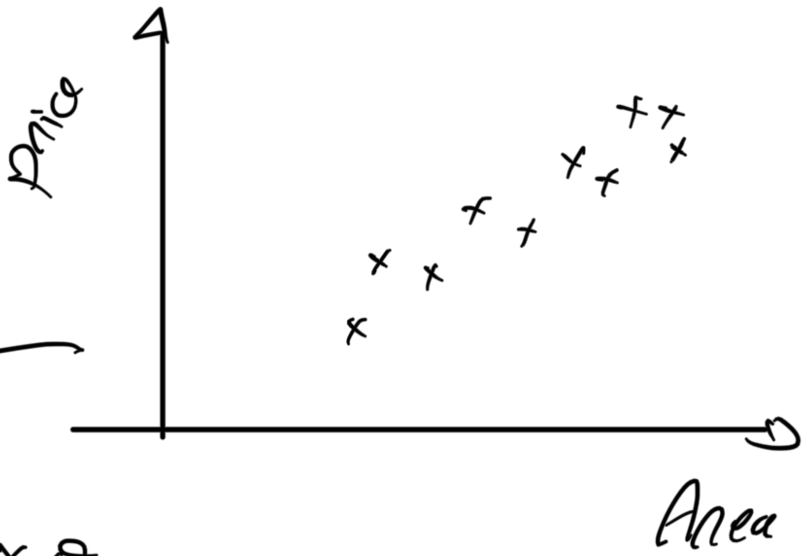
Neural Networks

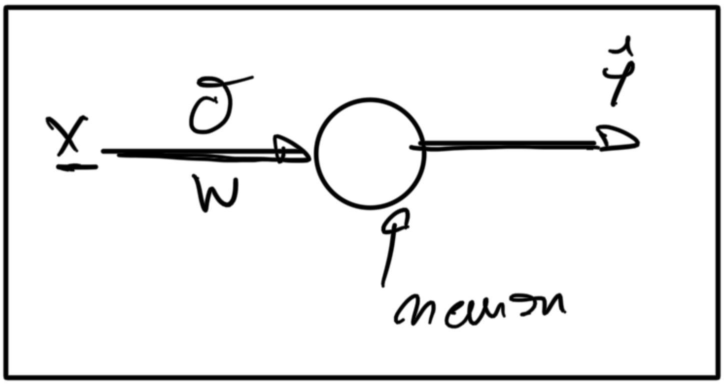
$$[\underline{x_1}, \underline{x_2}, \underline{x_3}, \underline{x_4}, \underline{x_5}]$$

