Session 1&2

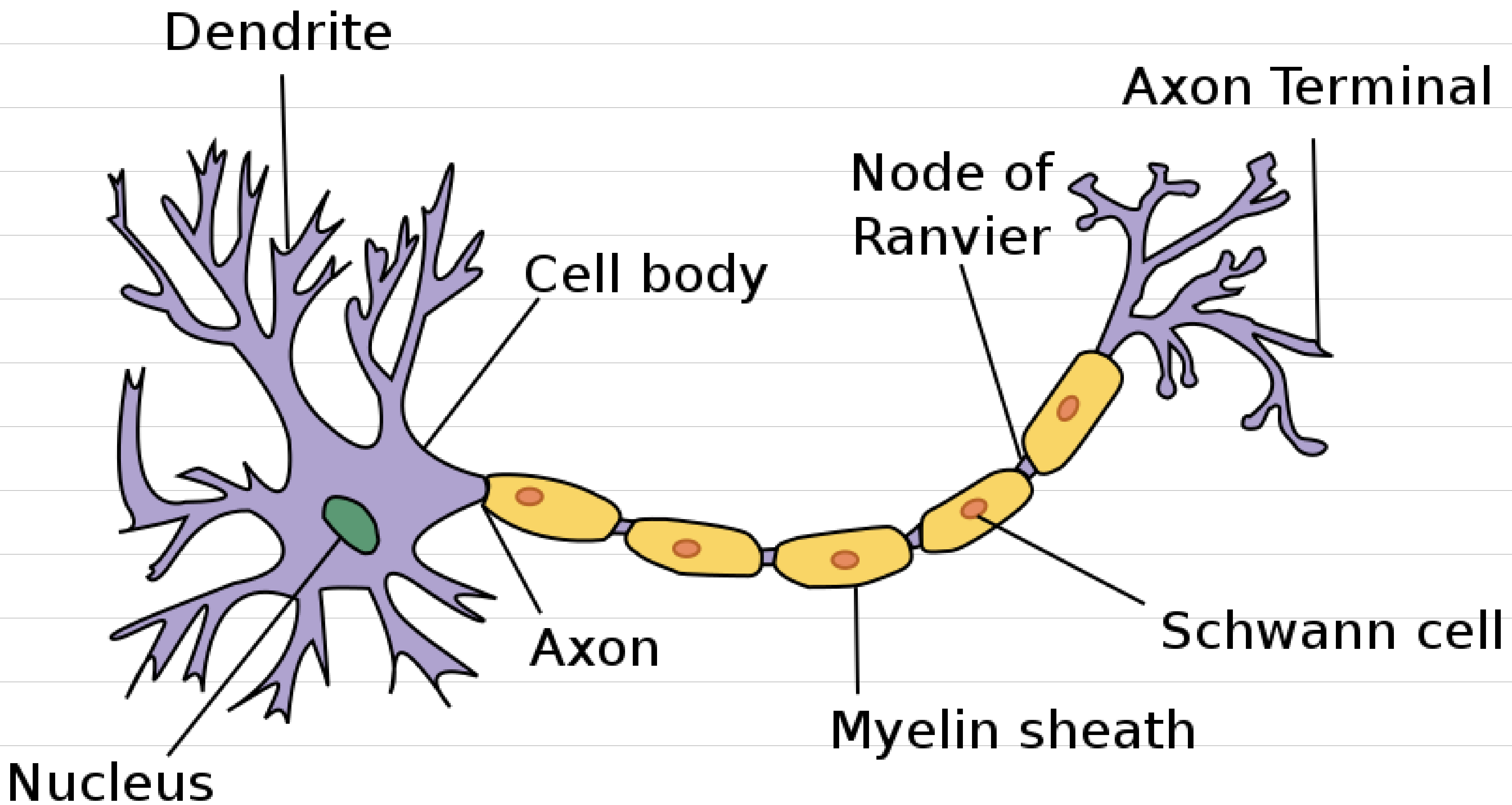
Feed Forward Neural Networks (FFNN)

Deep Learning | Zahra Amini

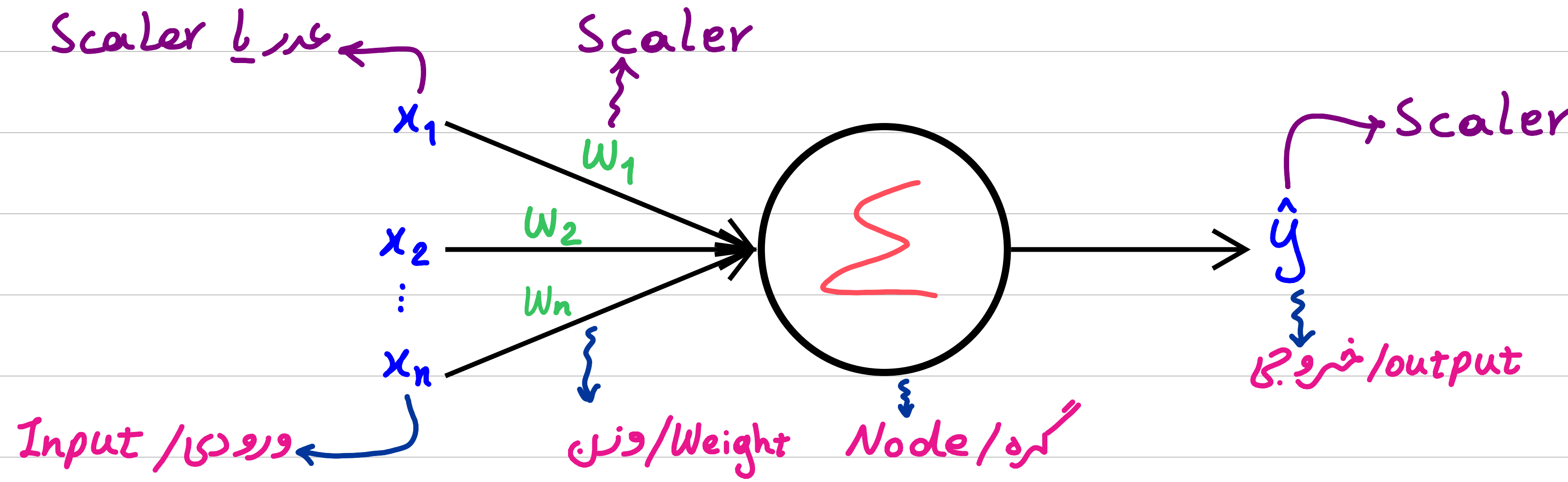
Telegram: @zahraamini_ai & Instagram:@zahraamini_ai & LinkedIn: @zahraamini-ai

