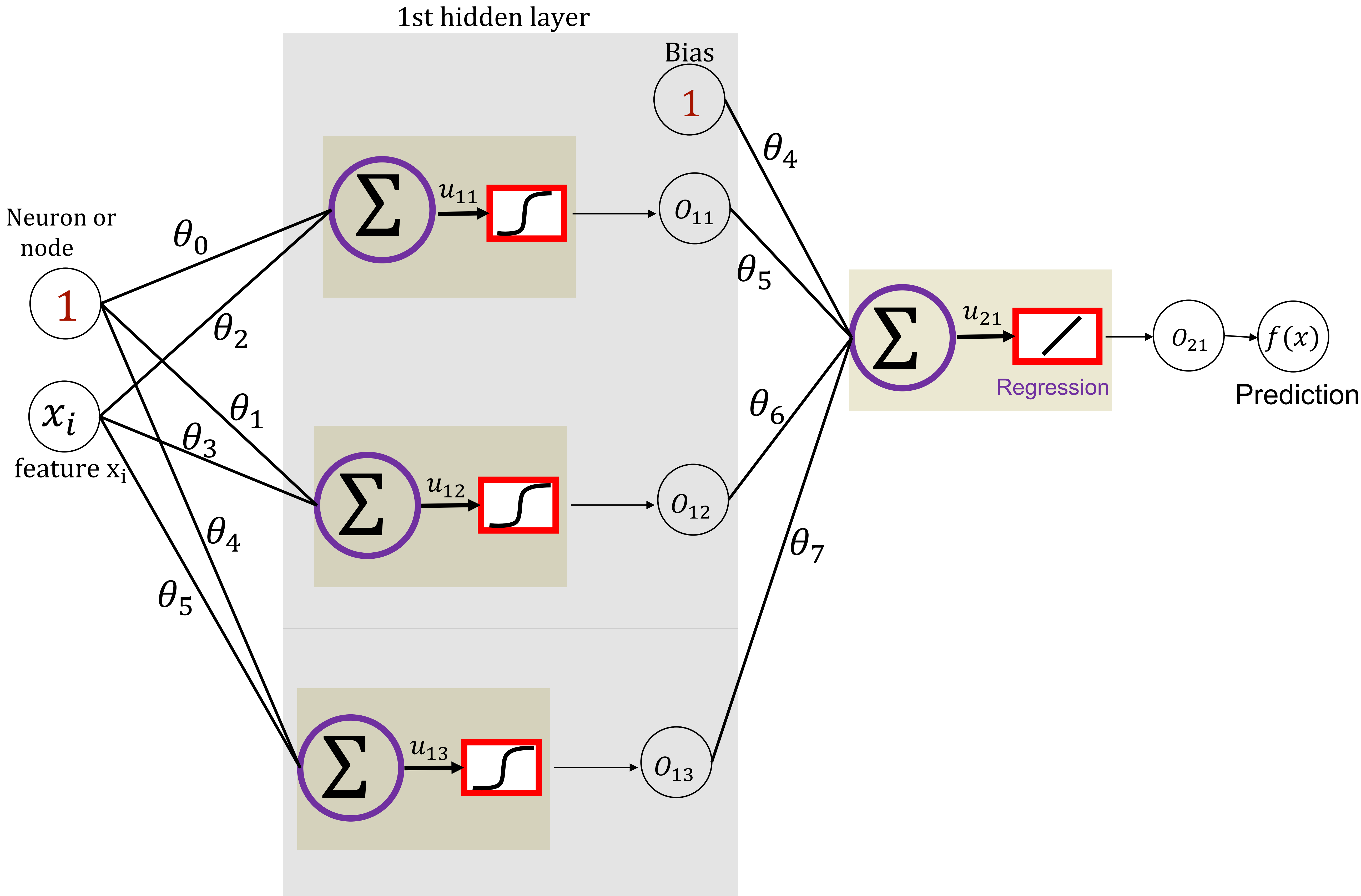
Fc layer = Fully connected layer



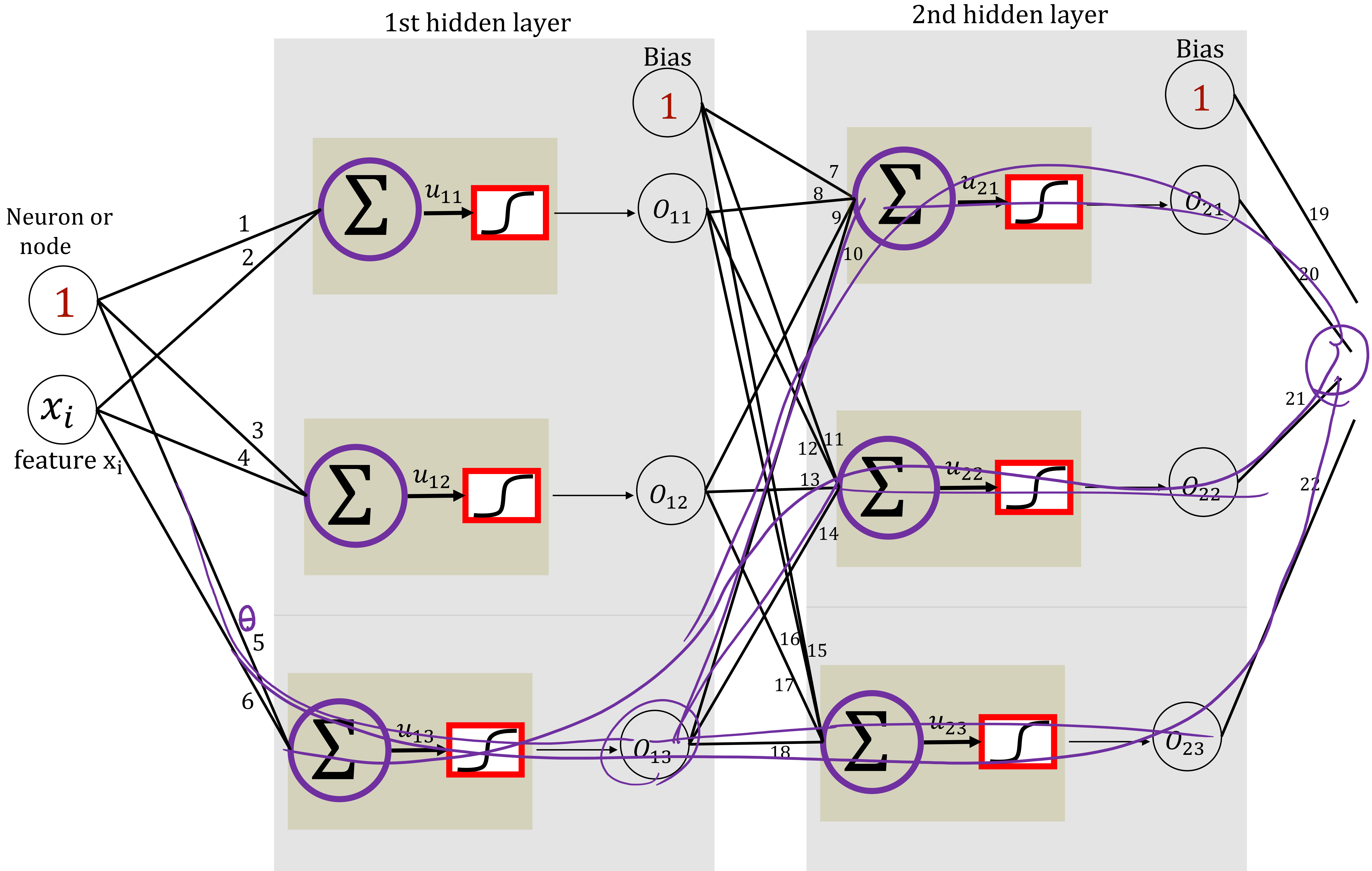


cat

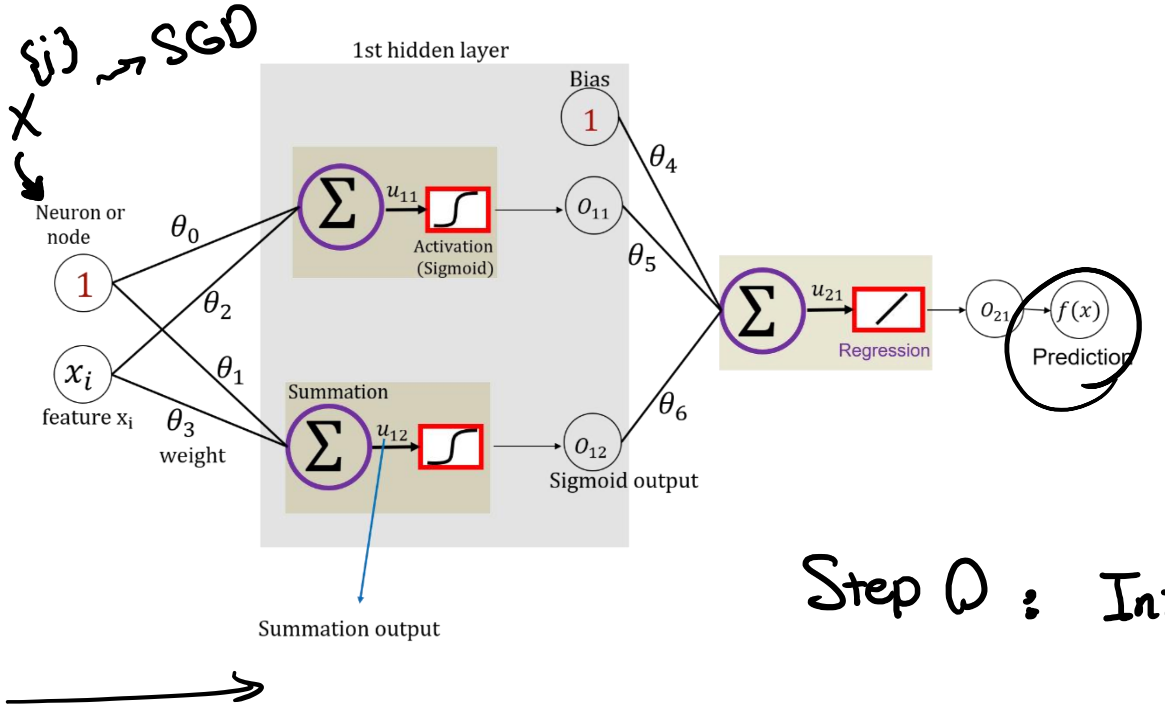
