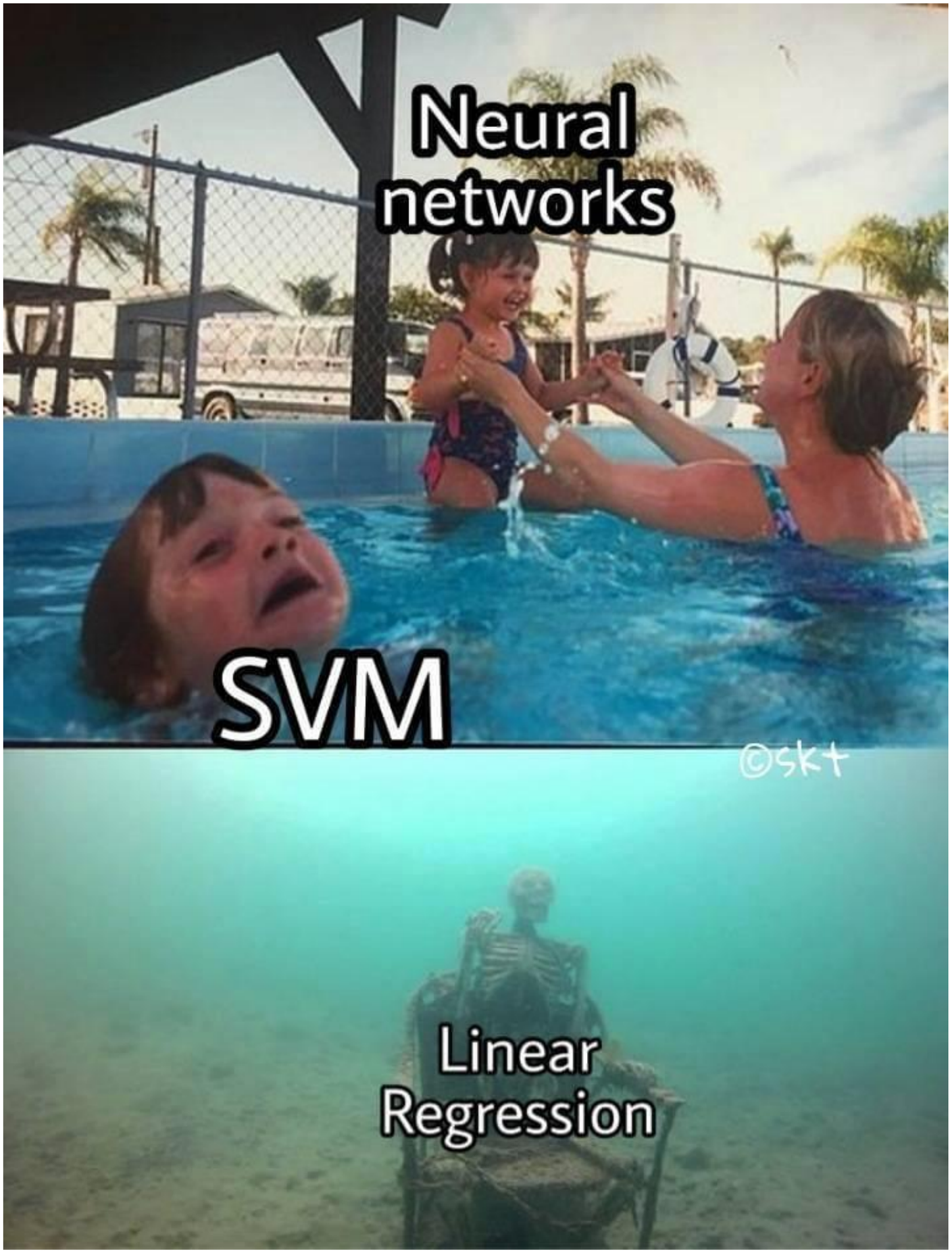
Georgia Tech

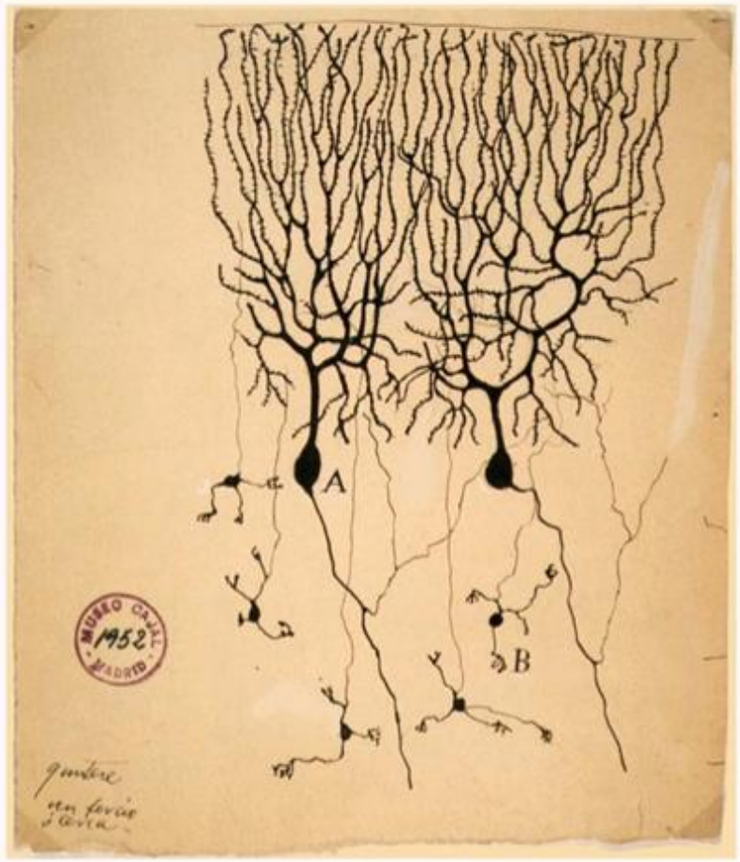
Neural Networks
Forward Pass and Back Propagation