https://zil.ink/zahraamini

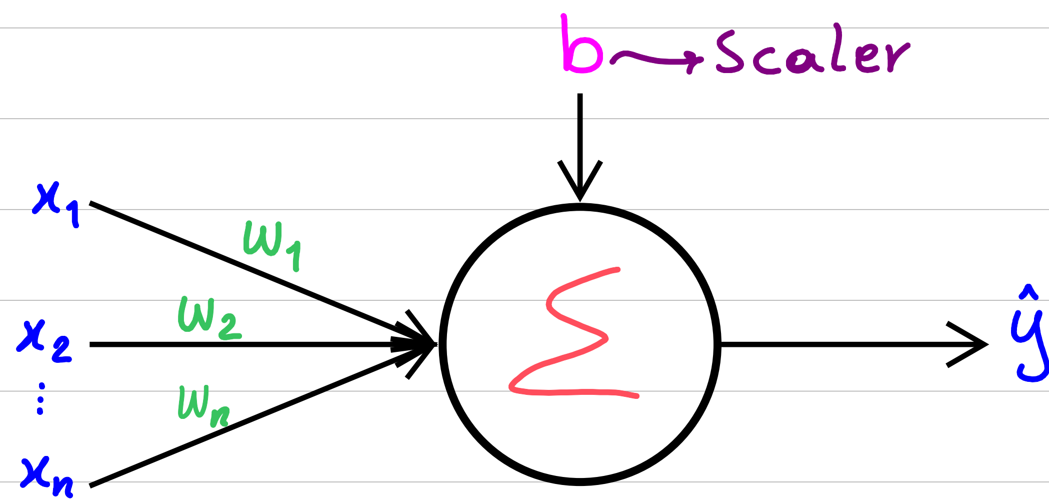
Neuron


Dendrite
Axon Terminal
Node of Ranvier
Cell body
Axon
Schwann cell
Myelin sheath
Nucleus

Artificial Neuron

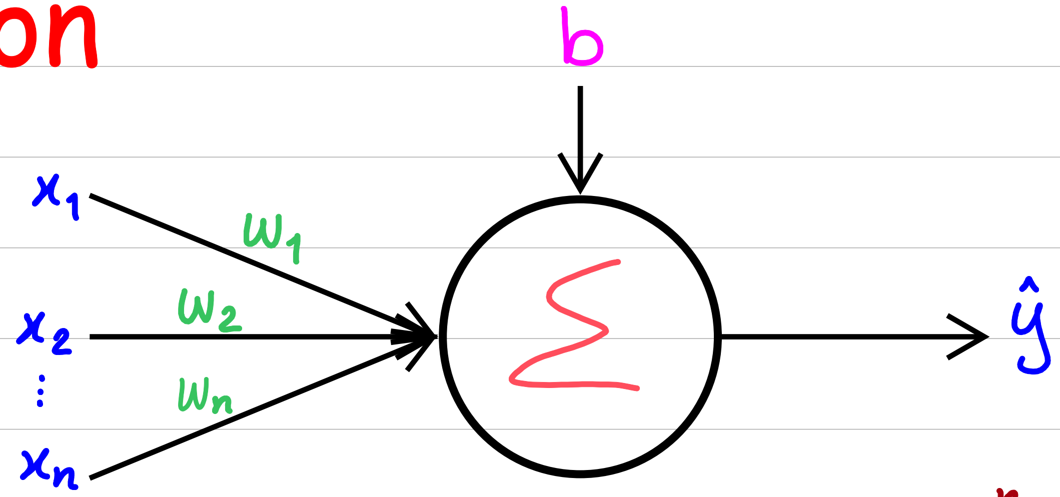


$$\hat{y} = x_1 \cdot w_1 + x_2 w_2 + \cdots + x_n w_n = \sum_{i=1}^{n} x_i w_i$$



$$\hat{y} = x_1 \cdot w_1 + x_2 w_2 + \cdots + x_n w_n + b = \sum_{i=1}^{n} x_i w_i + b$$

