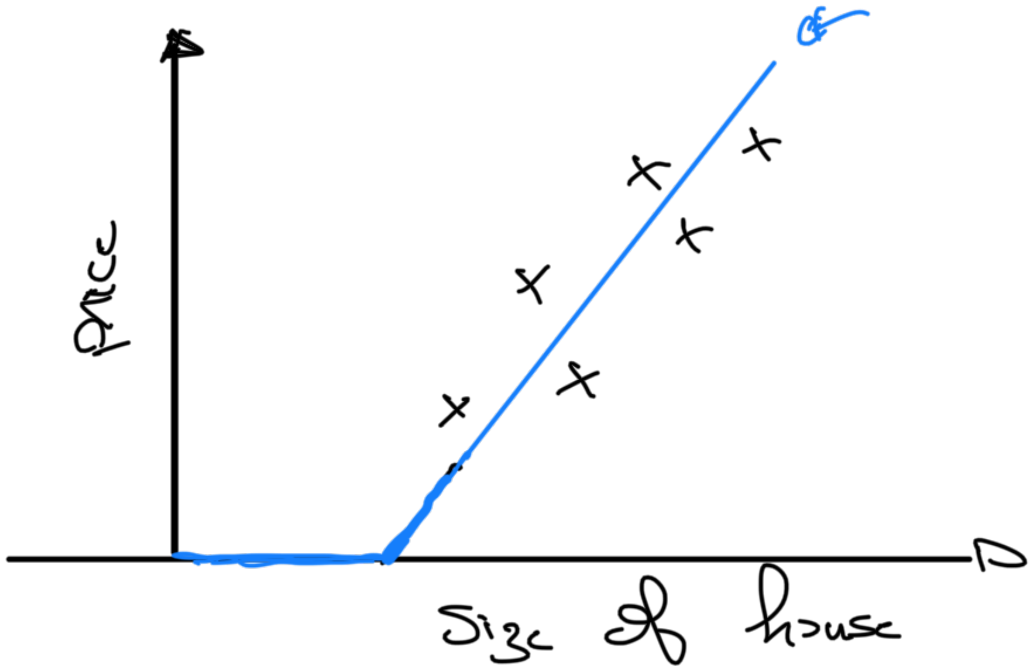
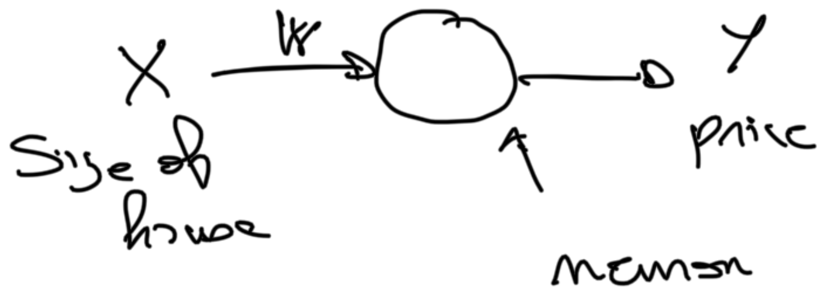
# Neural Networks

