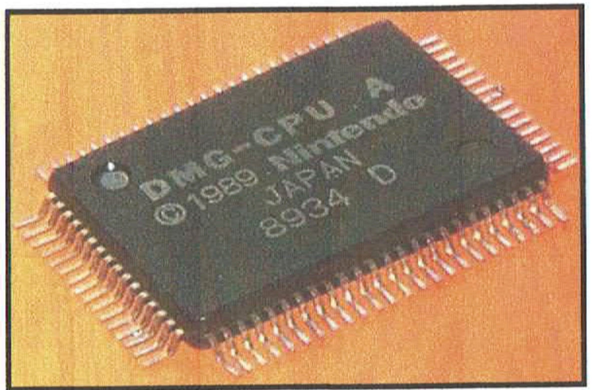
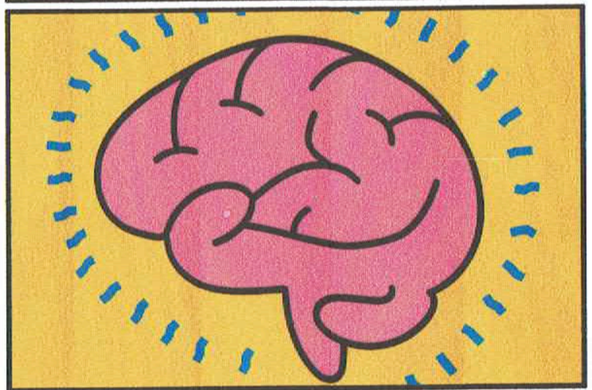
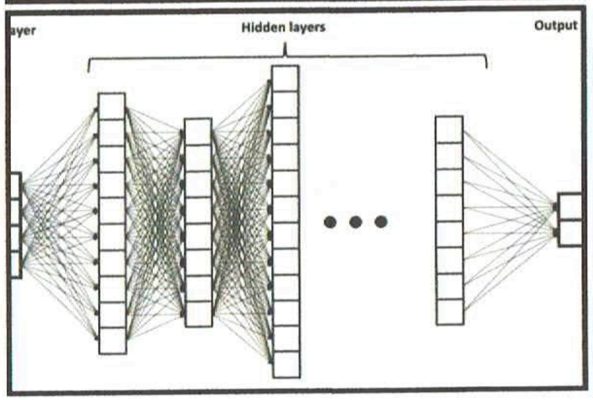
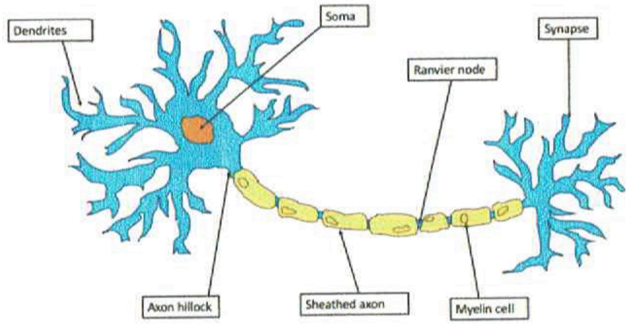
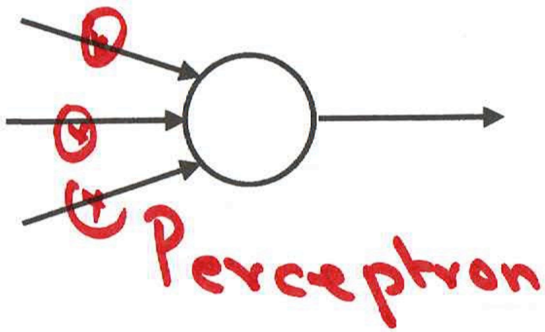
Building towards a complex task!









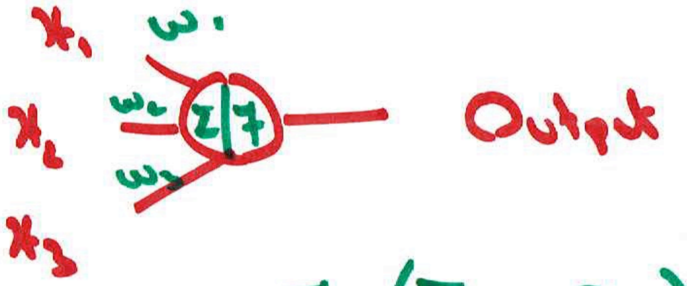


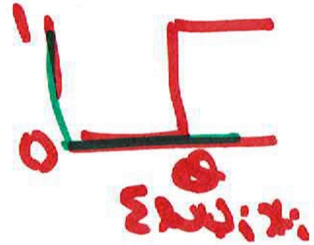
Perceptron!

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

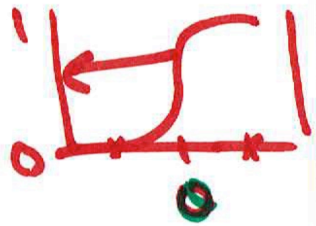
1958

- What does it do?

