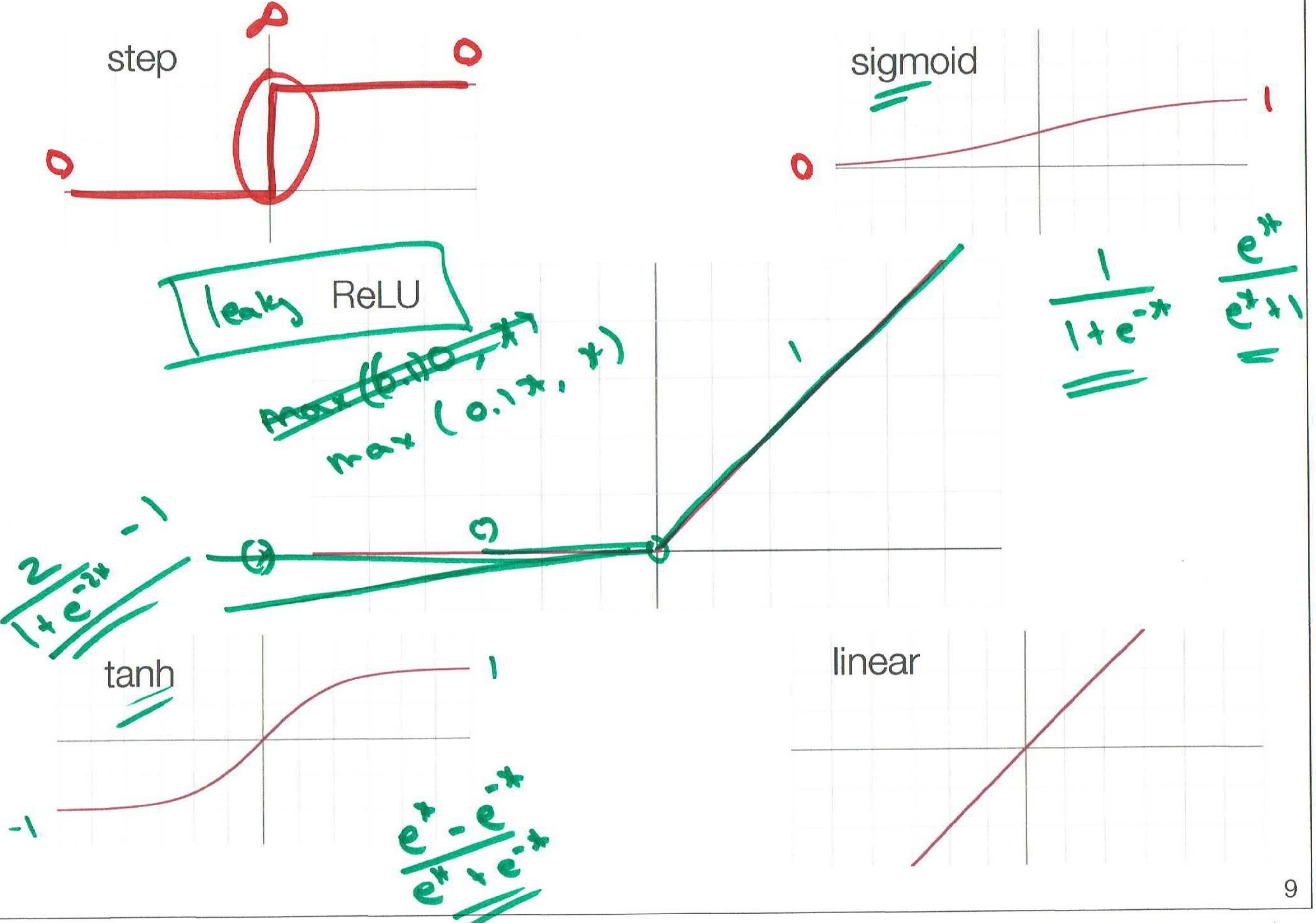
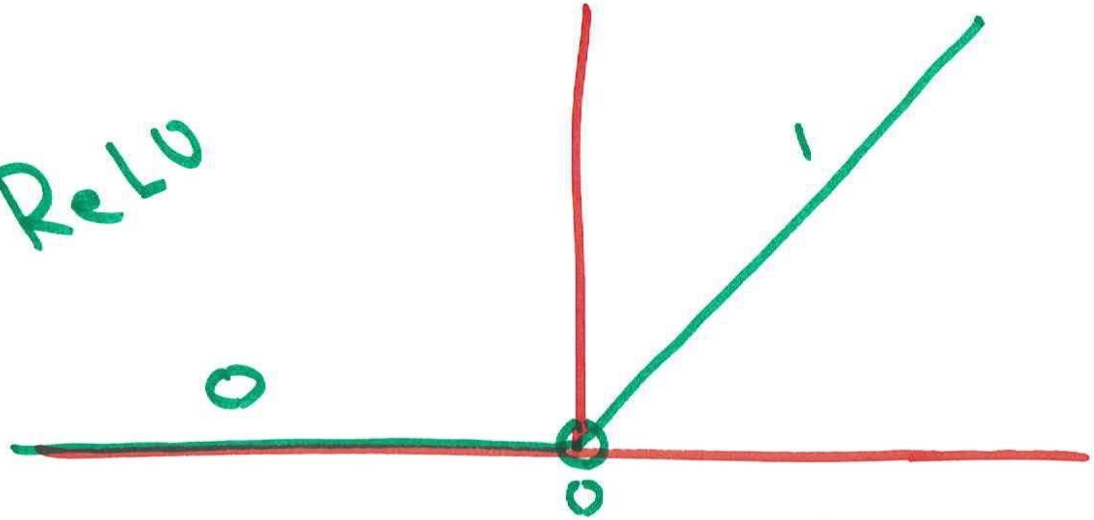
Activation Functions