Consider we have to predict the price of the home




$Y = XW$   Relu$(WX)$

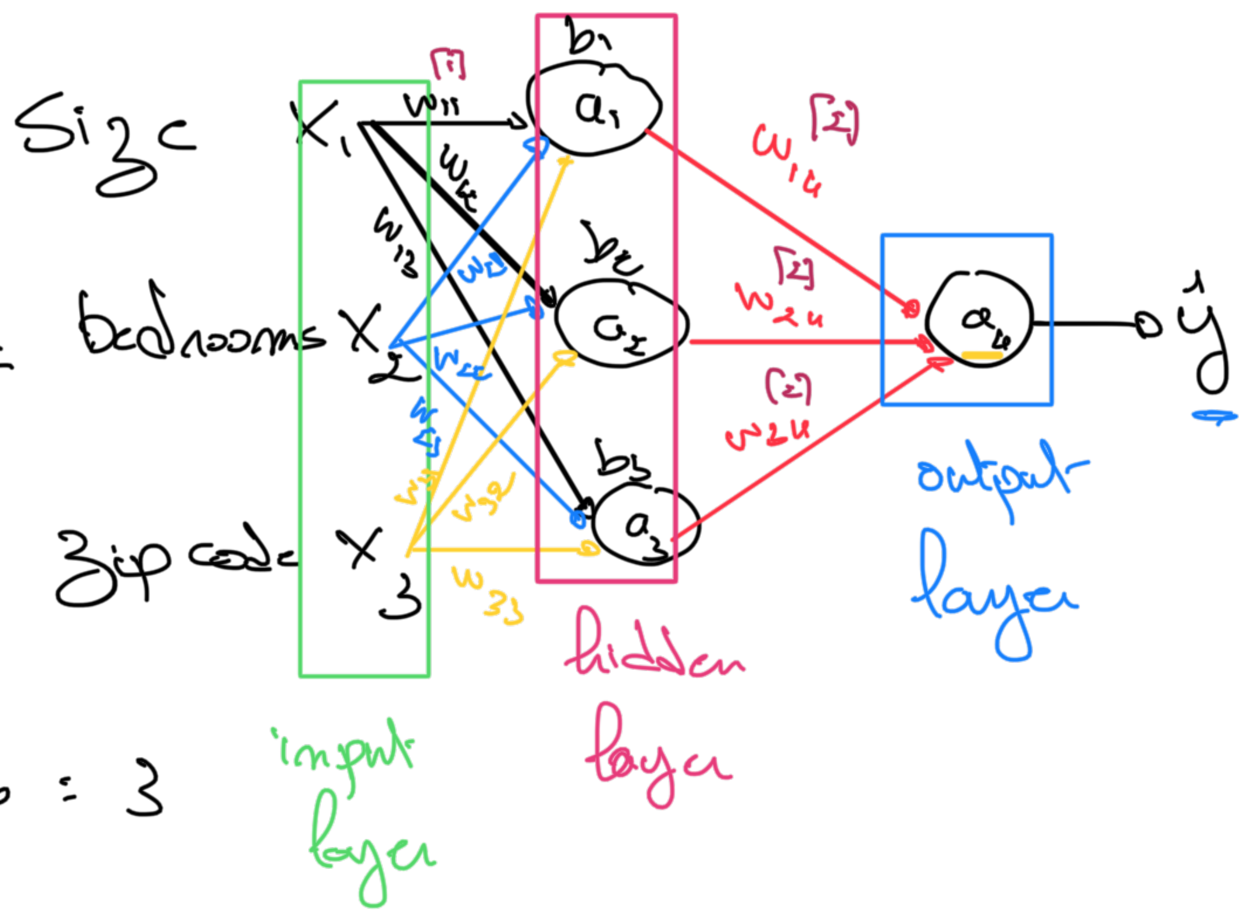
let's see when we have more features

first layer

$\boxed{W_{ij}^{[1]}, b_j^{[1]}}$

$\boxed{W_{ij}^{[2]}, b_j^{[2]}}$

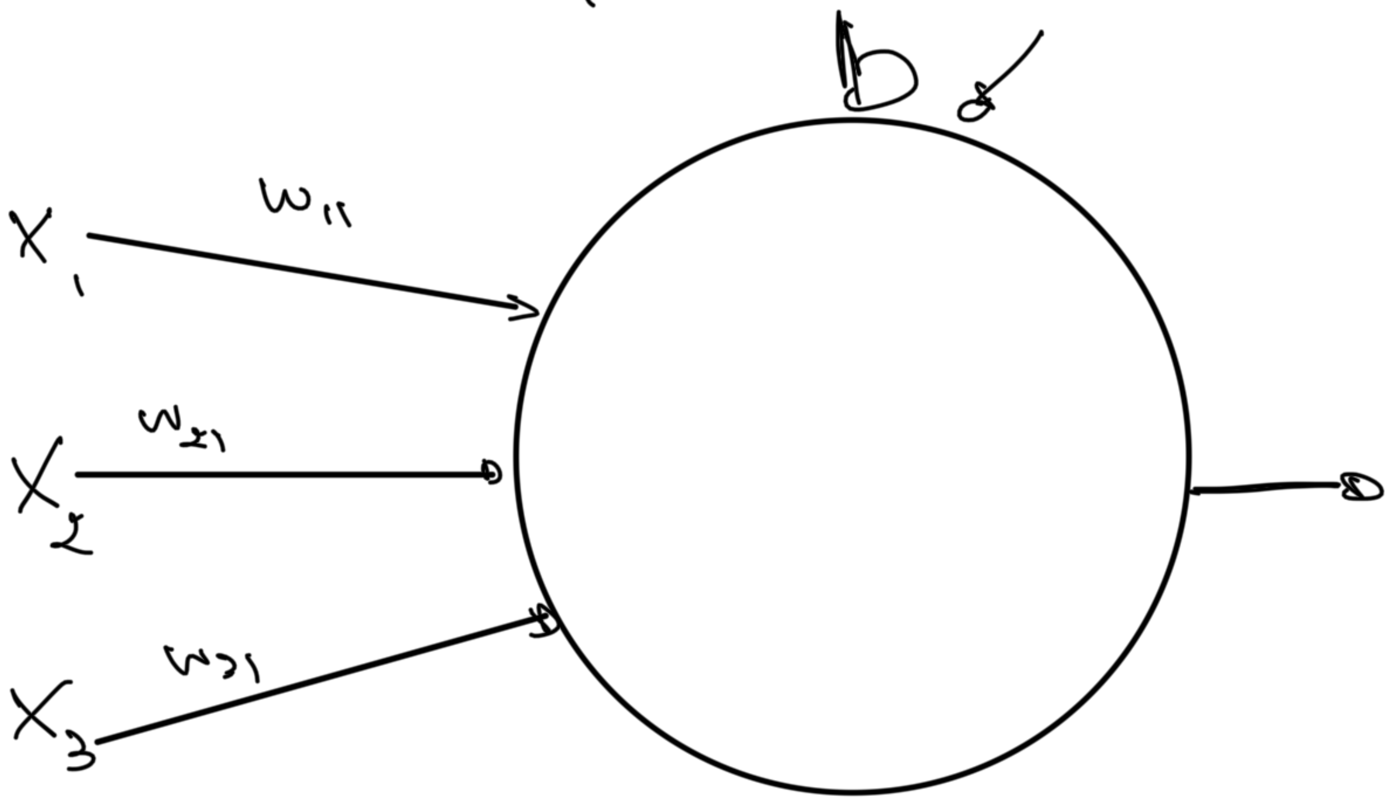
second layer




# features : 3

# Shallow Neural Network

learnable parameters W, b




$$z_1 = w_{11}X_1 + W_{21}X_2 + W_{31}X_3 + b$$

$$a_1 = \sigma(z_1)$$

sigmoid

in general an activation function

$$Z_1 = w_{11}^{[1]}X_1 + W_{21}^{[1]}X_2 + W_{31}^{[1]}X_3$$

$$a_1 = \sigma(z_1)$$

$$Z_2 = \underline{W_{12}^{[1]}X_1} + \underline{W_{22}^{[1]}X_2} + \underline{W_{32}^{[1]}X_3}$$