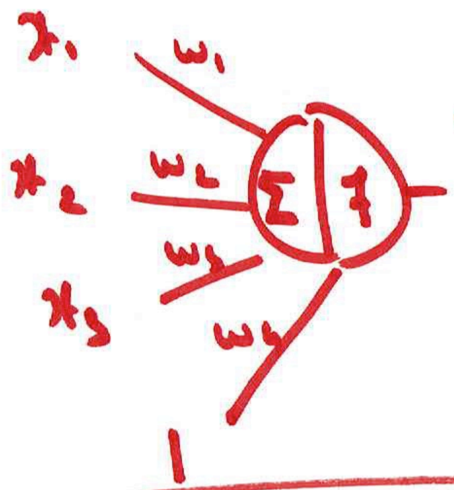




$f(\sum w_i x_i) = \text{Output}$

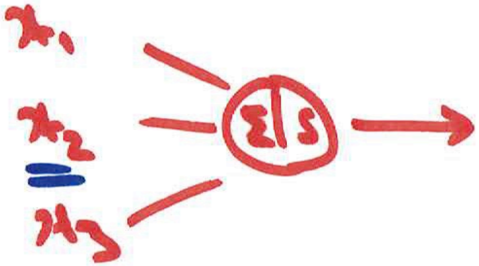


- Bias



$f(\sum w_i x_i + w_4)$

Perceptron

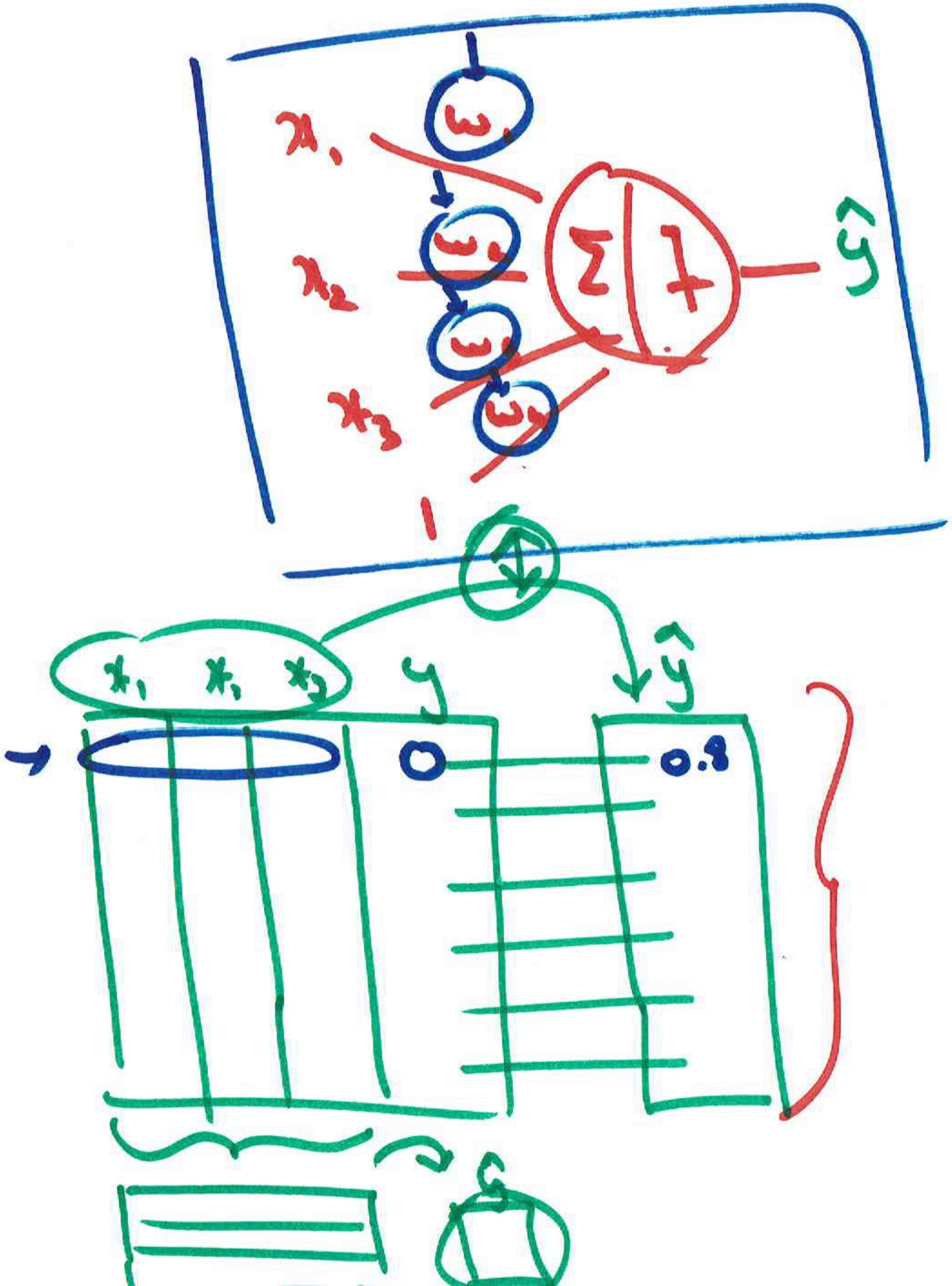


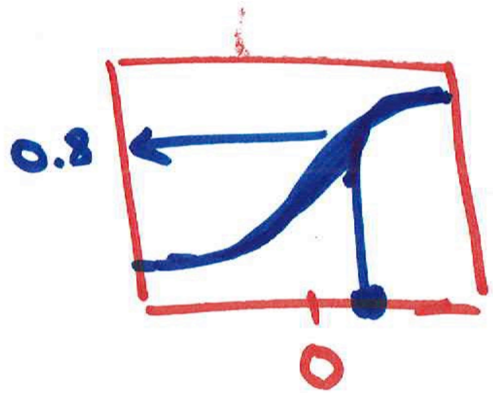
if $\sum x_i \geq \theta \Rightarrow 1$

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

$\theta = 1 \Rightarrow$ "OR"

$\theta = 3 \Rightarrow$ "AND"