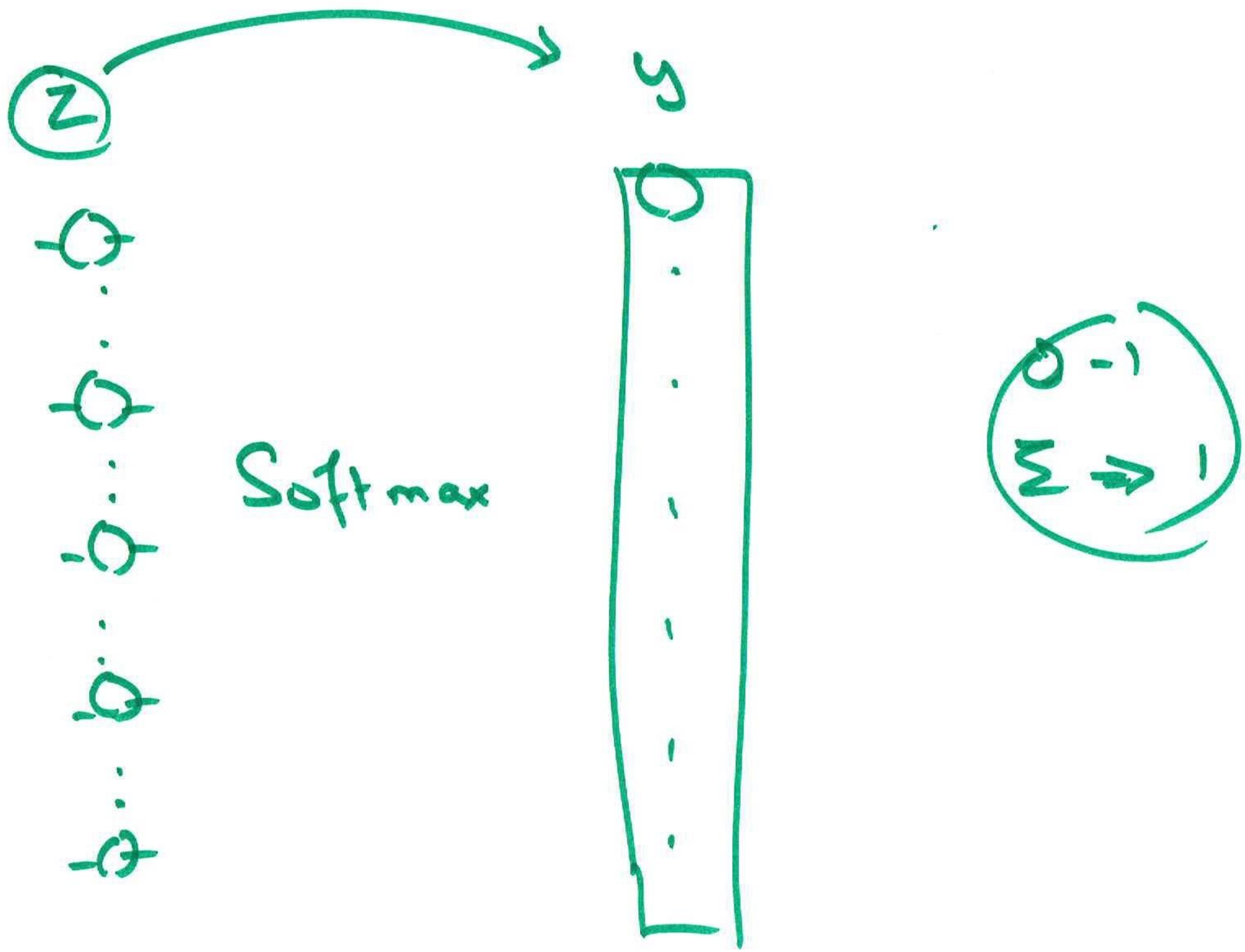


ReLU



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

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



z
y
Softmax
0 - 1
Σ ⇒ 1

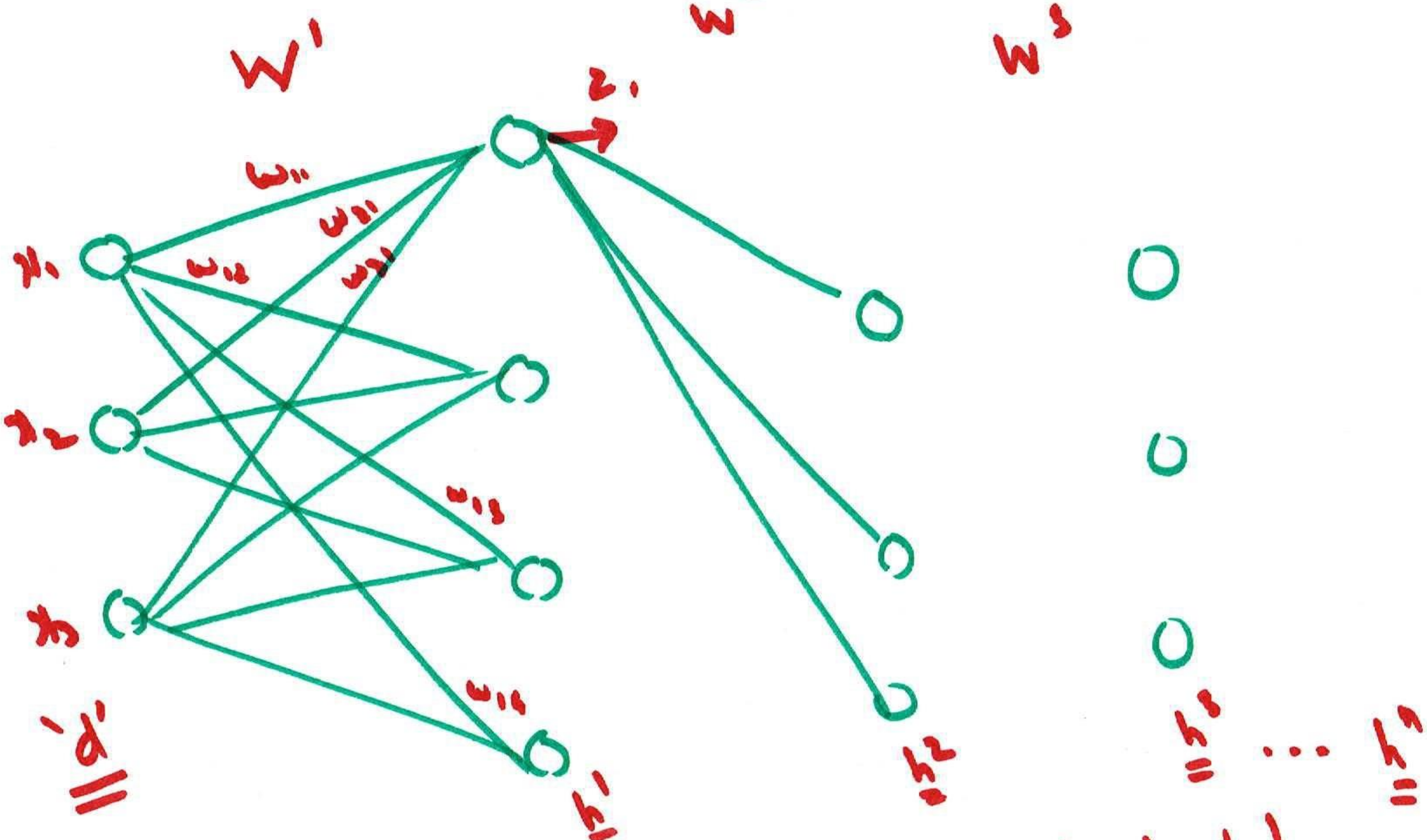
z

y

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

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

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



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

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