$$\rightarrow [x_1 x_2, \; x_1 x_2 x_3, \; x_5 x_2$$

$$4^5 = 1024$$

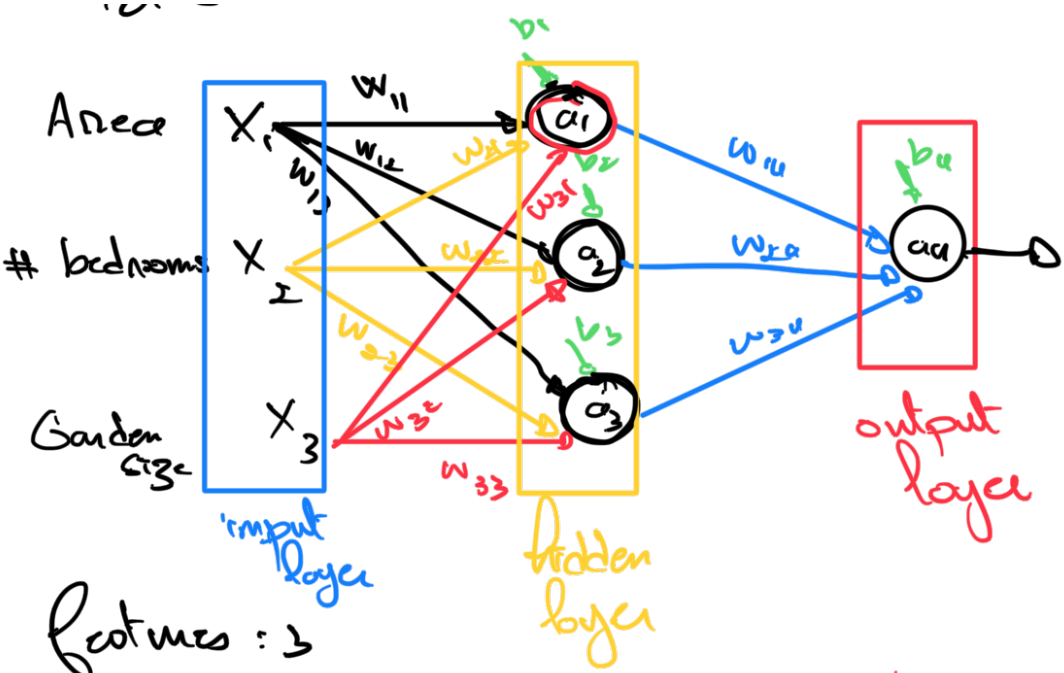


$$\hat{y} = x\theta$$



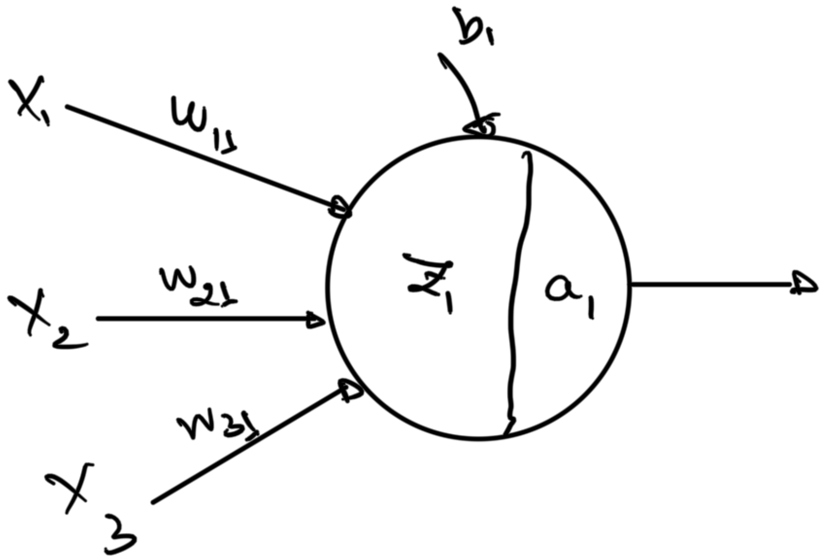
$$\hat{y} = x\theta$$

Let's consider more features



features : 3

Shallow NN



$$\rightarrow z_1 = w_{11} x_1 + w_{21} x_2 + w_{31} x_3 + b_1$$

$$\rightarrow a_1 = f(z_1) = \sigma(z_1)$$

$$z_2 = w_{12} x_1 + w_{22} x_2 + w_{32} x_3 + b_2$$

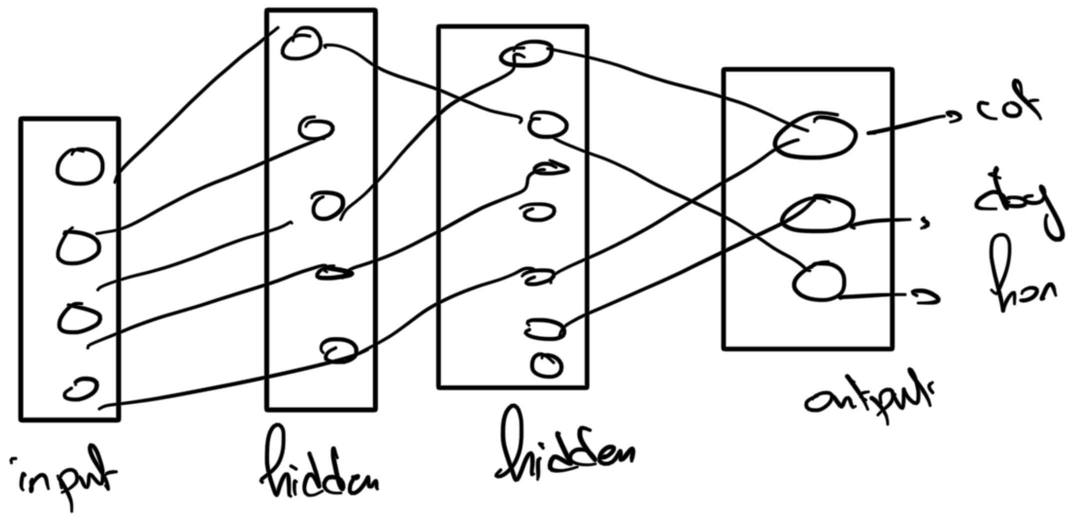
$$a_2 = \sigma(z_2)$$