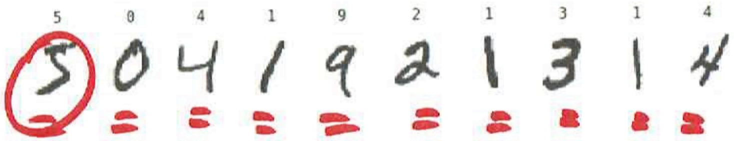




$$f(\underline{\sum w_i x_i + w_0})$$

An Example

MNIST



60000 train

10000

28

28

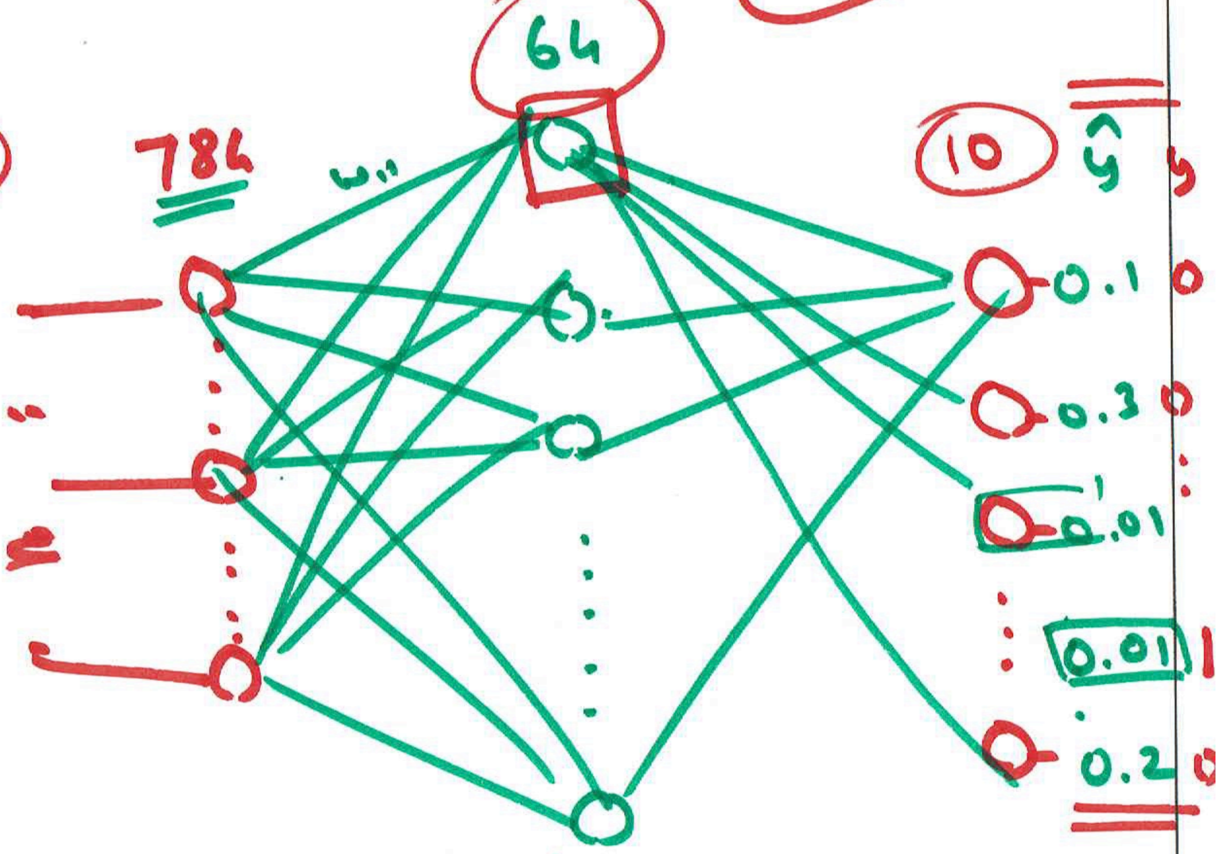
5

0

4

9

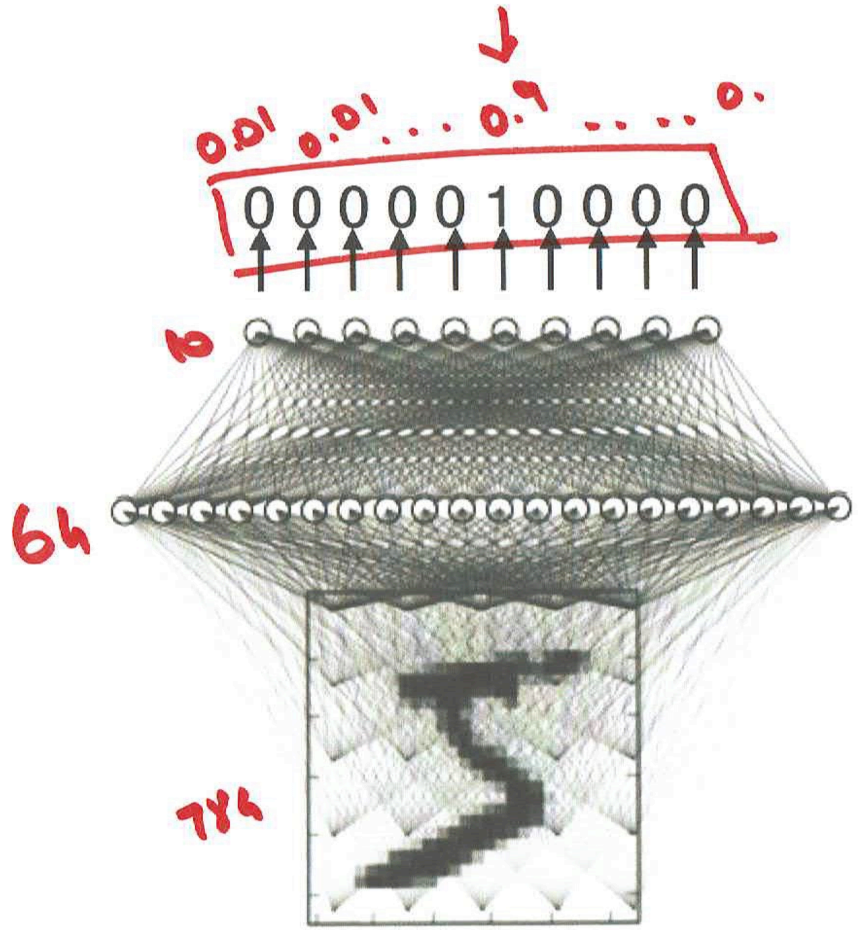
89%

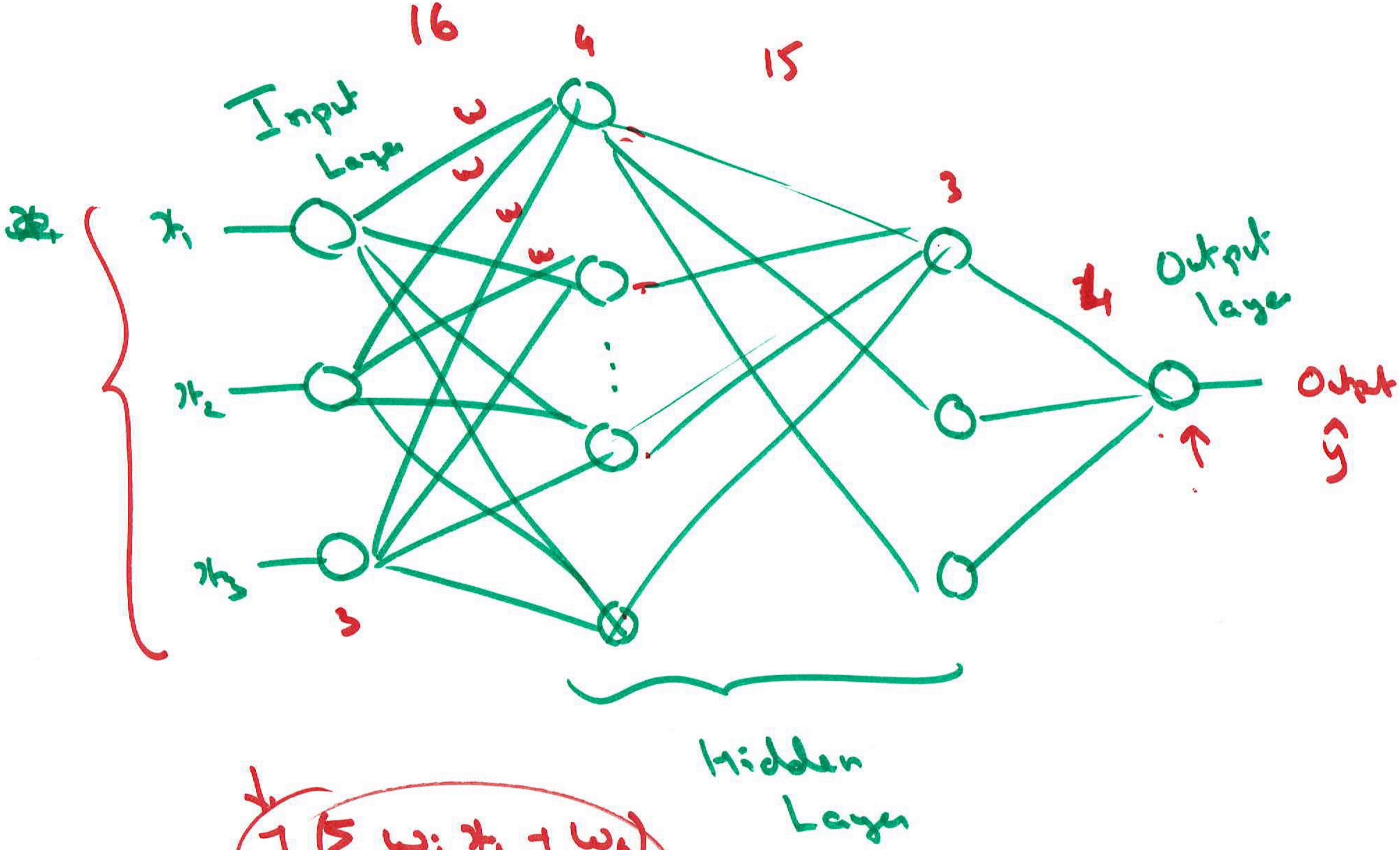


$(784 \times 64 + 64) + (64 \times 10 + 10)$

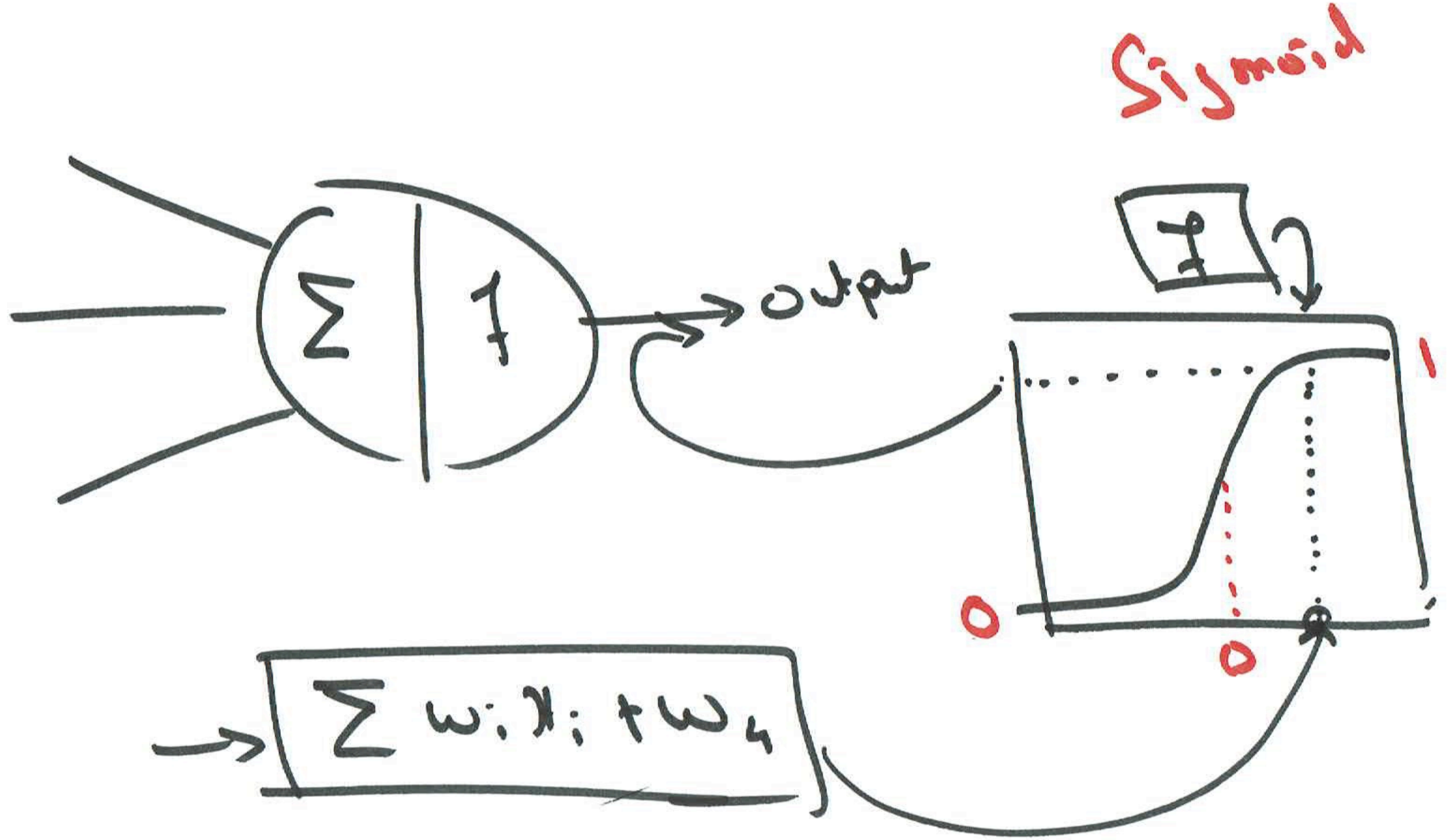
$= \boxed{50890}$

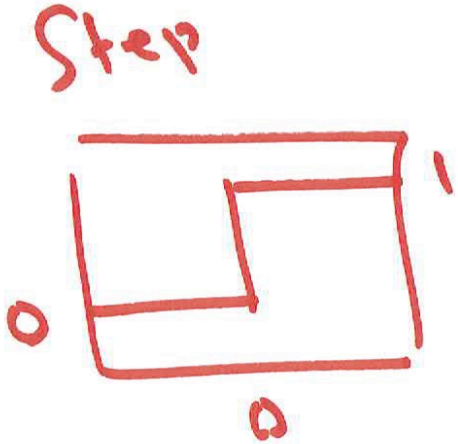
An Example





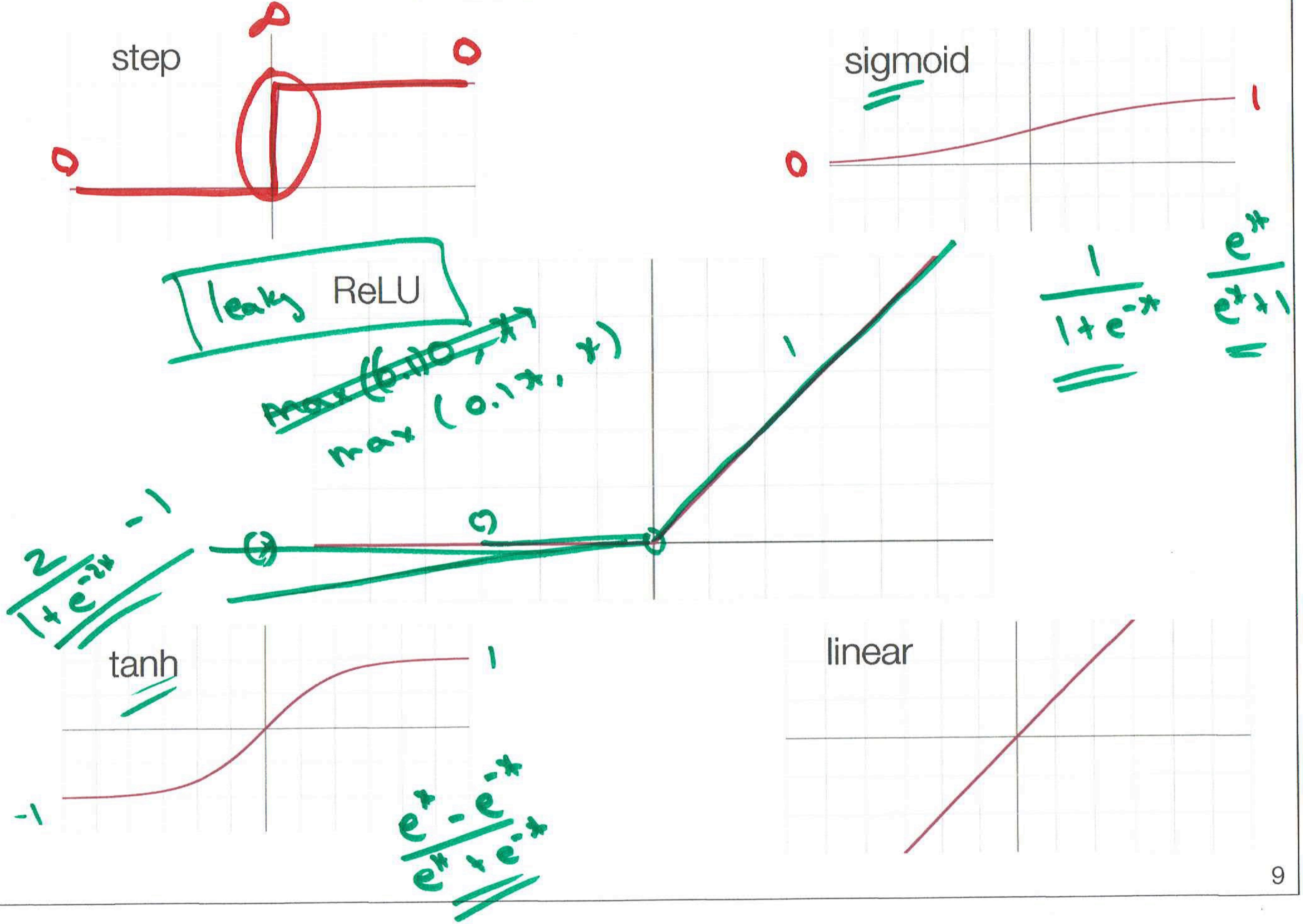
$f(\Sigma w_i x_i + w_0)$



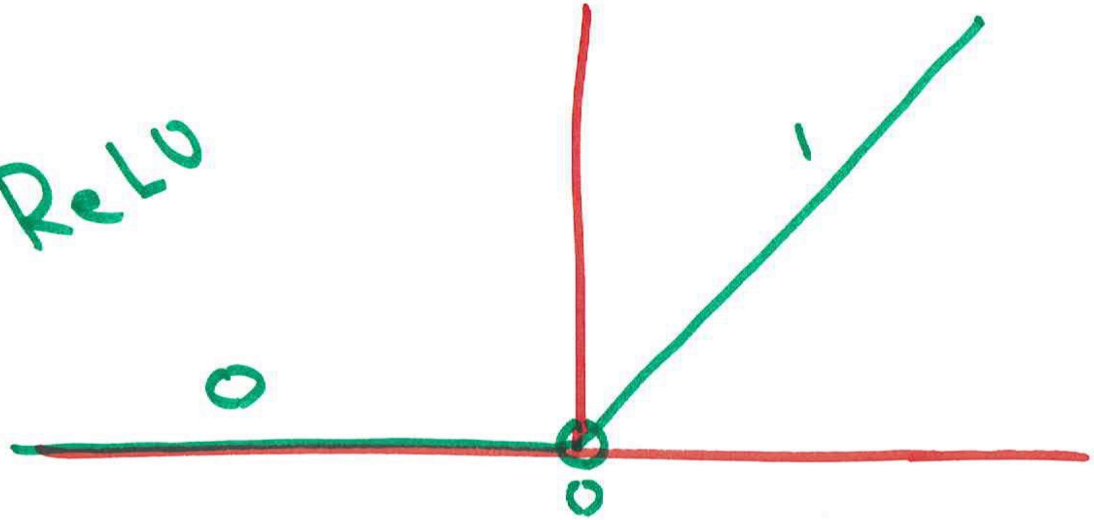