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

$h^1 \times d$ matrix

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

? ? ?

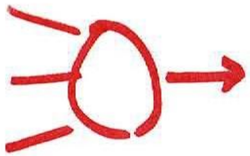
$\checkmark \sum w_{ij} x_i + b$

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

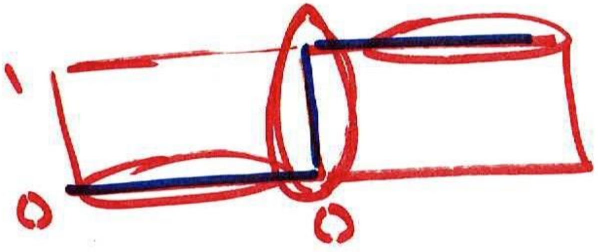
$\hat{y}$: $h^n \times 1$

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

b^2: $h^2 \times 1$



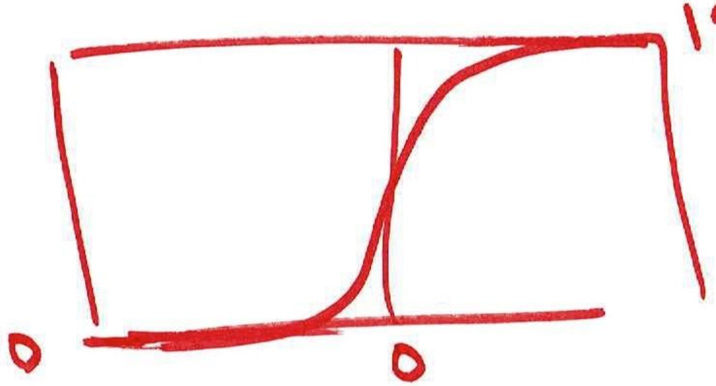
$$\underline{\mathrm{Step}\,(\overset{?}{w_1}x_1 + \overset{?}{w_2}x_1 + \overset{?}{w_3}x_3 + b)}$$