Vectorization



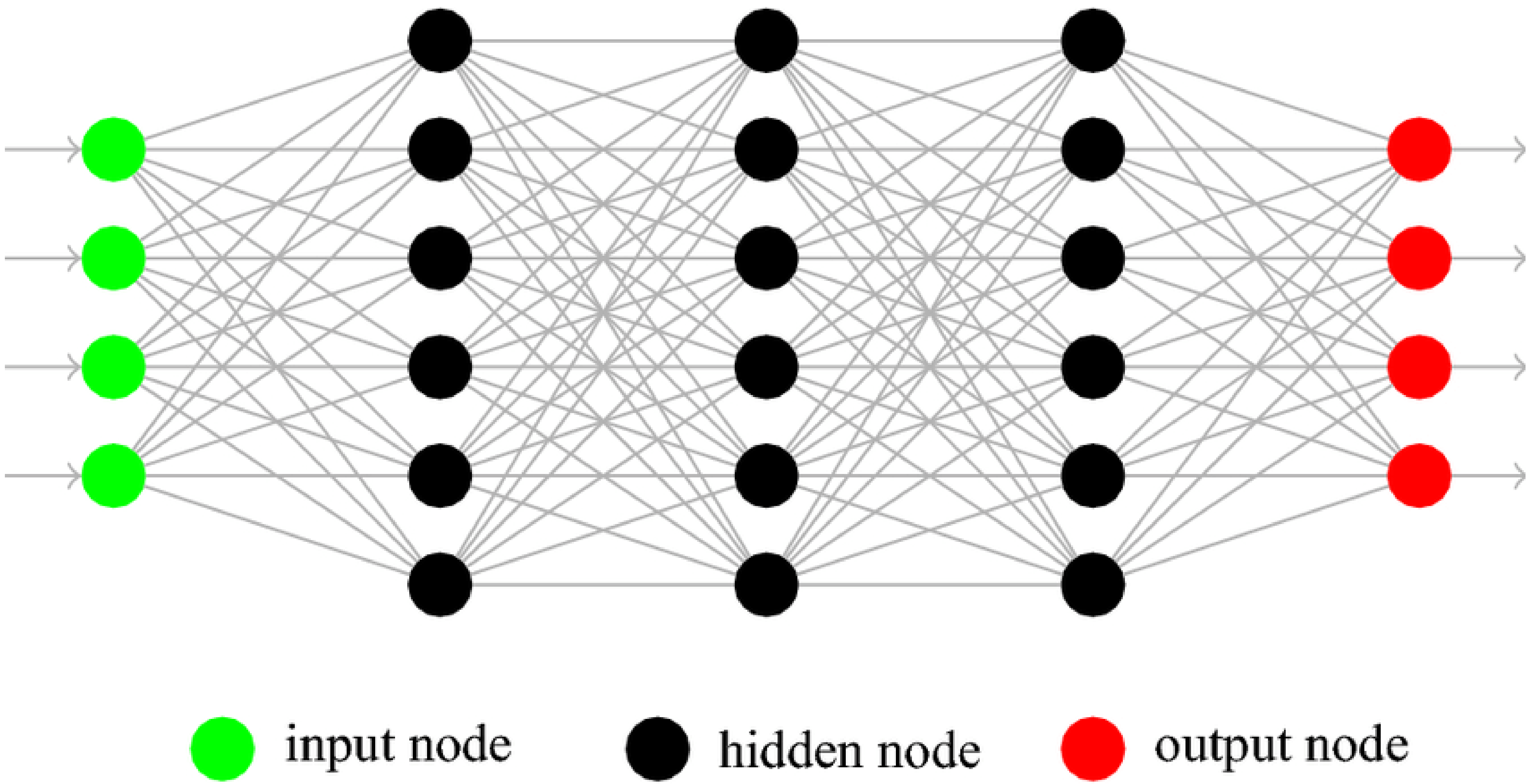
$$\hat{y} = x_1 \cdot w_1 + x_2 w_2 + \cdots + x_n w_n + b = \sum_{i=1}^{n} x_i w_i + b$$

$$\underset{W}{\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}_{n\times 1}} \quad \underset{X}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}} + b = \hat{y} \rightsquigarrow \underset{w^T}{\begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix}_{1\times n}} \underset{X}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}} + b = \hat{y}$$

$$\boxed{\checkmark}: \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0n} \\ a_{10} & a_{11} & \cdots & a_{1n} \\ \vdots & & & \\ a_{m0} & m_{m1} & \cdots & a_{mn} \end{bmatrix}_{m\times n} \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0p} \\ a_{10} & a_{11} & \cdots & a_{1p} \\ \vdots & & & \\ a_{n0} & m_{n1} & \cdots & a_{np} \end{bmatrix}_{n\times p} = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0p} \\ a_{10} & a_{11} & \cdots & a_{1p} \\ \vdots & & & \\ a_{m0} & m_{mm} & \cdots & a_{mp} \end{bmatrix}_{m\times p}$$

$$n = n$$

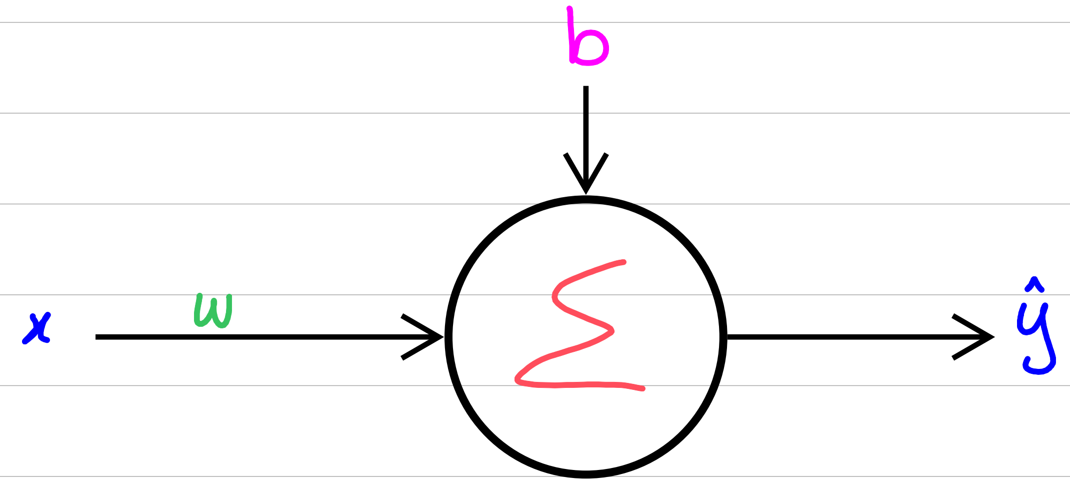
Neural Network Layers



☑: If #Hidden layers $\geq 2 \rightsquigarrow$ Deep Neural Network

Feed-Forward Neural Network (FFNN)

