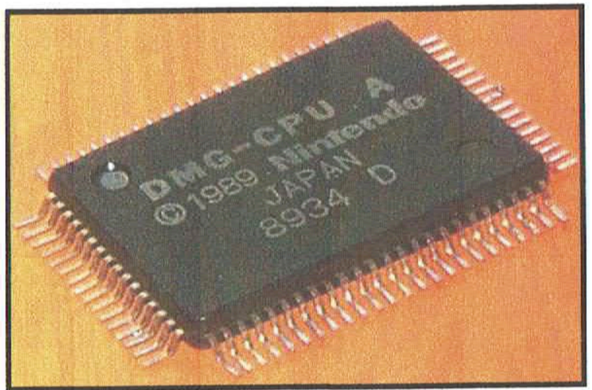
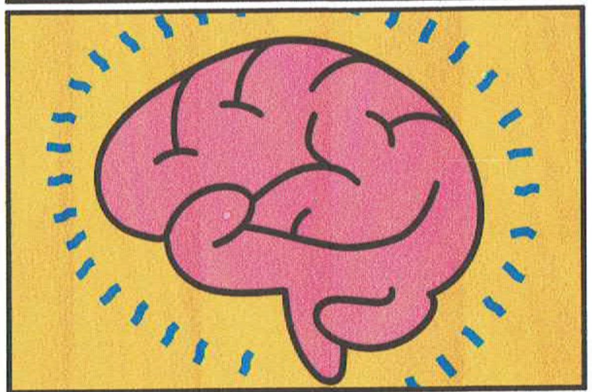
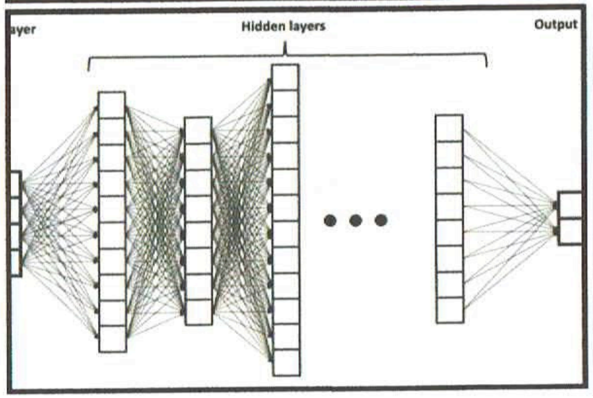
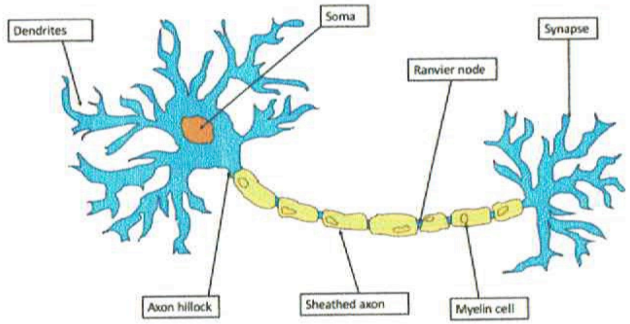
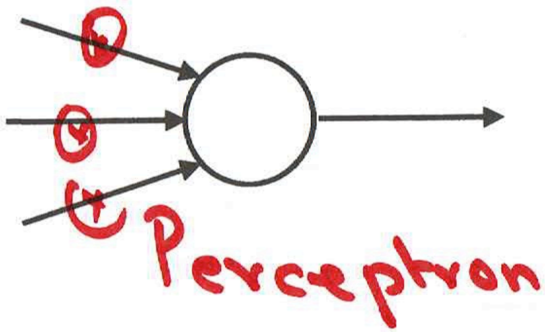
Building towards a complex task!







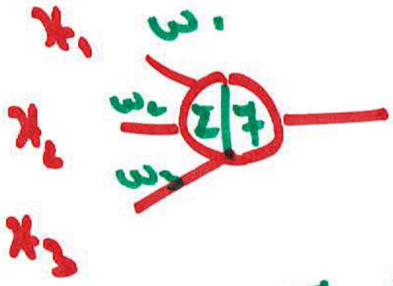




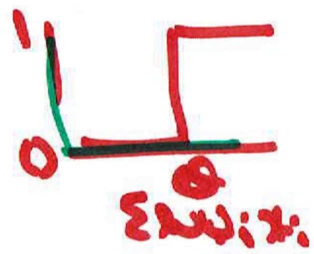
Perceptron!

- What does it do?

