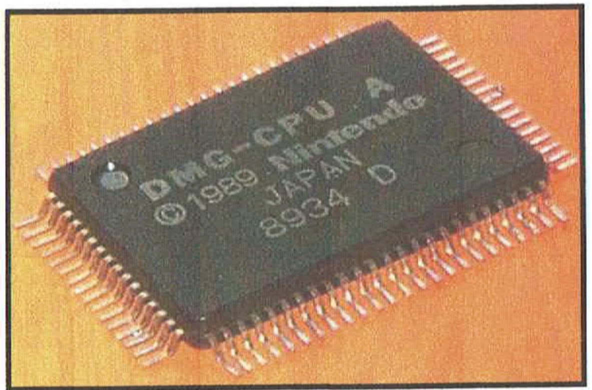
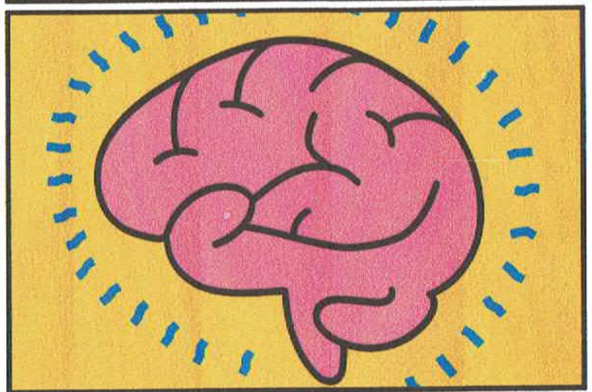
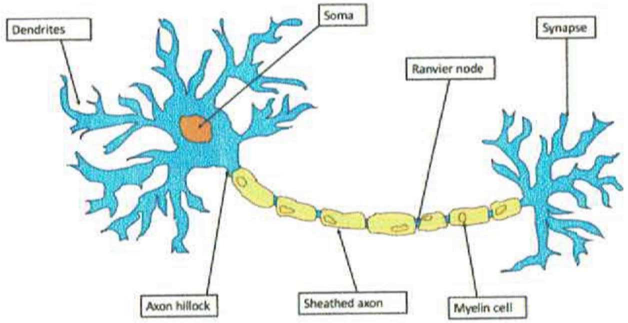
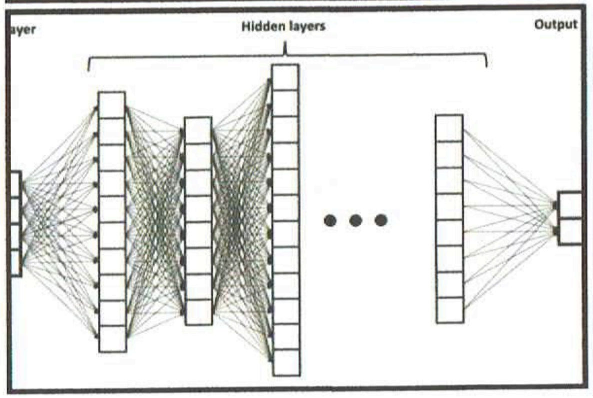
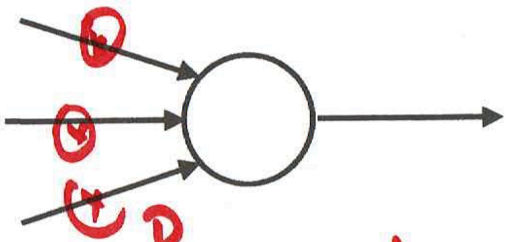
Building towards a complex task!











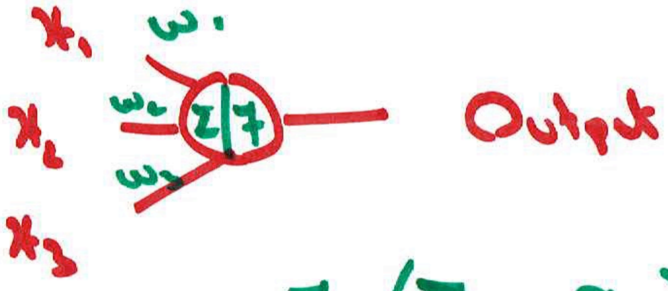
Perceptron

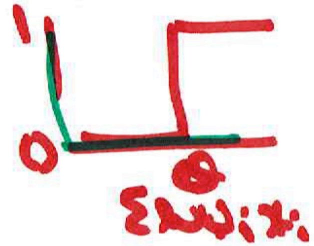
Perceptron!

$\cancel{f(x) = x^2}$

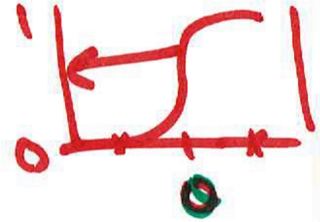
- What does it do?

1958

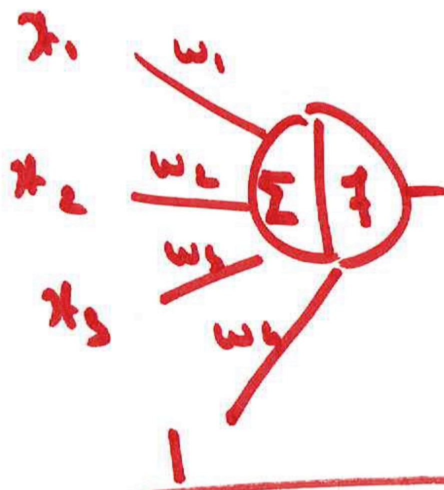




$f(\sum w_i x_i) = \text{Output}$

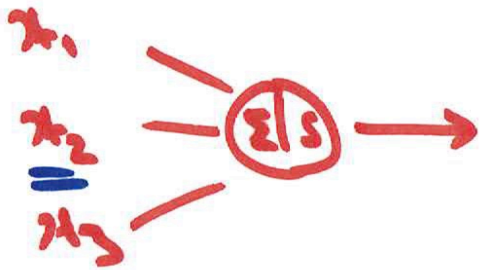


- Bias



$f\left(\sum w_i x_i + w_4\right)$

Perceptron

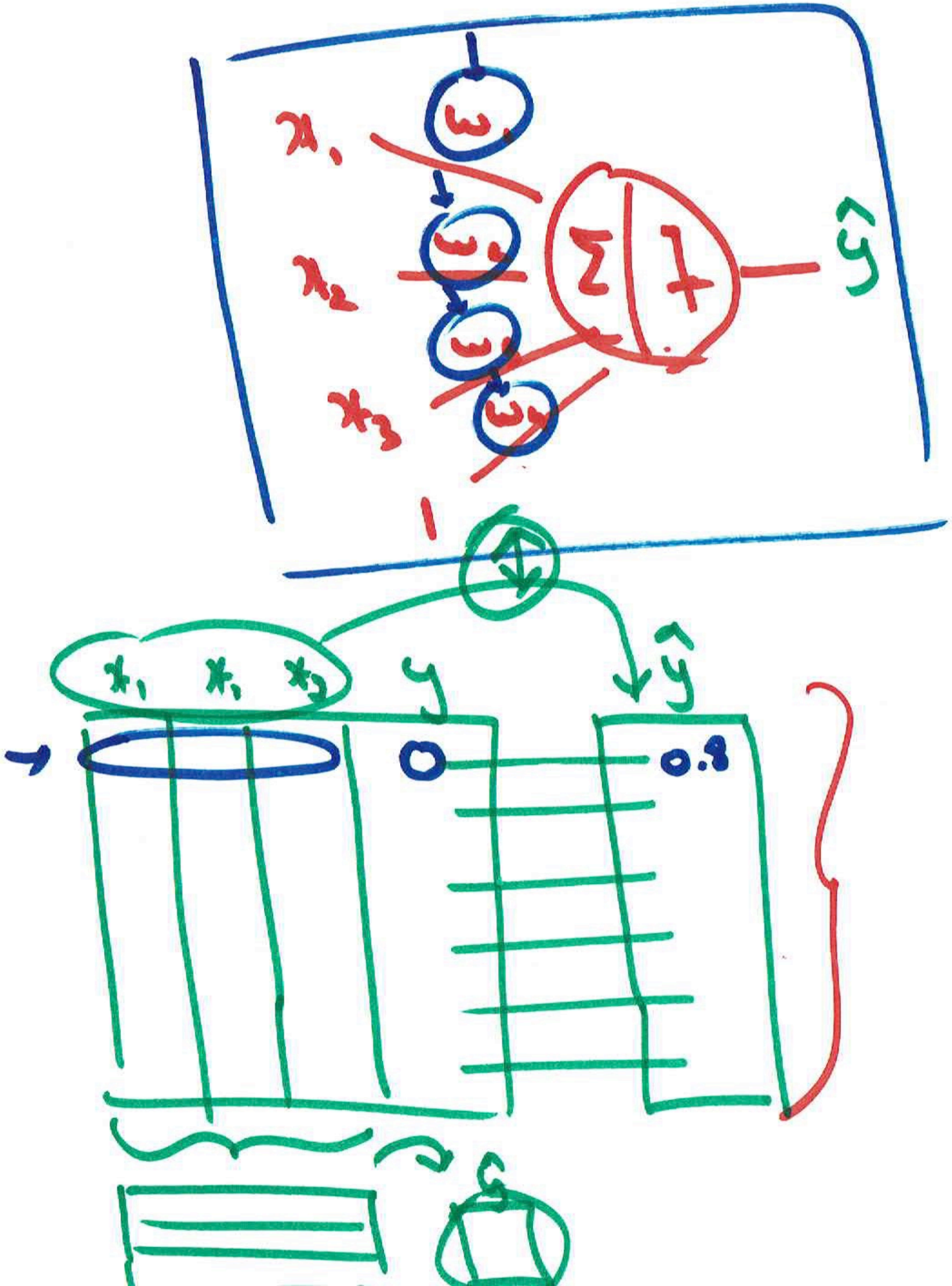


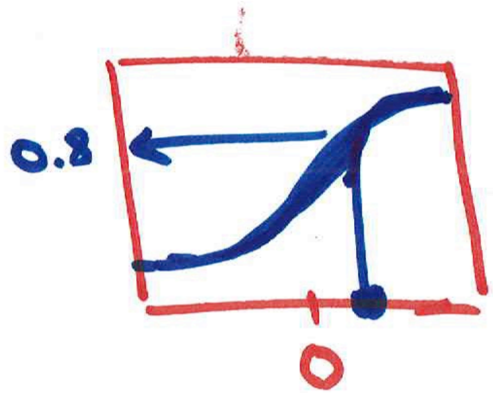
if $\sum x_i \geq \theta \Rightarrow 1$

if $\sum x_i < \theta \Rightarrow 0$

$\theta = 1 \Rightarrow$ "OR"