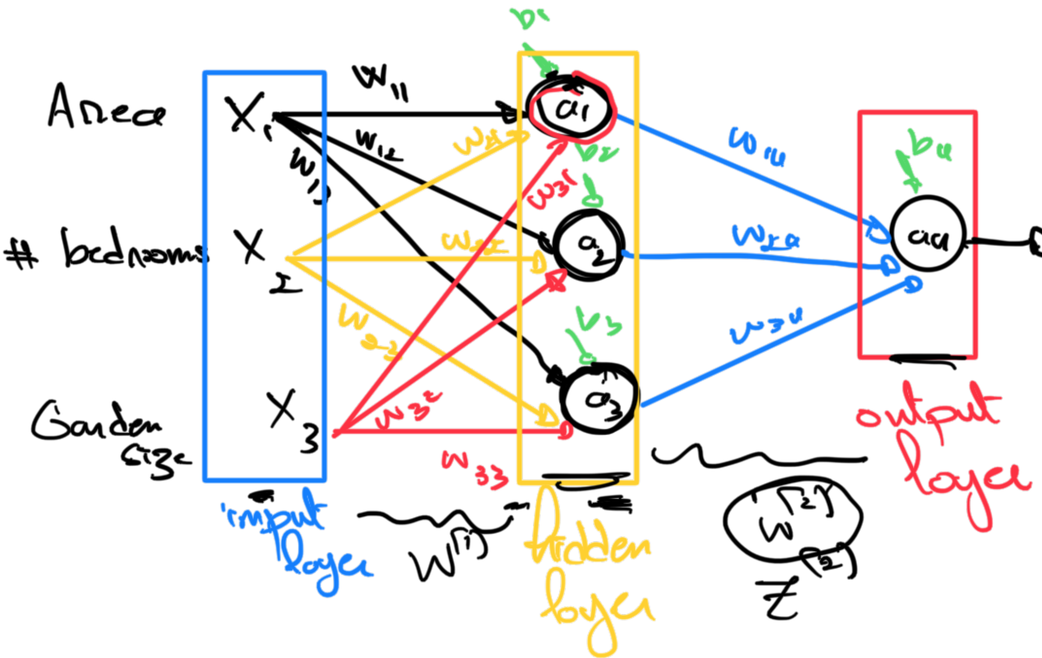
$$Z_3 = w_{13} x_1 + w_{23} x_2 + w_{33} x_3 + b_3$$

$$a_3 = \sigma(Z_3)$$

$$Z_4 = w_{14} a_1 + w_{24} a_2 + w_{34} a_3 + b_4$$

$$a_4 = Z_4 = \hat{y}$$





$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad \underset{(3,3)}{W^{[1]}} = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{pmatrix}$$

$$\underset{(3,1)}{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

$$Z^{[1]} = \underset{(3,3)}{W^{[1]}} \underset{(3,1)}{X} + \underset{(3,1)}{b^{[1]}} \qquad (3,1)$$