$$\text{Softmax} = \frac{\exp(-\overset{u_1}{s_1})}{\exp(-s_1) + \exp(-s_2) + \exp(-s_3)}$$

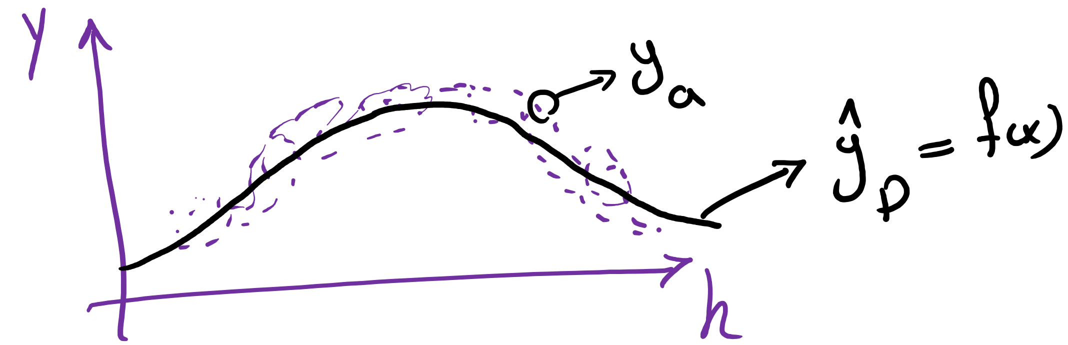


1st hidden layer
Bias
1
θ_4
u_{11}
o_{11}
Neuron or node
θ_0
1
θ_5
θ_2
u_{21}
o_{21}
$f(x)$
Regression
Prediction
x_i
θ_1
θ_6
θ_3
feature x_i
u_{12}
o_{12}
θ_4
θ_7
θ_5
u_{13}
o_{13}