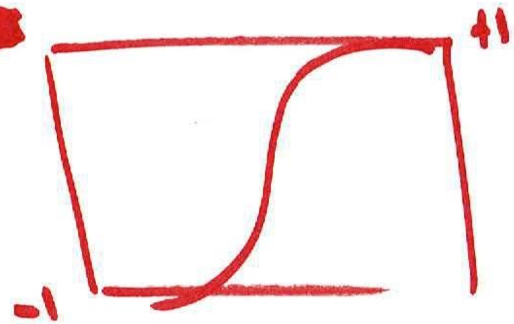


Sigmoid

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

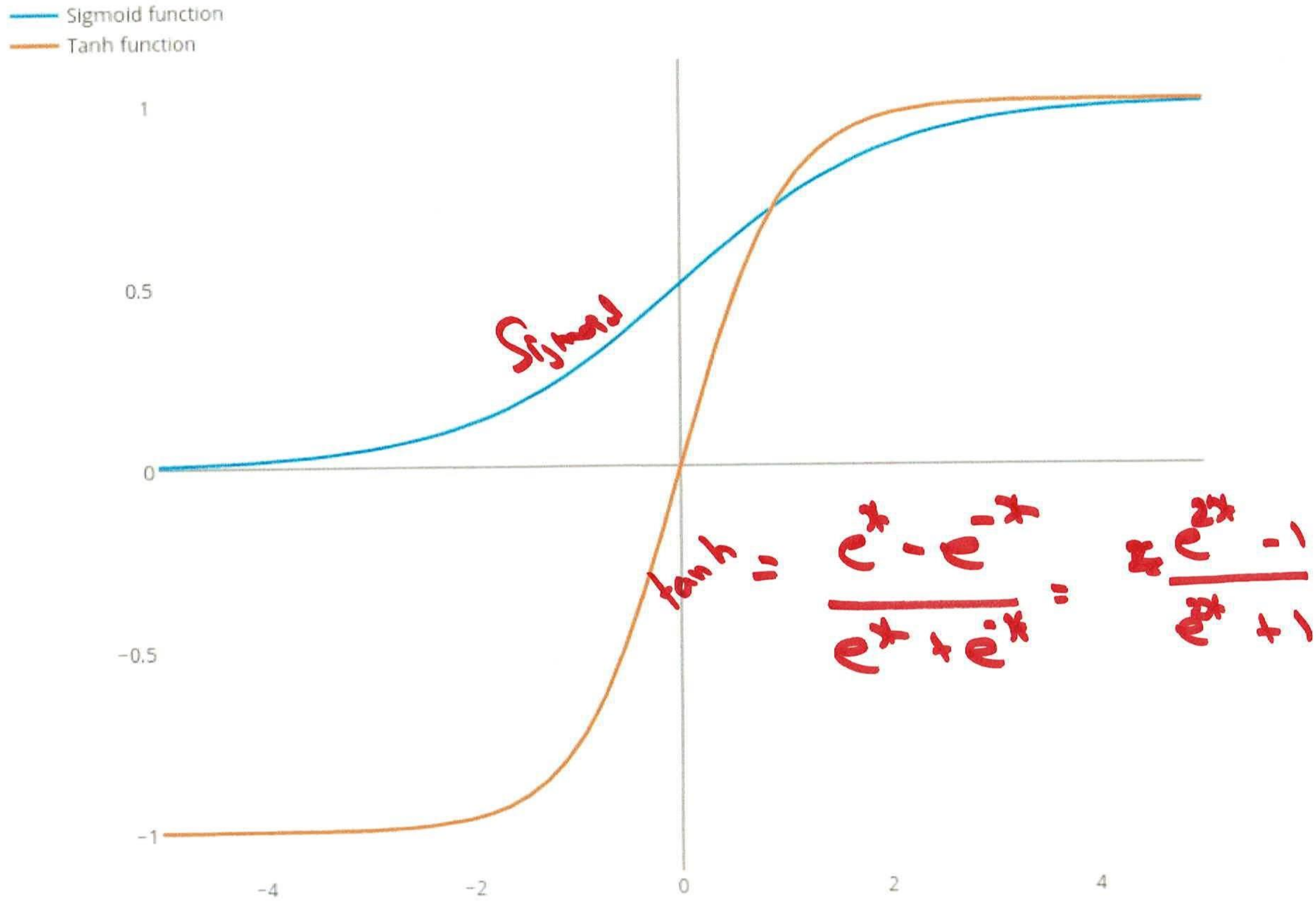
$$= \frac{e^x}{e^x+1}$$





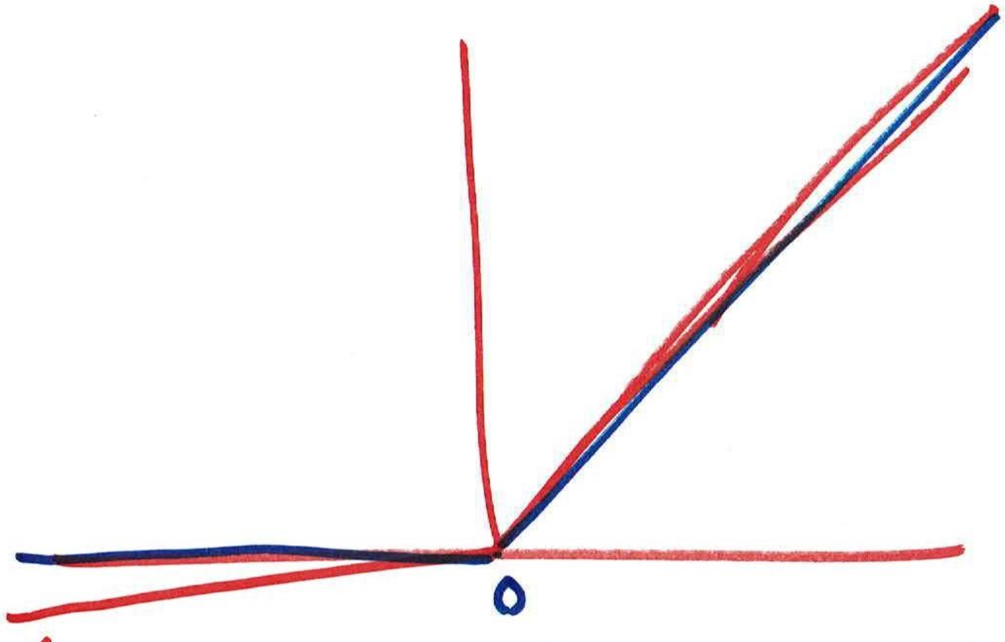
$$\boxed{\cancel{2\sigma(x)} - 1}$$

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



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

ReLU

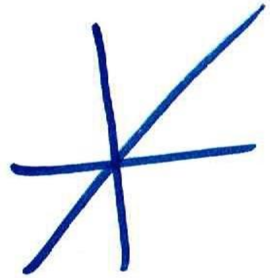