:



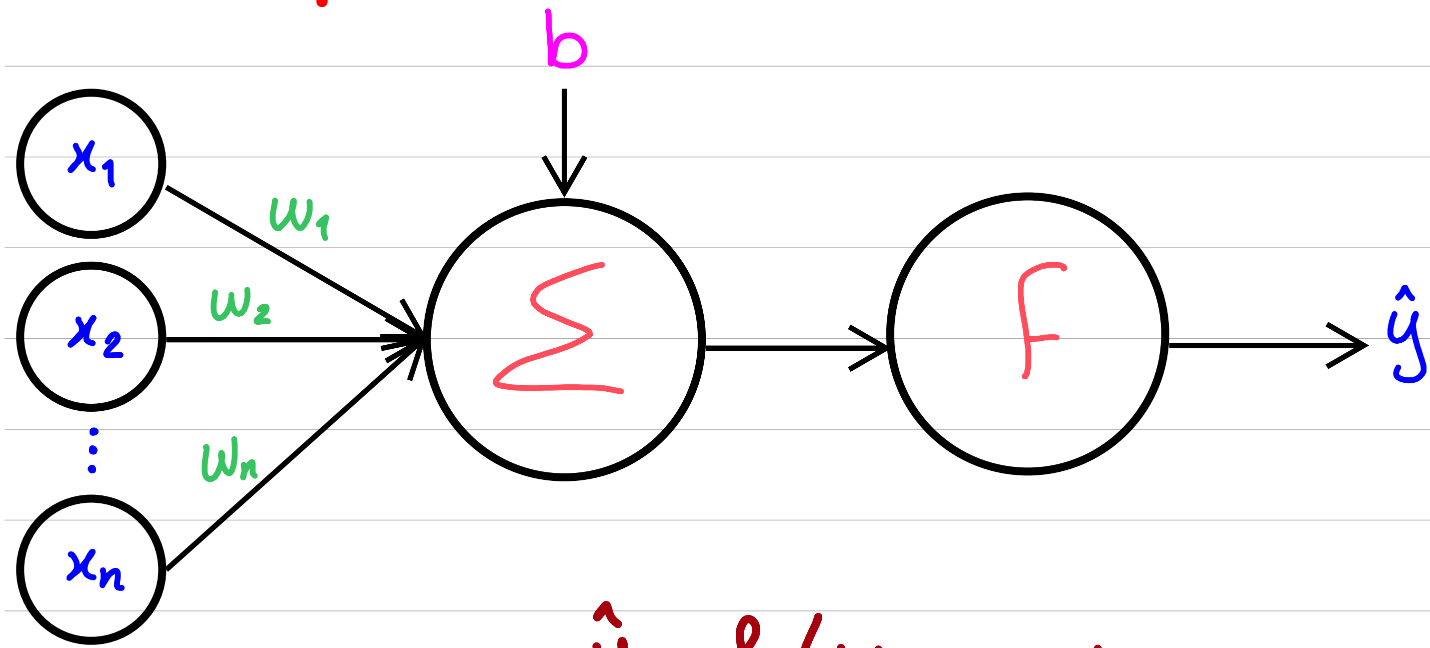
Size (m^2)	Price (k\$)	$\hat{y}$
800	200	208
1000	240	220
1200	280	232

$w = 60 \qquad b = 160000$

① $\hat{y} = wx = (60 \times 800) = 48{,}000 \qquad 48{,}000 + \overbrace{160{,}000}^{b} = 208{,}000$

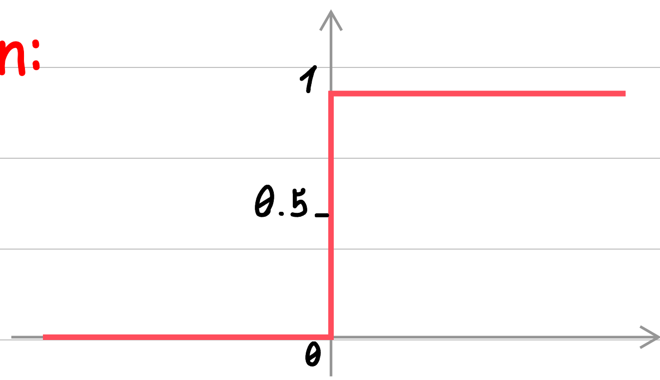
② $\hat{y} = wx + b = (60 \times 1000) + 160{,}000 = 220{,}000$ ③ $\hat{y} = wx + b = (1200 \times 60) + 160{,}000 = 232{,}000$