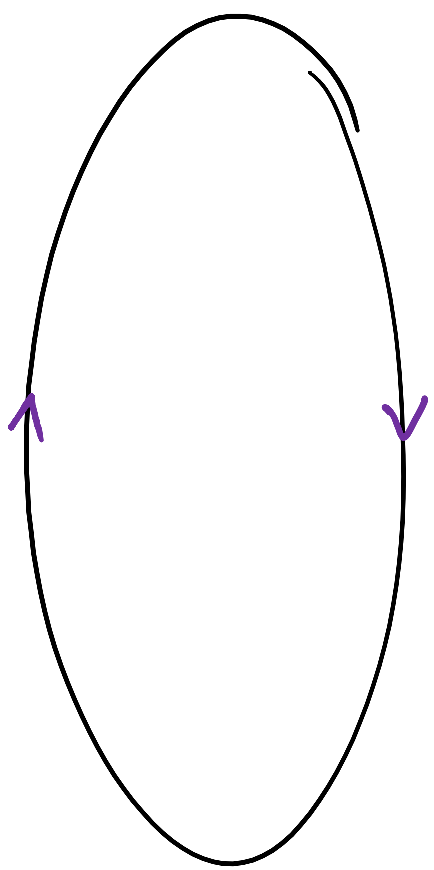
1st hidden layer
2nd hidden layer
Bias
1
Bias
1
Neuron or node
1
x_i
feature x_i
u_{11}
u_{12}
u_{13}
o_{11}
o_{12}
o_{13}
u_{21}
u_{22}
u_{23}
o_{21}
o_{22}
o_{23}
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
θ





Step 0 : Initialize all the parameters Θ_s randomly

Note : Do NOT initialize them with all zeros



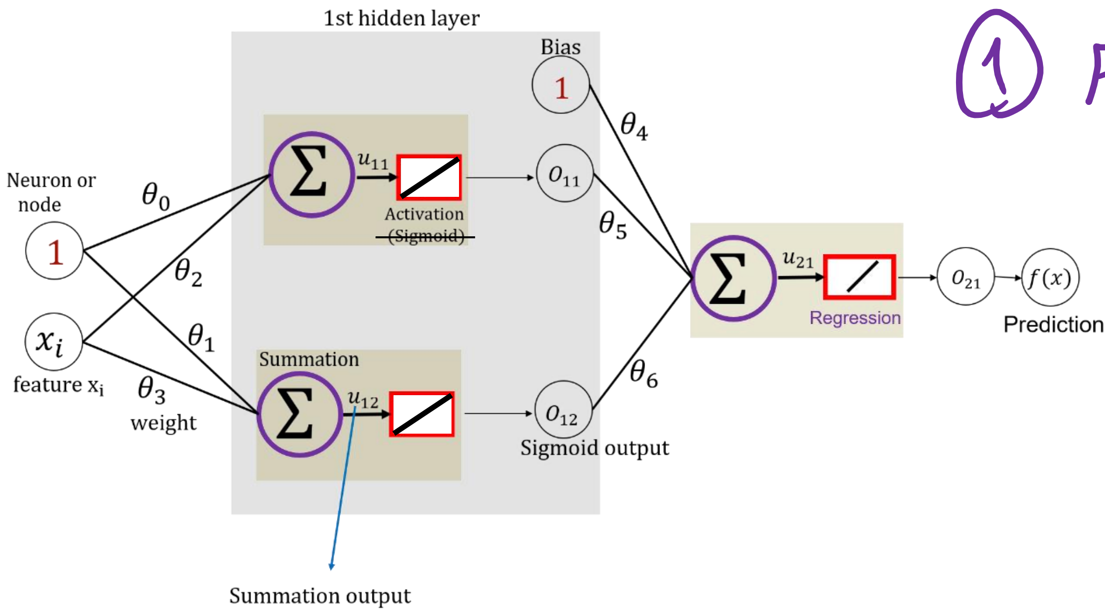
Step 1 : Calculate all the U_s & O_s using our current Θ_s

left ⟶ right Forward Pass

Step 2 : Optimize all Θ_s based on step 1 (O_s & U_s)

left ⟵ right Back Propagation

Step 3 : check for convergence (calculate your objective function)



1 Forward Pass (U_s & O_s)

$$U_{11} = \Theta_0 * 1 + \Theta_2 * x_i \Rightarrow O_{11} = U_{11} = \Theta_0 + \Theta_2 x_i$$