$$= \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

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

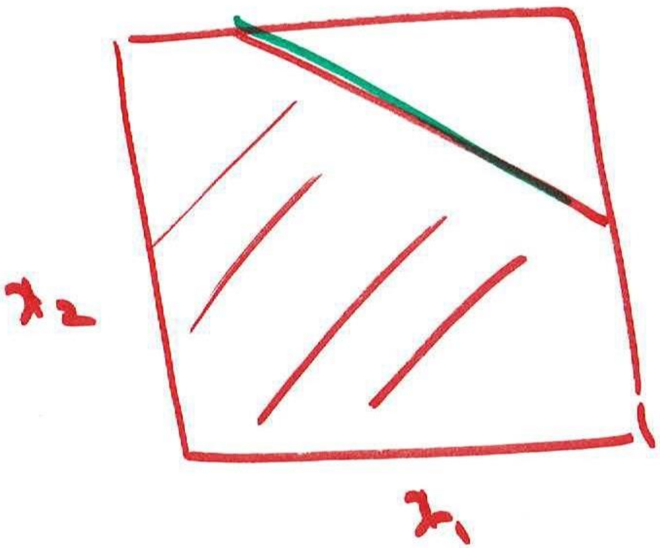
$\max(0.01x, x)$

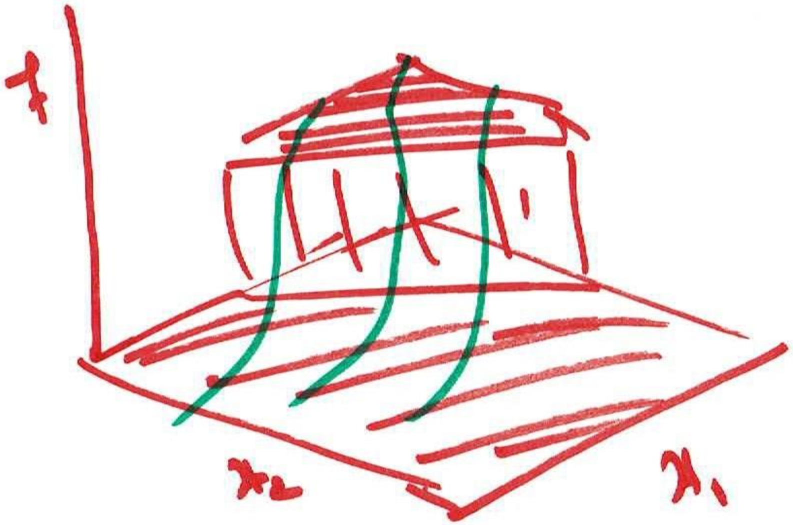
linear = x

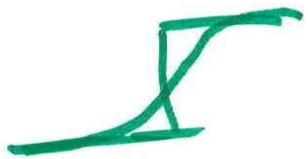


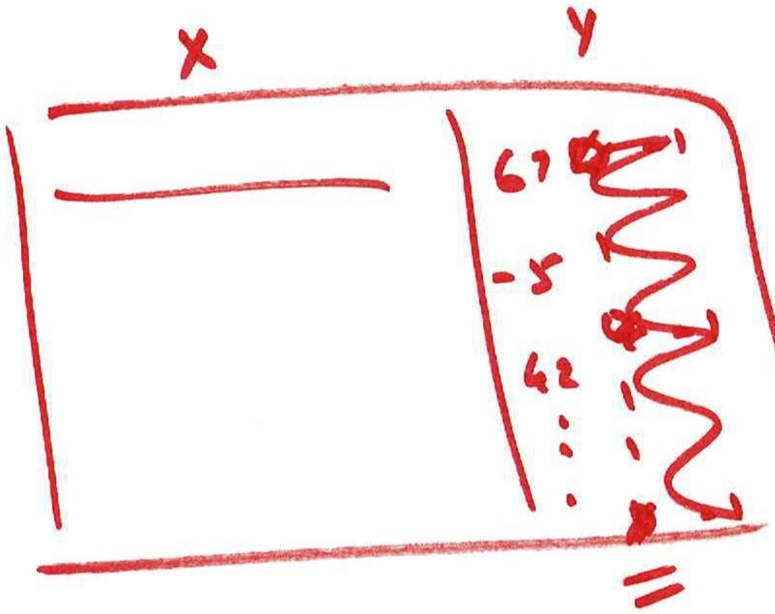
$$\text{Step} = f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