Mahdi Roozbahani

Georgia Tech


Neural networks
SVM
©skt
Linear Regression

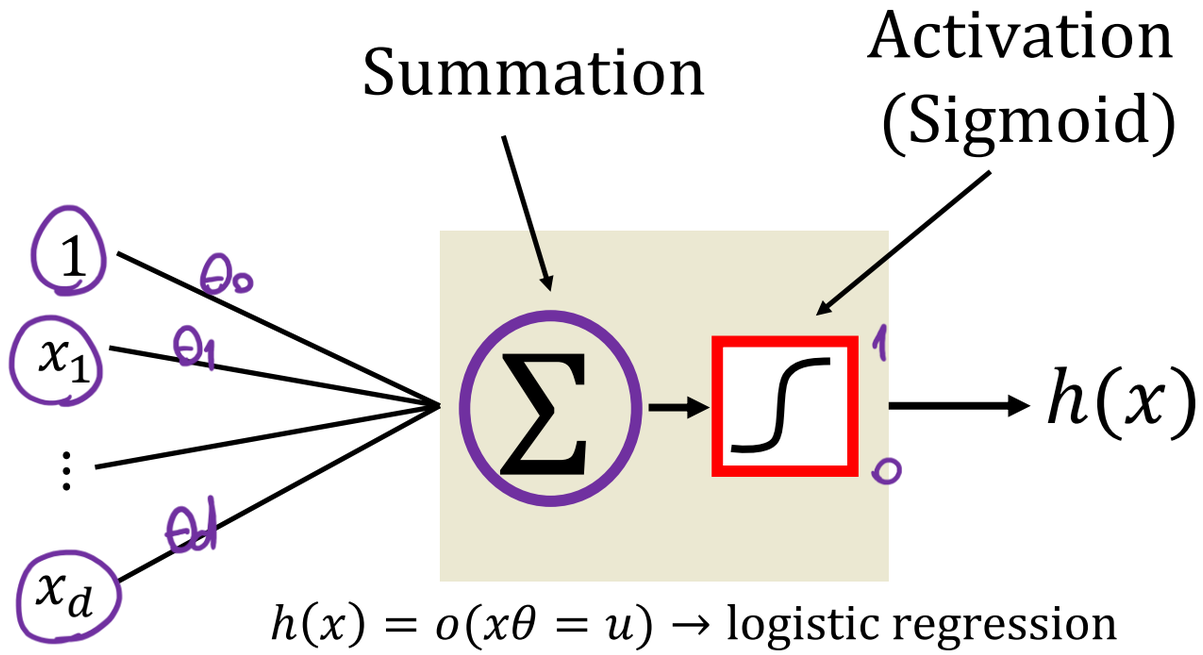
Inspiration from Biological Neurons



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

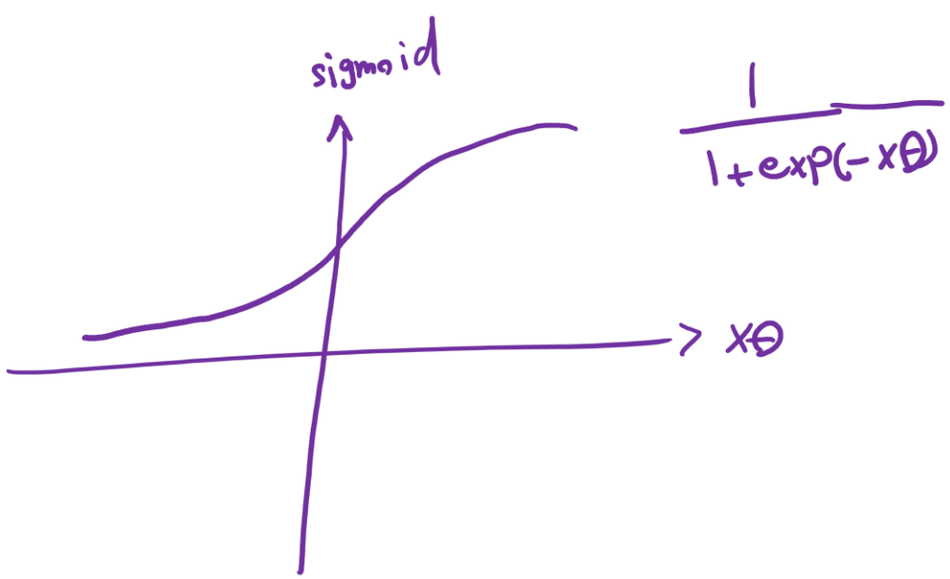
1. Dendrites that collect information from other neurons
2. An axon that generates outgoing spikes