$$\text{Output} = f(\Sigma w_i x_i + w_n)$$

Activation Functions

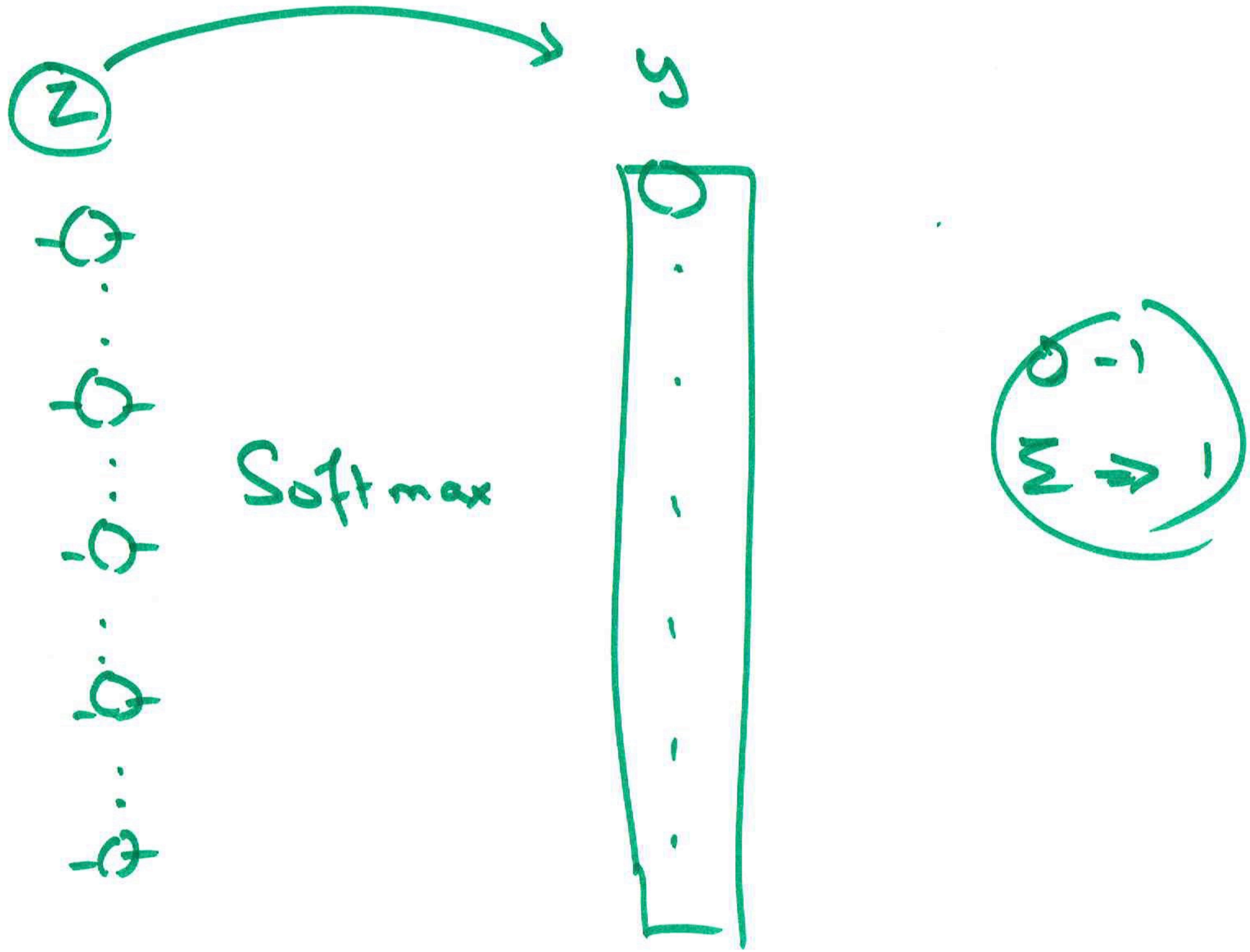


ReLU



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

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

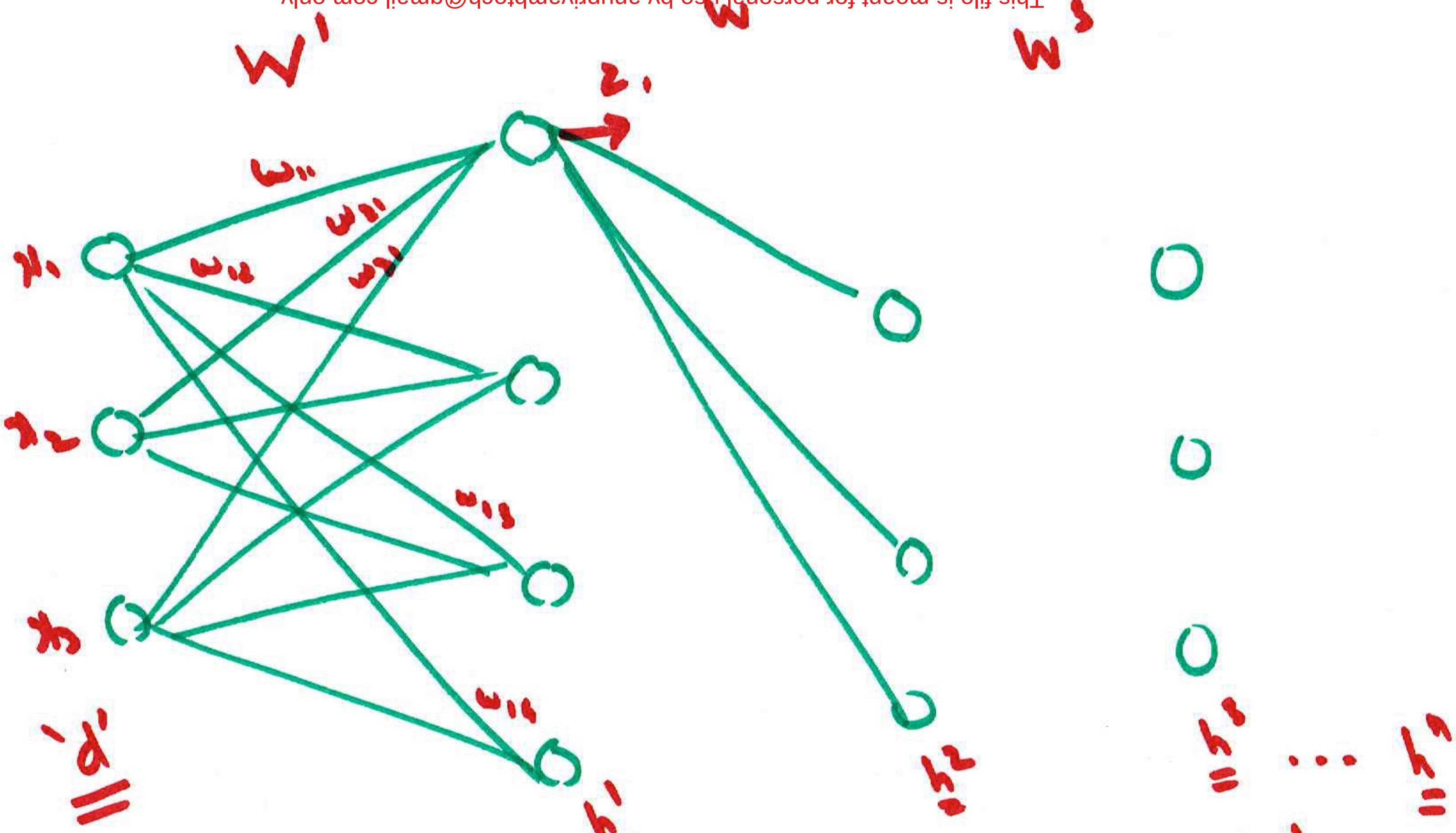

z
y
Softmax
0-1
Σ ⇒ 1

z y

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

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

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



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