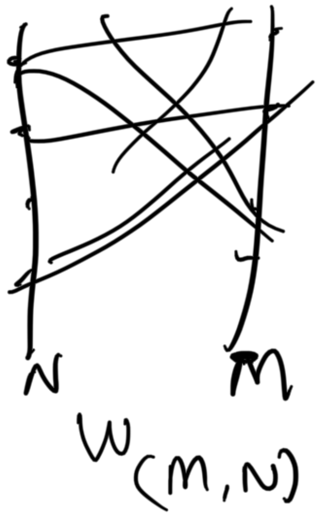
$$A^{[1]} = \text{Relu}(Z^{[1]}) \qquad (3,1)$$

$$W^{[2]} = \begin{pmatrix} w_{14} & w_{24} & w_{34} \end{pmatrix}$$

$$b^{[2]} = b_4$$

$$Z^{[2]} = \underset{(1,3)}{W^{[2]}} \underset{(3,1)}{A^{[1]}} + \underset{(1,1)}{b^{[2]}} \qquad (1,1)$$

$$A^{[2]} = Z^{[2]} \quad (1,1)$$



General Case

features : M

$W^{[1]}$	$W^{[2]}$	$W^{[3]}$
b	b	b

$W^{[1]} : (N, M)$

$b^{[1]} : (N, 1)$

$W^{[2]} : (H, N)$

$b^{[2]} : (H, 1)$

m N H L

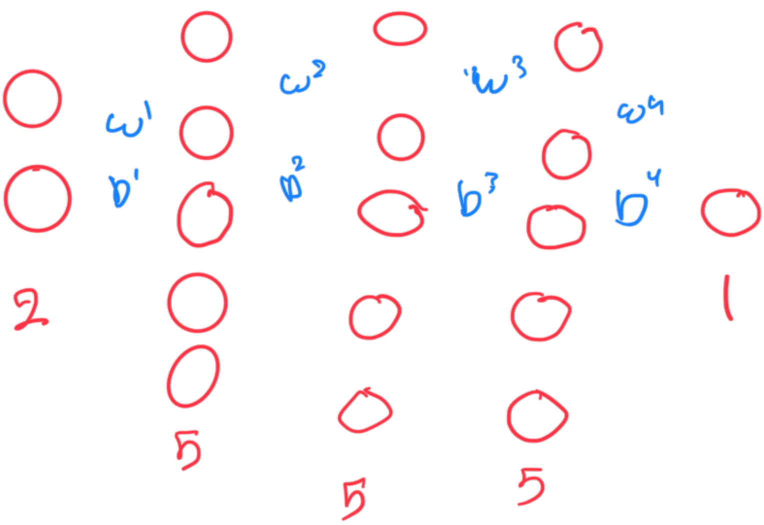
$W^{[3]} : (L, H)$

$b^{[3]} : (L, 1)$

Exercice

Neural Network with 3 hidden layer
each hidden layer has 5 neurons
with input 2 neurons.
with output 1 neuron.

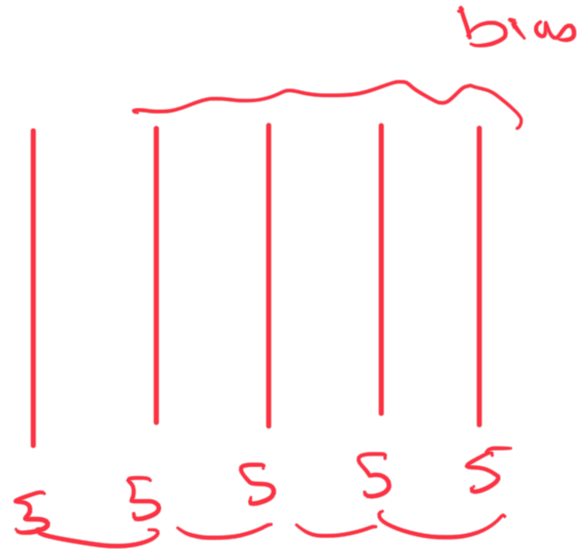
① Represent the NN
② count the learnable parameters.