$$a_2 = \sigma(z_2)$$

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

$$a_3 = \sigma(a_3)$$

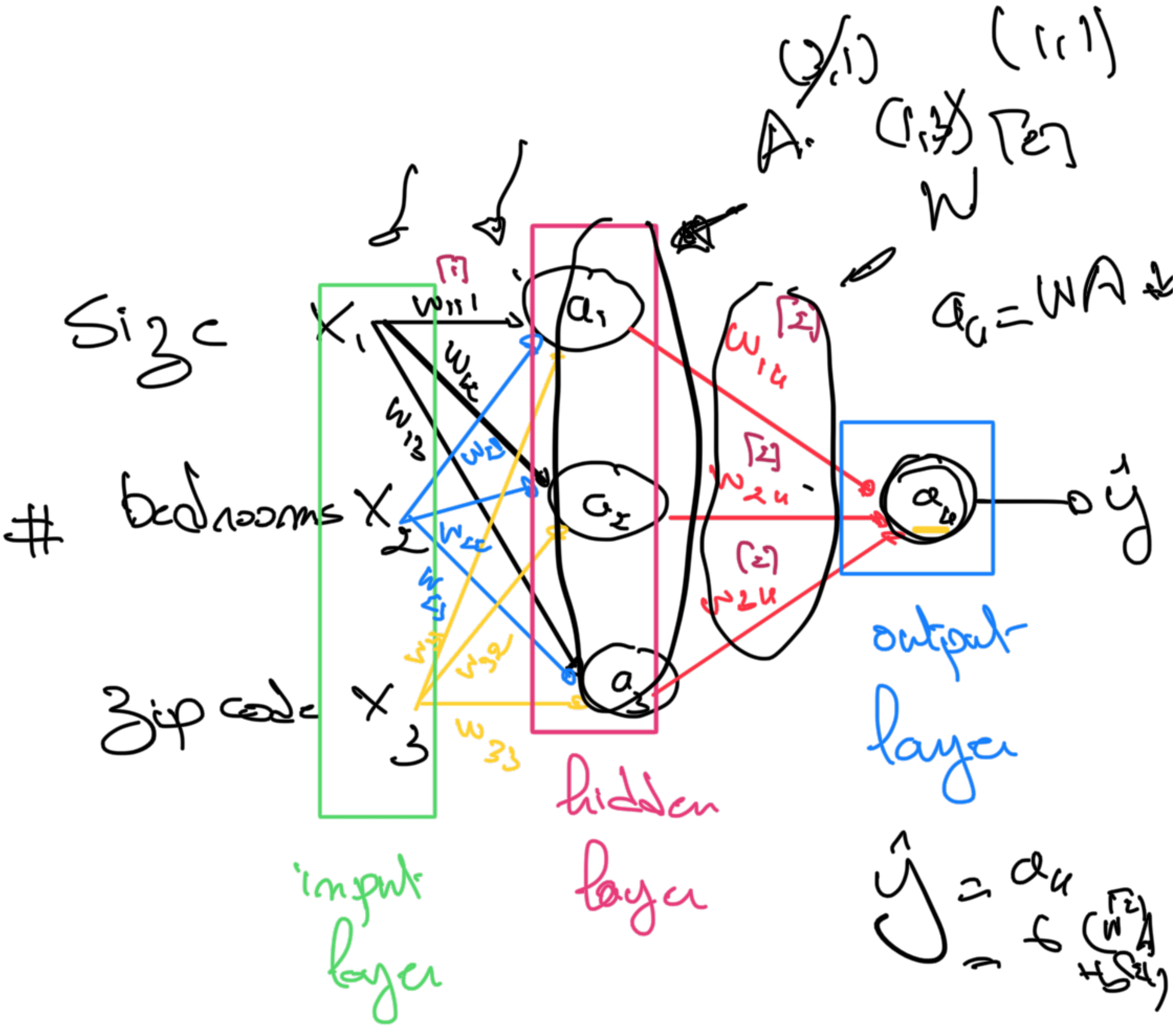
$$Z_4 = w_{14} a_1 + w_{24} a_2 + w_{34} a_3$$

$$a_4 = \sigma(Z_4)$$



$$\text{if } z < 0 \rightarrow \text{Relu}' = 0$$
$$\text{if } z \geq 0 \rightarrow \text{Relu}' = 1$$

$$\frac{1}{1+e^{-x}}$$




$X = \begin{pmatrix} X_1 \end{pmatrix}$ $\quad W^{[1]} = \begin{pmatrix} w_{11} & w_{12} & w_{13} \end{pmatrix}$

$$\underset{(3,1)}{} \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} \qquad \underset{(3,3)}{} \begin{pmatrix} w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{pmatrix}$$

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

$$\underset{(3,1)}{Z} = WX + b$$

$$\underset{(3,1)}{A} = \sigma(Z)$$

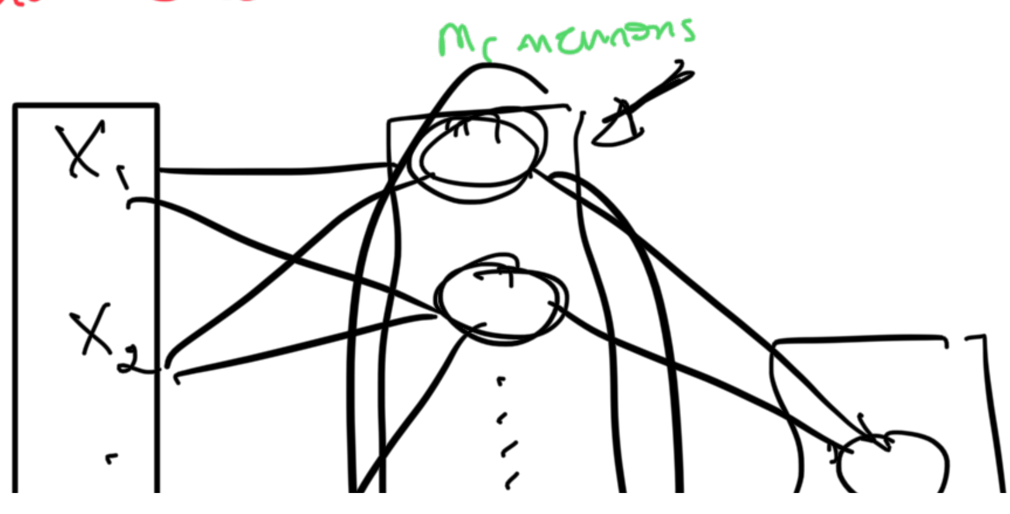
$$\underset{(1,3)}{W^{[2]}} = (w_{14}, w_{24}, w_{34})$$