$\theta = 3 \Rightarrow$ "AND"

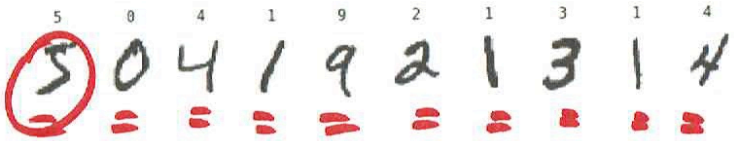




$$f\left(\underline{\sum w_i x_i + w_4}\right)$$

An Example

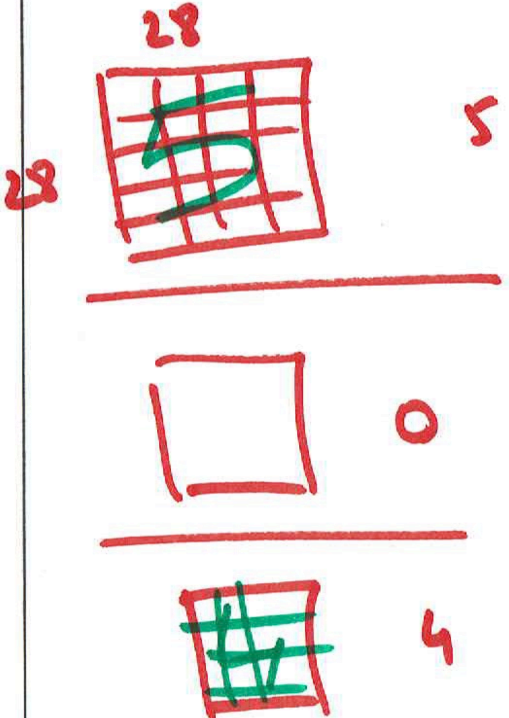
MNIST

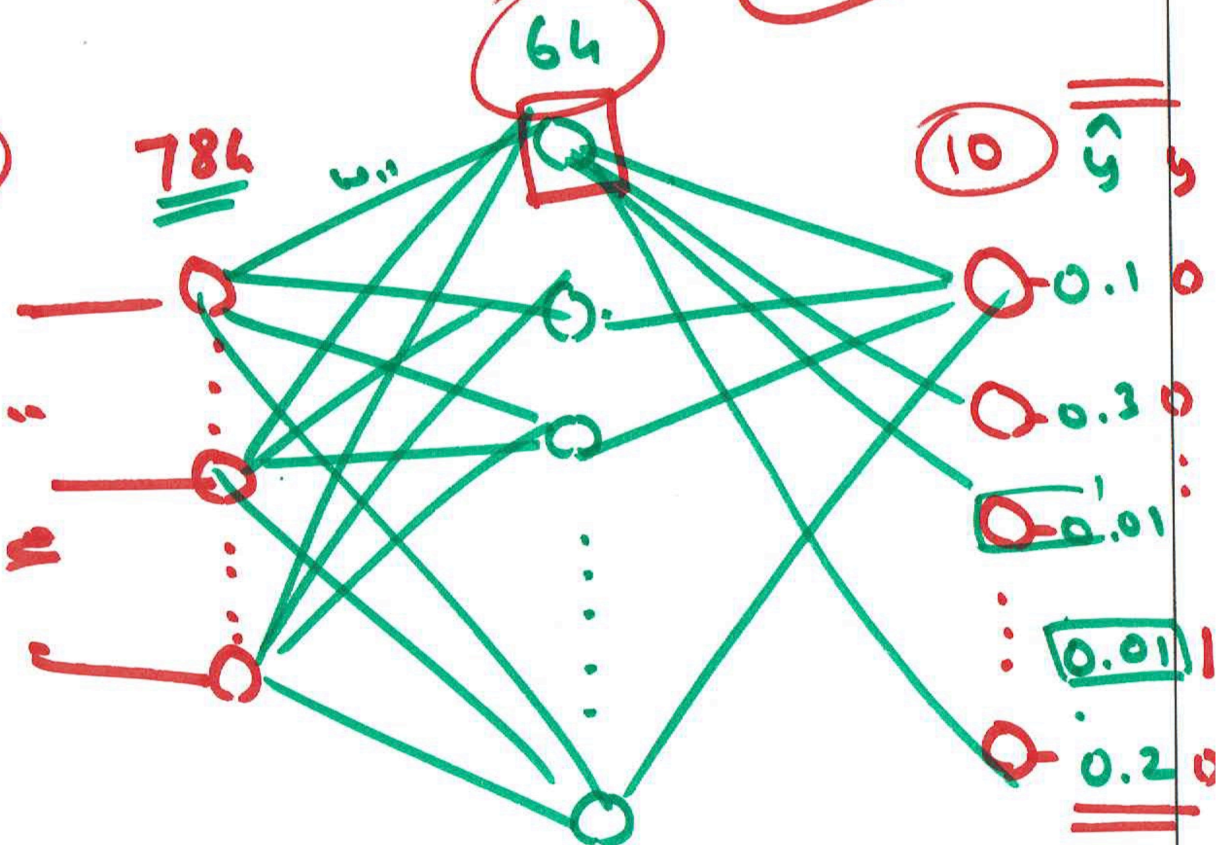


89%

60000 train

~~10000~~

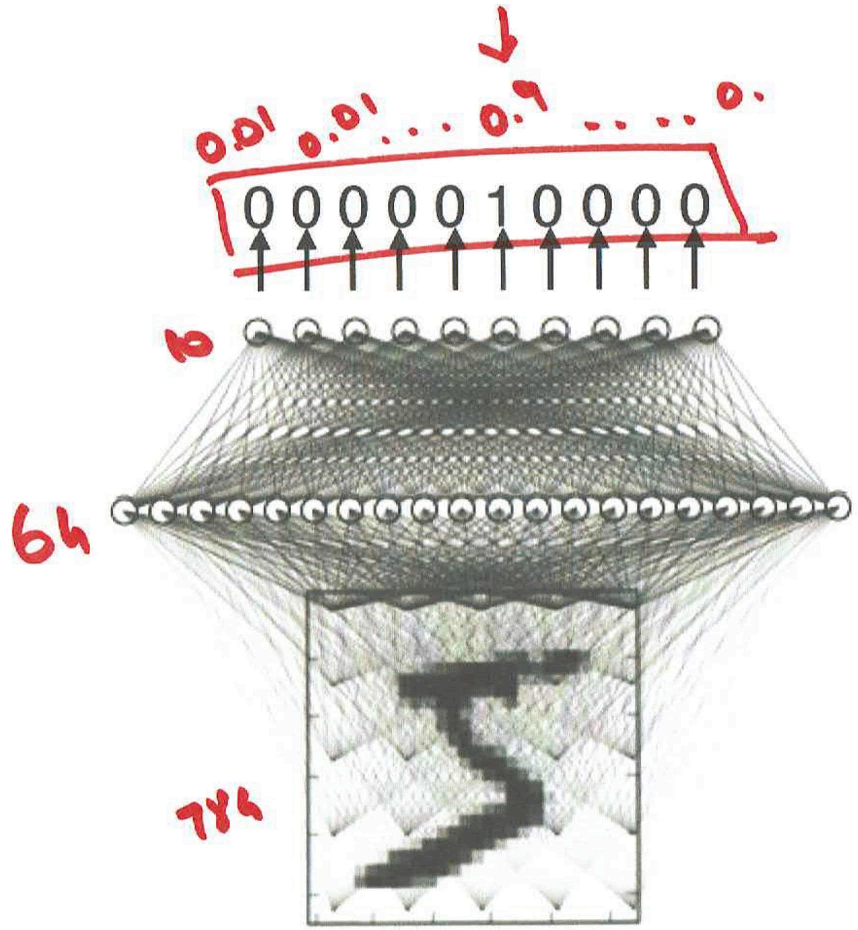


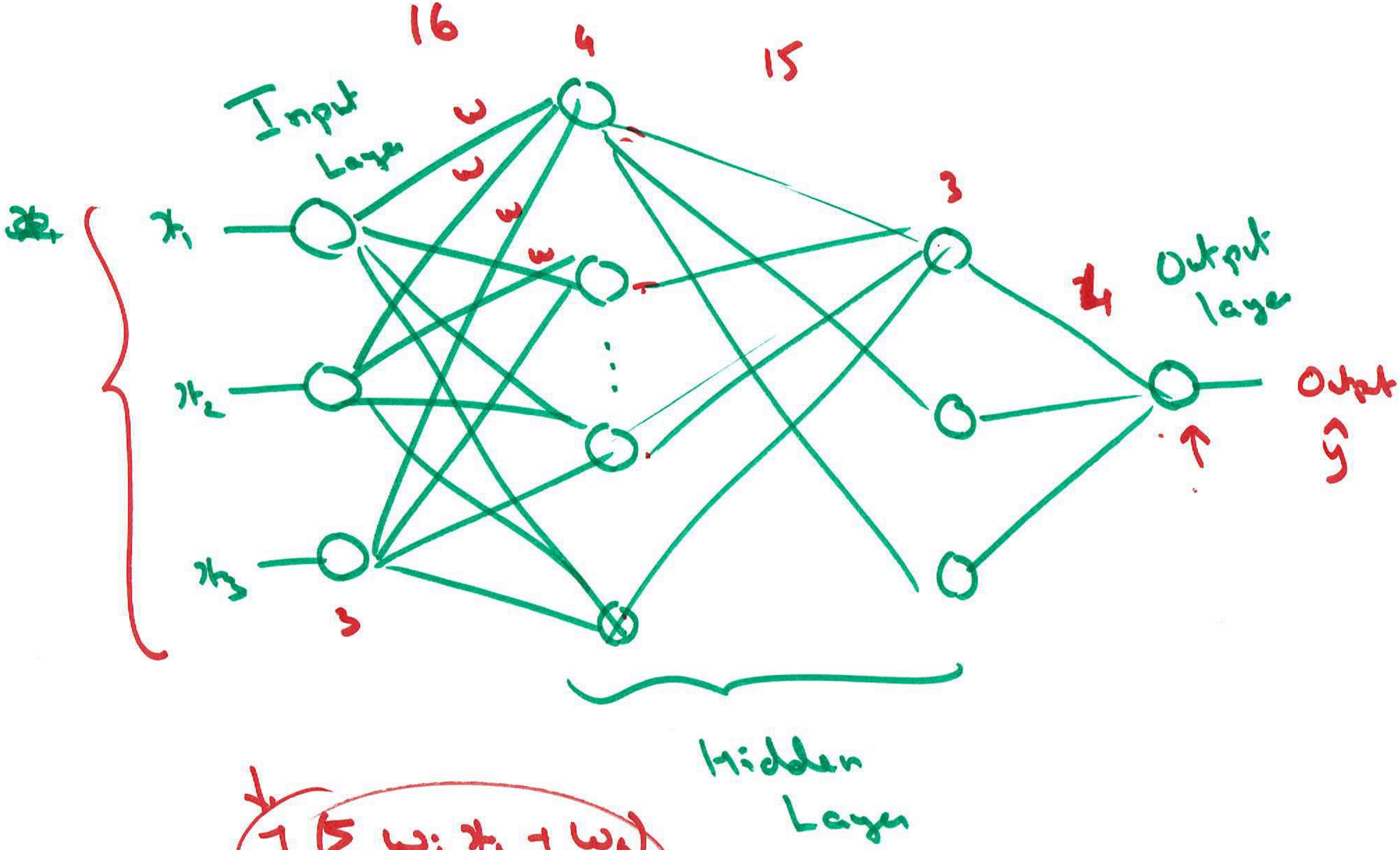


$(784 \times 64 + 64) + (64 \times 10 + 10)$

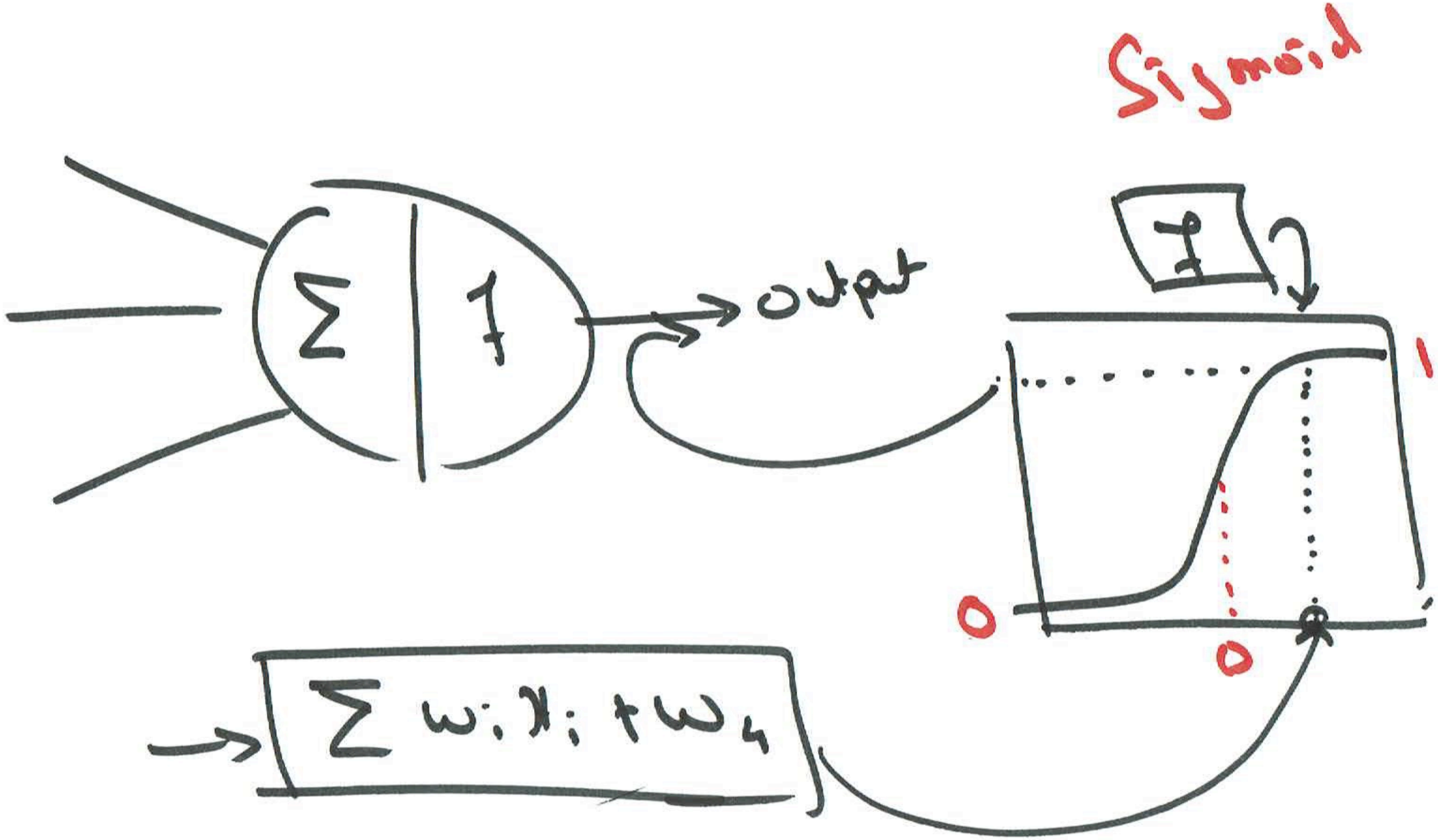
$= \boxed{50890}$