Perceptron



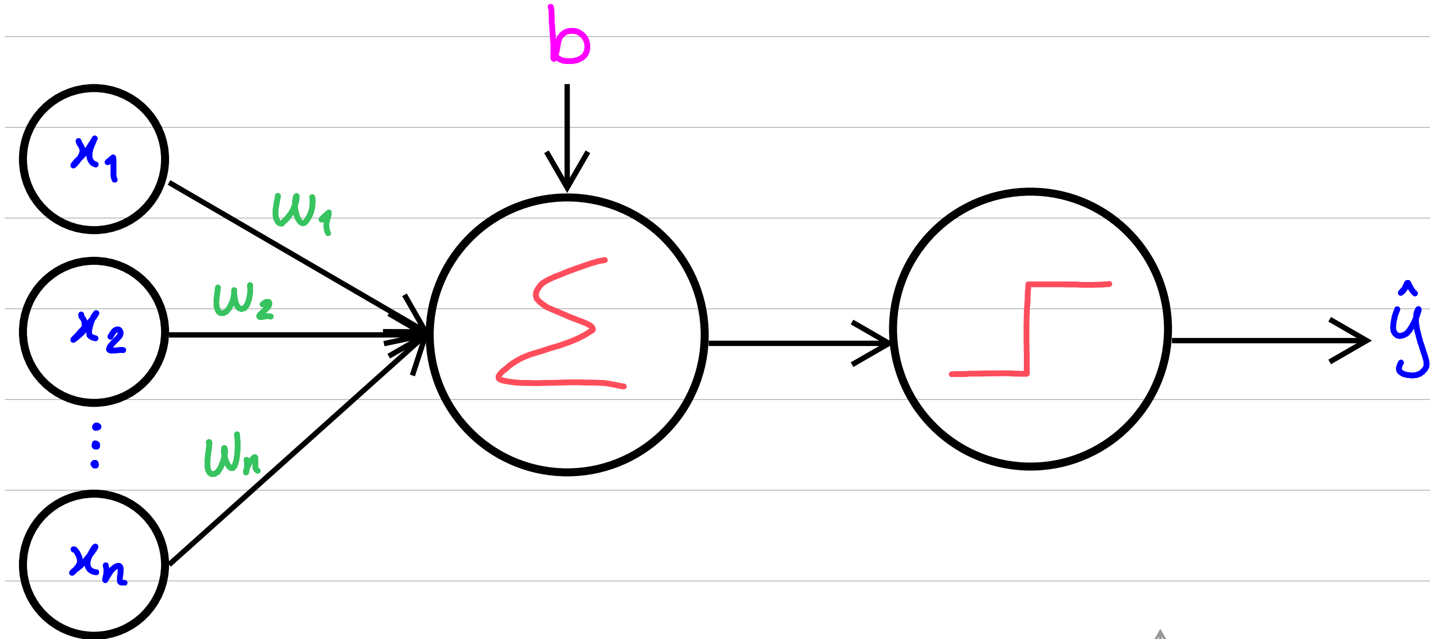
$$\hat{y} = f(w_1x_1 + w_2x_2 + \cdots + w_nx_n + b) = f\left(\sum_{i=1}^{n} w_ix_i + b\right)$$

Step Function:



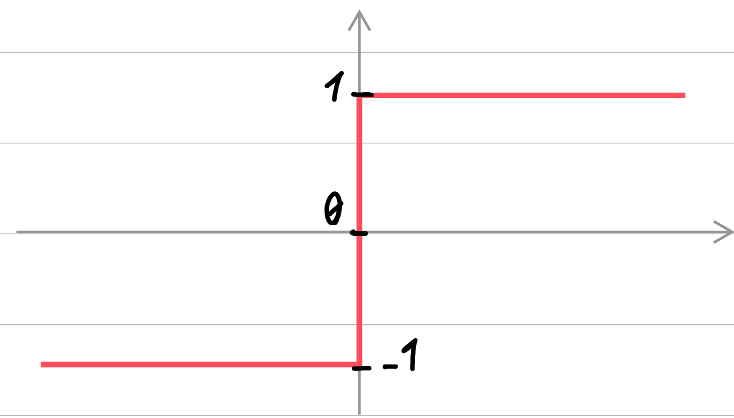
$$y = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$$

Perceptron



Sign Function:

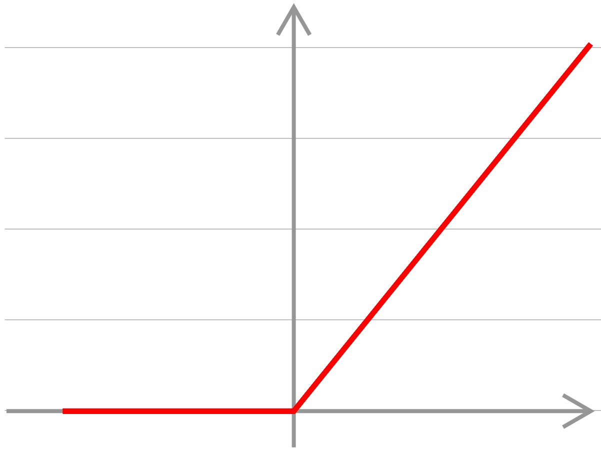
Heaviside



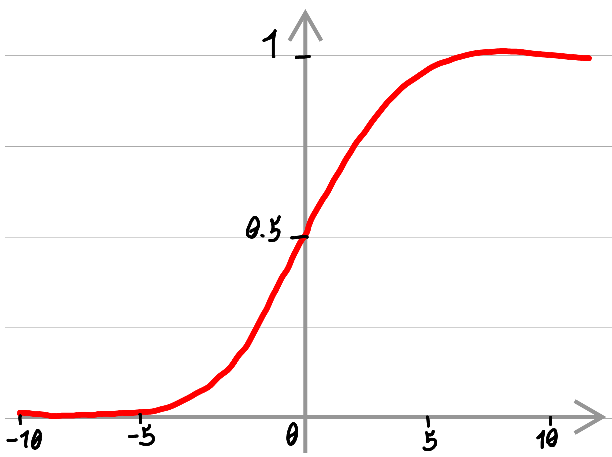
$$y = \begin{cases} 1 & x > \theta \\ -1 & \text{o.w} \end{cases}$$

$$\hat{y} = \text{Step}(w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b) = \text{Step}\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

Activation Function

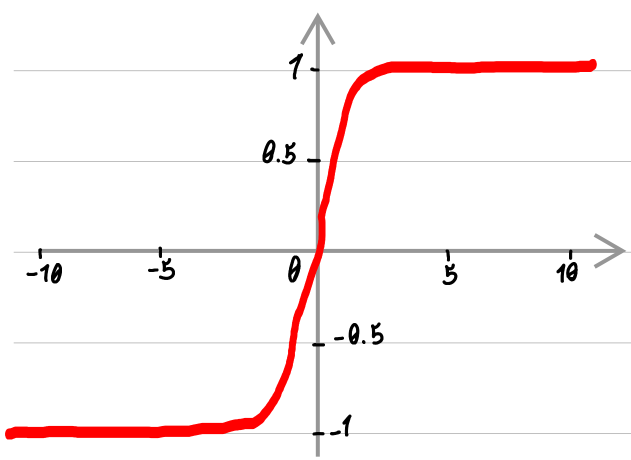


$$\mathrm{ReLU}(z) = \max(0, z)$$



Sigmoid $\rightsquigarrow \sigma(z) = \dfrac{1}{1 + e^{-z}}$

Softmax $\rightsquigarrow \sigma(z)_i = \dfrac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$