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









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

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

Softmax

z_1

z_2

z_3

z_4

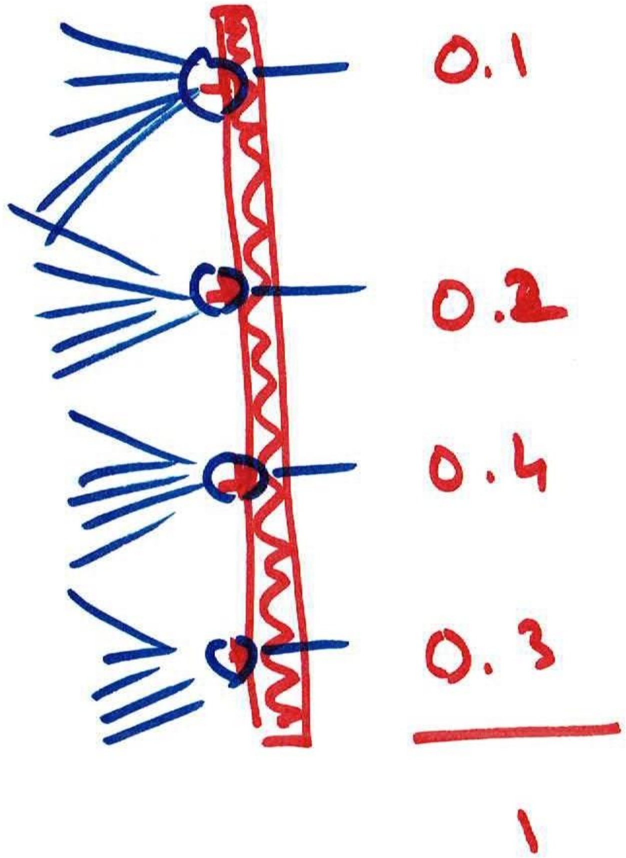
$$\frac{e^{z_i}}{\sum_j e^{z_j}}$$

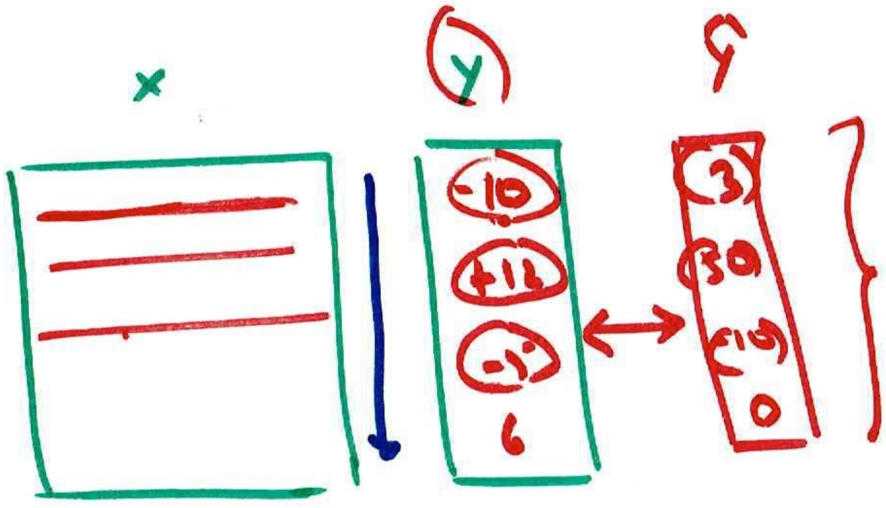
$\rightarrow \frac{e^{z_1}}{\sum e^{z_j}}$

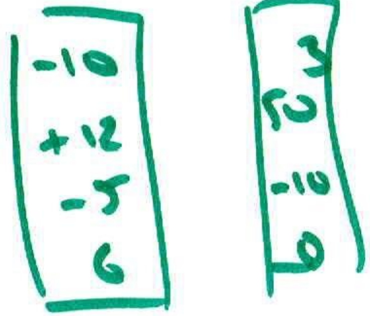
$\rightarrow \frac{e^{z_2}}{\sum e^{z_j}}$

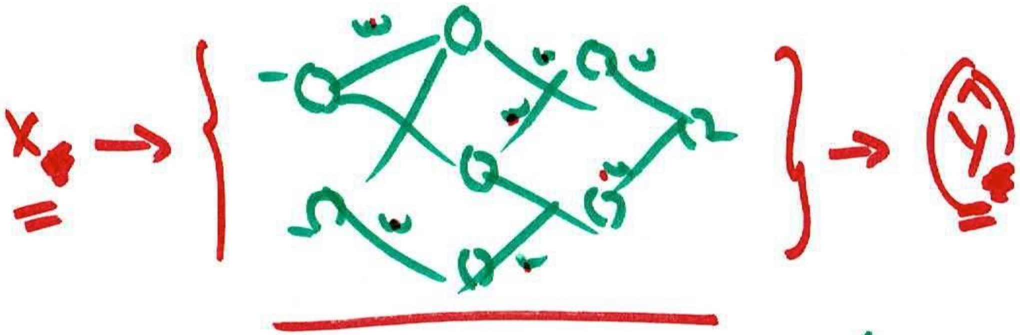
$\rightarrow \cdot$

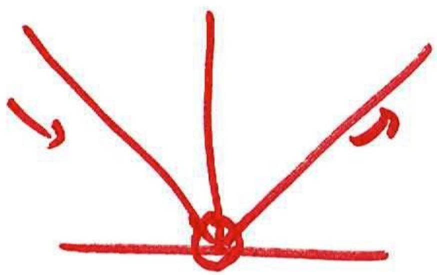
$\rightarrow \frac{e^{z_4}}{\sum e^{z_j}}$