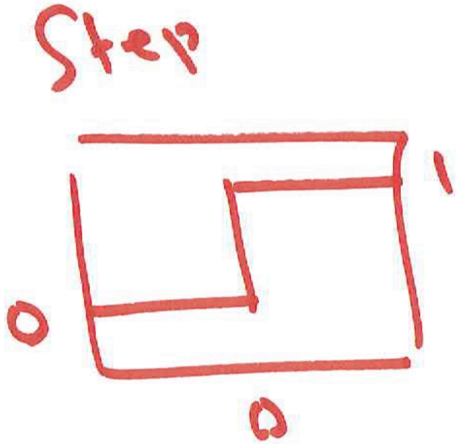
An Example





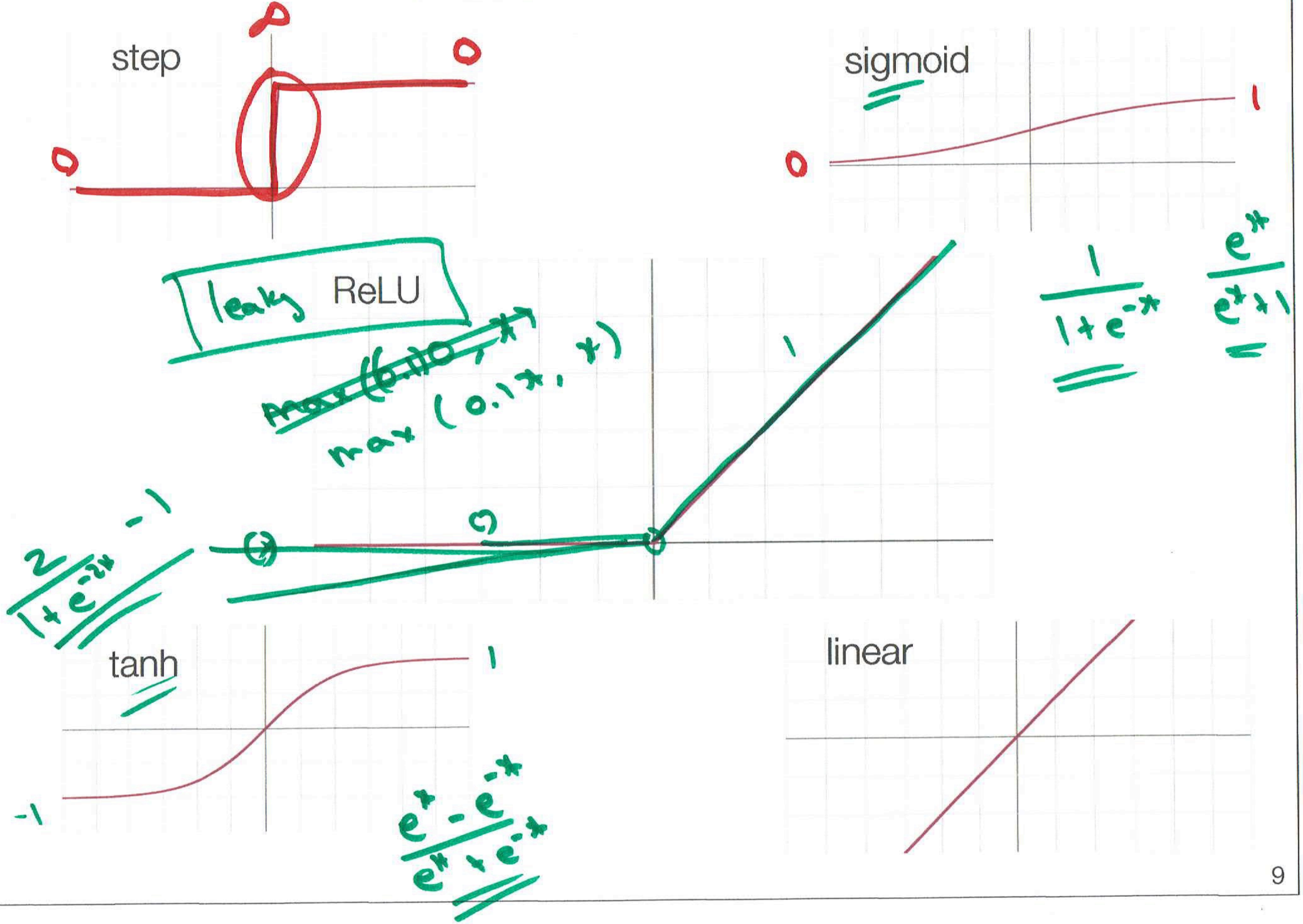
$$f\left(\sum w_i x_i + w_0\right)$$



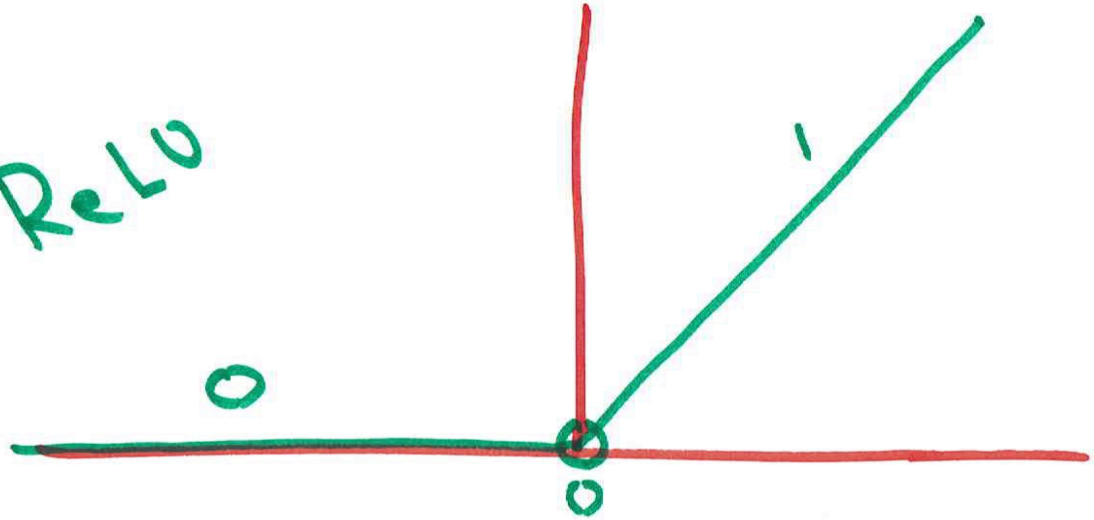


$$\text{Output} = f\left(\sum w_i x_i + w_n\right)$$

Activation Functions

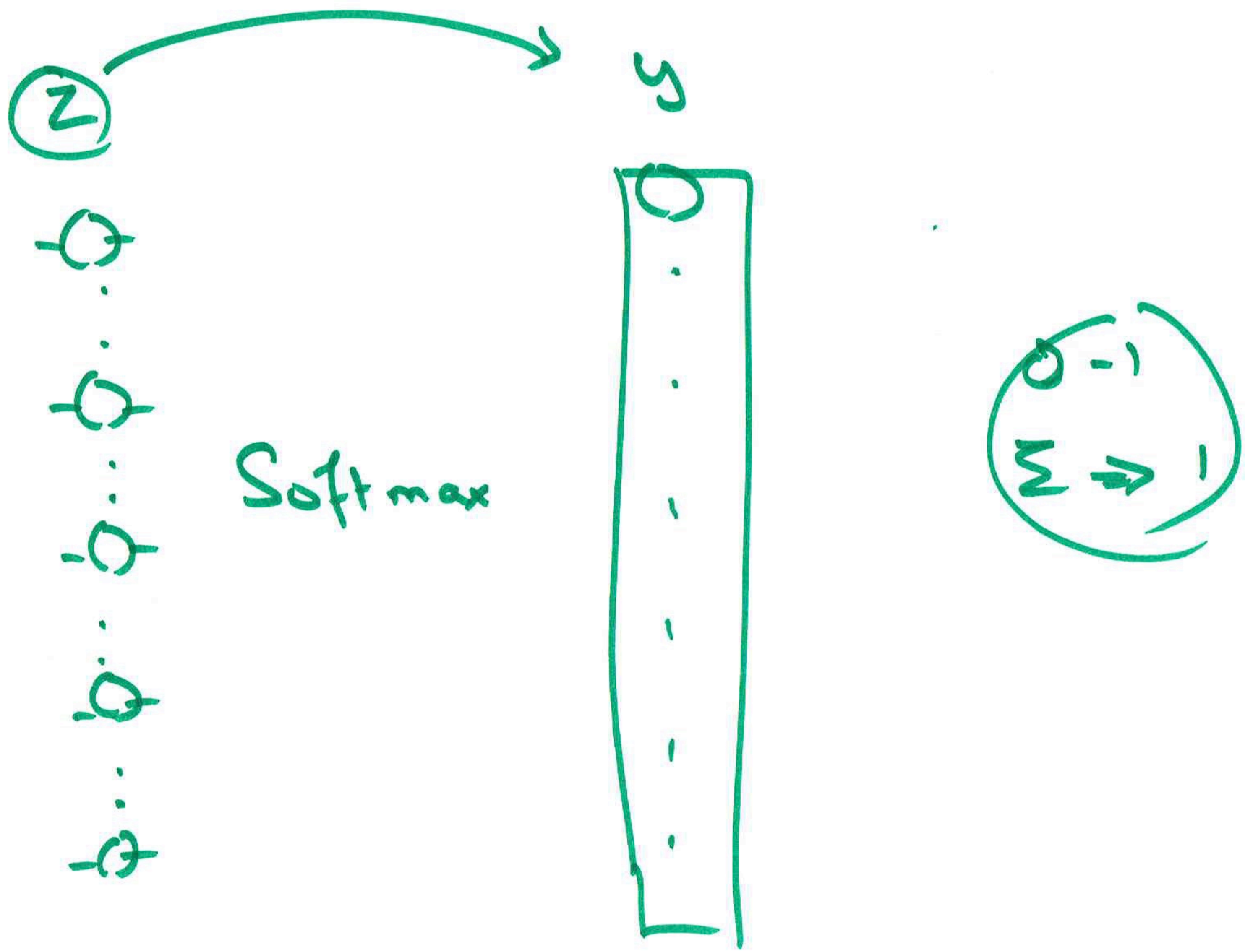


ReLU



$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$f(x) = \max(0, x)$$


z
y
Softmax
0-1
Σ ⇒ 1

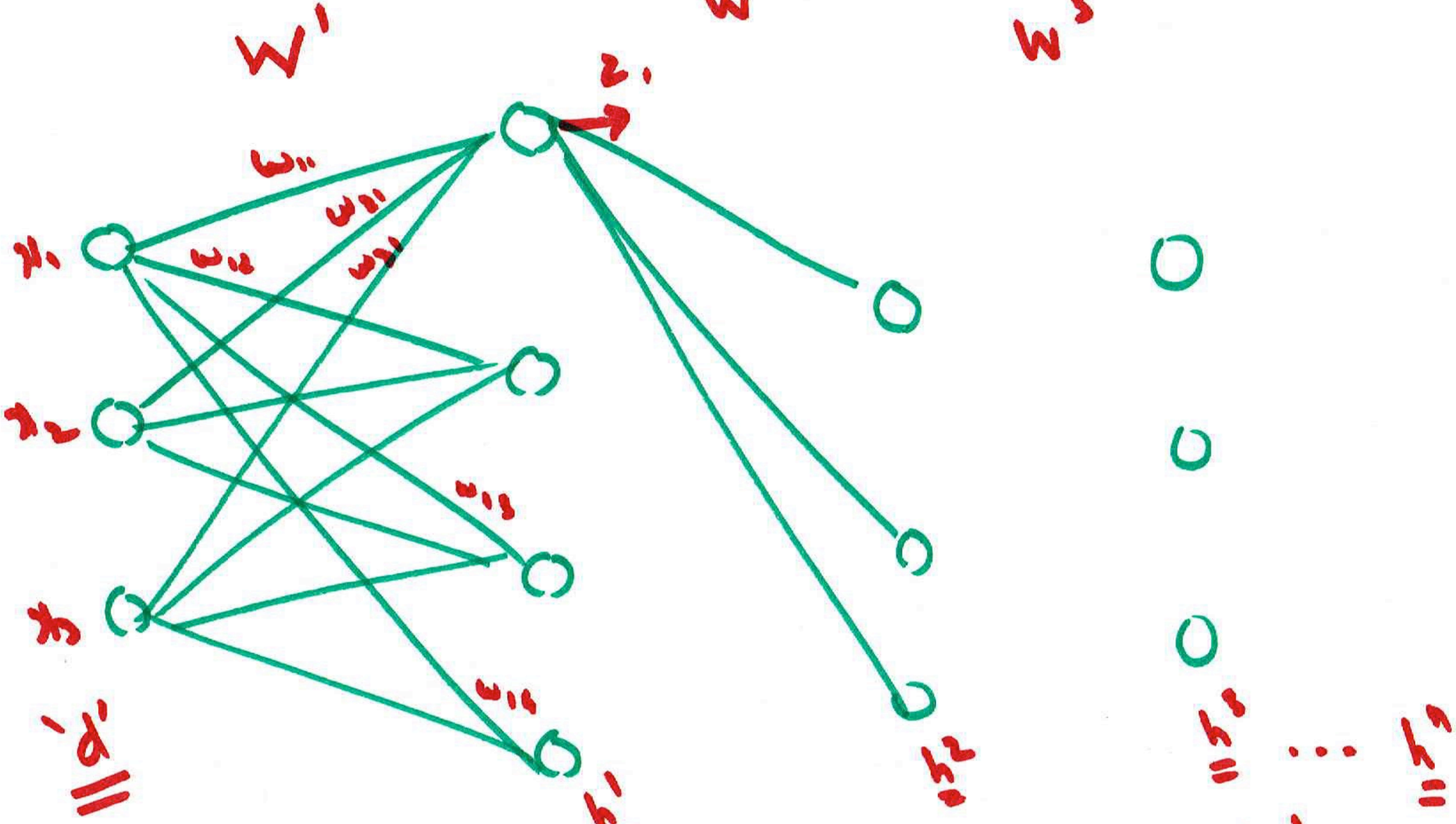
Z

y

$$\boxed{y_i = \frac{e^{z_i}}{\sum e^{z_i}}}$$