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

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

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

$h' \times d$ matrix

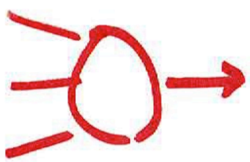
$$\sum w_{ij} x_i + b$$

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

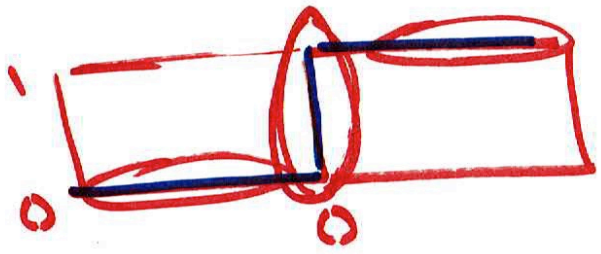
? ? ?

$\hat{y}$: $h^n \times 1$

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



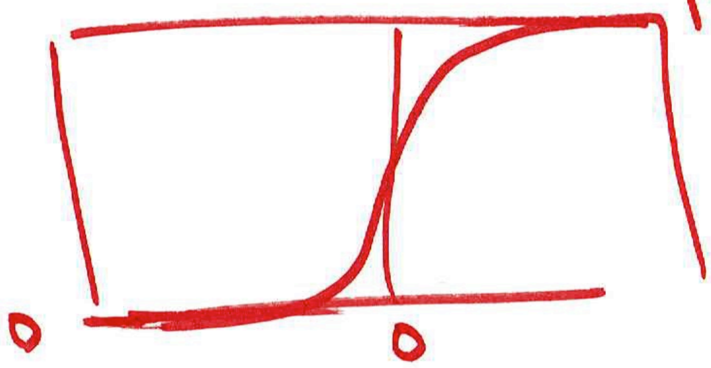
Step $(w_1 x_1 + w_2 x_2 + w_3 x_3 + b)$