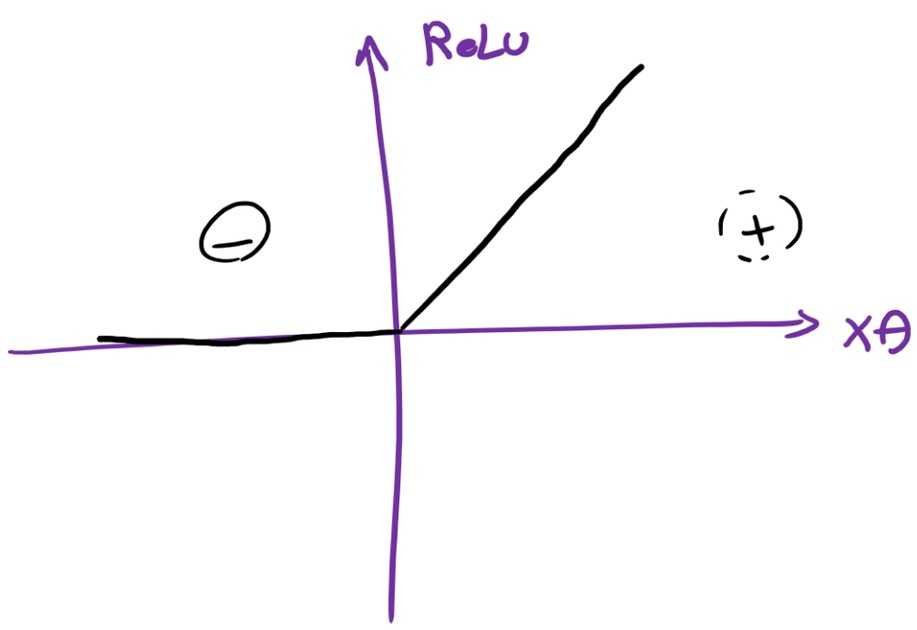


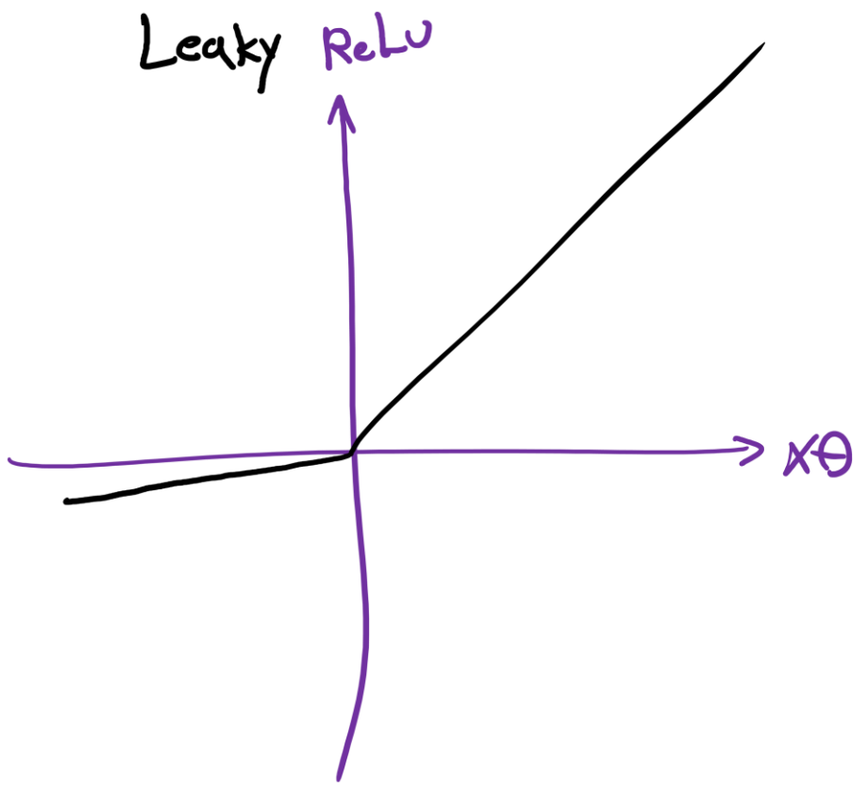
$h(x) = o(x\theta = u) \rightarrow$ logistic regression