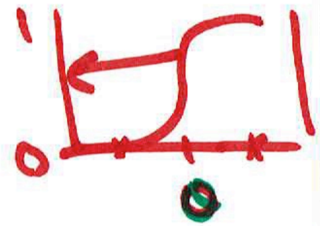
1958



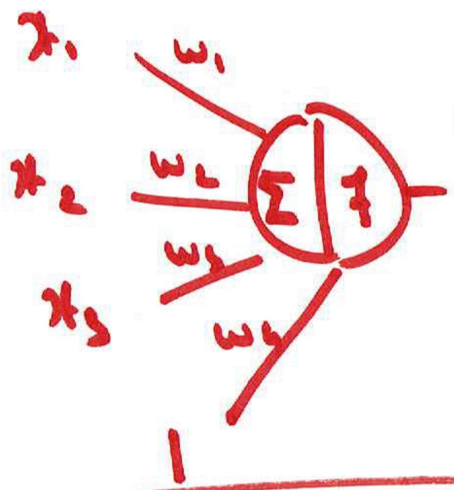
Output



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

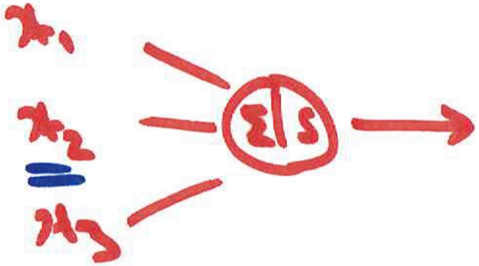


- Bias



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

Perceptron

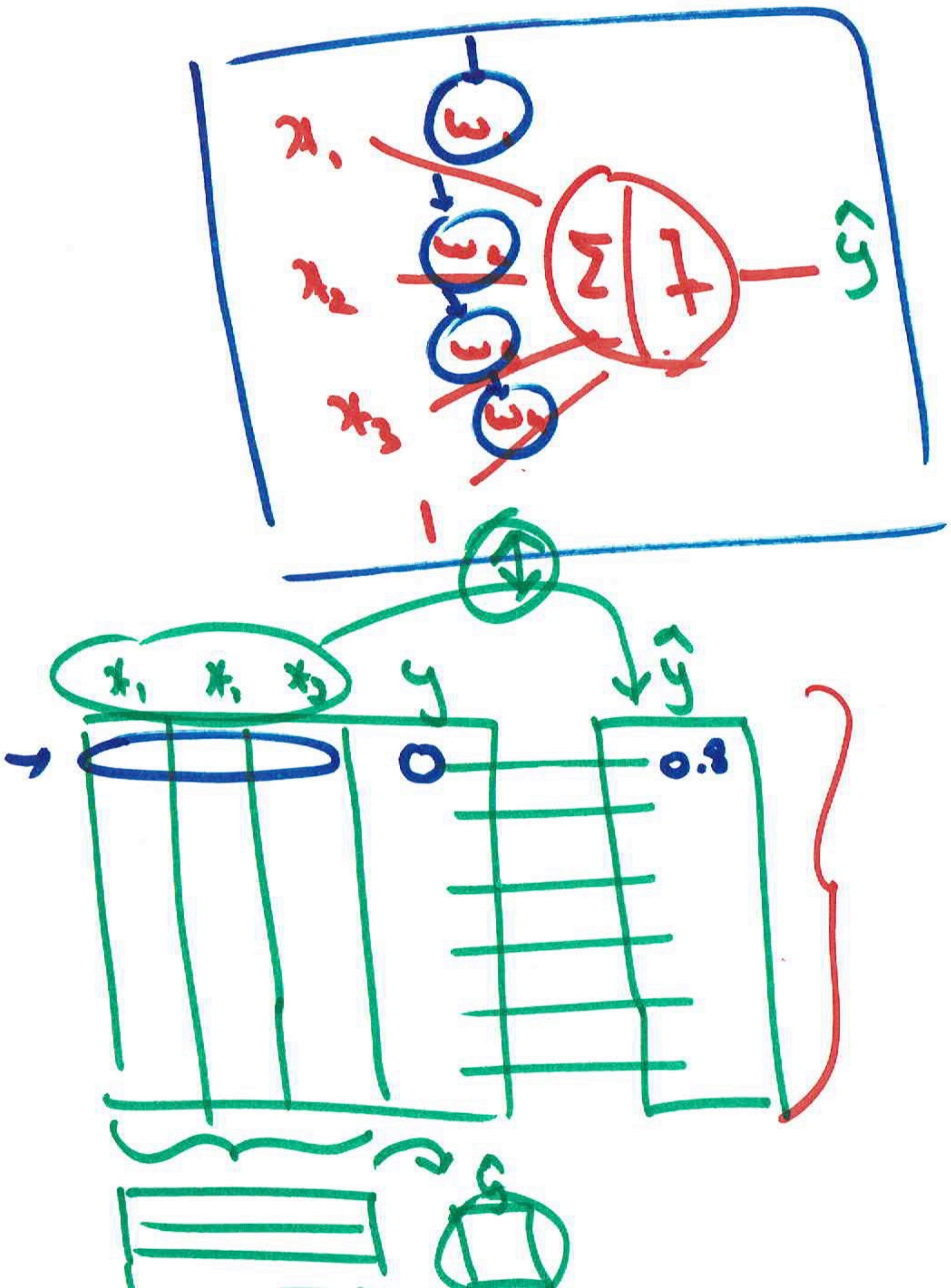


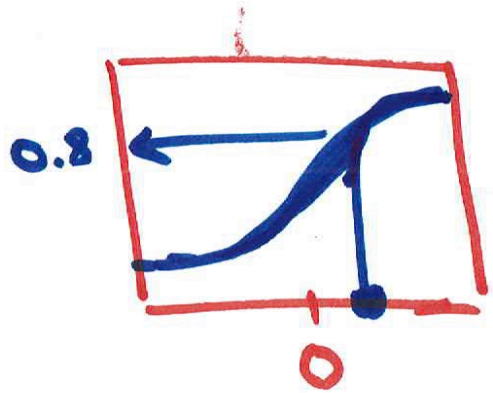
if $\sum x_i \geq \theta \Rightarrow 1$

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

$\theta = 1 \Rightarrow$ "OR"

$\theta = 3 \Rightarrow$ "AND"