$$w^{new} = w^{old} - \eta \nabla_w l(w)$$

$$= w^{old} - \frac{1}{N} \eta \sum \nabla_w l_i(w)$$



$$z_1 = f(w_{11}x_1 + w_{21}x_2 + w_{31}x_3 + b)$$

$$\boxed{z_j = f\left(\sum_i w_{ij}x_i + b_j\right)}$$

$$W^1 = \begin{pmatrix} w_{11} & w_{21} & w_{31} & \cdots & w_{d1} \\ w_{12} & w_{22} & w_{32} & & \\ \vdots & & & & \\ w_{1j} & w_{2j} & w_{3j} & \cdots & w_{dj} \\ \\ w_{1h'} & \cdots & \cdots & \cdots & w_{dh'} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

$h' \times d$ matrix

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

$$\sum w_{ij} x_i + b$$

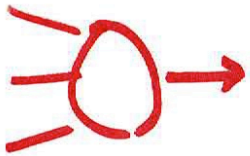
$$\hat{y} = f^n(\cdots f^3(W^3 f^2(W^2 f^1(W^1 X + b^1) + b^2) + b^3)) \cdots$$

$$\underbrace{\underbrace{h' \times d \quad d \times 1}_{h' \times 1} \quad h' \times 1}_{h' \times 1}$$

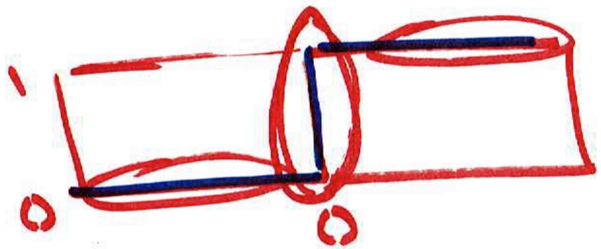
$$\underbrace{h^2 \times h' \quad h^1 \times 1}_{h^2 \times 1}$$