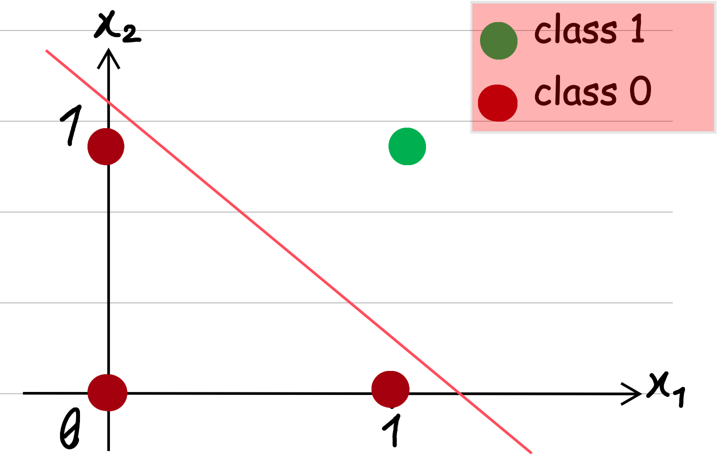


tanh $\rightsquigarrow \sigma(z) = \dfrac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$

Logic Gates

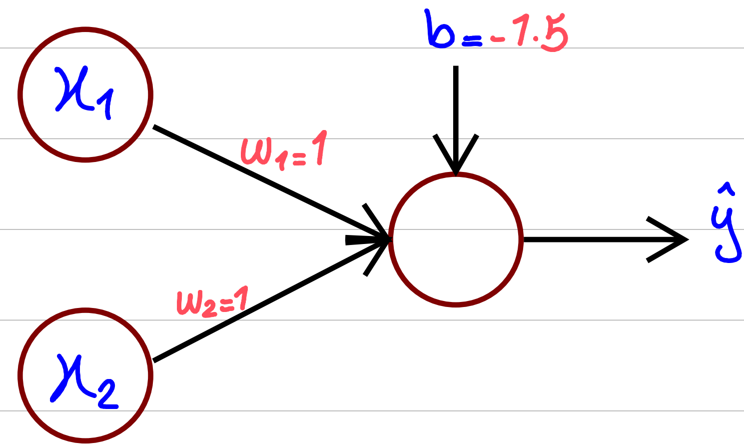
1. And &

$$\text{Step} = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$$



:

x_1	x_2	x_1 and x_2
T 1	T 1	T 1
F 0	T 1	F 0
T 1	F 0	F 0
F 0	F 0	F 0



① : $\hat{y} = \text{Step}\left(\sum wx+b\right) = \text{Sign}\left(\overset{0.5}{\cancel{(1\times1)+(1\times1)+(-1.5)}}\right) = 1$

② : $\hat{y} = \text{Step}\left(\overset{-0.5}{\cancel{(1\times0)+(1\times1)+(-1.5)}}\right) = 0$

③ : $\hat{y} = \text{Step}\left(\overset{-0.5}{\cancel{(1\times1)+(0\times1)+(-1.5)}}\right) = 0$

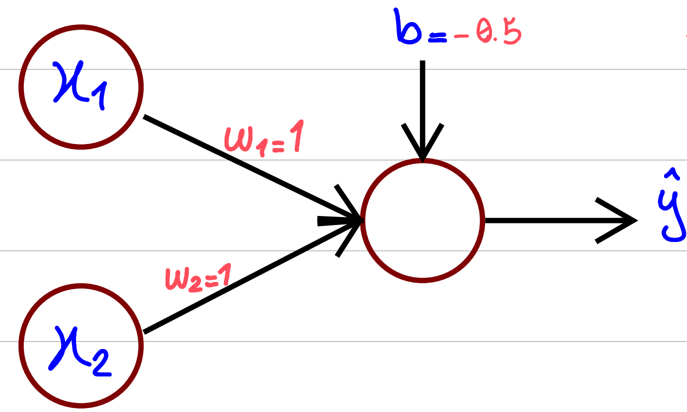
④ : $\hat{y} = \text{Step}\left(\overset{-1.5}{\cancel{(0\times1)+(0\times1)+(-1.5)}}\right) = 0$

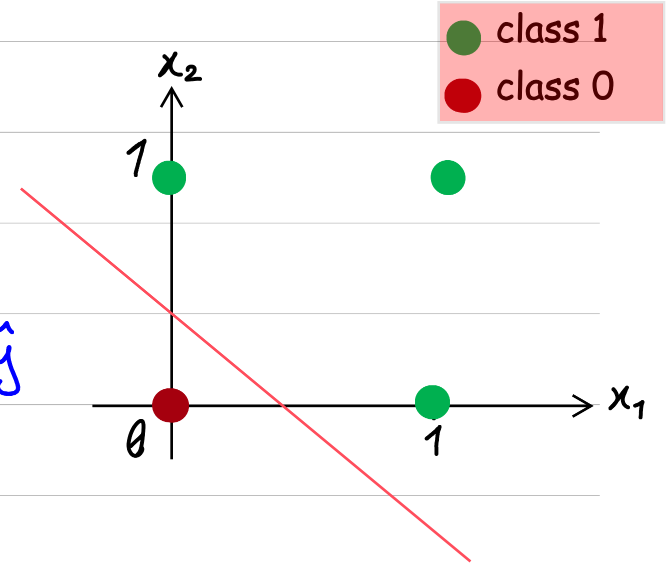
Logic Gates

2. Or |

:

x_1	x_2	x_1 or x_2
T 1	T 1	T 1
F 0	T 1	T 1
T 1	F 0	T 1
F 0	F 0	F 0





①: $\hat{y} = \text{Step}\left(\overset{1.5}{\cancel{(1\times1)+(1\times1)+(-0.5)}}\right) = 1$