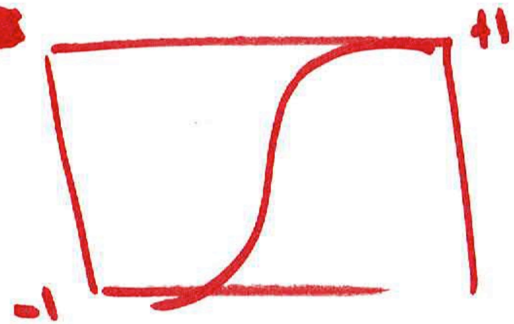


Sigmoid

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

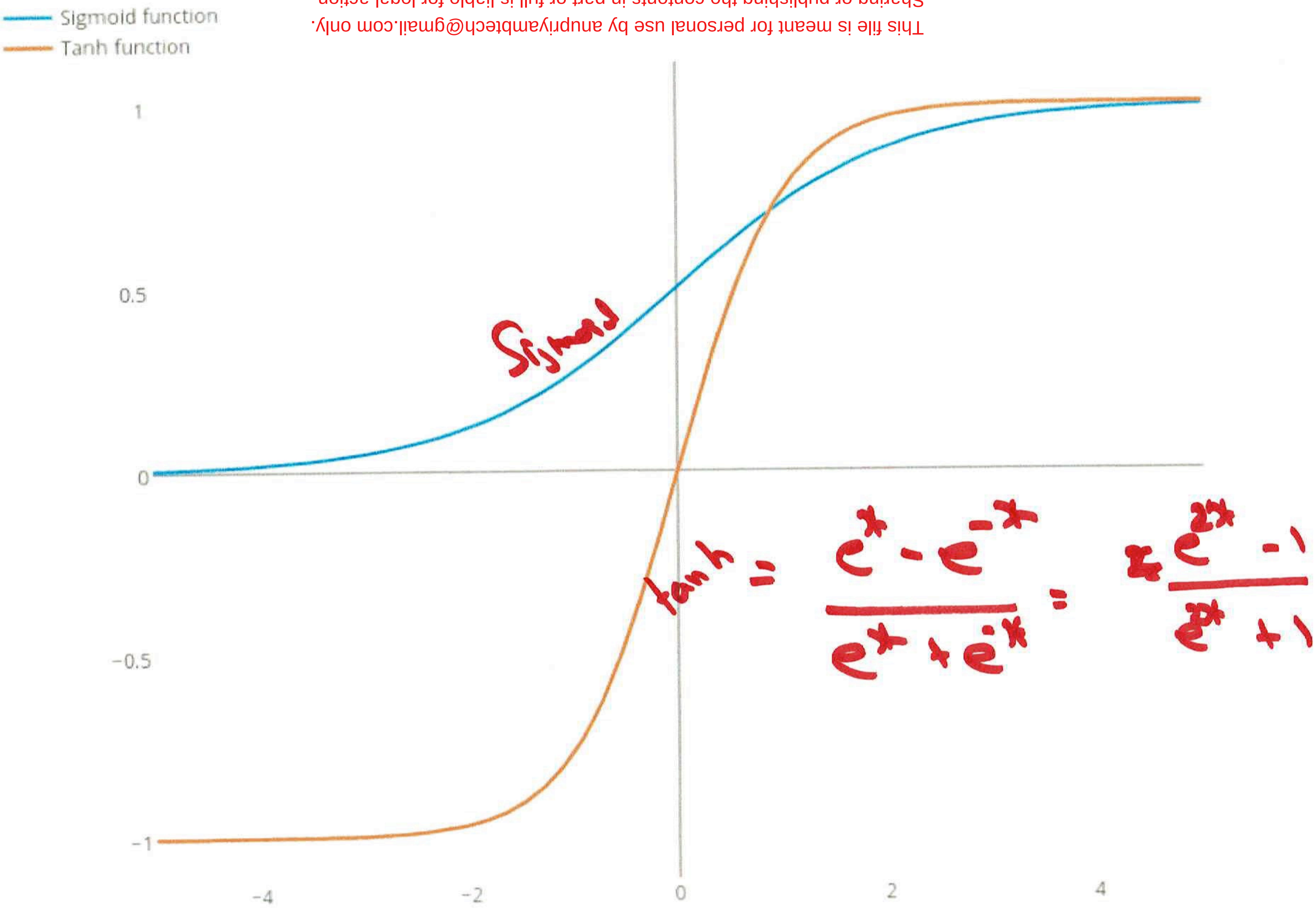
$$= \frac{e^x}{e^x + 1}$$





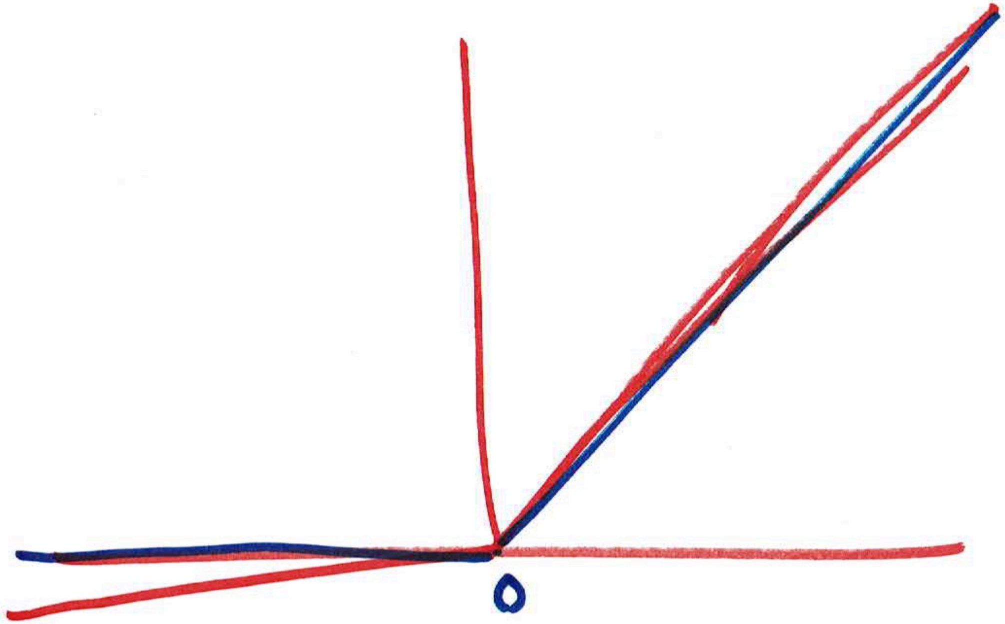
$$\boxed{\cancel{2\sigma(x)} - 1}$$

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



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

ReLU



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

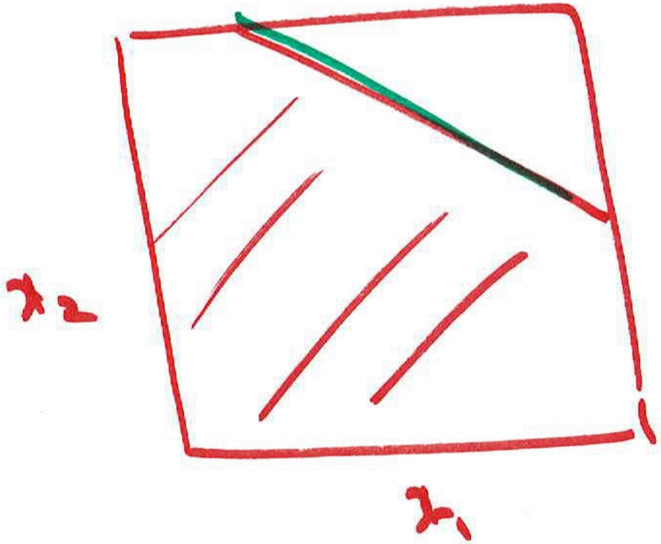
$$= \max(0, x)$$

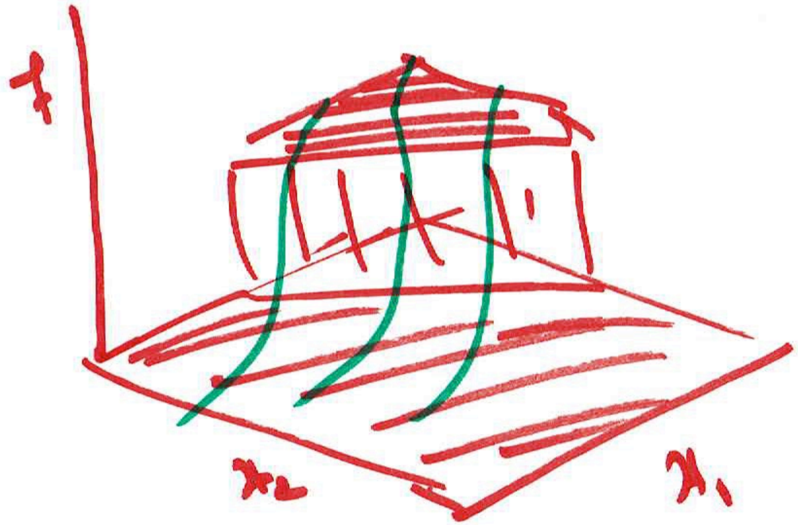
$\max(0.01x, x)$

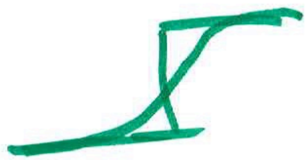
linear

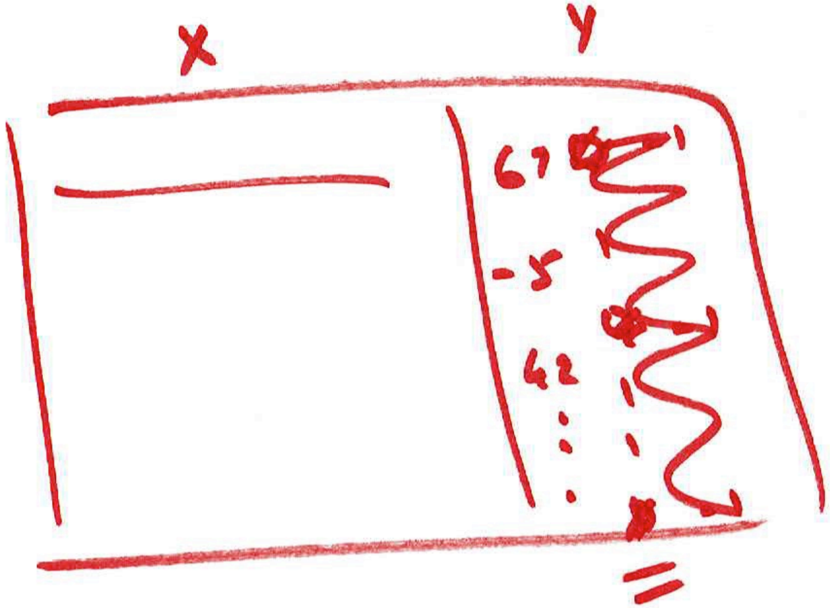
$$\text{Step} = f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

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









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