$$\underset{(1,1)}{b^{[2]}} = b_1 \qquad (1,\cancel{3})(\cancel{3},1) \to (1,1)$$

$$\underset{(1,1)}{Z_4} = W^{[2]} A + b^{[2]}$$

$$\underset{(1,1)}{A_4} = \sigma(Z_4)$$

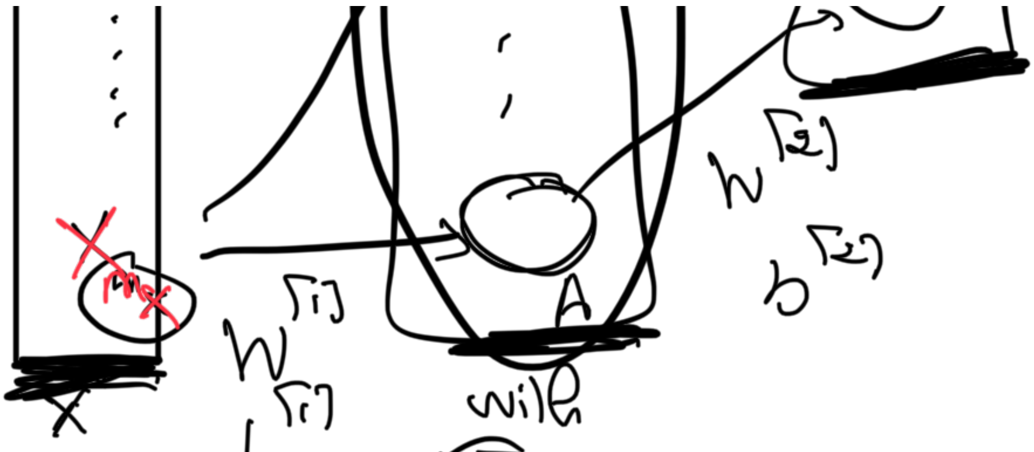
* General case




$n_x$ size of

features

$n_1$ nbr of neurons in hidden layer




# features $n_x$ $\quad$ $n_1$ neurons

$$X_{(n_x,1)} \qquad W^{[1]}_{(n_1,n_x)} \qquad Z^{[1]} = W^{[1]}X + b$$

$$(n_1 \times n_x)(n_x,1) + (n_1,1)$$

$$b^{[1]}_{(n_1,1)} \qquad \boxed{Z^{[1]}_{\sigma}(n_1,1)}$$

$$W^{[2]}_{(1,n_1)} \qquad b^{[2]}_{(1,1)} \qquad A^{[1]}_{(n_1,1)}$$

$$Z^{[2]}_{(1,1)} = \underline{W^{[2]}A^{[1]} + b^{[2]}}$$

$$Y = A^2 = \sigma(W^{[2]}\underline{A^{[1]}} + b^{[2]})$$

$$= \sigma(W^{[2]}\,\sigma(W^{[1]}X) + b^{[2]})$$

$$= (\underbrace{W^{[2]}W^{[1]}}\,X$$

$$WX$$

$$\frac{e^x}{e^x} \cdot \frac{1}{1+e^{-x}} = \sigma(x)$$

$$\frac{f}{g} = \frac{f'g - fg'}{g^2}$$

$$\sigma(x) = \frac{e^x}{e^x+1} \leftarrow \frac{f}{g}$$

$$\frac{f'g - g'f}{g^2}$$

$$\frac{e^x(e^x+1) - e^x(e^x)}{(e^x+1)^2}$$

$$\frac{e^x\left[(\cancel{e^x}+1) - \cancel{e^x}\right]}{\left[e^x+1\right]^2}$$

$$\sigma(x)' = \frac{e^x}{\left[e^x+1\right]^2} \cdot \frac{e^{-2x}}{e^{-2x}} = (e^{-x})^2$$