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

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

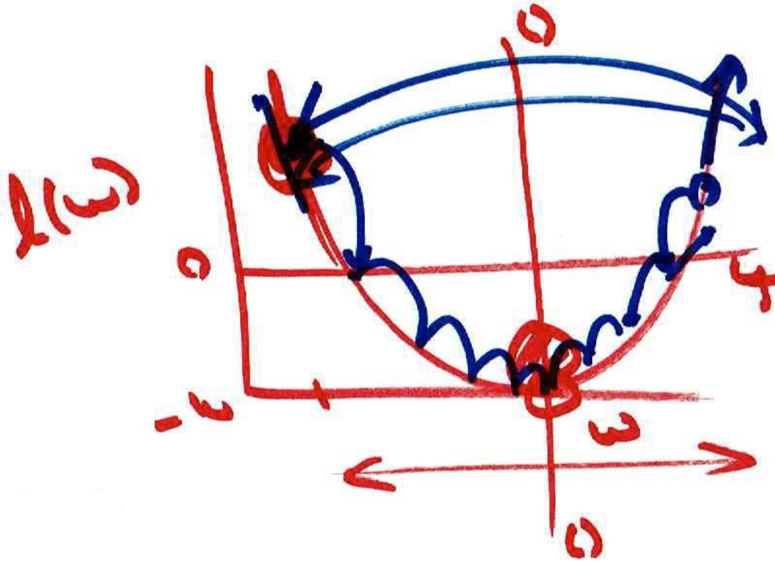
Classificati $L(y, \hat{y}) = -\left(y_i \log(\hat{y}_i) + (1-y_i)(\log(1-\hat{y}_i))\right)$