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

Softmax

z_1
z_2
z_3
z_4

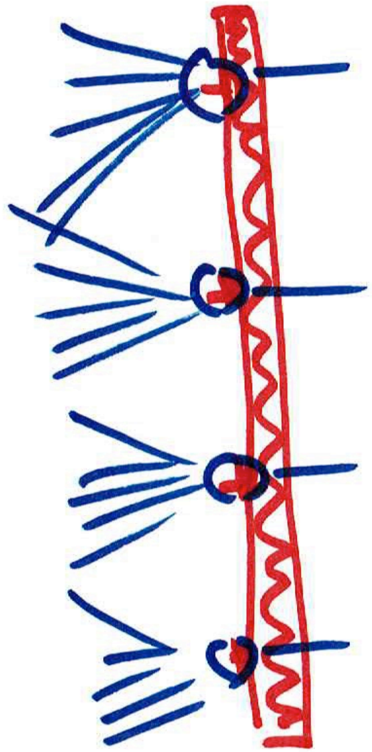
$$\frac{e^{z_i}}{\sum_j e^{z_j}}$$

$\rightarrow \frac{e^{z_1}}{\sum e^{z_j}}$

$\rightarrow \frac{e^{z_2}}{\sum e^{z_j}}$

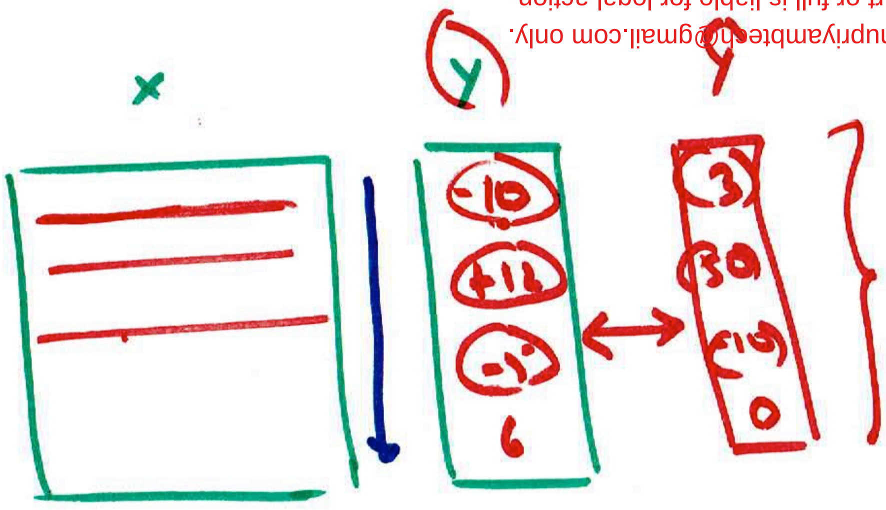
$\rightarrow \vdots$

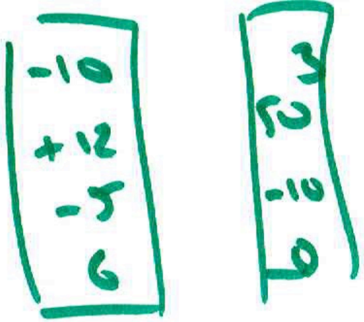
$\rightarrow \frac{e^{z_4}}{\sum e^{z_j}}$

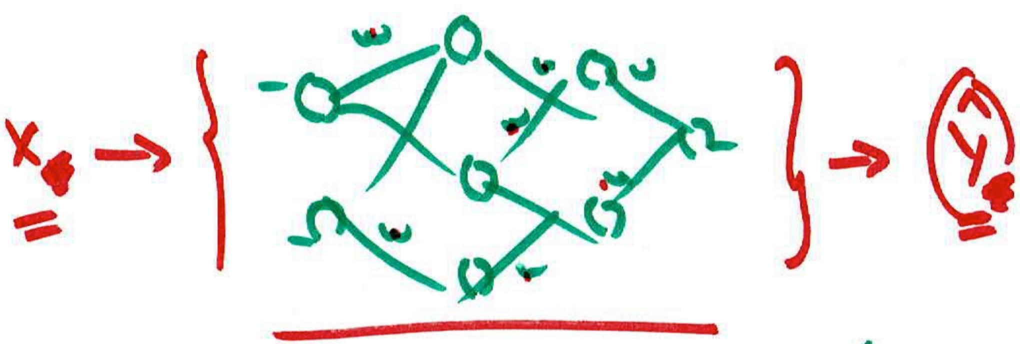


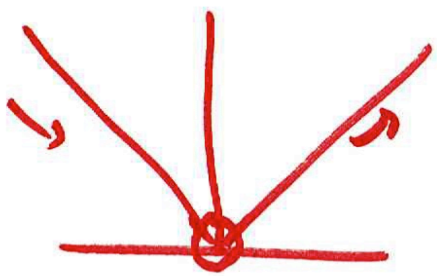
0.1
0.2
0.4
0.3

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

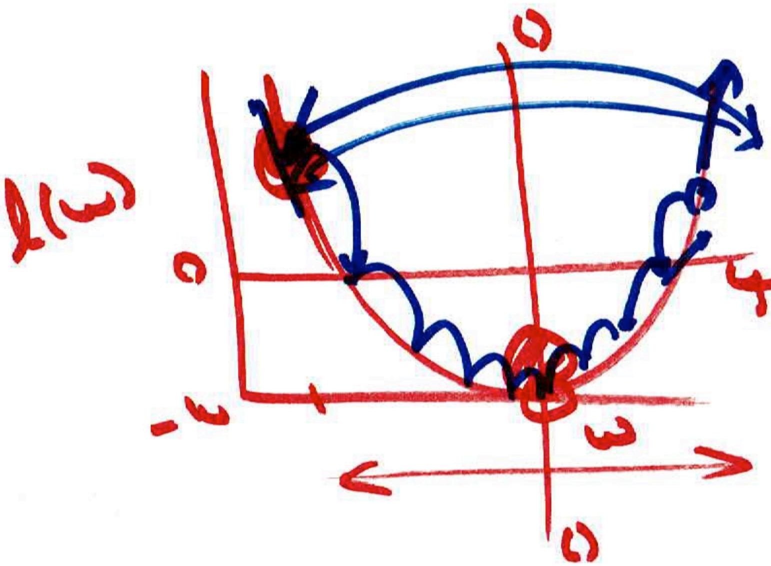
$$L(y, \hat{y}) = l(w)$$

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 $L(y, \hat{y})$ by ~~Chng~~ Changing ~~tex~~ my $W^1, W^2, \dots W^n$

$$\begin{matrix} w_{11} & w_{12} & \dots \\ \vdots & & \\ \dots & & \end{matrix}$$