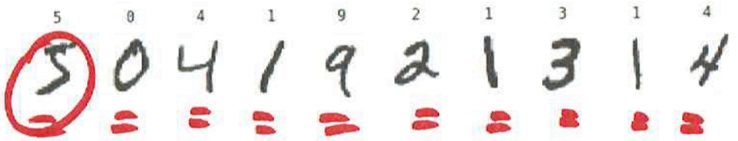




$$f(\sum w_i x_i + w_0)$$

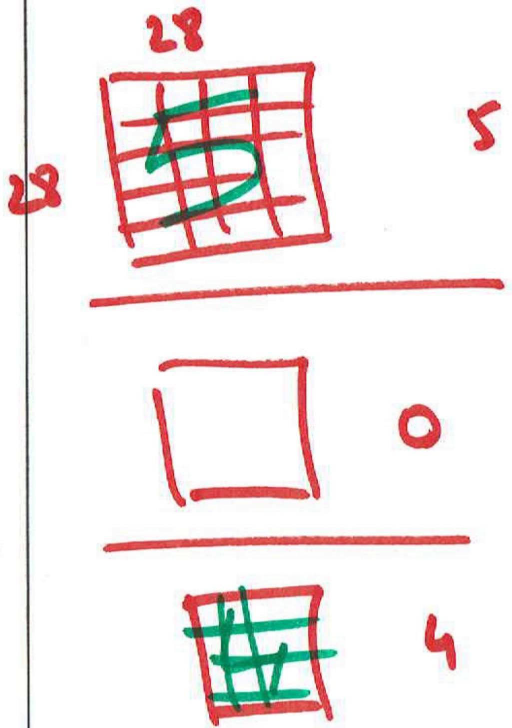
MNIST

An Example

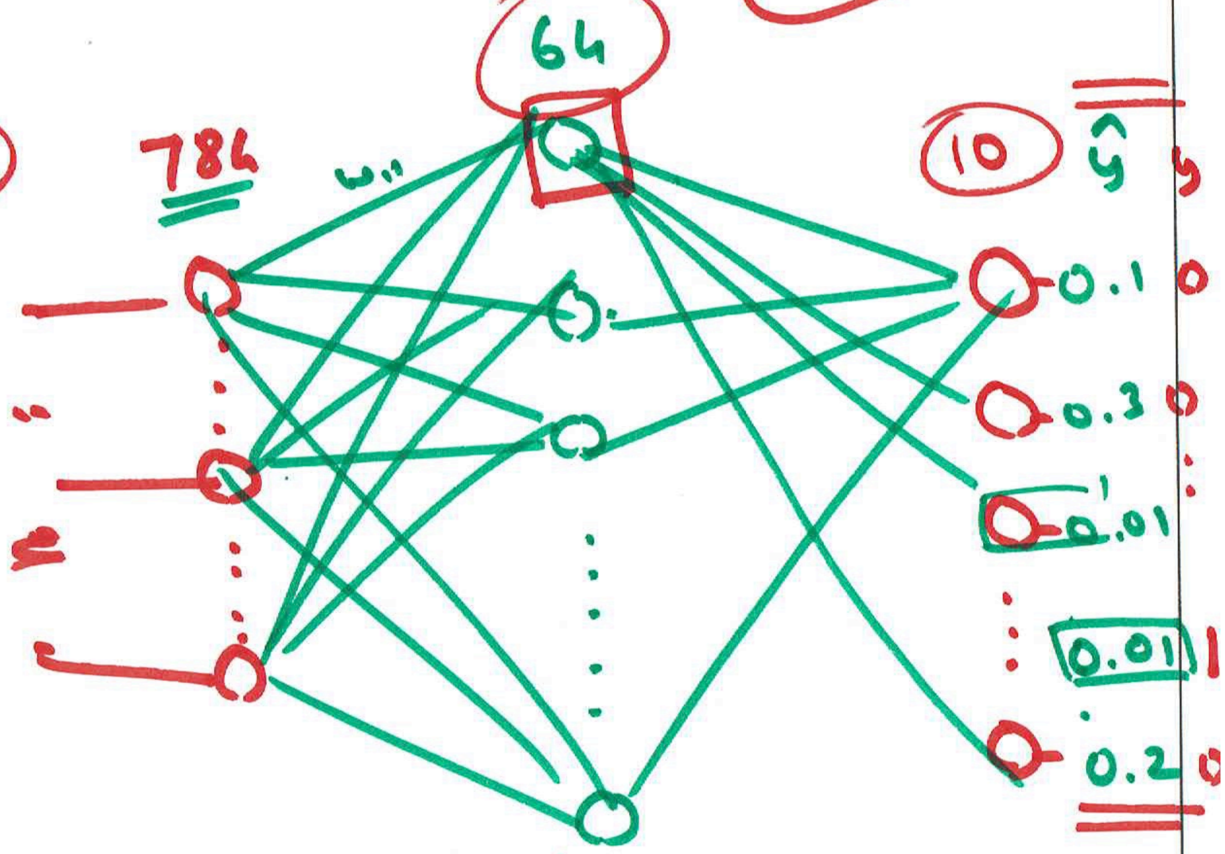


$\boxed{60000}$ train

~~10000~~



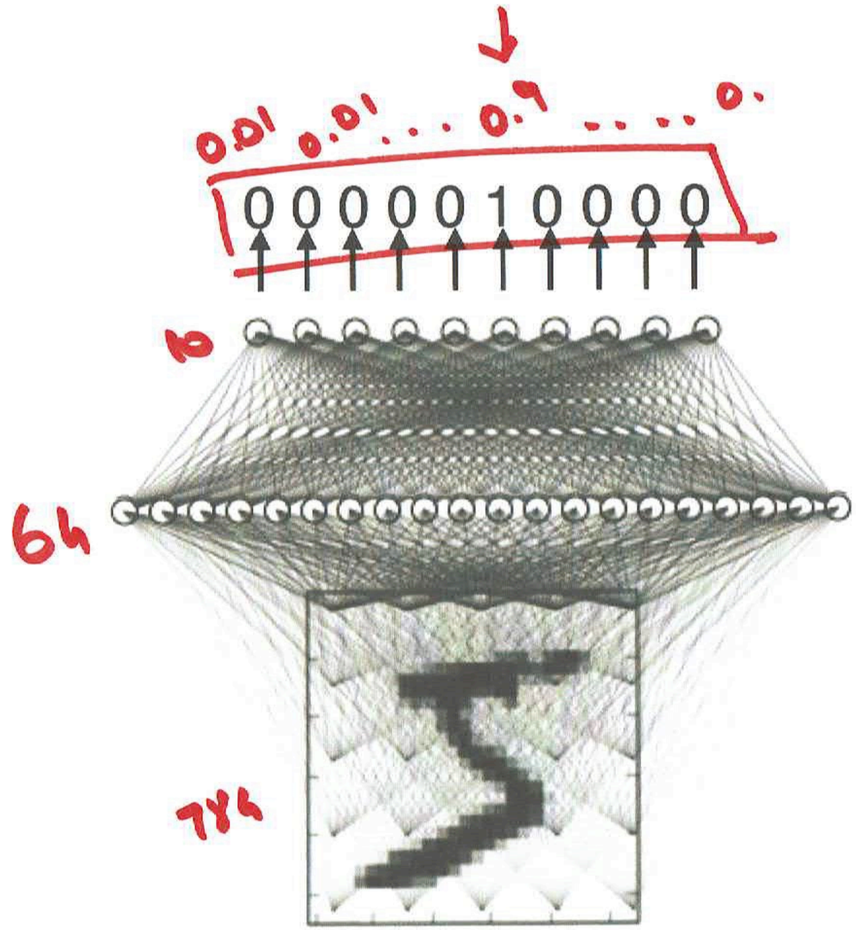
89%

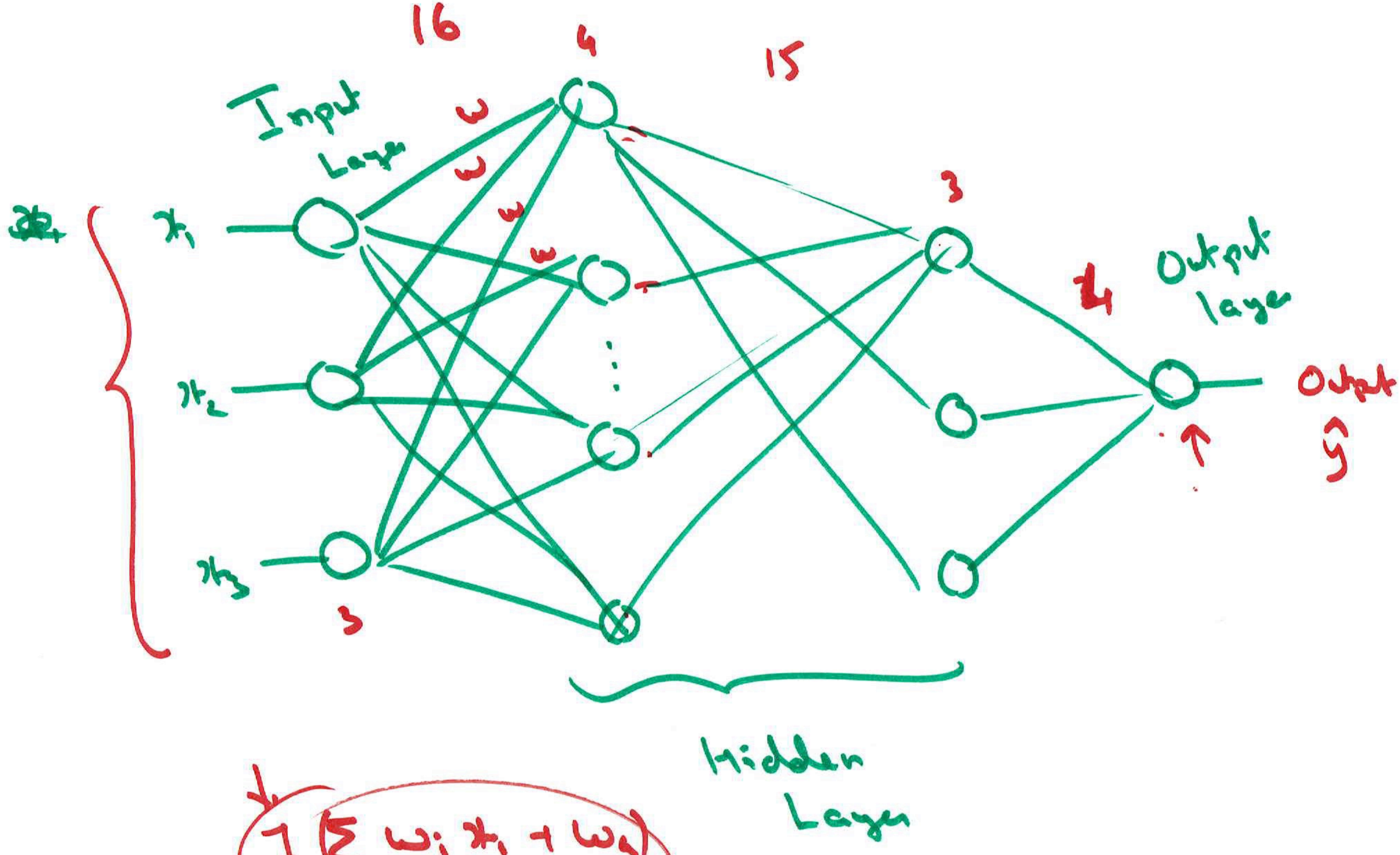


$$(784 \times 64 + 64) + (64 \times 10 + 10)$$

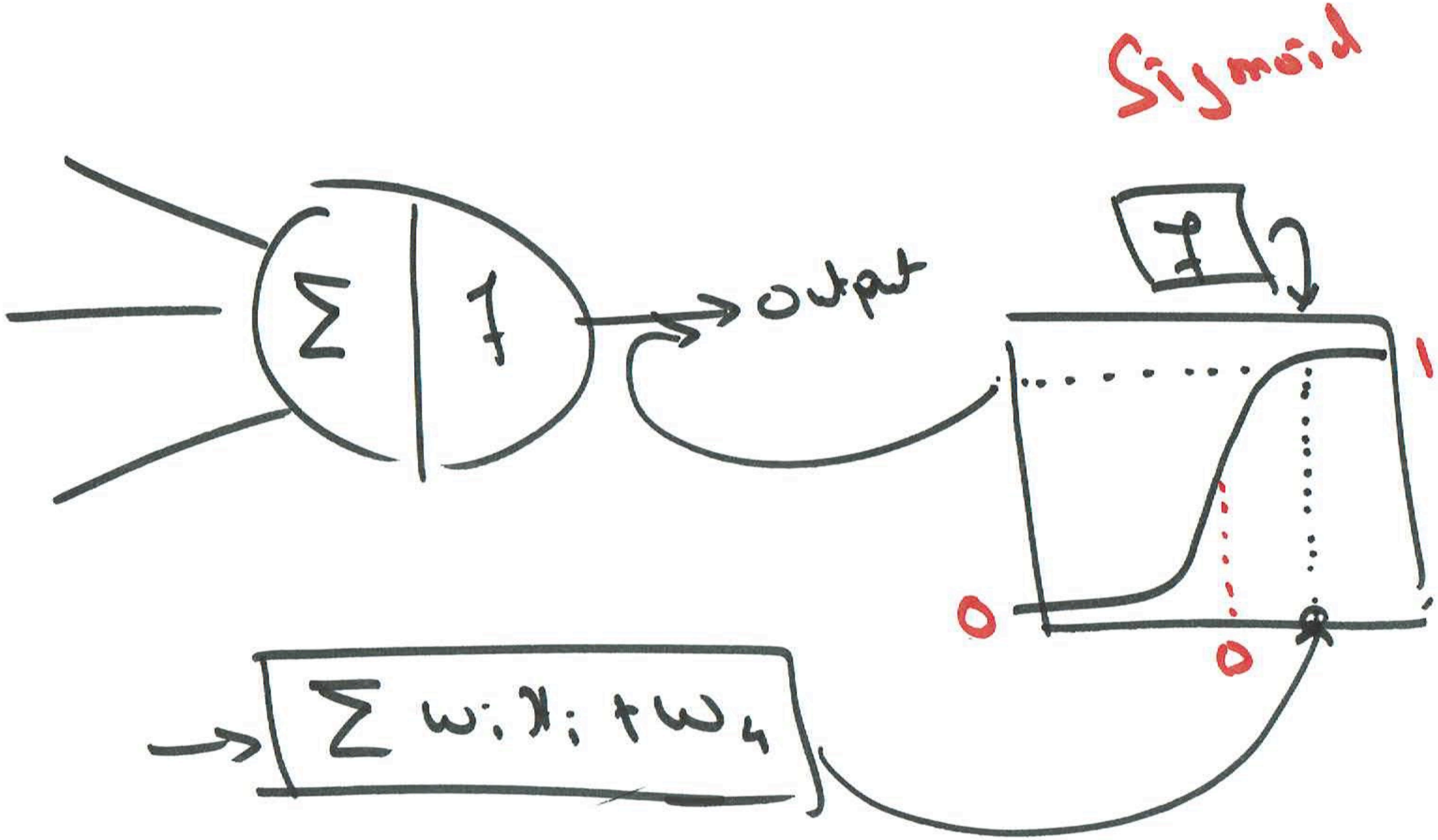
$$= \boxed{50890}$$

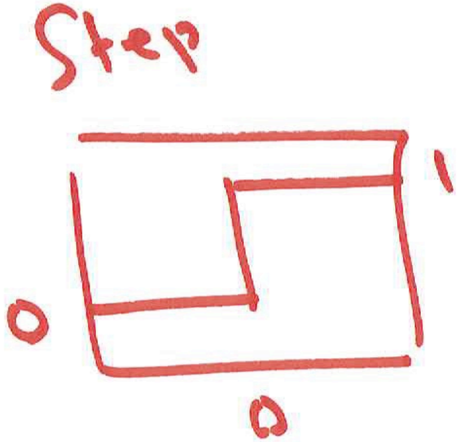
An Example





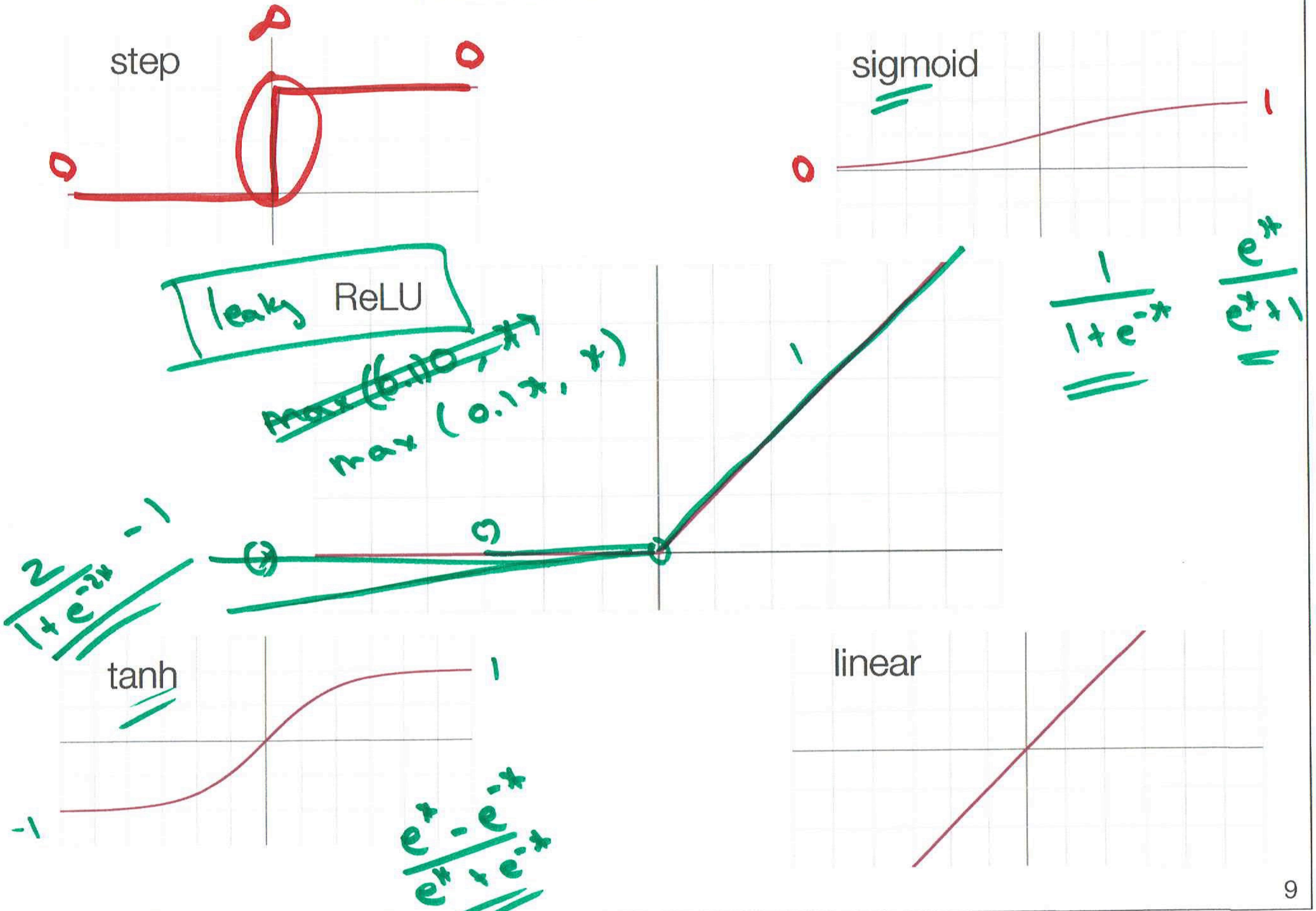
$$f\left(\sum w_i x_i + w_0\right)$$



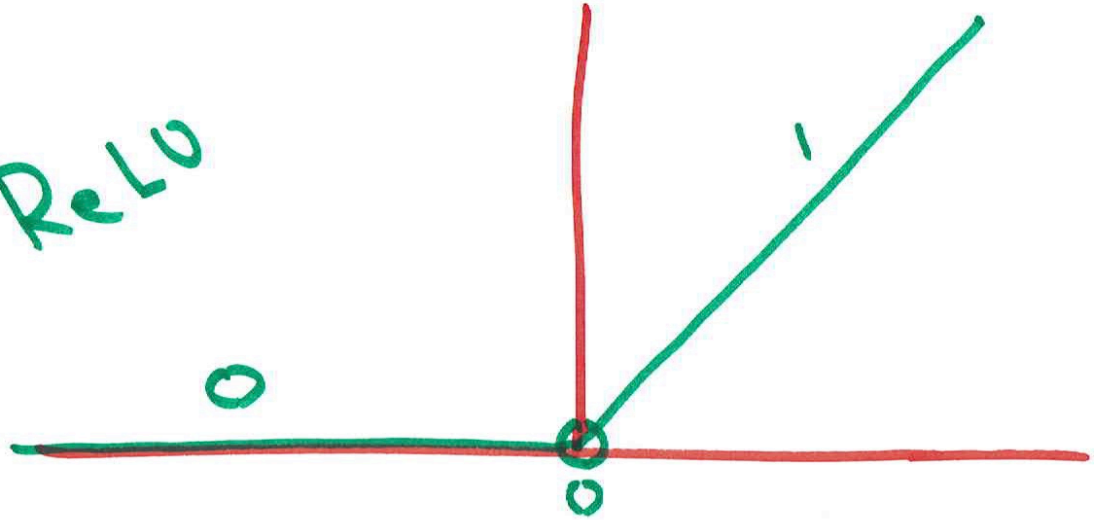