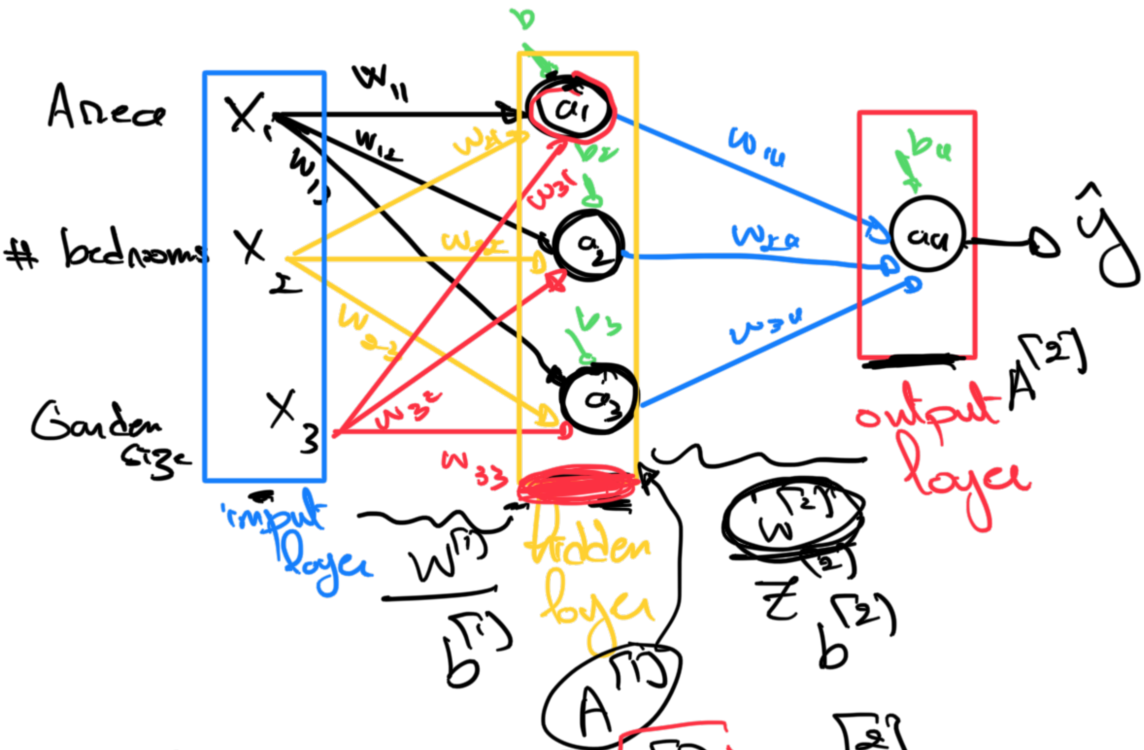
Exercice 2

ℓ neuron each?

5 layer with



$$LP: 5\times5 + 5\times5 + 5\times5 + 5\times5 + \underbrace{5 + 5 + 5 + 5}_{\text{bias}} = 120$$



① $A^{[2]} = \hat{y} = W^{[2]} A^{[1]} + b^{[2]}$

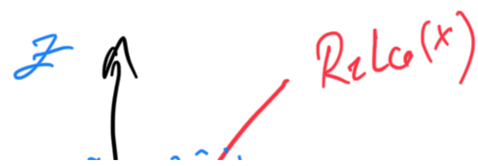
$$= W^{[2]}\left(\text{Relu}\left(\boxed{W^{[1]} X + b^{[1]}}\right)\right) + b^{[2]}$$

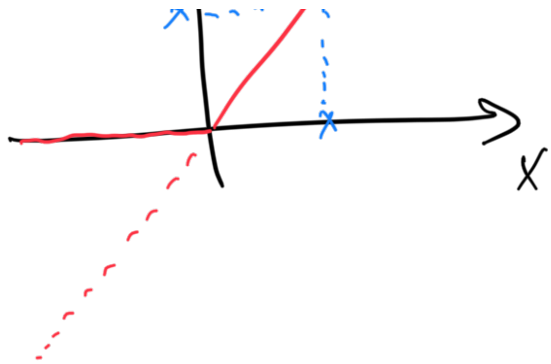
$$= \underbrace{W^{[2]} W^{[1]}} X + \boxed{W^{[2]} b^{[1]} + b^{[2]}}$$