$h^2 \times 1$

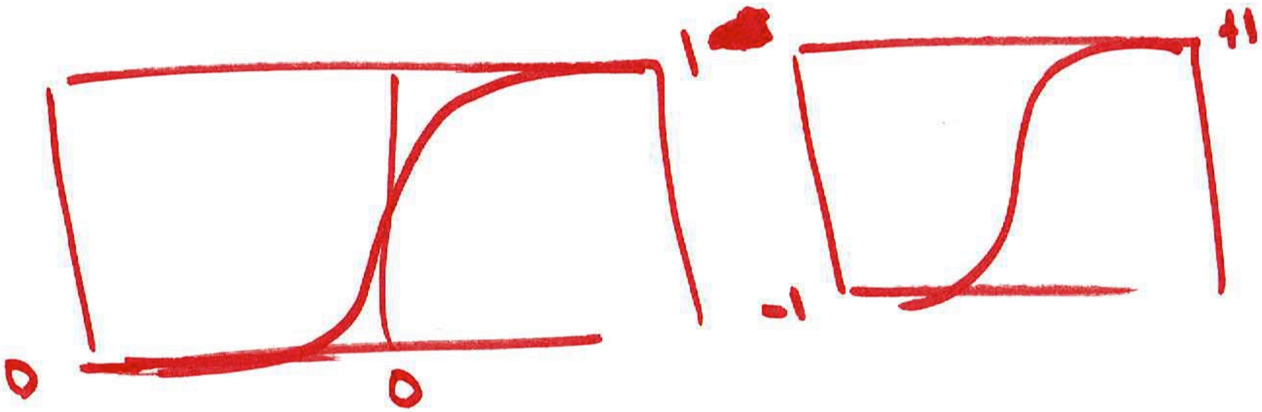
$\underline{\underline{h^n \times 1}}$



$\underline{\text{Step}(w_1x_1 + w_2x_2 + w_3x_3 + b)}$



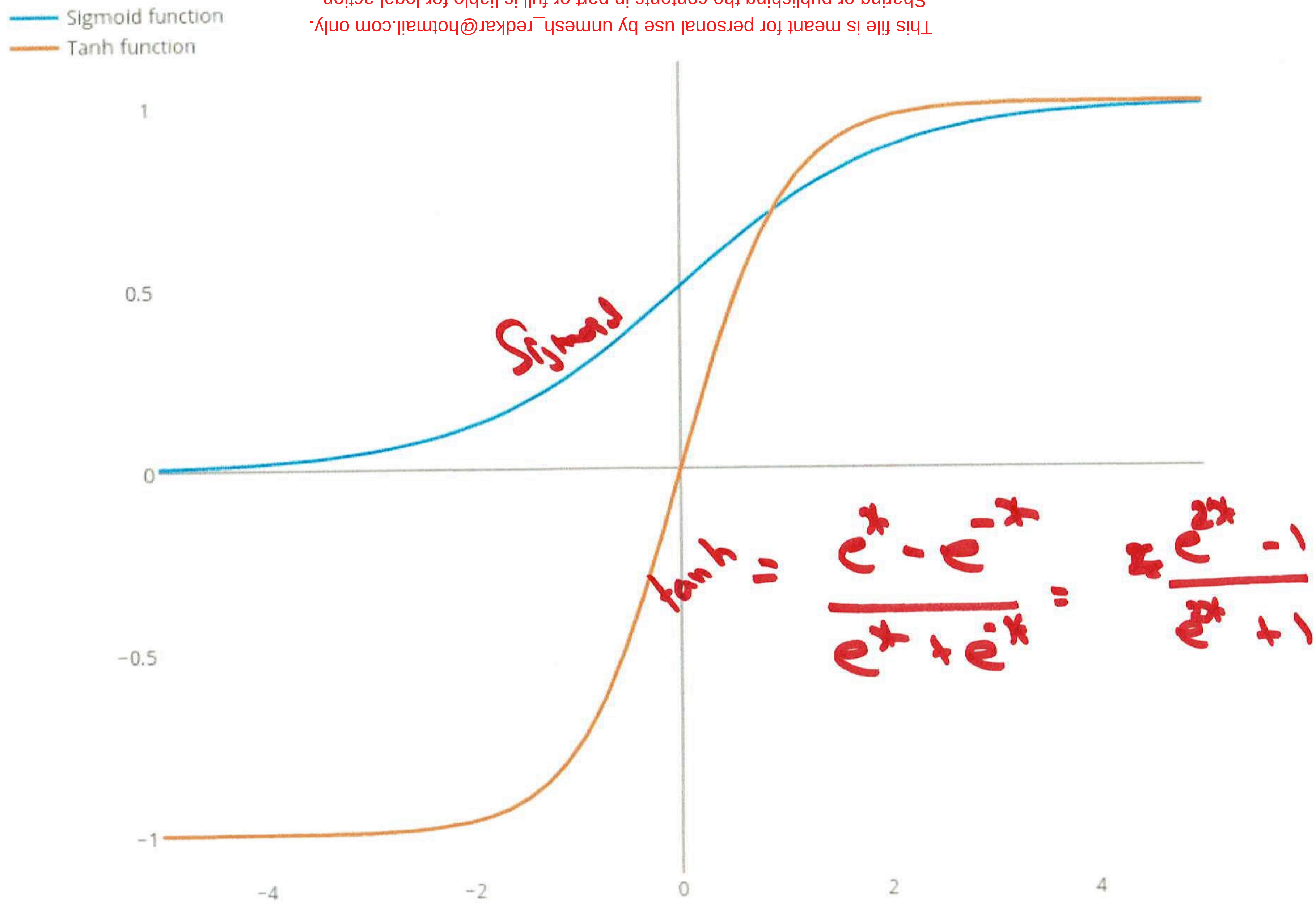
Sigmoid



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

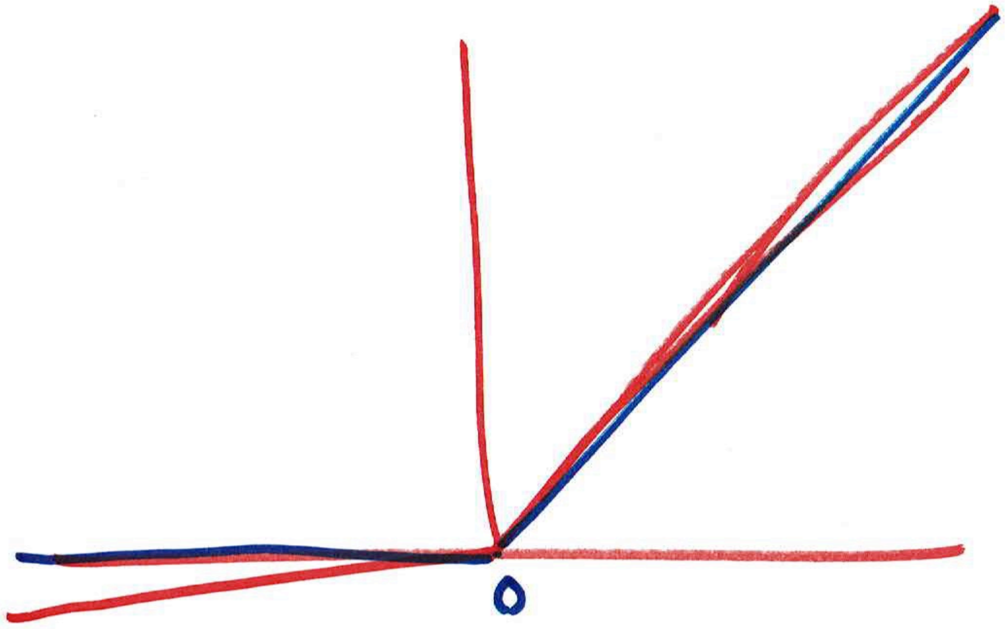
$$\boxed{\cancel{2\sigma(x)} - 1}$$

$$\boxed{\tanh = \underbrace{2\sigma(2x)} - 1}$$



$$\tanh = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}$$

ReLU



$$= \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$= \max(0, x)$$

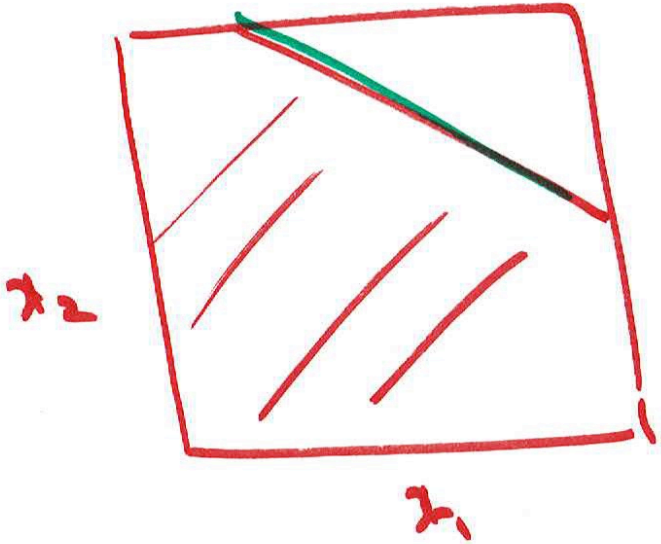
$$\max(0.01x, x)$$

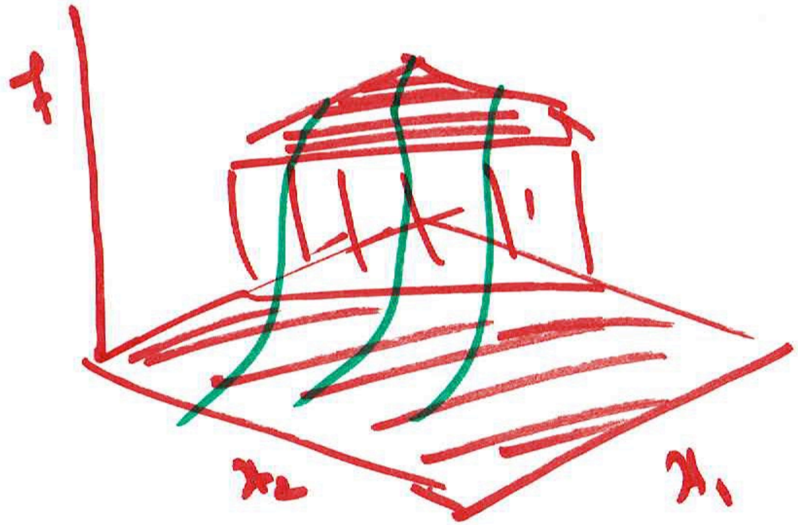
linear

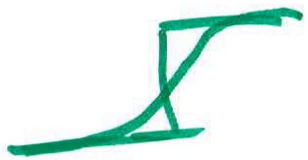
= *

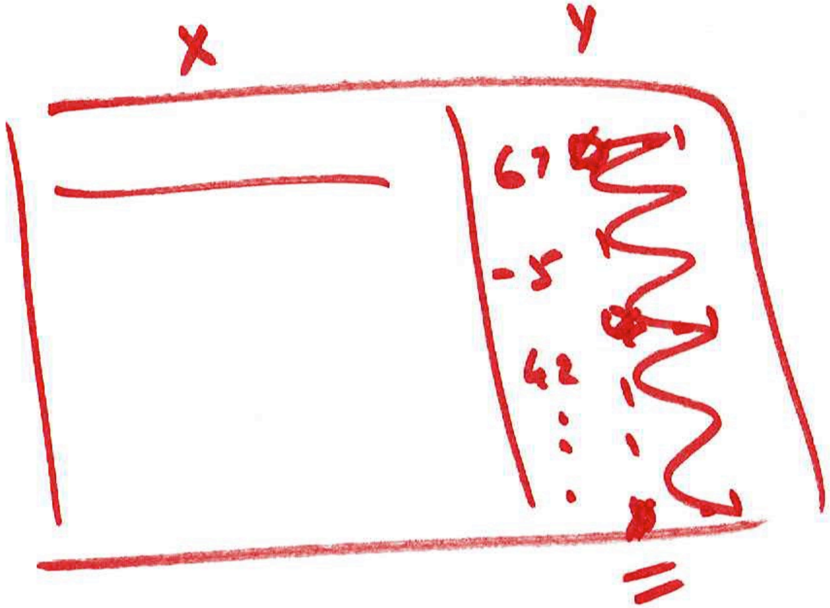
$$\text{Step} = f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