$$\text{Output} = f(\Sigma w_i x_i + w_n)$$

Activation Functions

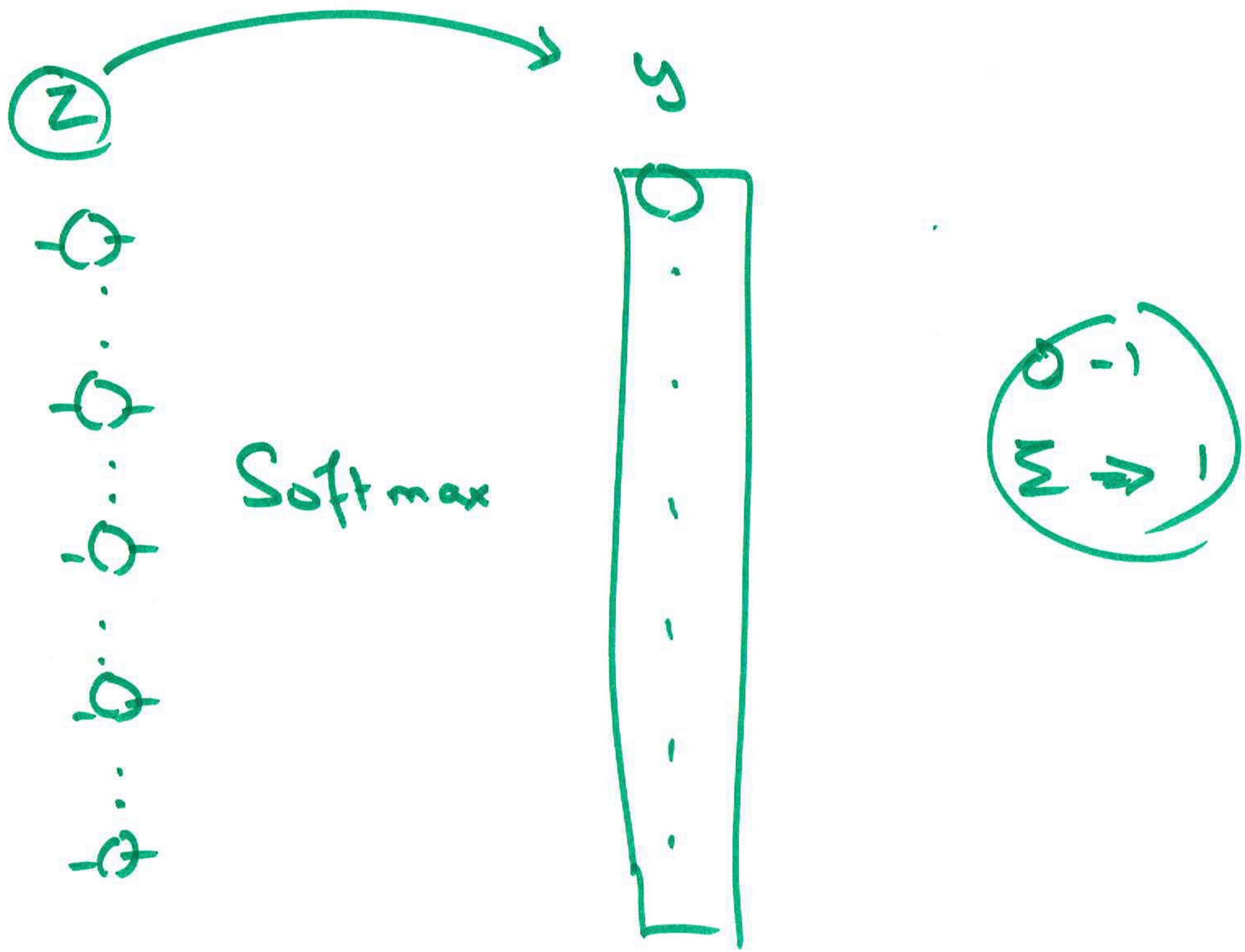


ReLU



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

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


z
y
Softmax
0-1
Σ ⇒ 1

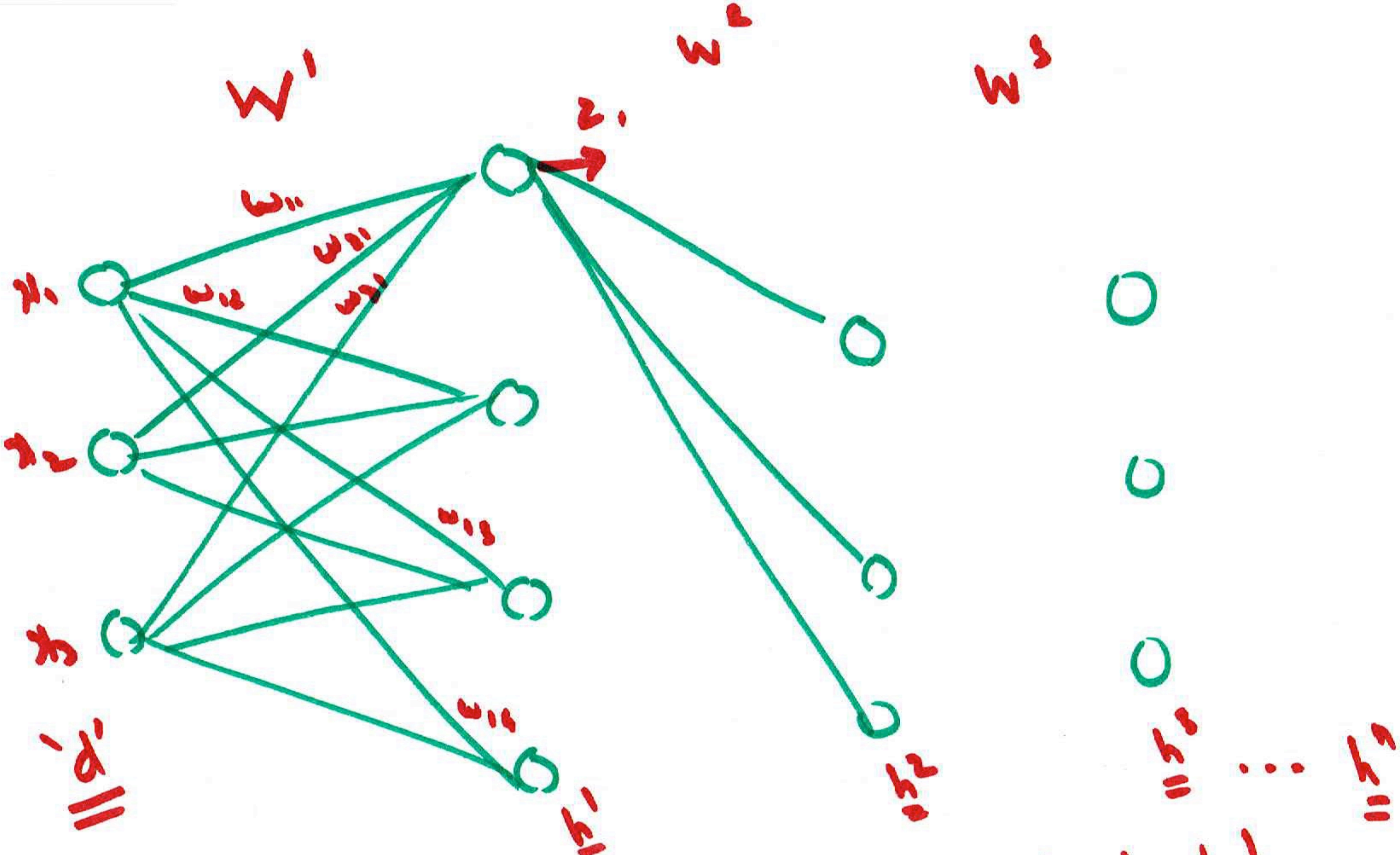
z

y

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

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

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



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

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