②: $\hat{y} = \text{Step}\left(\overset{0.5}{\cancel{(0\times1)+(1\times1)+(-0.5)}}\right) = 1$

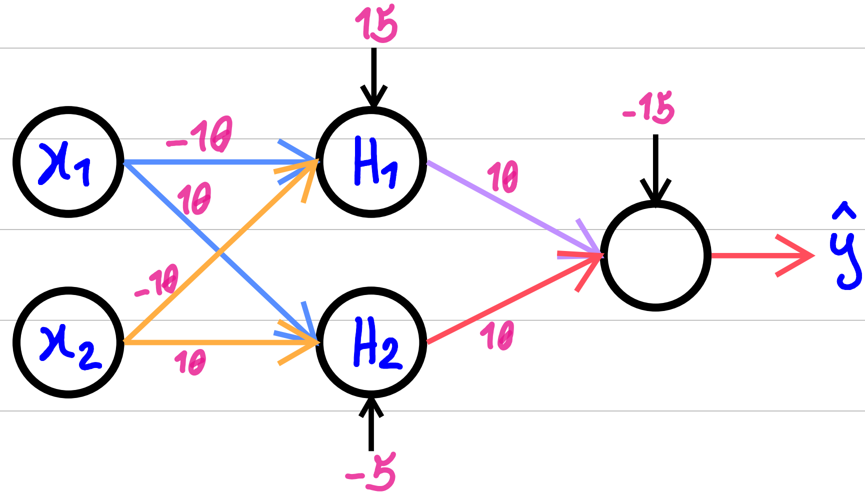
③: $\hat{y} = \text{Step}\left(\overset{0.5}{\cancel{(1\times1)+(0\times1)+(-0.5)}}\right) = 1$

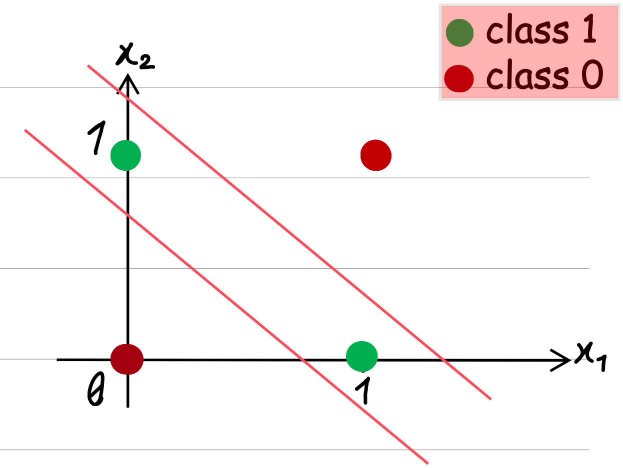
④: $\hat{y} = \text{Step}\left(\overset{-0.5}{\cancel{(0\times1)+(0\times1)+(-0.5)}}\right) = 0$

Multilayer Perceptron (MLP)

3. XOR

x_1	x_2	x_1 xor x_2
T 1	T 1	F 0
F 0	T 1	T 1
T 1	F 0	T 1
F 0	F 0	F 0





① $H_1 = \text{Sigmoid}\left(\sum Wx + b\right) = \text{Sigmoid}\left((-10 \times 1) + (-10 \times 1) + 15\right) = \text{Sigmoid}(-5) = 0.006$

$H_2 = \text{Sigmoid}\left(\sum Wx + b\right) = \text{Sigmoid}\left((10 \times 1) + (10 \times 1) - 5\right) = \text{Sigmoid}(15) = 0.999$

$\hat{y} = \text{Sigmoid}\left((10 \times 0.006) + (10 \times 0.999) - 15\right) = \text{Sigmoid}(-4{,}88) = 0.0007 \simeq 0 \rightsquigarrow \hat{y} = 0$

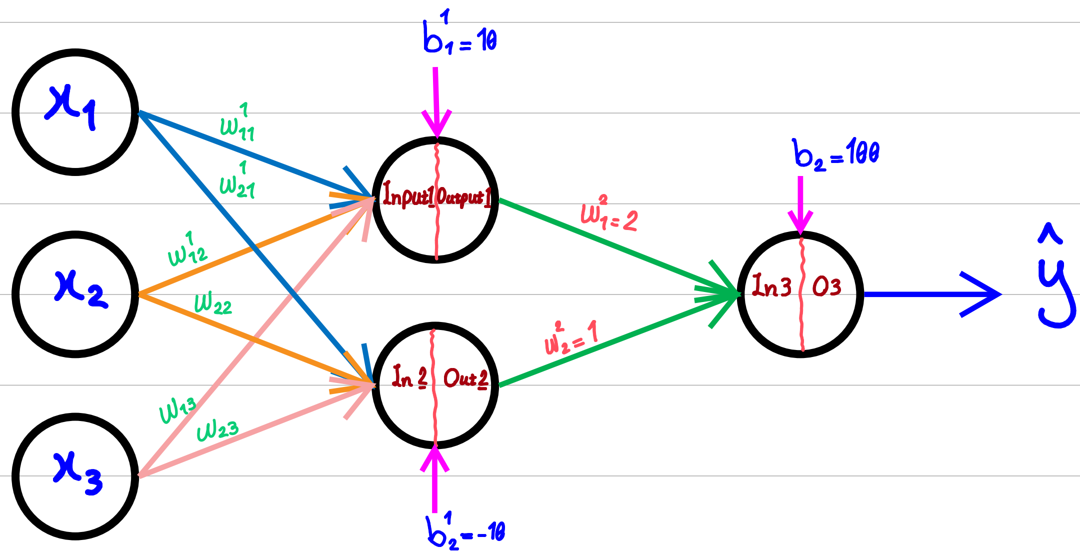
② $H_1 = \text{Sigmoid}\left((-10 \times 0) + (-10 \times 1) + 15\right) = \text{Sigmoid}(5) = 0.993$, $H_2 = \text{Sigmoid}\left((10 \times 0) + (10 \times 1) - 5\right) = \text{Sigmoid}(5) = 0.993$