$$f(w_1 x_1 + w_2 x_2 + b)$$



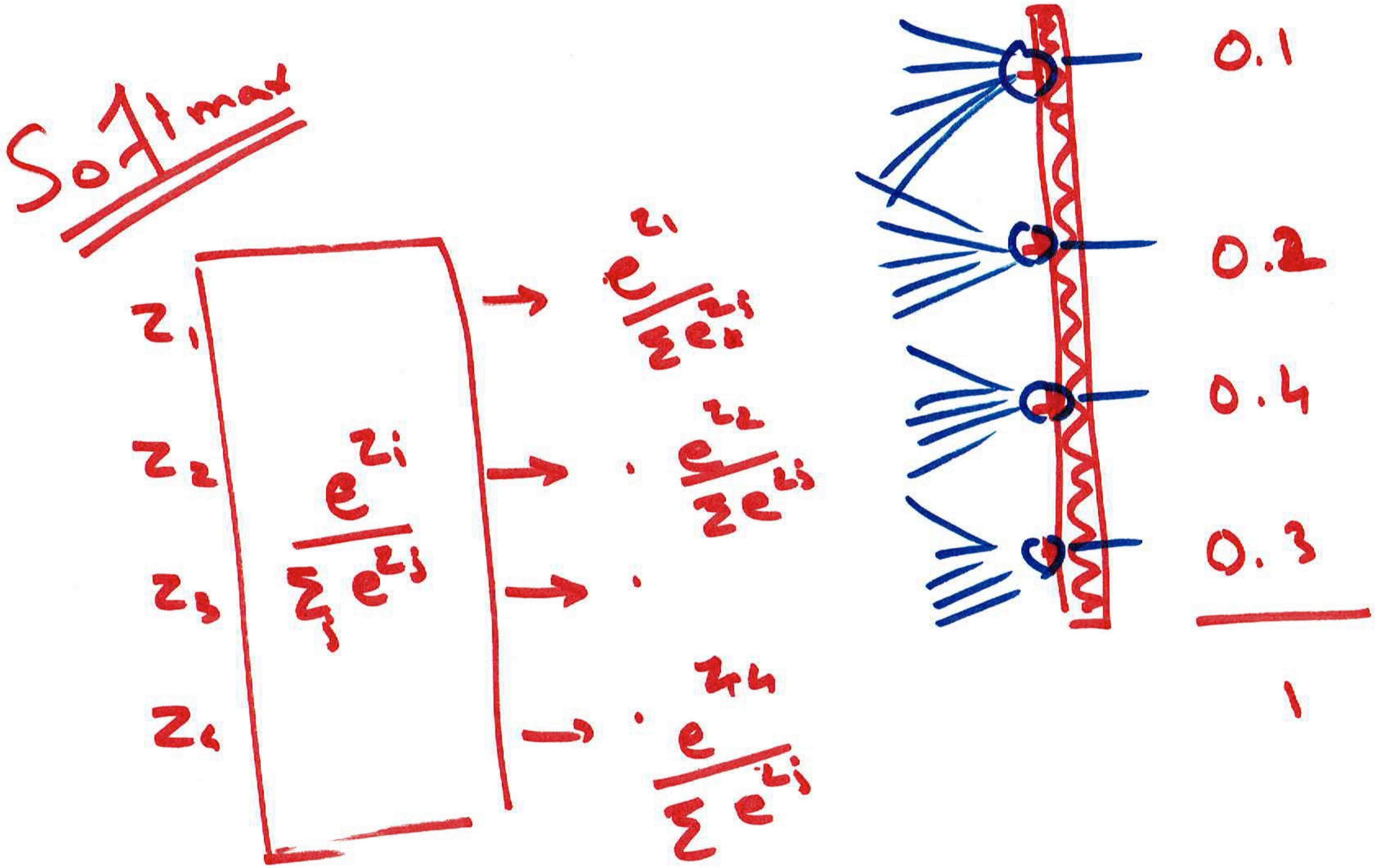


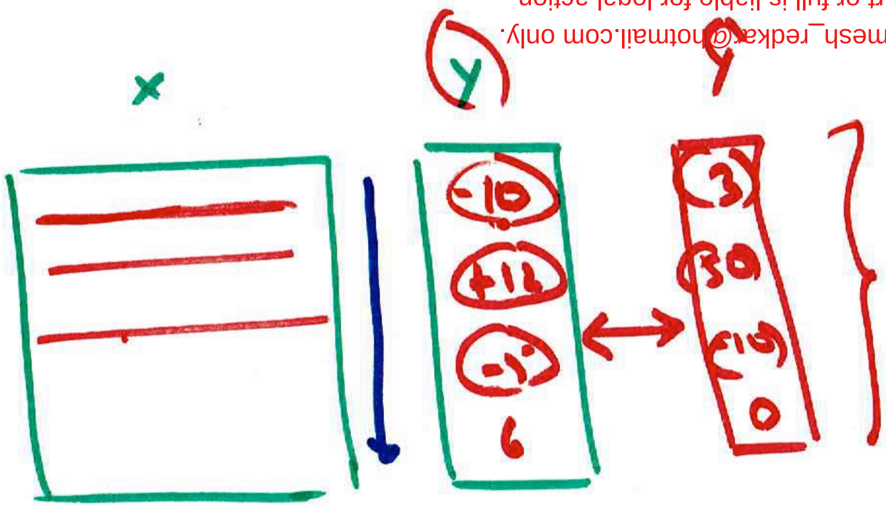


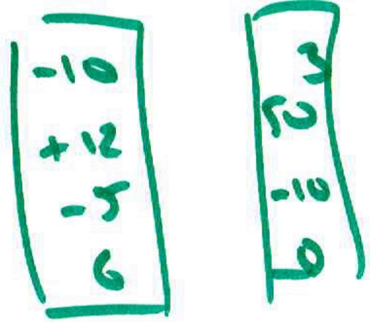


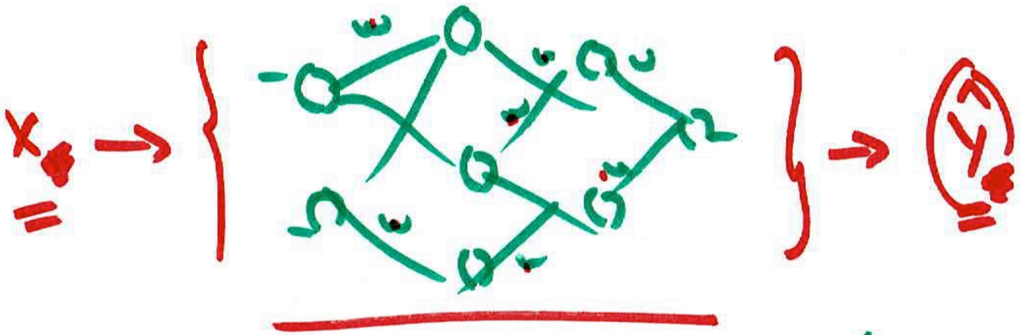
	Output nodes	hidden layer
Classification →	Sigmoid, tanh, Softmax	Sigmoid ✓ tanh ✓ $\boxed{\text{ReLU ✓}}$ ~~linear~~
Reg →	linear	

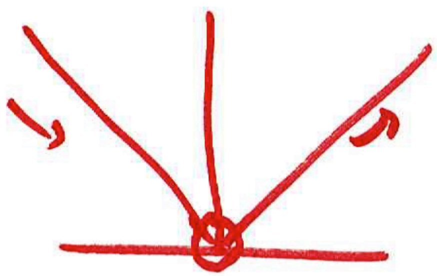
$$\boxed{\hat{a} + \hat{b}\,(a + bx)}$$


Softmax
z_1
z_2
z_3
z_4
$\frac{e^{z_i}}{\sum_j e^{z_j}}$
$\frac{e^{z_1}}{\sum e^{z_j}}$
$\frac{e^{z_2}}{\sum e^{z_j}}$
$\frac{e^{z_4}}{\sum e^{z_j}}$
0.1
0.2
0.4
0.3
1









Loss Function $L(y, \hat{y})$

$$\boxed{\frac{1}{N}\sum_i \left(y_i - \hat{y}_i\right)^2}$$

L_2 loss
MSE
SSE

Reg

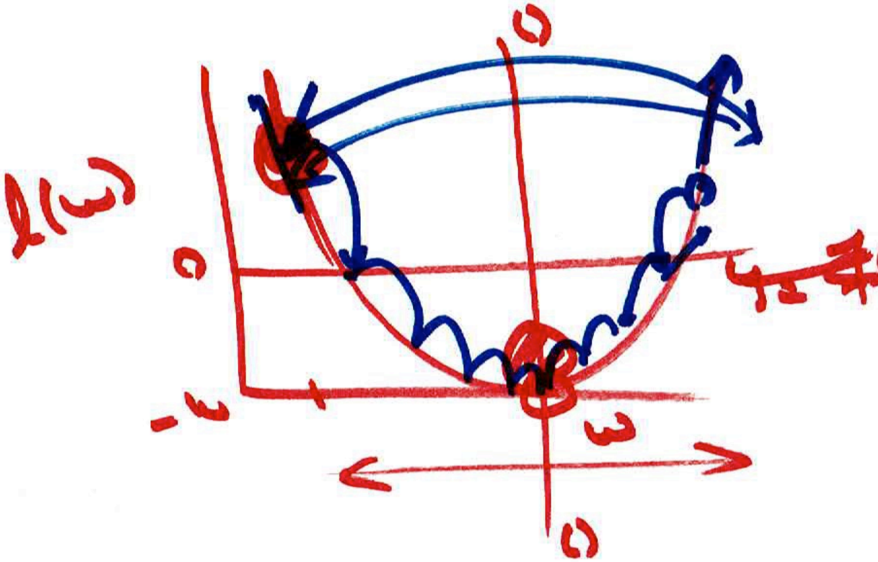
$$L(y, \hat{y}) = l(w)$$

Classification $L(y, \hat{y}) = -\left(y_i \log(\hat{y}_i) + (1-y_i)(\log(1-\hat{y}_i))\right)$