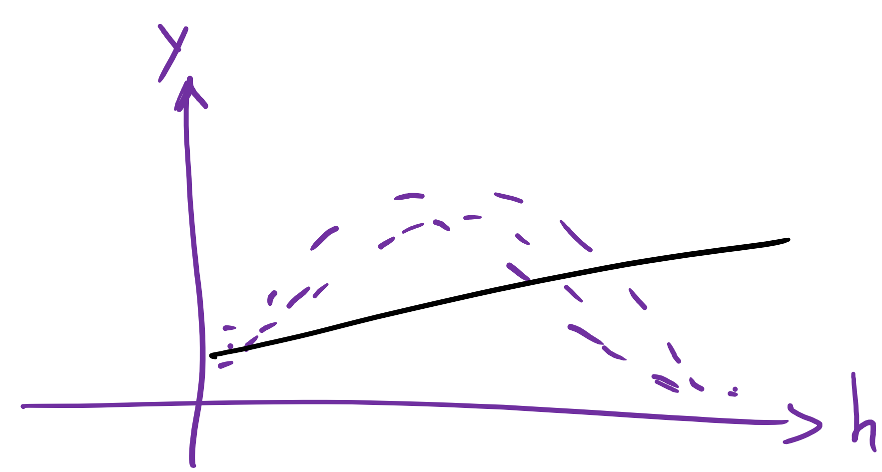
$$U_{12} = \Theta_1 + \Theta_3 x_i \Rightarrow O_{12} = U_{12} = \Theta_1 + \Theta_3 x_i$$

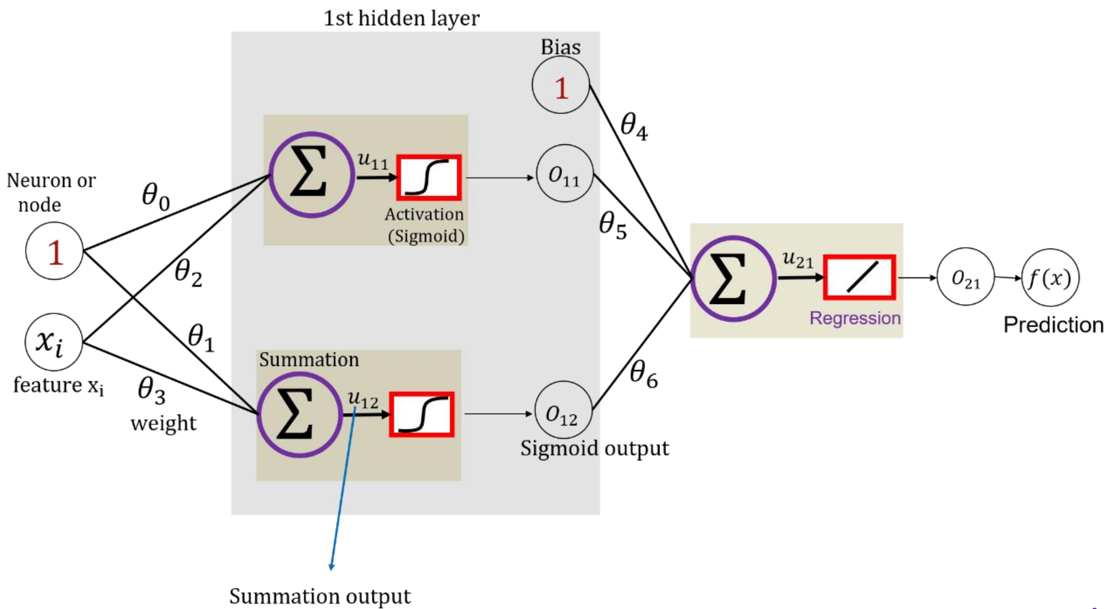
$$U_{21} = \Theta_4 + \Theta_5 O_{11} + \Theta_6 O_{12} \Rightarrow O_{21} = U_{21} = f(x) \Rightarrow f(x) = O_{21}$$

$$f(x) = \Theta_4 + \Theta_5 (\Theta_0 + \Theta_2 x_i) + \Theta_6 (\Theta_1 + \Theta_3 x_i)$$