Cross entropy loss

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

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

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



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

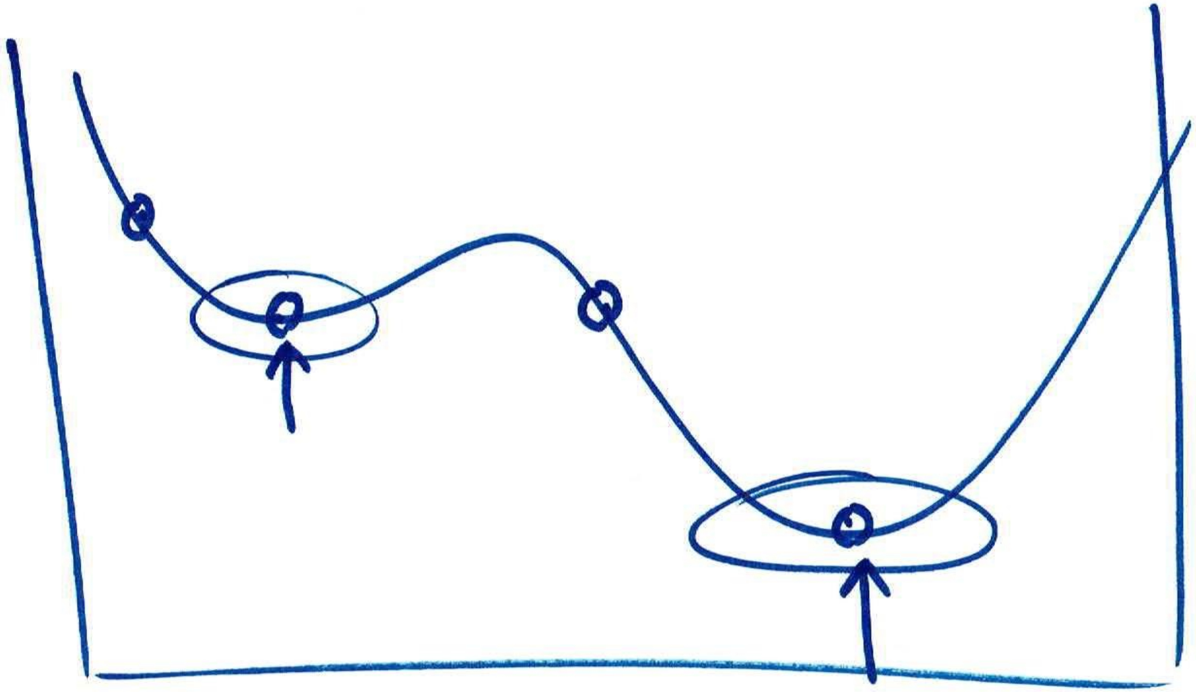
$$\frac{dy}{dx} = 2x = 0 \qquad x = 0$$

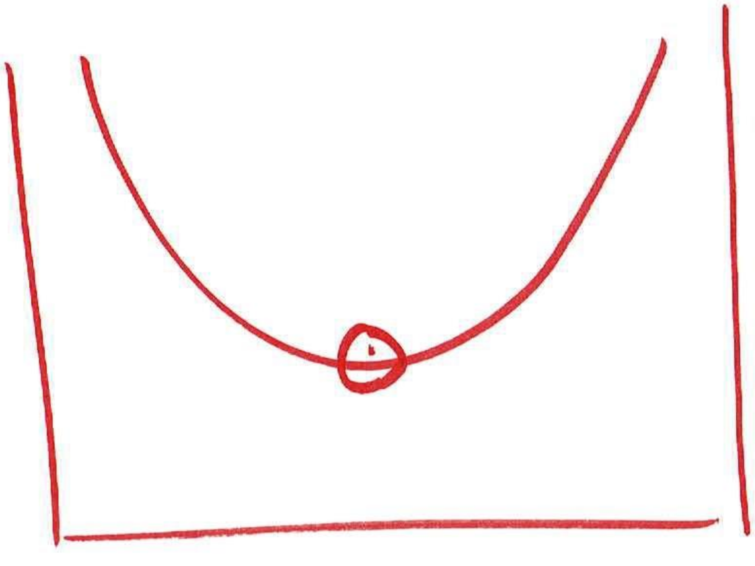
$$\frac{dL}{dw} = \boxed{\quad} = 0$$

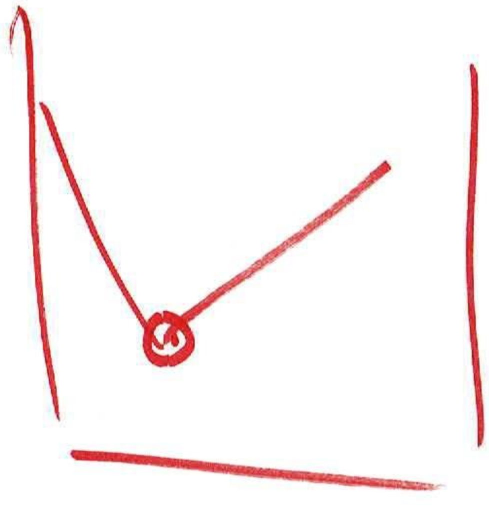
$$\frac{dl}{dw}$$

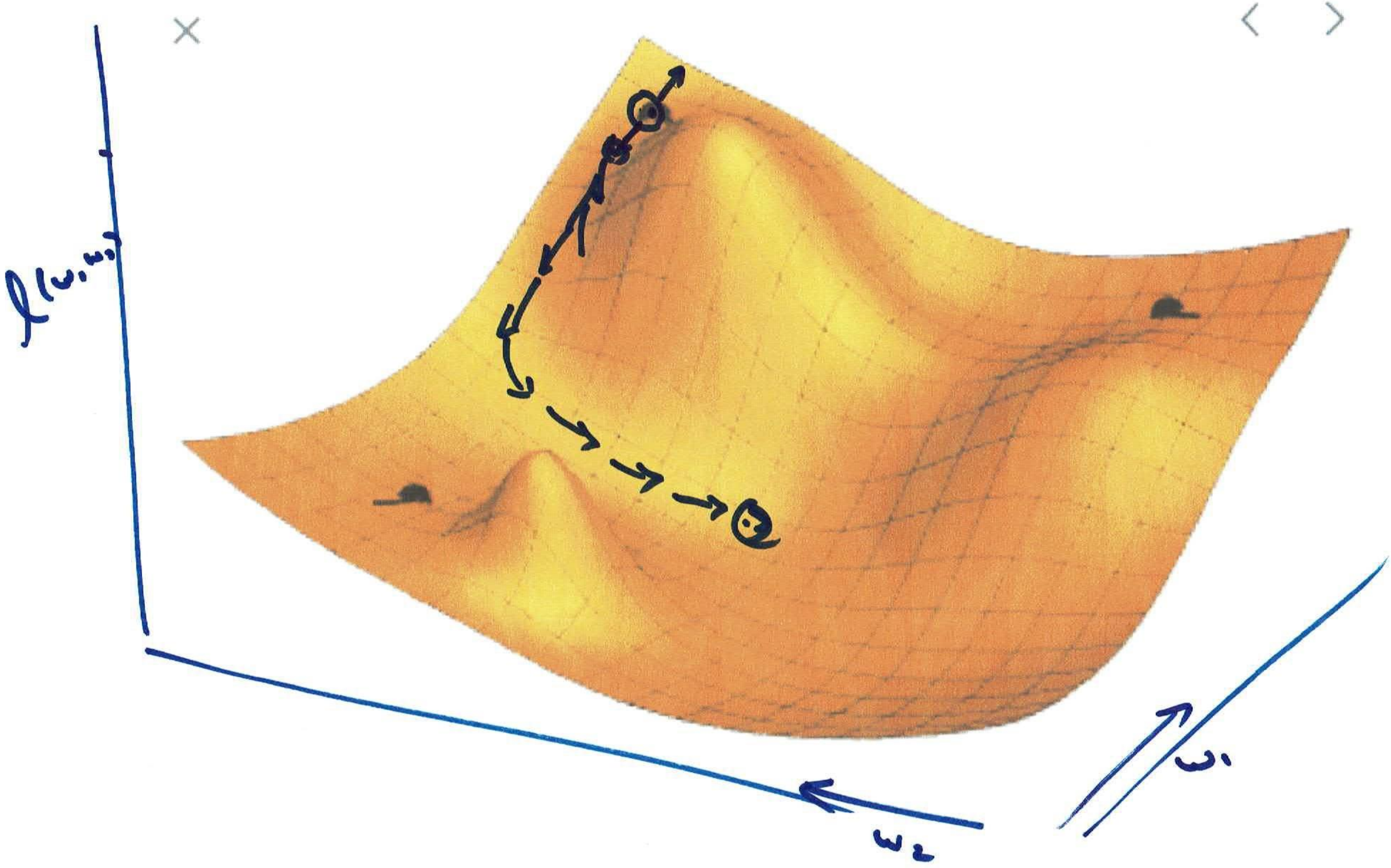
$$w^{new} = w - \eta \nabla_w l$$

learning rate









l(w₁,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

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

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

Loss

Function of w $(l(w))$

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

Chain Rule

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

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

Back propagation