$y = x^2 - 10 = -10$

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

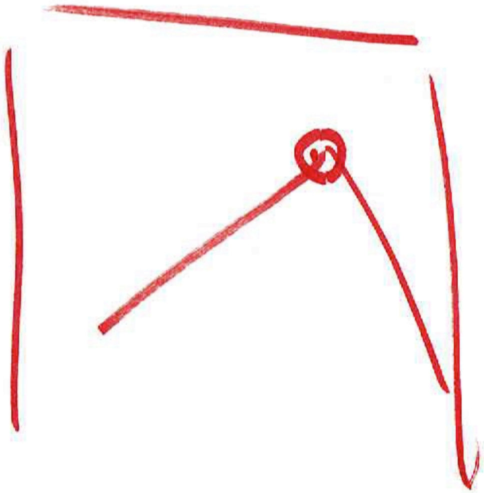
$x = 0$

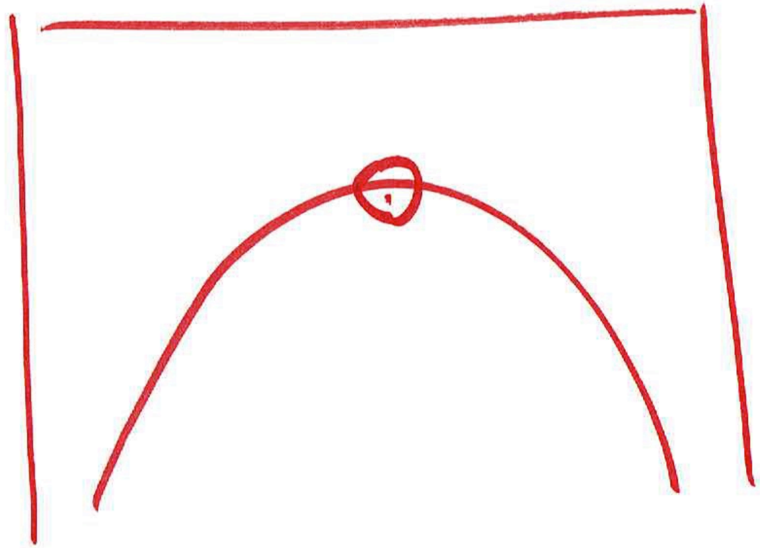
$\frac{dL}{dw} = \boxed{\quad} = 0$

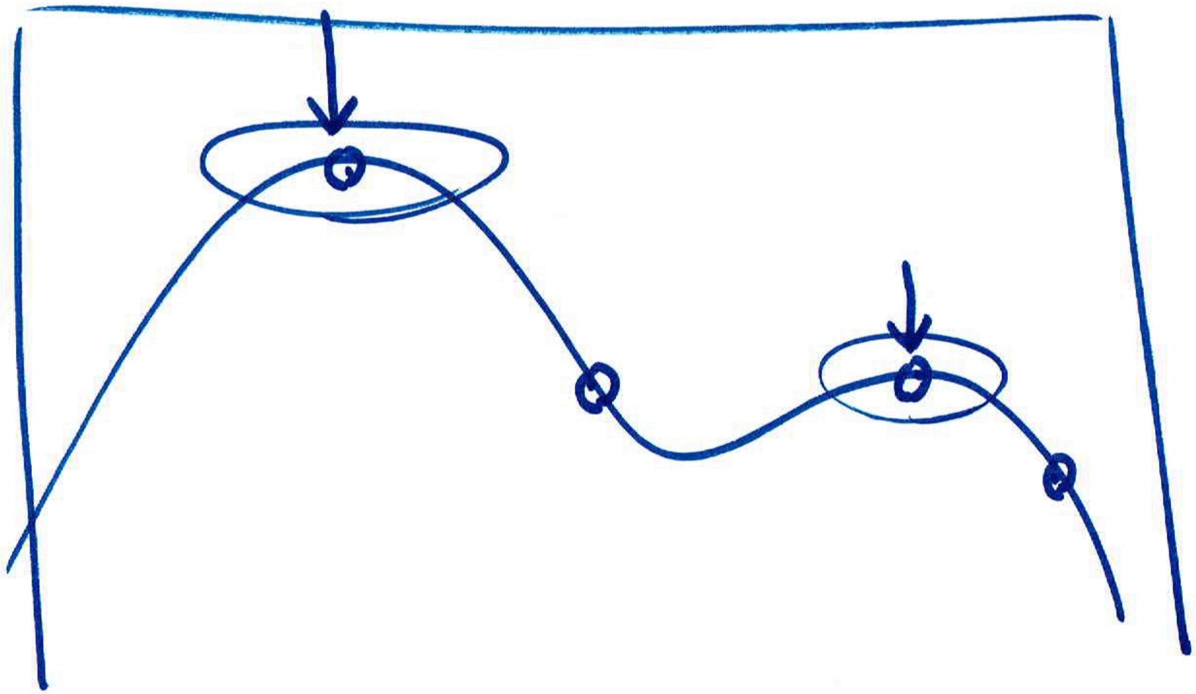
$\frac{dl}{dw}$

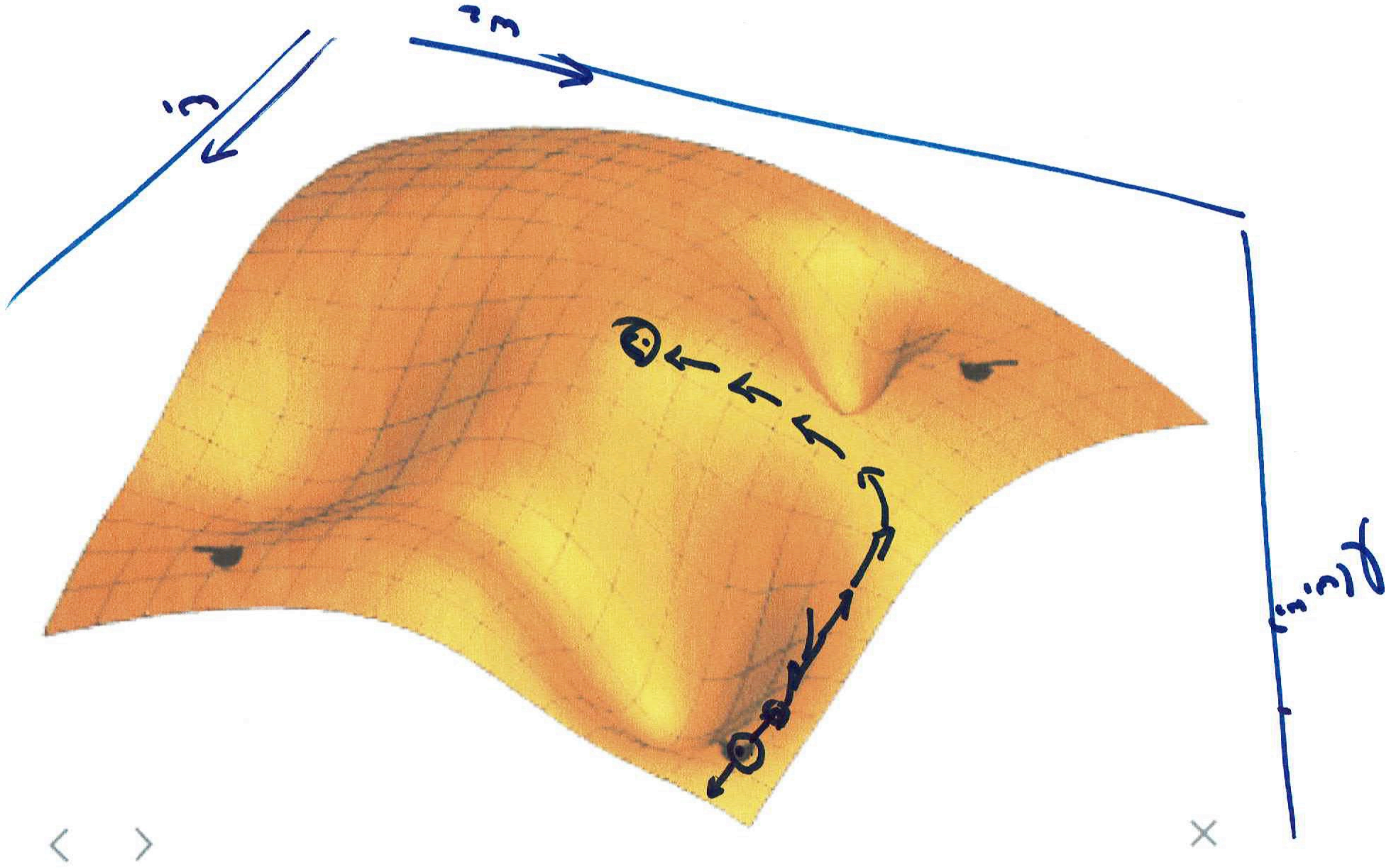
$w^{new} = w - \eta \nabla_w l$

learning rate








w_1
w_2
$l(w_1,w_2)$

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

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

SGD

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

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

function of w $(l(w))$

Loss

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

Chain Rule

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

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

Back propagation