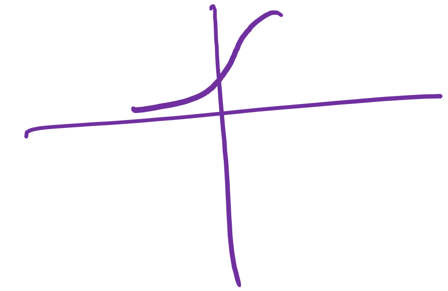
$$f(x) = \Theta_4 + \Theta_5 \Theta_0 + \Theta_6 \Theta_1 + (\Theta_5 \Theta_2 + \Theta_6 \Theta_3) x_i$$

$$f(x) = \Theta_0 + \Theta_1 x_i$$

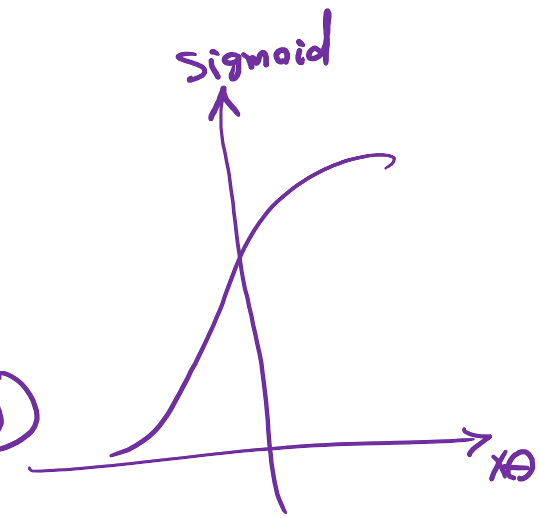




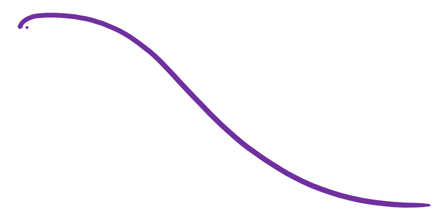
$$O(U) = \frac{1}{1+\exp(-U)} * C$$



$$U_{11} = \theta_0 + \theta_2 x_i \Rightarrow O_{11}(U_{11}) = \frac{1}{1+\exp(-U_{11})} = \frac{1}{1+\exp(-[\theta_0 + \theta_2 x_i])}$$



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



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

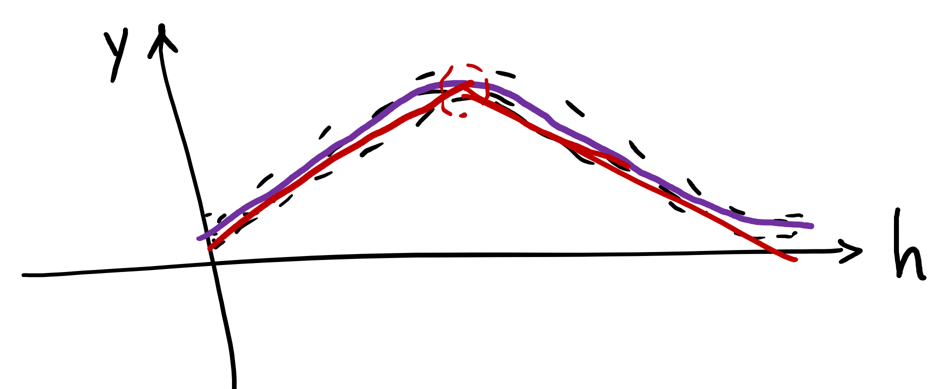
$$f(x) = \theta_4 + \theta_5 \frac{1}{1+\exp(-[\theta_0 + \theta_2 x_i])} + \theta_6 \frac{1}{1+\exp(-[\theta_1 + \theta_3 x_i])}$$