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

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

$h^1 \times d$ matrix

$\sum w_{ij} x_i + b$

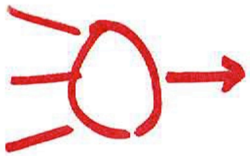
? ? ?

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

$h^n \times 1$

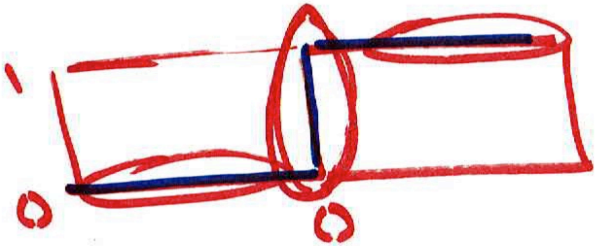
$\underbrace{h^1 \times d \quad d \times 1}_{h^1 \times 1}$ $\quad h^1 \times 1$

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



? ? ?

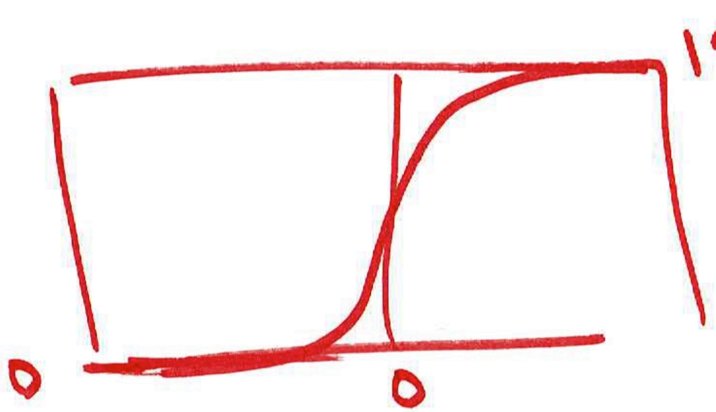
$\underline{\text{Step}\,(w_1 x_1 + w_2 x_1 + w_3 x_3 + b)}$