Cross entropy loss

How min $\underline{L(y, \hat{y})}$ by ~~chag~~ Changing ~~to~~ my $\underbrace{W^1, W^2, \dots W^n}$

$$\begin{Bmatrix} w_{11} & w_{12} & \dots \\ \dots & & \\ \dots & & \end{Bmatrix}$$



$$y_2 \neq y = x^2 - 10 = -10$$

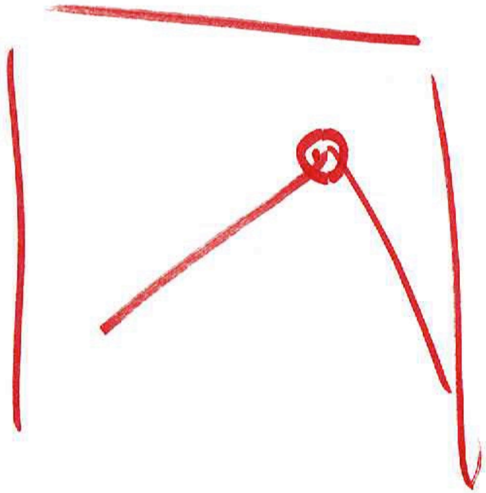
$$\frac{dy}{dx} = 2x = 0 \qquad x = 0$$

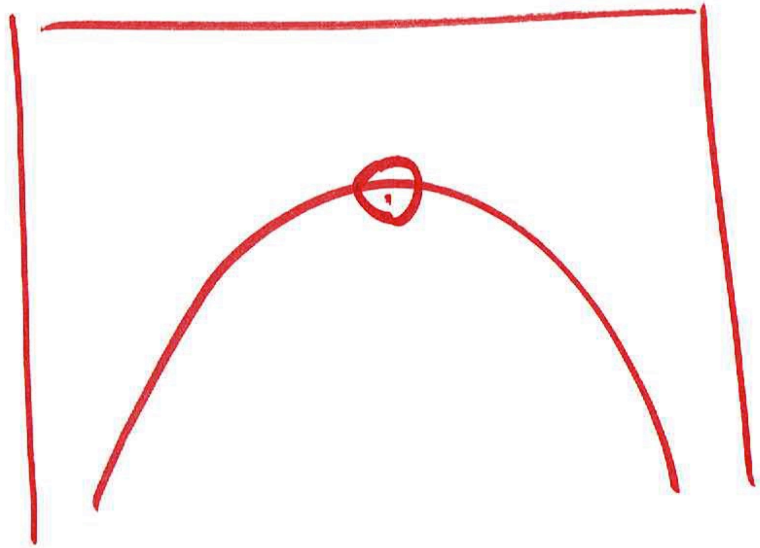
$$\frac{dL}{dw} = \underbrace{\boxed{\quad}}_{} = 0$$

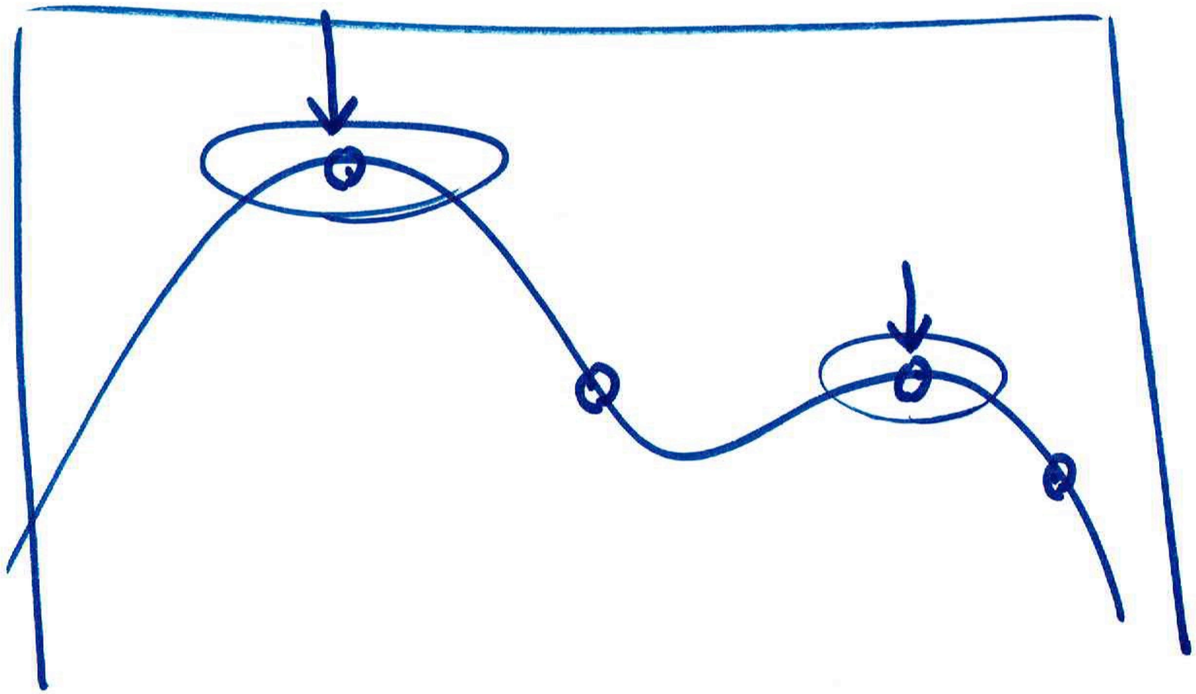
$$\frac{dl}{dw}$$

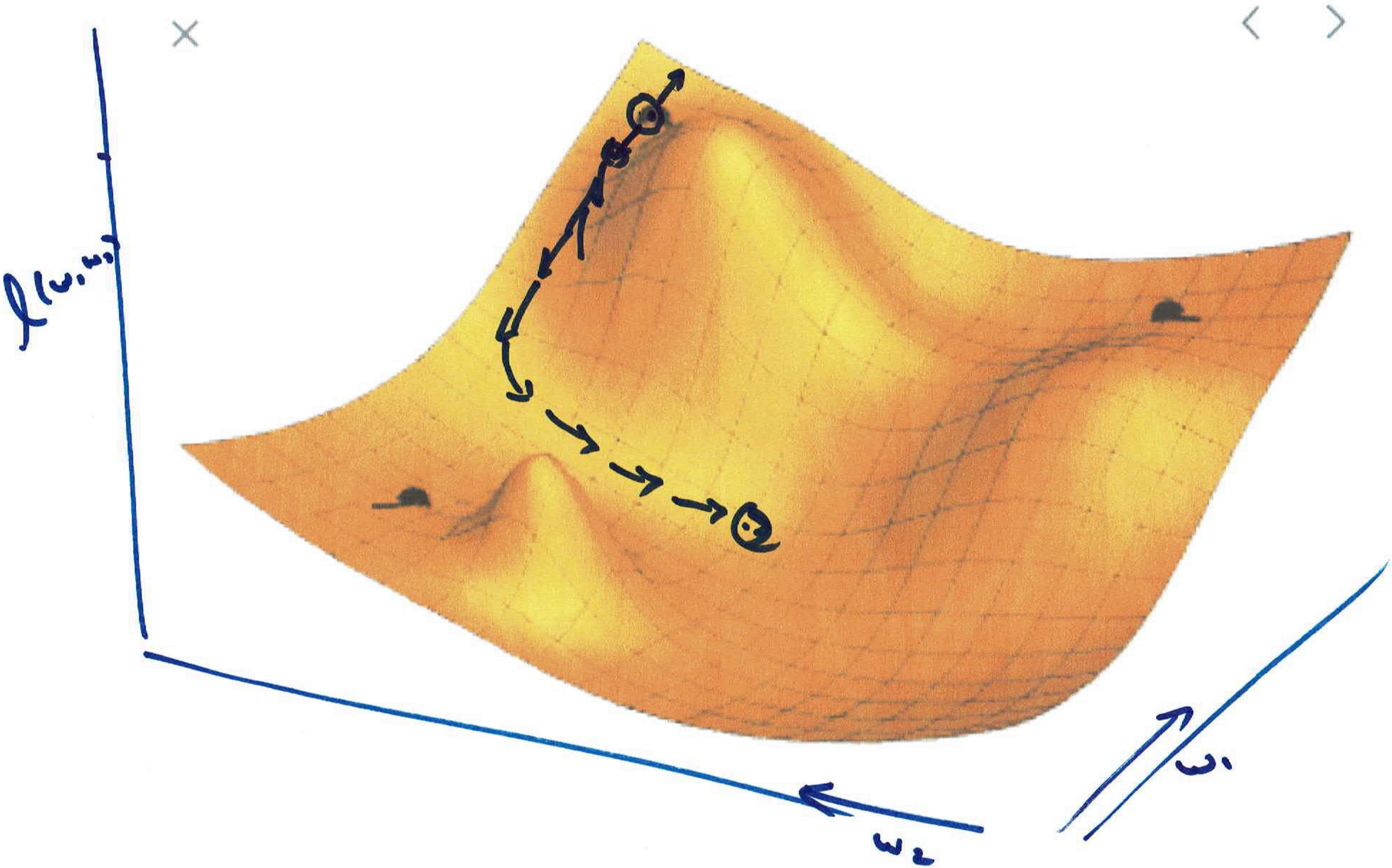
$$w^{new} = w - \eta \nabla_w l$$

learning rate








$l(w_1, w_2)$
w_1
w_2

Batch

$$w^{new} = w^{old} - \eta \nabla_w l(w)$$

$$= w^{old} - \frac{1}{N} \eta \sum \nabla_w l_i(w)$$

SGD

$$w^{new} = w^{old} - \eta \nabla_w l_i(w)$$

$$\boxed{w^{new} = w^{old} - \frac{1}{N} \eta \sum_{\substack{\text{over} \\ \text{a min} \\ \text{batch}}} \nabla_w l_i(w)}$$

function of w $(l(w))$

Loss

$$L = \frac{1}{N} \sum \left(y_i - f^n\left(\ldots f\left(w^2 f^1\left(w^1 x + b^1\right) + b^2\right)\right)_i \right)^2$$

Chain Rule

$$f(g(h(.x.)))$$

$$\frac{df}{dx} = \boxed{\frac{df}{dg} \cdot \frac{dg}{dh}} \cdot \frac{dh}{dx}$$

Back propagation