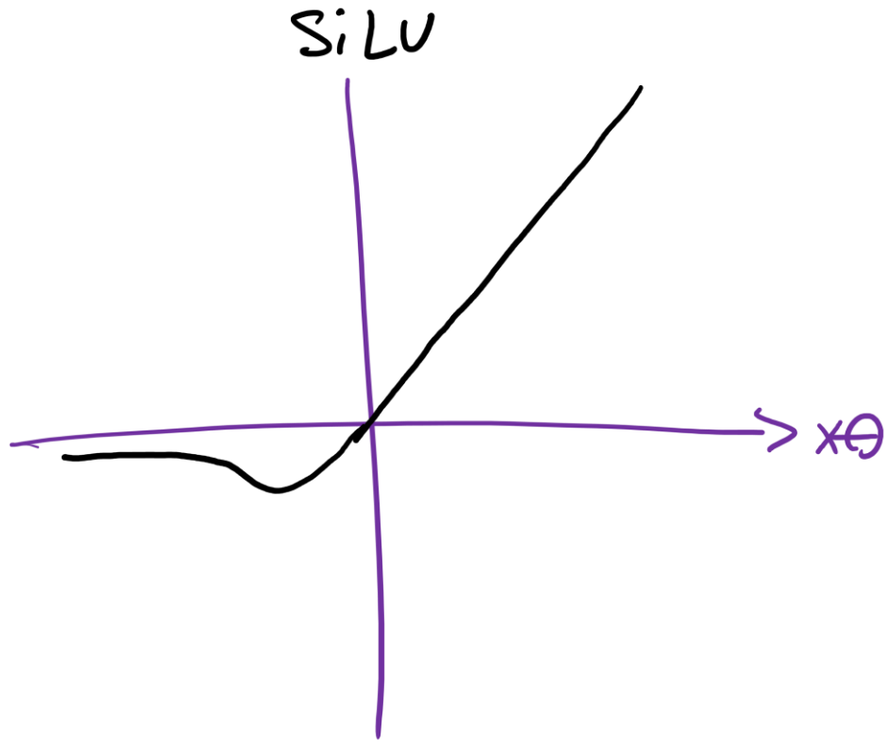
$$output = activation(x\theta \ + b)$$

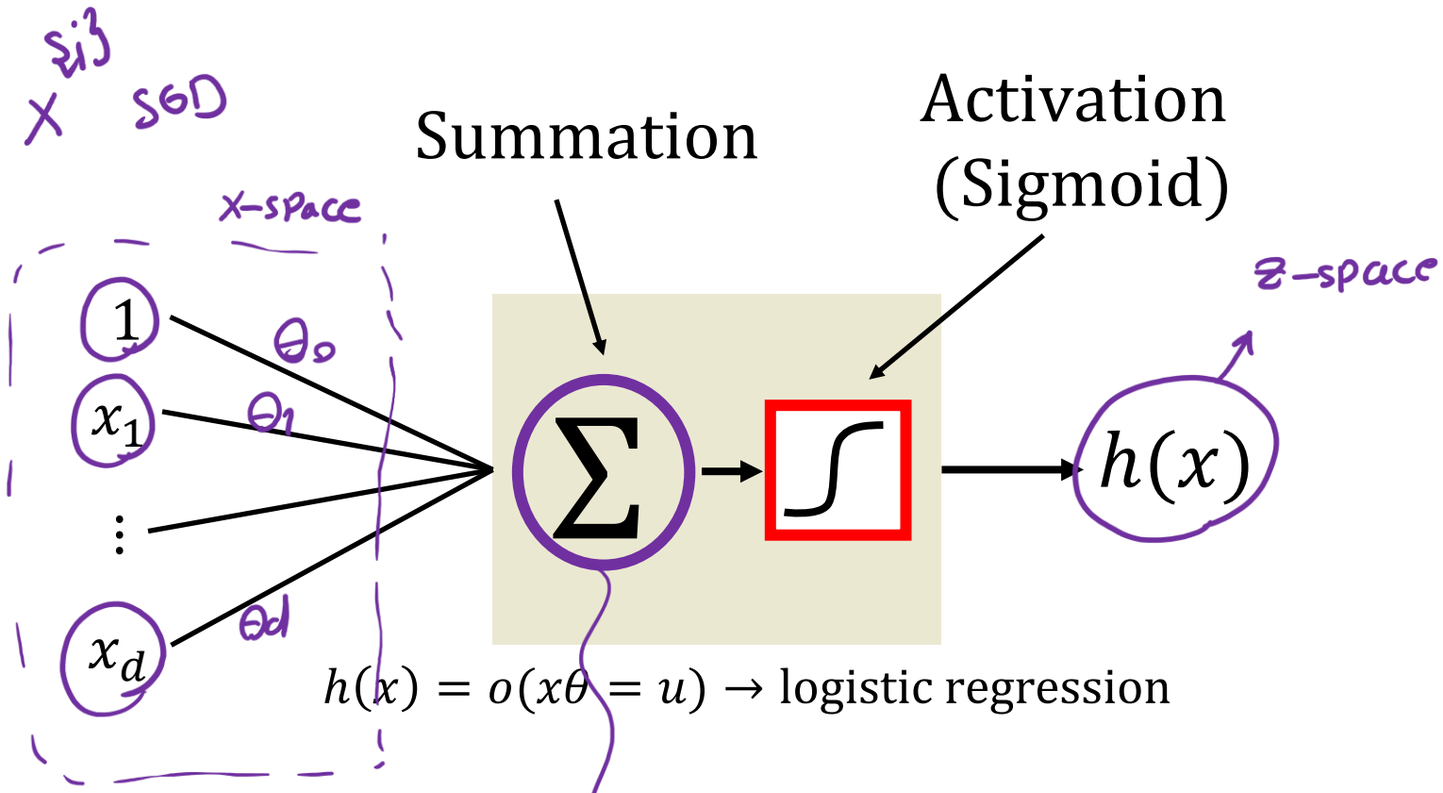
Name of the neuron	Activation function: $activation(z)$
Linear unit	z
Threshold/sign unit	$\text{sgn}(z)$
Sigmoid unit	$\dfrac{1}{1+\exp(-z)}$
Rectified linear unit (ReLU)	$\max(0, z)$
Tanh unit	$\tanh(z)$


sigmoid
1
1+exp(-xθ)
xθ


ReLU
xθ


Leaky ReLU
xθ


SiLU
xθ



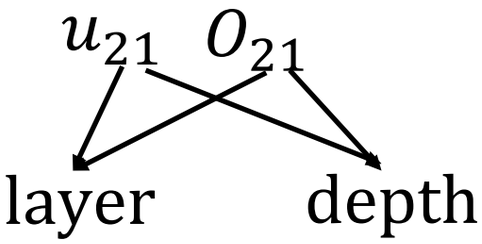
$h(x) = o(x\theta = u) \rightarrow$ logistic regression

$z = LCF = \theta_0 + \cdots + \theta_d x_d = x\theta \in \mathbb{R} \rightarrow$ ∫ $\rightarrow$ ○

Fully connected operaction

FC layer

NN Regression



Layer 0

Input neuron

X-space

1st hidden layer

Z-space

Output layer

Neuron or node

Bias

Activation (Sigmoid)