$\hat{y} = \text{Sigmoid}\left((10 \times 0.993) + (10 \times 0.993) - 15\right) = \text{Sigmoid}(4.86) = 0.992 \rightsquigarrow \hat{y} = 1$

③ $\hat{y} = 1$ ④ $\hat{y} = 0$

Feed-Forward Neural Network (FFNN)

:



Size	#bedrooms	Age	Price
800	3	20	200

$$W_1 = \begin{bmatrix} 0.01 & 50 & 5 \\ 0.02 & 2.5 & -2.5 \end{bmatrix}, \quad b_1 = \begin{bmatrix} 10 \\ -10 \end{bmatrix},$$

$$W_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad b_2 = 100$$

$$\underbrace{\text{Activation Function}}_{\text{Hidden Layer}} = \text{ReLU} = \max(0, \text{number})$$

$$\text{Input } 1 = \sum wx + b = (0.01 \times 800) + (50 \times 3) + (5 \times 20) + 10 = 268 \xrightarrow{AF} \text{Output1} = \text{ReLU}(268) = 268$$

$$\text{Input } 2 = (0.02 \times 800) + (2.5 \times 3) + (-2.5 \times 20) + (-10) = 31 \xrightarrow{AF} \text{Output } 2 = \text{ReLU}(31) = 31$$

$$\text{In } 3 = (2 \times 268) + (1 \times 31) + 100 = 667 \longrightarrow \text{Output } 3 = \hat{y} = 667$$

Feed-Forward Neural Network (FFNN)

:

: Normalization $\rightsquigarrow$ Value/max_value

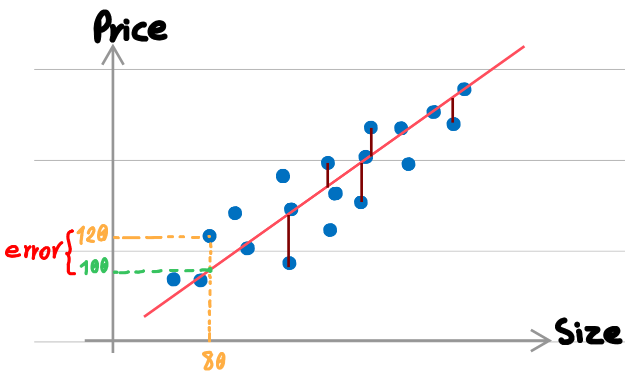
$Size = 800/1000 = 0.8$ $Bedrooms = 3/5 = 0.6$ $Age = 20/50 = 0.4$

$In1 = (0.01 \times 0.8) + (50 \times 0.6) + (5 \times 0.4) + 10 = 42.008 \xrightarrow{AF} Output1 = ReLU(42.008) = 42.008$

$In2 = (0.02 \times 0.8) + (25 \times 0.6) + (-2.5 \times 0.4) + (-10) = 4.016 \xrightarrow{AF} Output2 = ReLU(4.016) = 4.16$

$In3 = (2 \times 42.008) + (1 \times 4.016) + 100 = 188.032 \rightsquigarrow Output3 = \hat{y} = 188.032$

Loss Function $\rightsquigarrow$ Error $\rightsquigarrow$ به اختلاف بین مقدار واقعی و پیشبینی مدل خطا می گویند.



Regression Loss

1. $MSE = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2$?
2. $MAE = \frac{1}{m} \sum_{i=1}^{m} |\hat{y}_i - y_i|$
3. $RMSE = \sqrt{MSE}$
4. $R2\text{-}Score = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$

$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \rightarrow SS_R$ (SUM of Squares for Regression)

$\sum_{i=1}^{n} (y_i - \bar{y})^2 \rightarrow SS_M$ (Sum of Squares for Model)

x	y	$\hat{y}$
1	2	1
2	4	2
3	6	3
4	8	4

$$x \xrightarrow{w=1} \bigcirc \rightarrow \hat{y}$$

$$MSE = \frac{1}{m} \sum_{i=1}^{m} (\hat{y} - y)^2 = \frac{1}{4}\left[\overset{1}{\cancel{(1-2)^2}} + \overset{4}{\cancel{(2-4)^2}} + \overset{9}{\cancel{(3-6)^2}} + \overset{16}{\cancel{(4-8)^2}}\right] = \frac{30}{4} = 7.5$$

$$R2 = 1 - \frac{SS_R}{SS_m} = 1 - \frac{30}{20} = -0.5$$?

-0.5 خوبه یا بد ؟!

$$SS_R = (2-1)^2 + (4-2)^2 + (6-3)^2 + (8-4)^2 = 30$$

$$SS_m = (2-5)^2 + (4-5)^2 + (6-5)^2 + (8-5)^2 = 20$$

$$\bar{y} = \frac{2+4+6+8}{4} = 5$$

R2 چه عددی باشد عملکرد مدل خوبه؟

R2 =1: Perfect fit

0<R2 <1: Good fit

*R*2=0: equivalent to using the mean

*R*2<0: Worse than using the mean

Thanks for
your attention