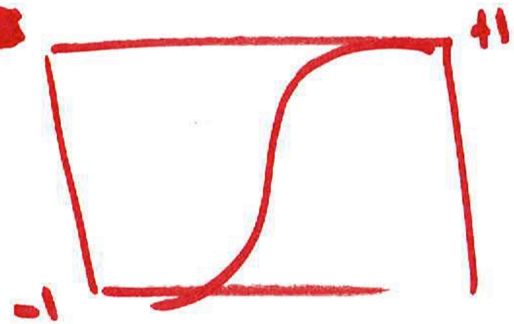


Sigmoid

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

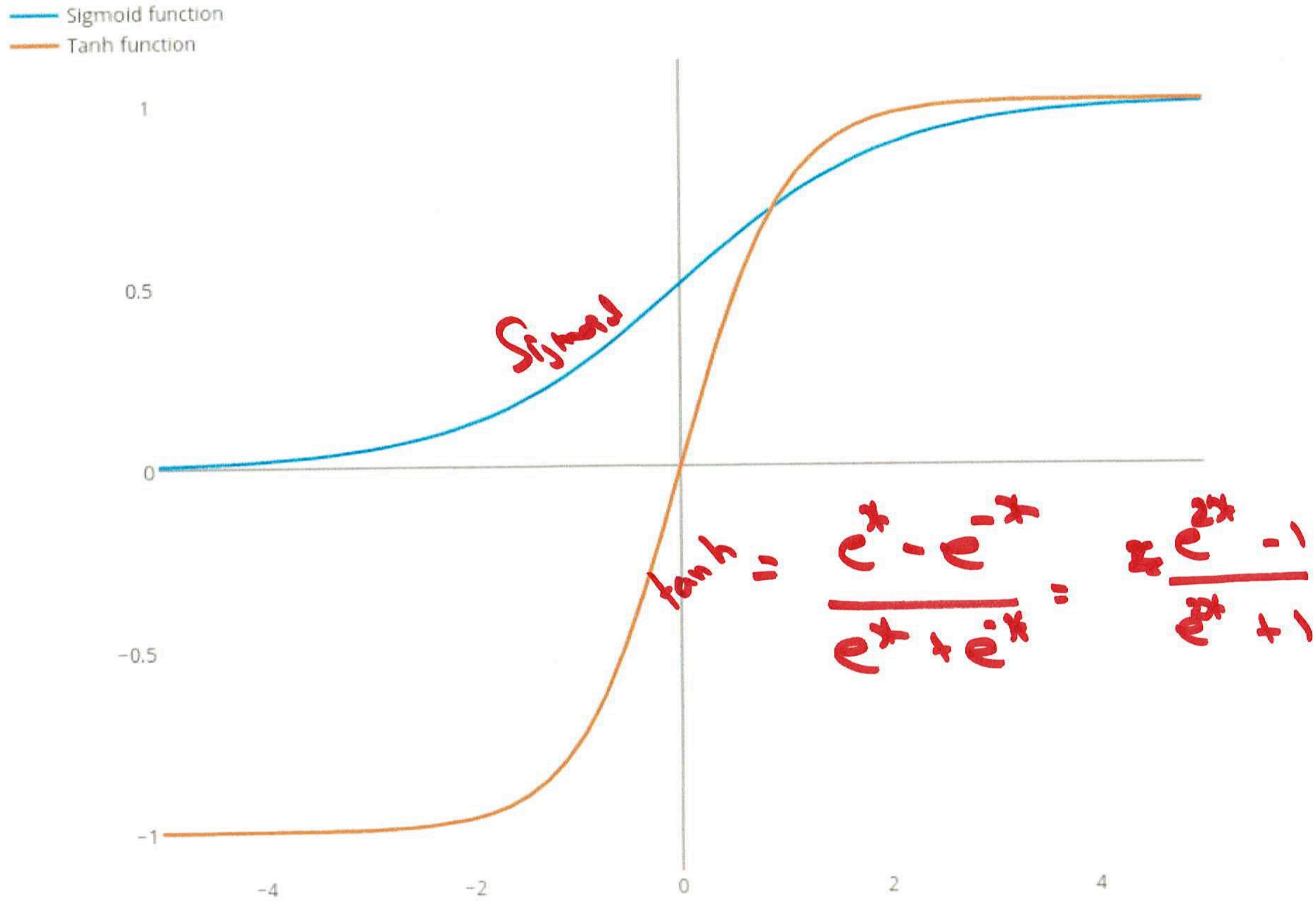
$$= \frac{e^{x}}{e^{x}+1}$$





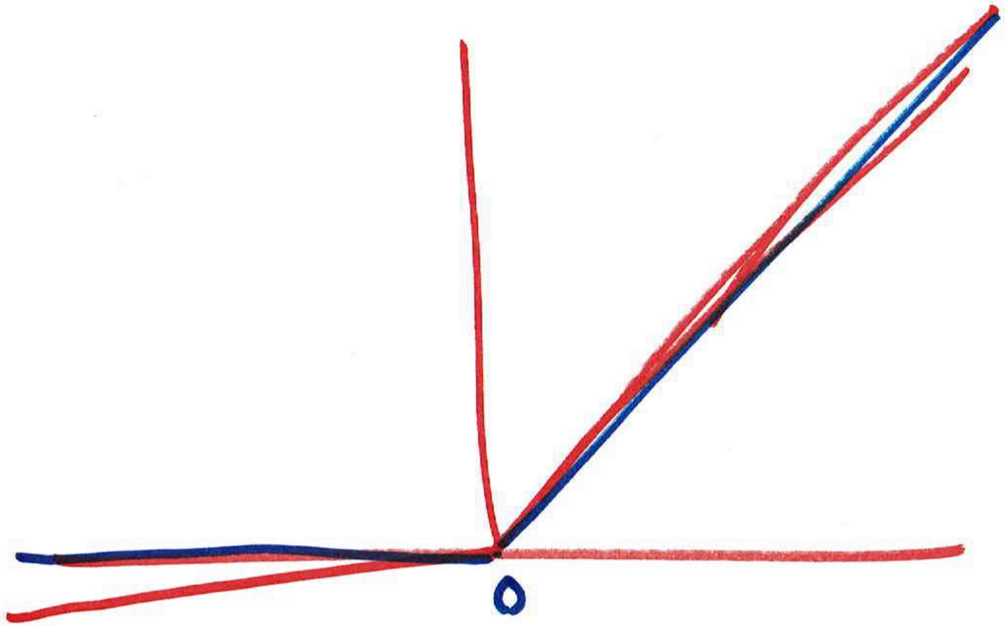
$$\boxed{\cancel{2\sigma(x)} = 1}$$

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



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

ReLU

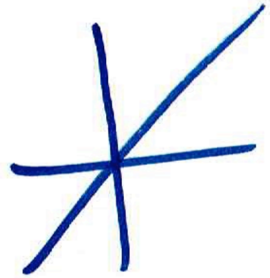


$$= \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

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

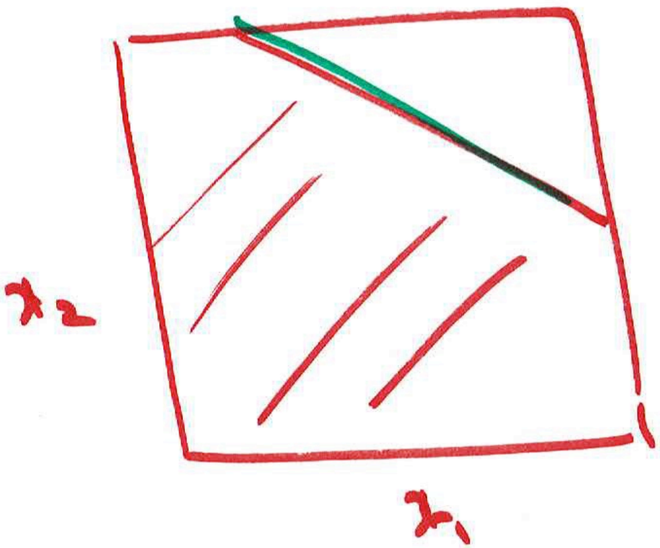
$\max(0.01x, x)$

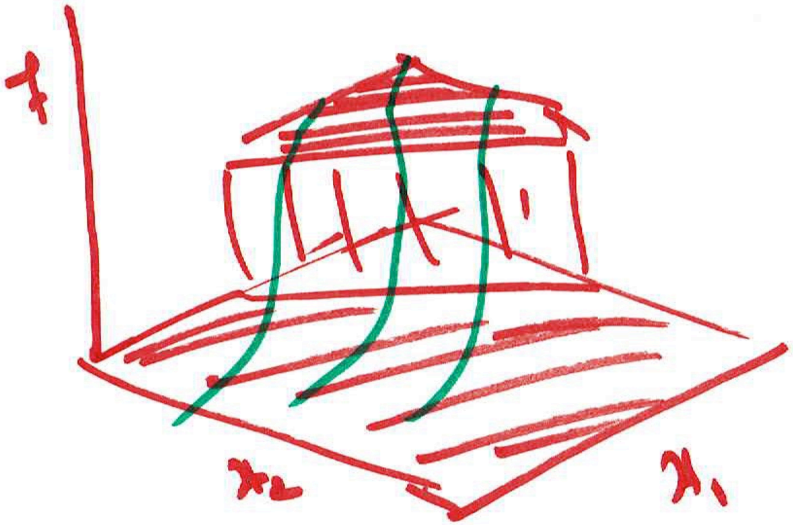
linear = x

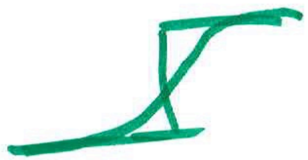


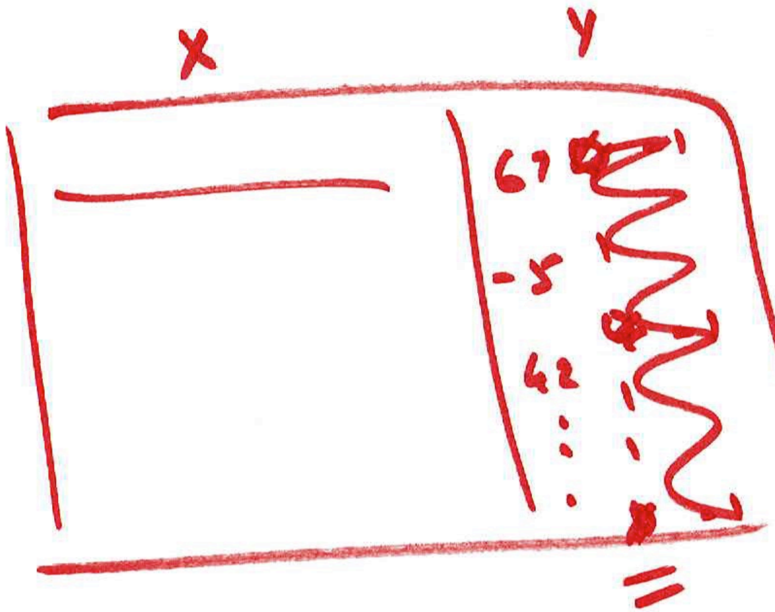
$$\text{Step} = f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

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









	Output nodes	hidden layer