Summation

Regression

Prediction

feature x_i

weight

Sigmoid output

Summation output

θ_0 θ_1 θ_2 θ_3 θ_4 θ_5 θ_6

u_{11} u_{12} u_{21}

O_{11} O_{12} O_{21} $f(x)$

1st hidden layer

Bias

Neuron or node

feature x_i

θ_0 θ_1 θ_2 θ_3 θ_4 θ_5 θ_6 θ_7

u_{11} u_{12} u_{13} u_{21}

o_{11} o_{12} o_{13} o_{21}

$f(x)$

Regression

Prediction

Cat

dog

fish

1st hidden layer

2nd hidden layer

Bias

Bias

Neuron or node

1

x_i

feature x_i

$\sum$ u_{11}

$\sum$ u_{12}

$\sum$ u_{13}

$\sum$ u_{21}

$\sum$ u_{22}

$\sum$ u_{23}

O_{11}

O_{12}

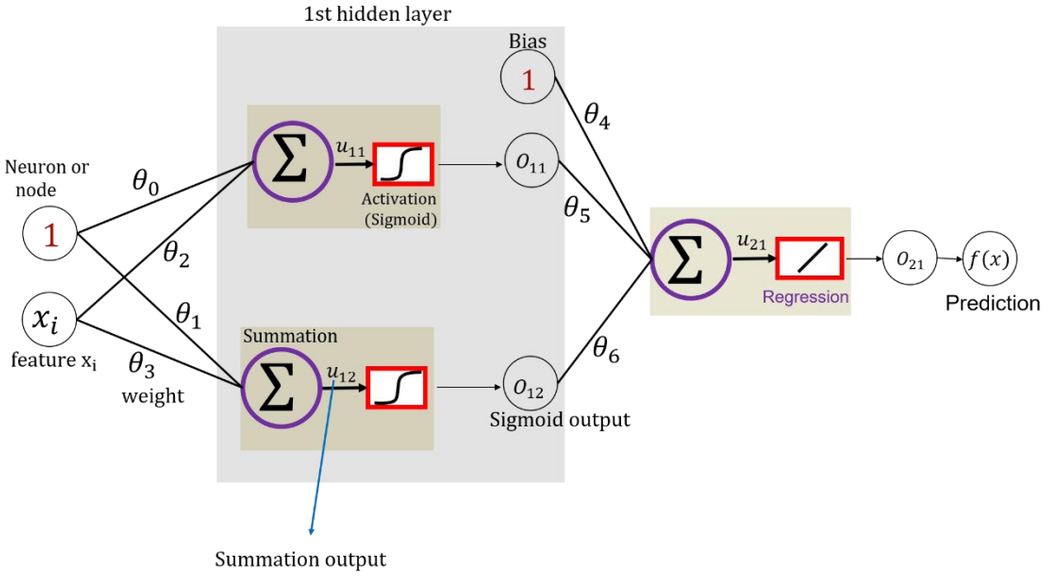
O_{13}

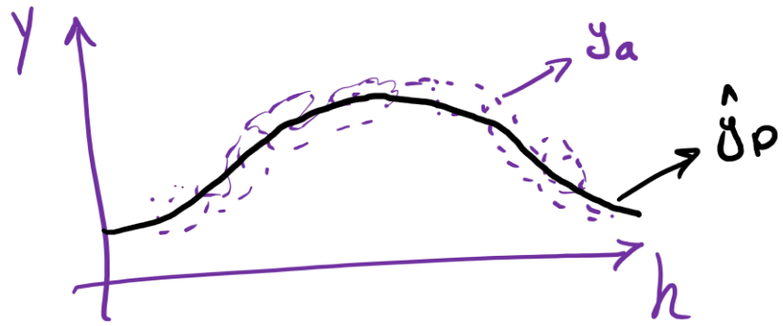
O_{21}

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





Step 0: Randomly initialize Parameters (θ_s)

⇒ do not use zeros for random initialization

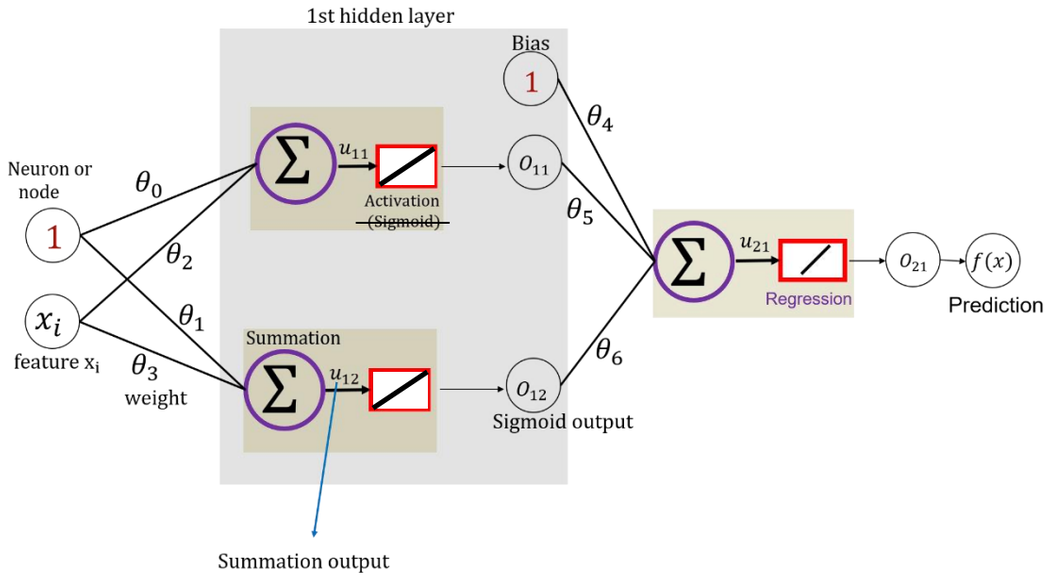
Step 1: Forward Pass:

Calculate all the U_s & O_s

Step 2: Back Propagation

optimize Parameters (θ_s)

Step 3: Check for convergence



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

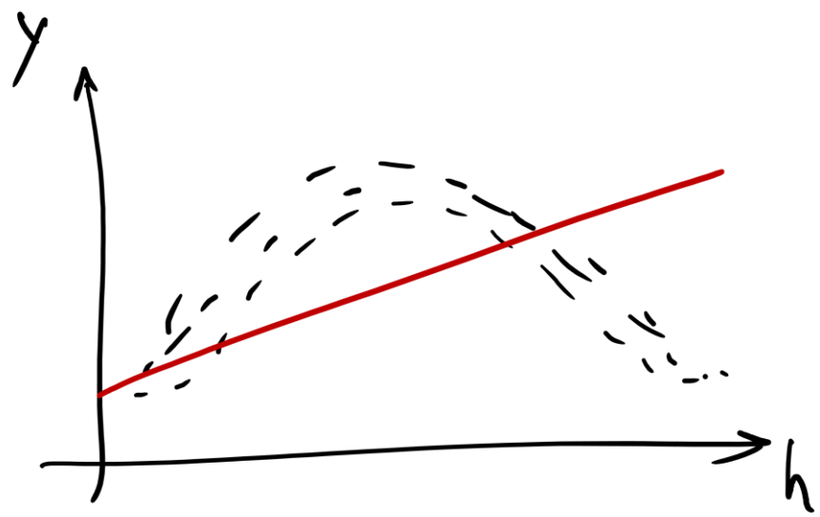
$u_{12} = \theta_1 + \theta_3 x_i \qquad o_{12} = u_{12}$

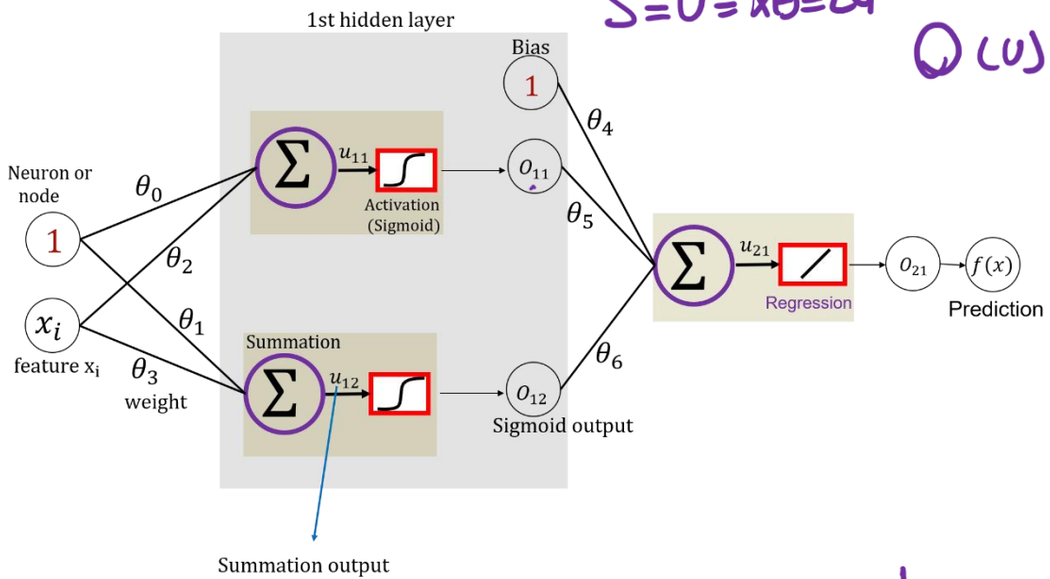
$u_{21} = \theta_4 + \theta_5 o_{11} + \theta_6 o_{12} \qquad o_{21} = u_{21} = f(x) = \hat{y}_p$

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

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

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

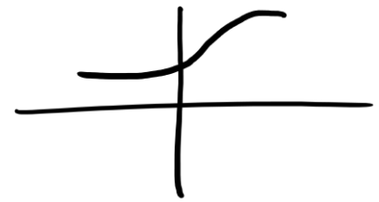




$S = U = X\Theta = LF$

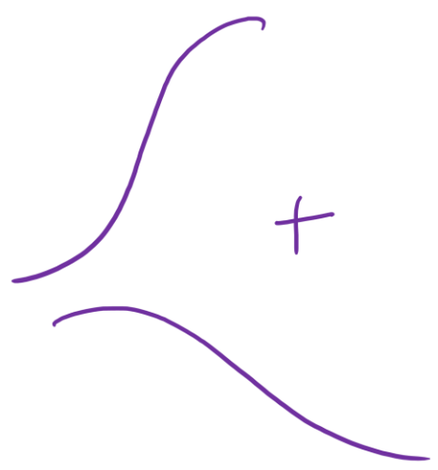
$$O(U) = \left(\frac{1}{1+\exp(-U)}\right)$$

$$\frac{1}{1+\exp(-u)} * C$$



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

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



+

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

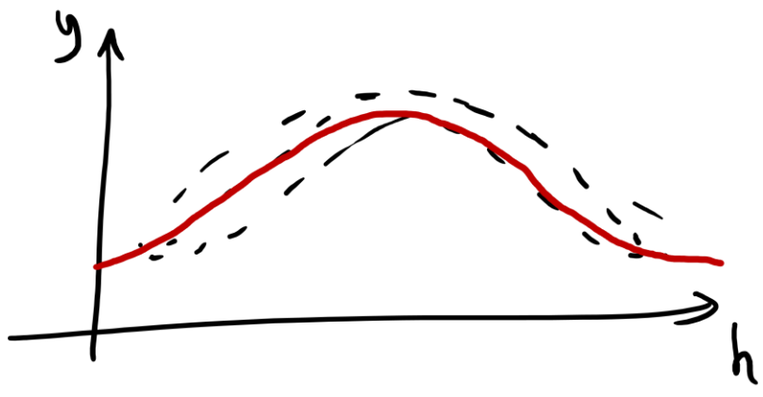
$$f(x) = \theta_4 + \theta_5 \frac{1}{1+\exp(-[\theta_0+\theta_2 x_i])} + \theta_6 \frac{1}{1+\exp(-[\theta_1+\theta_3 x_i])}$$