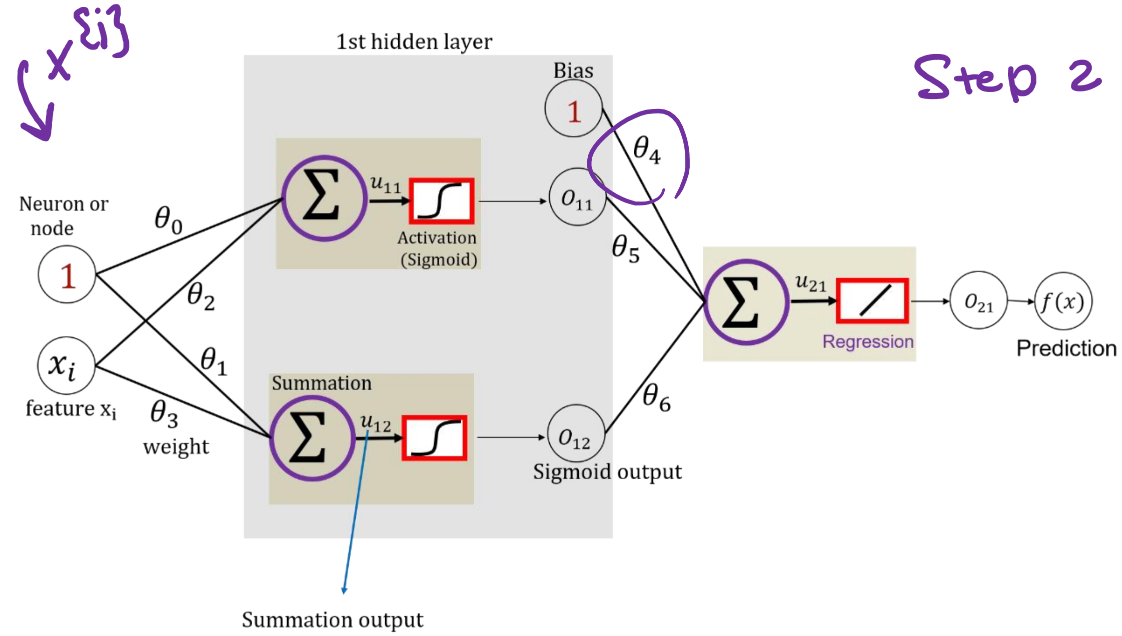
translation in vertical direction

Scaling in vertical

horizonal shifting

horizontal Scaling





Step 2 → Back Propagation $\underset{}{\text{Min}}\ E(\Theta) = \frac{1}{N}\sum_{i=1}^{N}\left(y_a^{\{i\}} - \underline{f(x)^{\{i\}}}\right)^2$

$$E(\Theta) = \frac{1}{2}\left(y_a^{\{i\}} - f(x)^{\{i\}}\right)^2 \qquad \frac{\partial (U^2)}{\partial U} = 2U\,U'$$

$$\nabla_\Theta E(\Theta) = 0 \qquad \underbrace{\left(f(x)^{\{i\}} - y_a^{\{i\}}\right)}\frac{\partial f(x)}{\partial \Theta} = \nabla_\Theta E(\Theta)$$

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

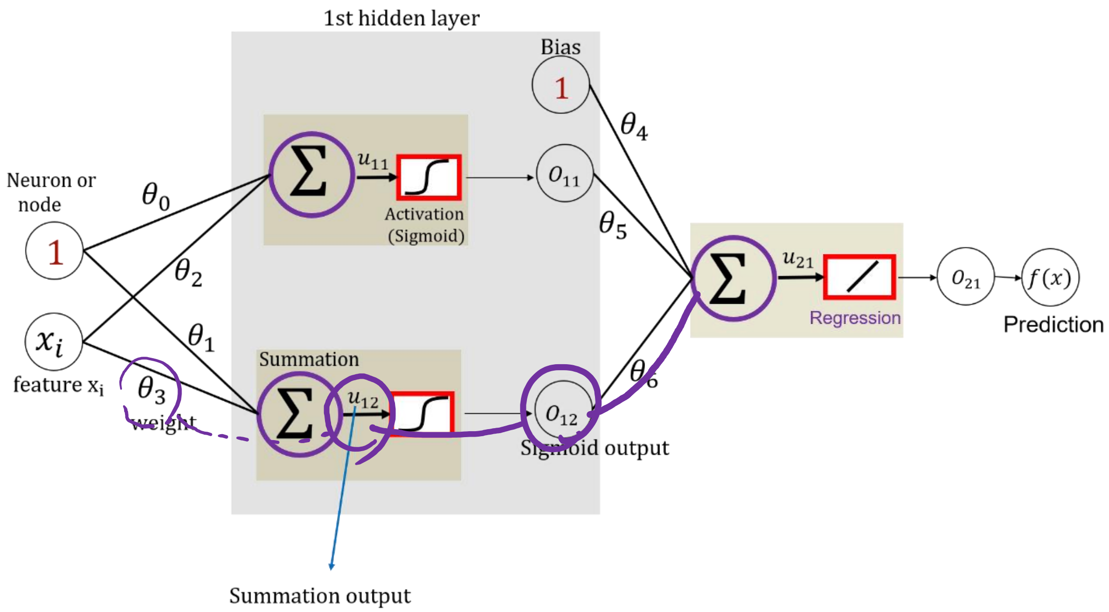
$$\Theta_6, \Theta_5, \Theta_4, \Theta_3, \Theta_2, \Theta_1, \Theta_0$$