$$= \frac{e^{-x}}{\left[1+e^{-x}\right]^2}$$

$(e^{-x} + 1) - 1$

$$= \frac{\dots}{[1+e^{-x}]^2}$$

$$\sigma'(x) = \frac{1}{1+e^{-x}} - \frac{1}{(1+e^{-x})^2}$$

$$\sigma'(x) = \sigma(x) - \sigma^2(x)$$
$$= \sigma(x)(1 - \sigma(x))$$

$$A(Wx)$$

Why we add an activation function?

$$\Rightarrow \quad Y = \sigma(W^{[2]}\underline{A} + b^{[2]})$$

if $\boxed{\sigma(x) = x}$

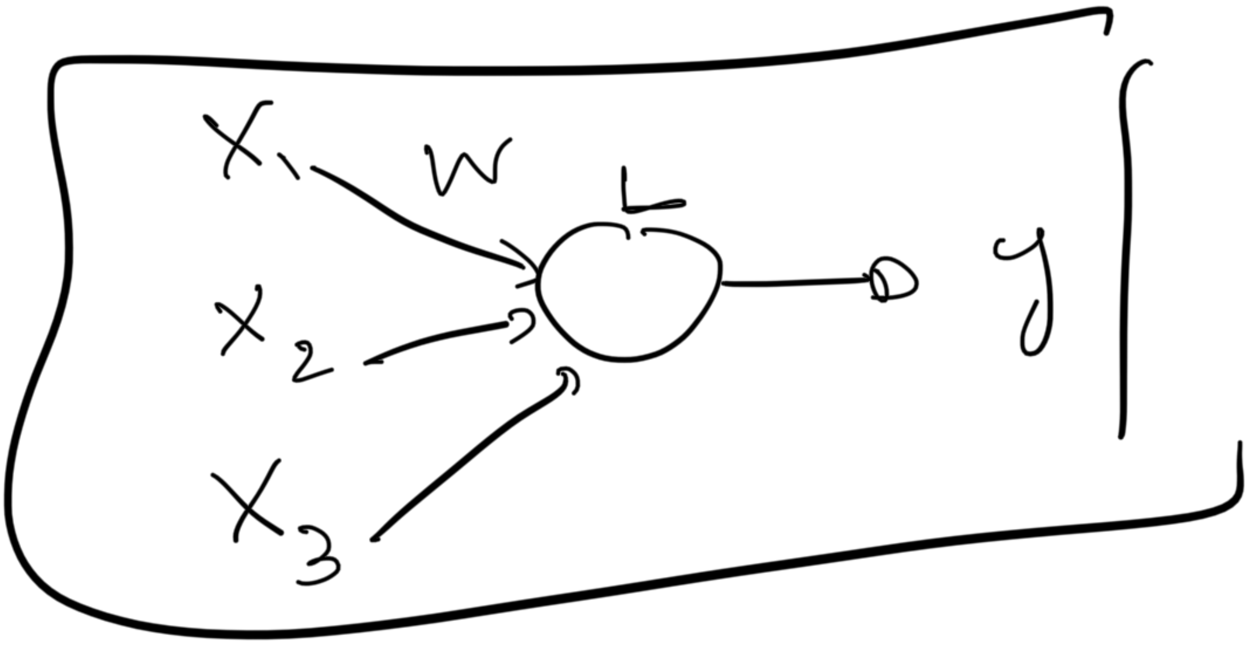
$$= \sigma(W^{[2]}\sigma(W^{[1]}X + b_1) + b_2)$$