Classification →	Sigmoid, tanh Softmax	Sigmoid ✓ tanh ✓ ReLU ✓ ~~linear~~
Reg →	linear	

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

Softmax

z_1

z_2

z_3

z_4

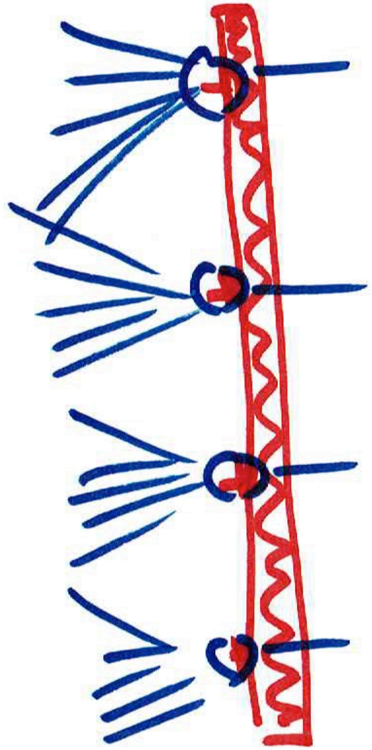
$$\frac{e^{z_i}}{\sum_j e^{z_j}}$$

$\rightarrow \dfrac{e^{z_1}}{\sum e^{z_k}}$

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

$\rightarrow$

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



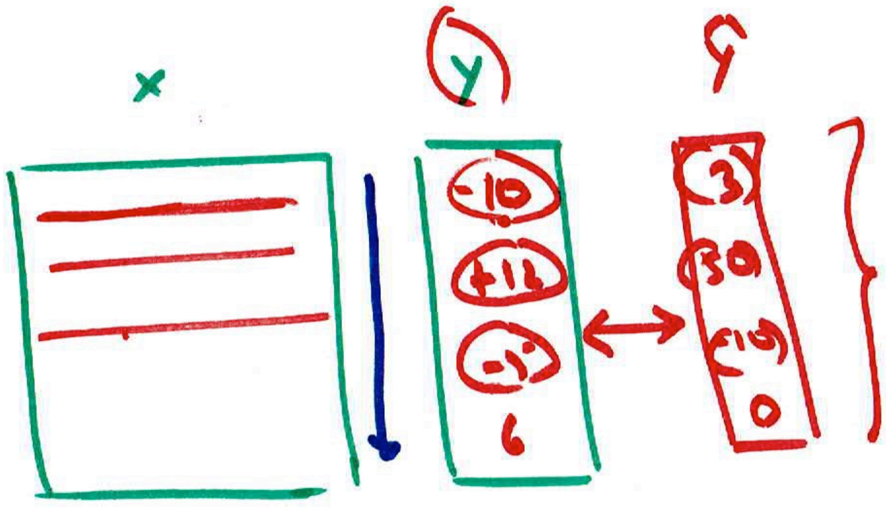
0.1

0.2

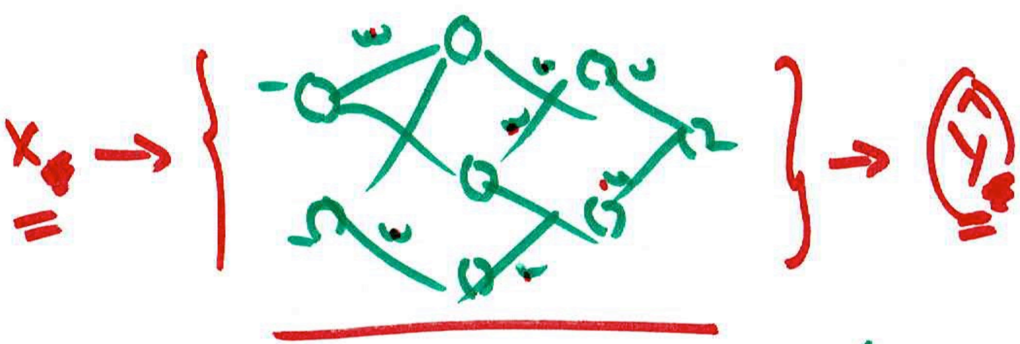
0.4

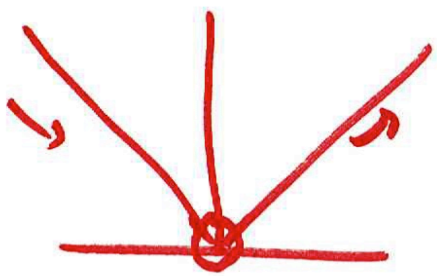
0.3

1



$$\begin{bmatrix} -10 \\ +12 \\ -5 \\ 6 \end{bmatrix} \quad \begin{bmatrix} 3 \\ 30 \\ -10 \\ 0 \end{bmatrix}$$





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

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

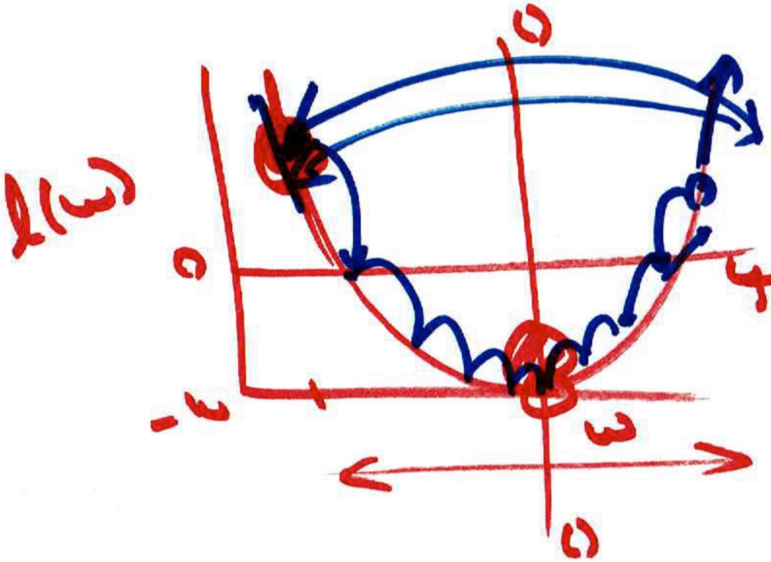
Classificati $L(y, \hat{y}) = -\left(y_i \log(\hat{y}_i) + (1-y_i)(\log(1-\hat{y}_i))\right)$