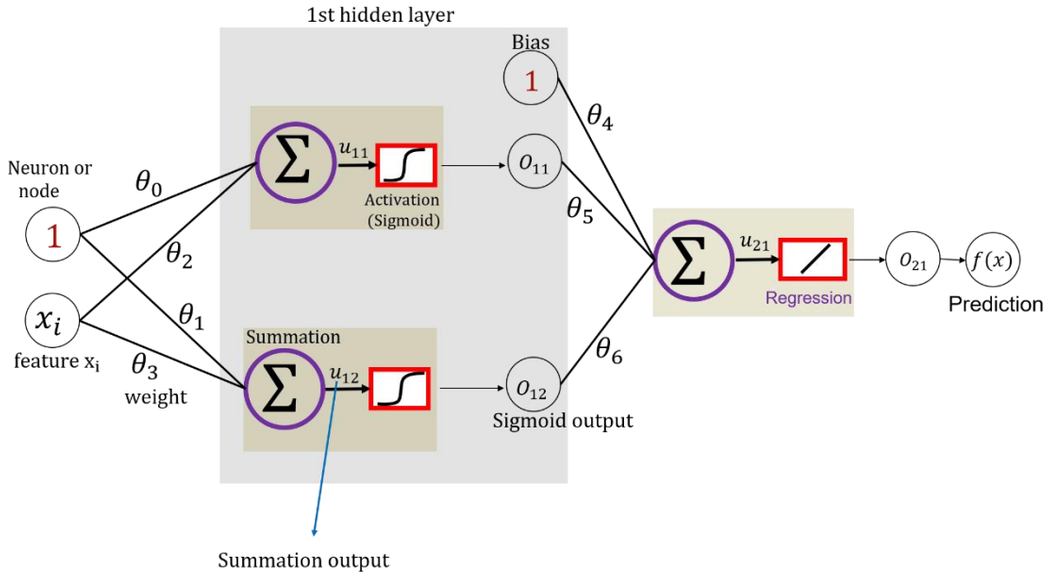
Vertical translation

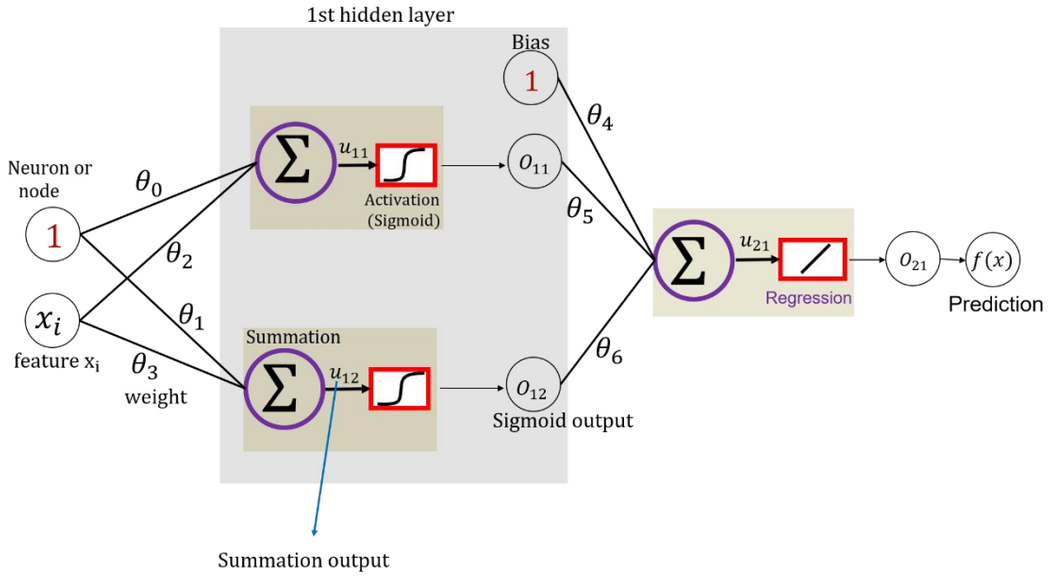
Stretching / Squashing in vertical

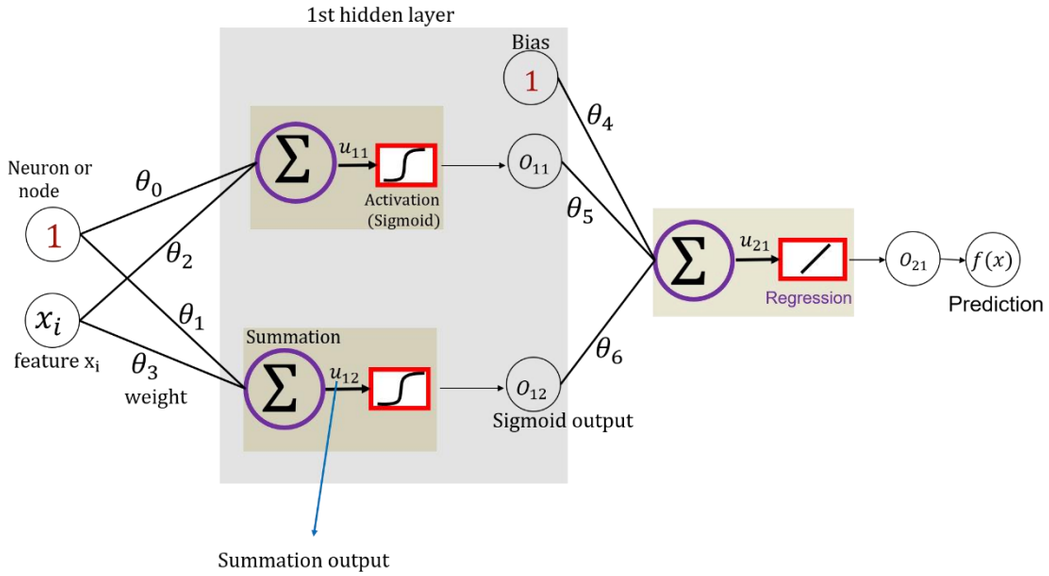
Translation in horizontal direction

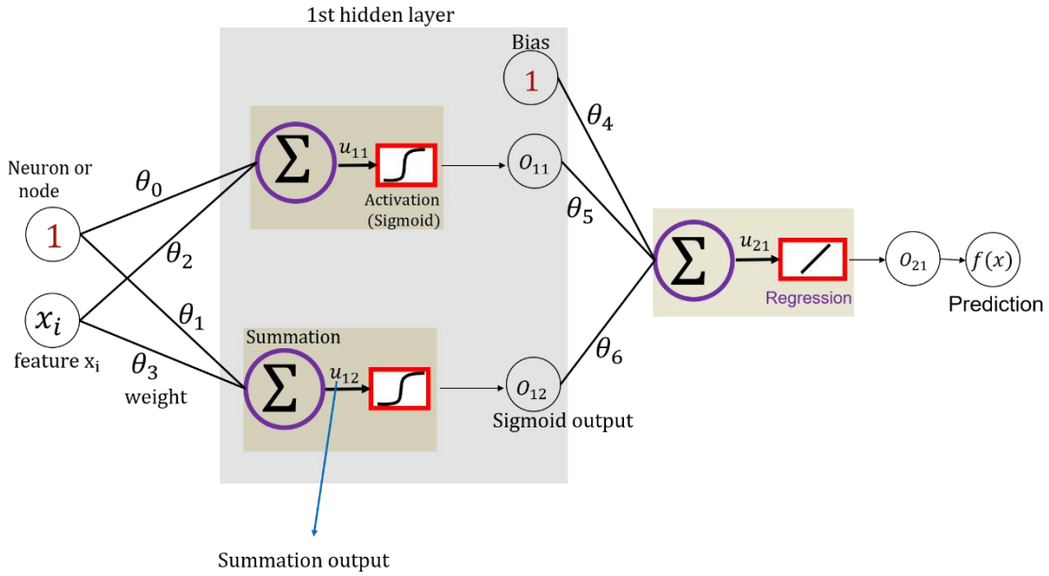
Stretch or squash in horizontal

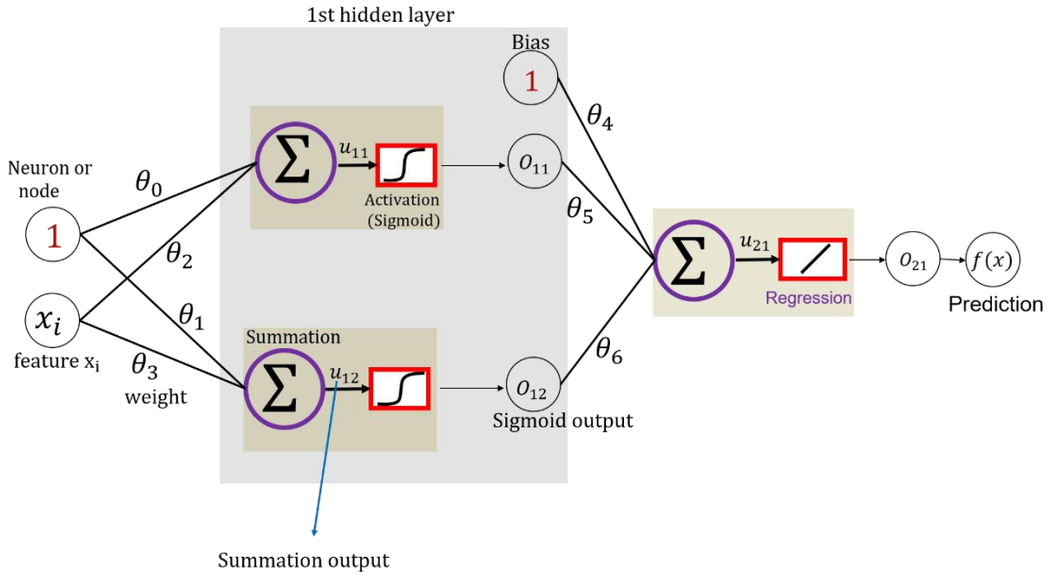



1st hidden layer
Bias
1
Neuron or node
1
x_i
feature x_i
θ_0
θ_1
θ_2
θ_3
weight
u_{11}
Activation (Sigmoid)
o_{11}
Summation
u_{12}
o_{12}
Sigmoid output
Summation output
θ_4
θ_5
θ_6
u_{21}
Regression
o_{21}
$f(x)$
Prediction


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


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


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


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