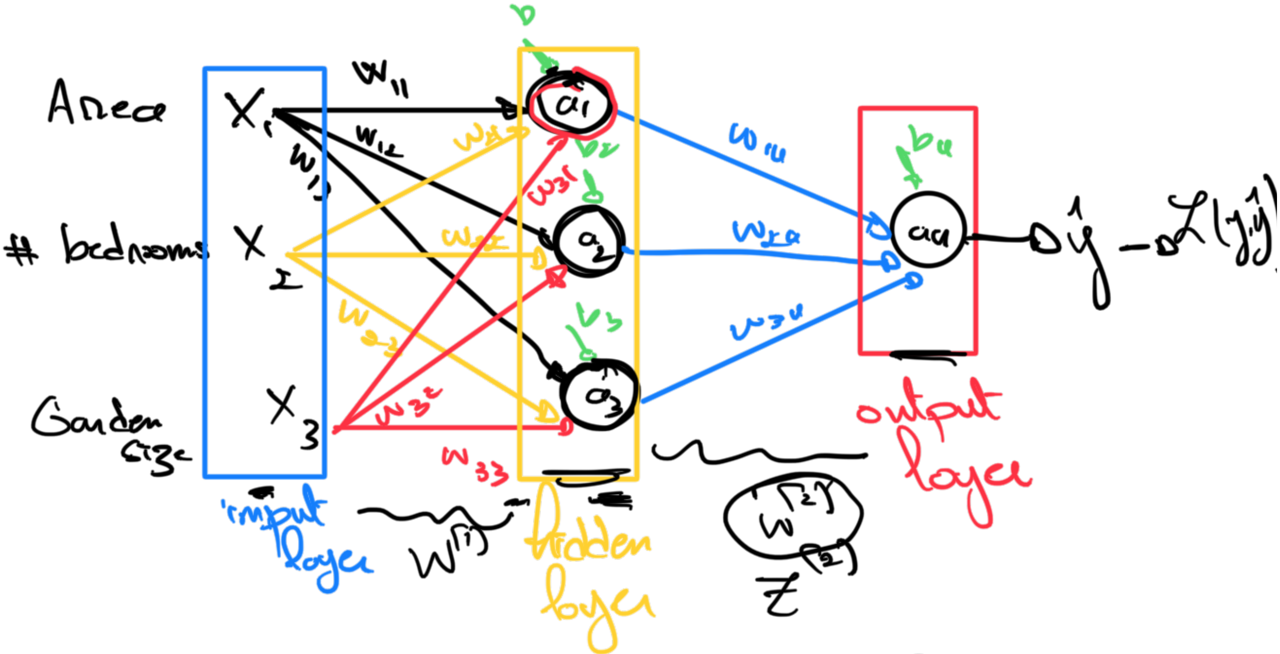
$$\hat{y} = \theta X + B$$

$$A^{[2]} = \hat{y} = \sigma\left(W^{[2]} A^{[1]} + b^{[2]}\right)$$

$$= W^{[2]} \text{Relu}\left(W^{(1)} X + b^{(1)}\right) + b^{(2)}$$




X



GD:

forward pass $\longrightarrow$

$\longleftarrow$ backward pass

$$\frac{\partial \mathcal{L}}{\partial b^{[2]}} \quad \frac{\partial \mathcal{L}}{\partial b^{[1]}} \quad \frac{\partial \mathcal{L}}{\partial w^{[1]}} \quad \frac{\partial \mathcal{L}}{\partial w^{[2]}}$$

$$W^{[1]} \longleftarrow W^{[1]} - \alpha \nabla_{W^{[1]}} \mathcal{L}$$

$$b^{[1]} \longleftarrow b^{[1]} - \alpha \nabla_{b^{[1]}} \mathcal{L}$$

$$W^{[2]} \longleftarrow$$

$$b^{[2]} \longleftarrow$$