Cross entropy loss

How min $L(y, \hat{y})$ by ~~chag~~ Changing

~~tx~~ my $W^1, W^2, \dots W^n$

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



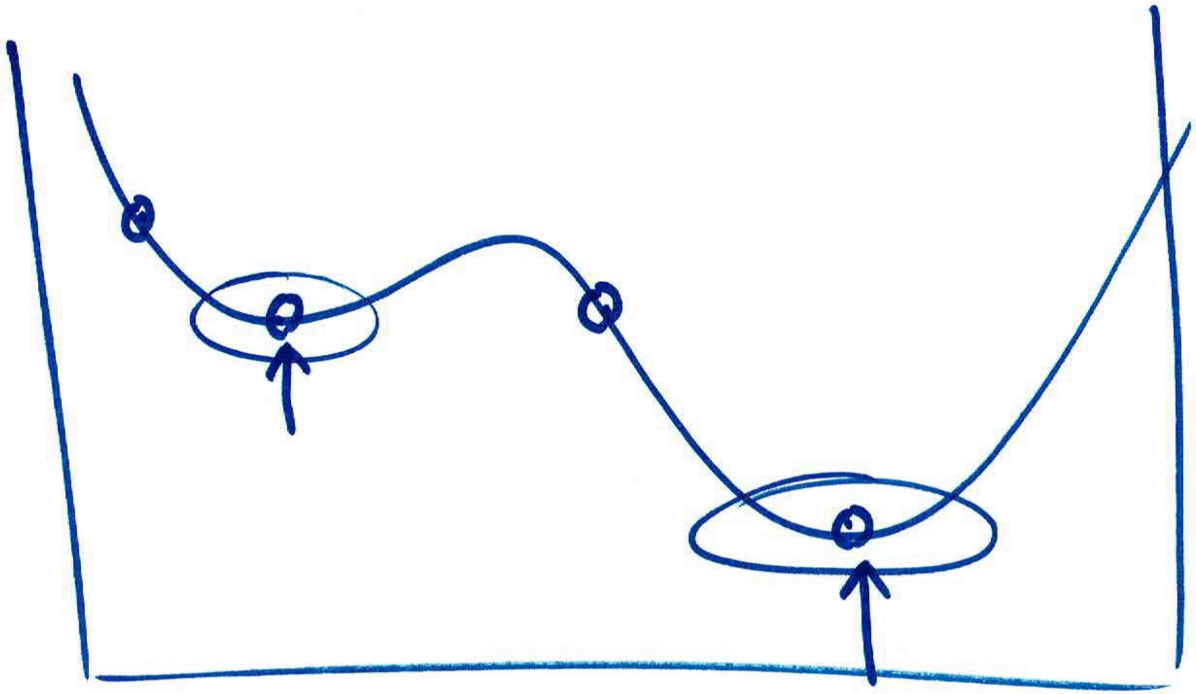
$$y = x^2 - 10 = -10$$

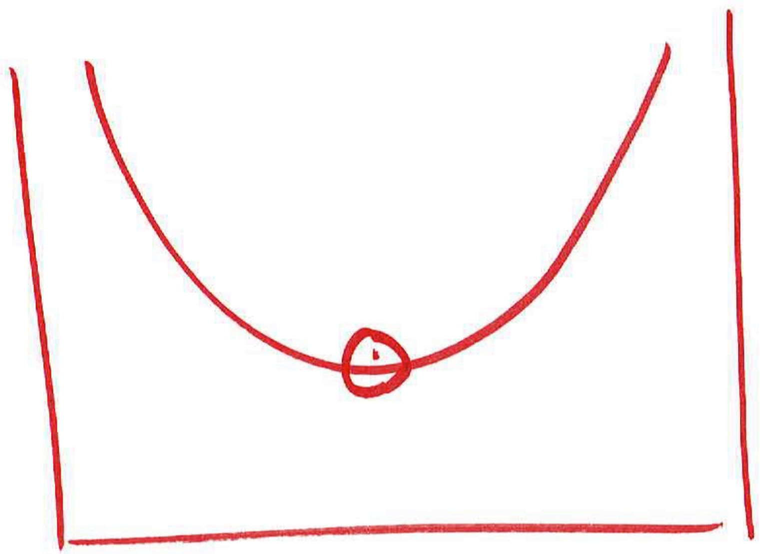
$$\frac{dy}{dx} = 2x = 0 \qquad x = 0$$

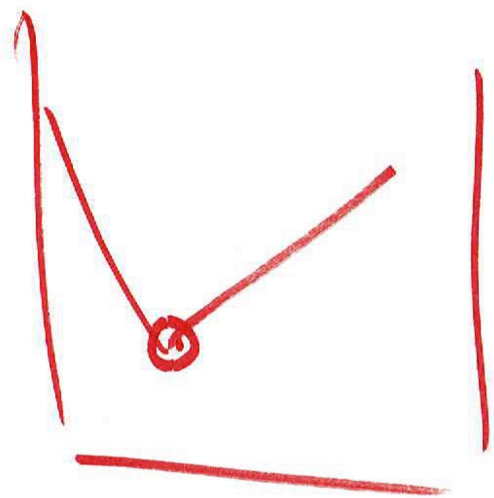
$$\frac{dL}{dw} = \boxed{\quad} = 0 \qquad \frac{dl}{dw}$$

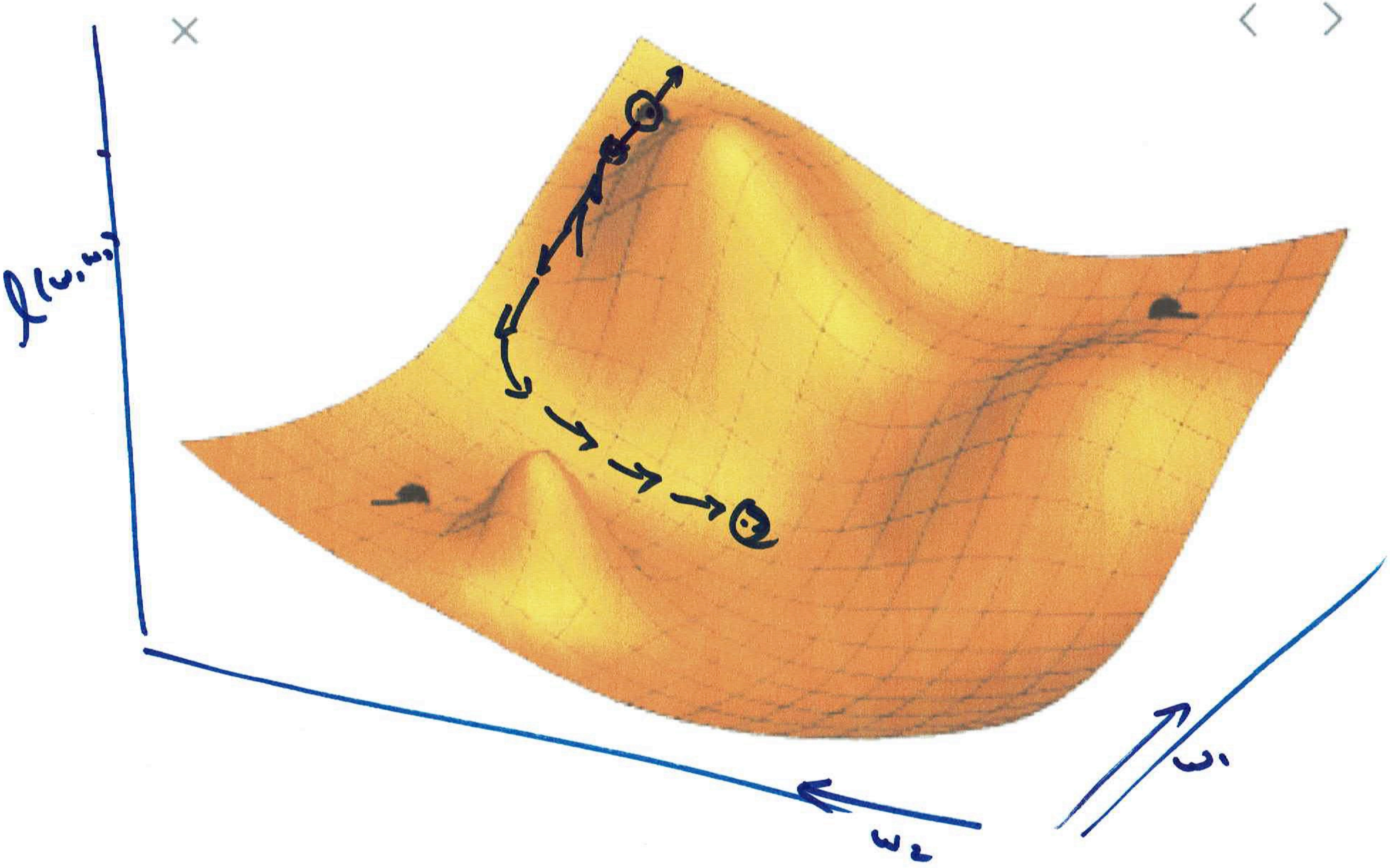
$$w^{new} = w - \eta \nabla_w l$$

learning rate








l(w,w₂)
w₂
w₁

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

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

SGD

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

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

Loss

Function of w (l(w))

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

Chain Rule

$$f(g(h(\ldots x \ldots)))$$

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

Back propagation