$$= W^{[2]}W^{[1]}X + W^{[2]}b_1 + b_2$$

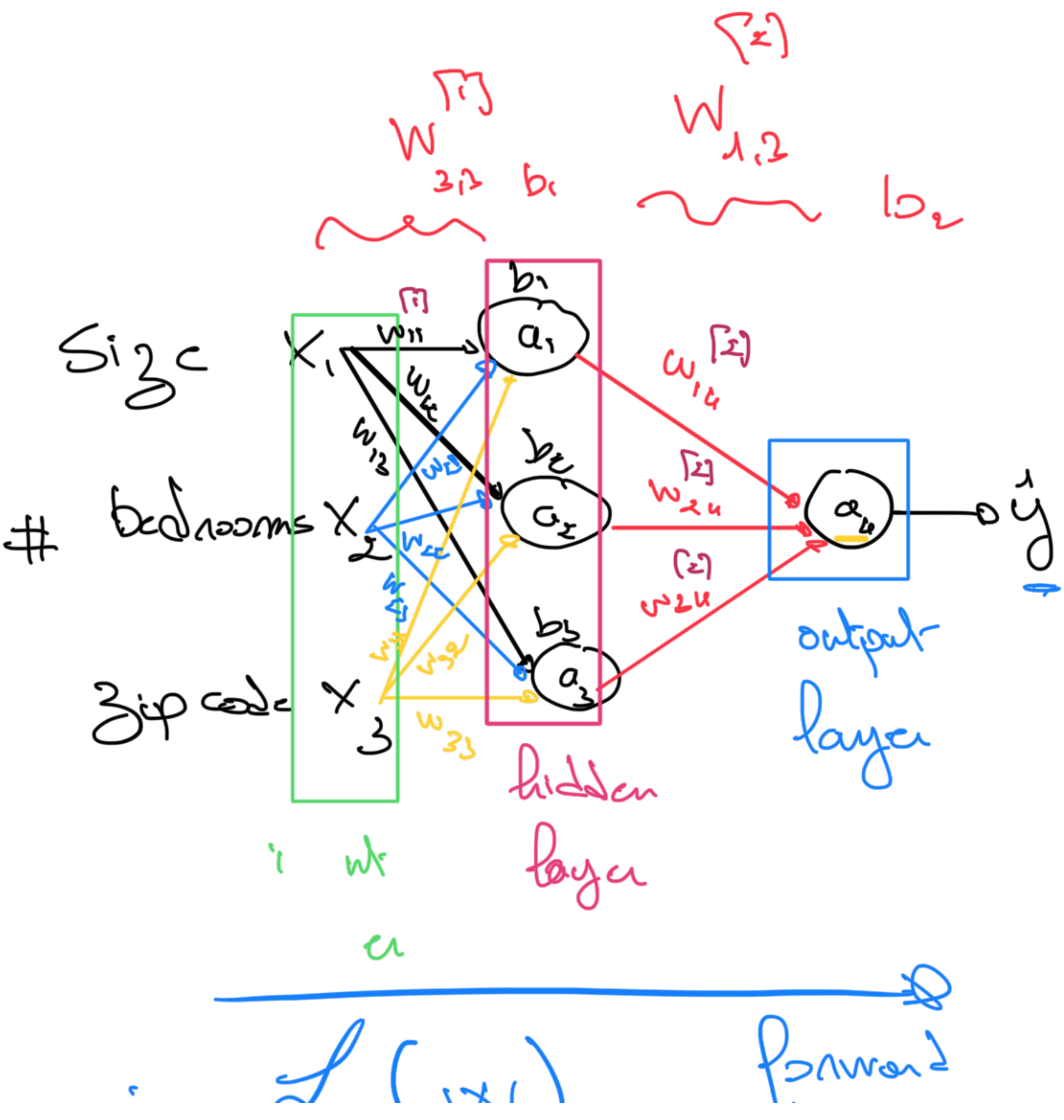
$$= \underline{WX + B}$$

equivalent!

eq




$$y = WX + b$$



W [1] 3,2
b1
W [2] 1,2
b2
Size
# bedrooms
zip code
x1
x2
x3
w11
w13
w33
a1
a2
a3
a4
b1
b2
b3
w [2] 14
w [2] 24
ŷ
hidden layer
output layer
forward

$\min_{w} \mathcal{L}(w)$ pass

backward pass

$\frac{\partial \mathcal{L}}{\partial w^{[1]}}$ $\frac{\partial \mathcal{L}}{\partial w^{[2]}}$ $\frac{\partial \mathcal{L}}{\partial b^{[1]}}$ $\frac{\partial \mathcal{L}}{\partial b^{[3]}}$

learning rate

$$W^{[1]} \leftarrow W^{[1]} - \eta \frac{\partial \mathcal{L}}{\partial W^{[1]}}$$

$b^{[1]}$

$W^{[2]}$

$b^{[2]}$