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

$$f(x) = O_{21} = U_{21} = \Theta_4 + \Theta_5 O_{11} + \Theta_6 O_{12} \Rightarrow \frac{\partial f(x)}{\partial \Theta_4} = 1$$

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

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



$$u_{12} = \theta_1 + \theta_3 x_i \qquad o_{12} = \frac{1}{1+\exp(-u_{12})}$$

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

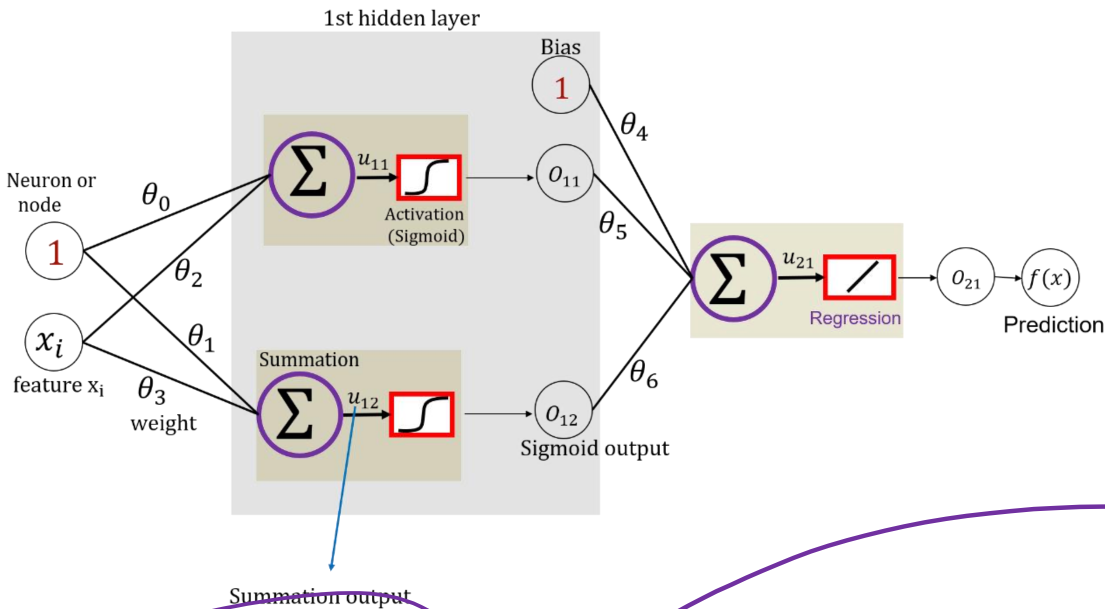
$$\theta_3^{\{t+1\}} \leftarrow \theta_3^{\{t\}} - \alpha \Delta \frac{\partial f(x)}{\partial \theta_3}$$

$$\frac{\partial f(x)}{\partial \theta_3} = \frac{\partial f(x)}{\partial o_{12}} \cdot \frac{\partial o_{12}}{\partial u_{12}} \cdot \frac{\partial u_{12}}{\partial \theta_3} = \theta_6 \; o_{12}[1-o_{12}] \quad x_i$$

$$o = \frac{1}{1+\exp(-u)} = (1+\exp(-u))^{-1} \qquad \frac{\partial o}{\partial u} = -1 * -1 * \exp(-u)\,(1+\exp(-u))^{-2} = \frac{\exp(-u)}{[1+\exp(-u)]^2}$$

$$\frac{\partial o}{\partial u} = \frac{1+\exp(-u)-1}{[1+\exp(-u)]^2} = \frac{1}{1+\exp(-u)}\left[\frac{1+\exp(-u)}{1+\exp(-u)} - \frac{1}{1+\exp(-u)}\right]$$

$$\frac{\partial o}{\partial u} = o\,[1-o]$$



0

$0 < \quad < 1$

$$\theta_3^{\{t+1\}} \leftarrow \theta_3^{\{t\}} - \cancel{\alpha \Delta \theta_6 O_{12} \underbrace{[1 - O_{12}]}_{0 < \quad < 1} x_i}$$

Vanishing gradient

Exploding gradient