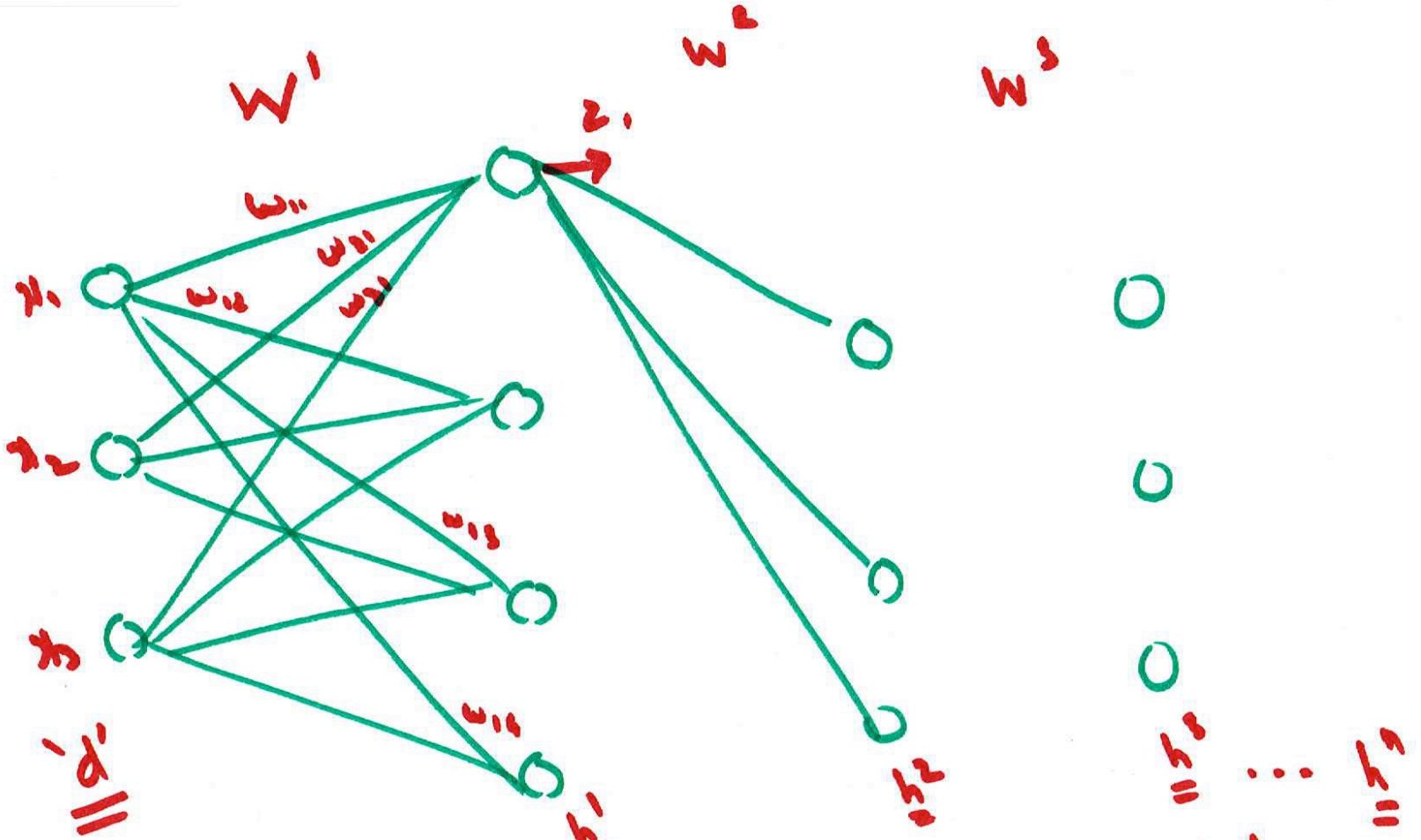
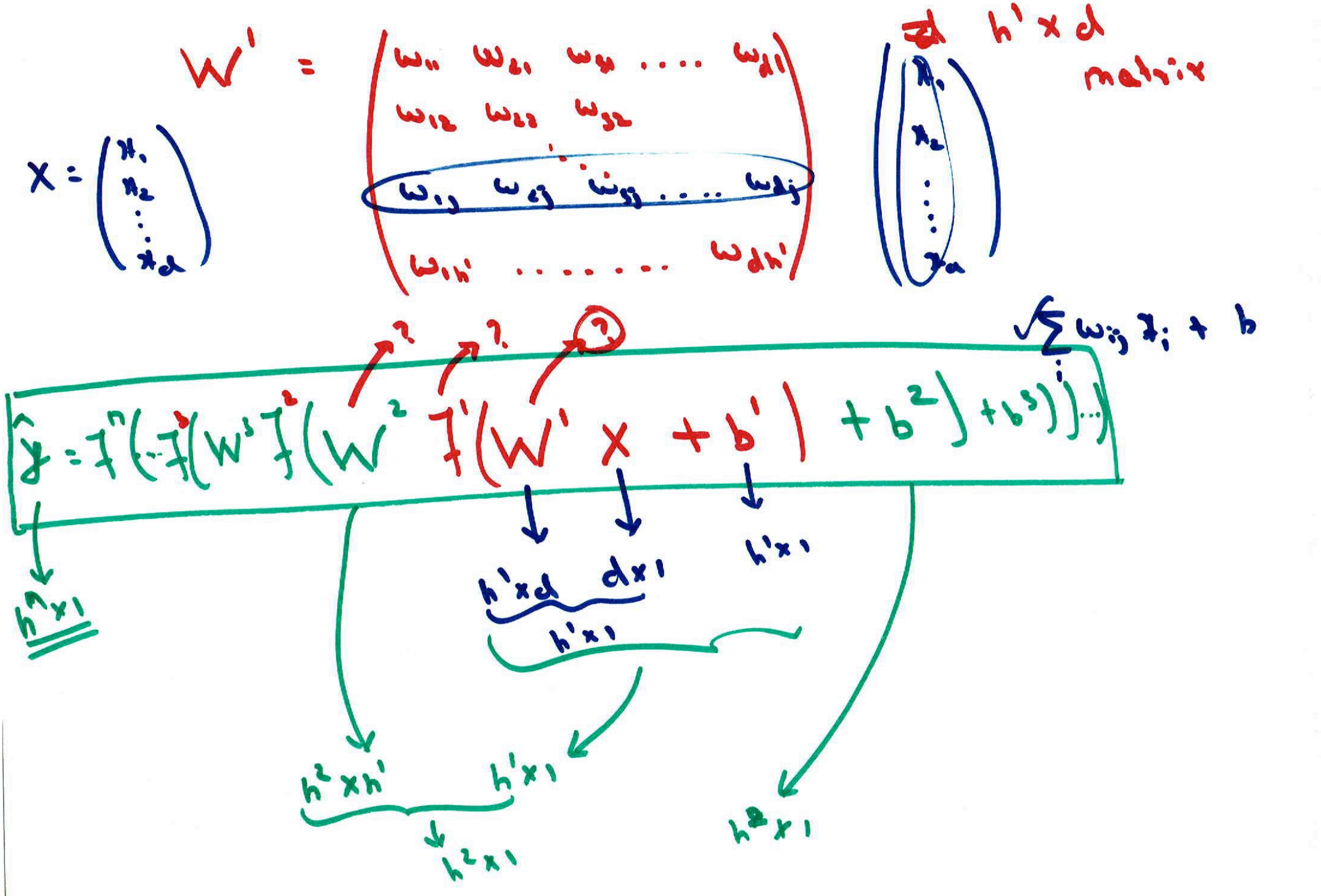
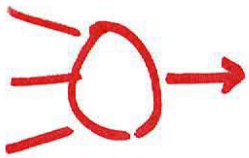


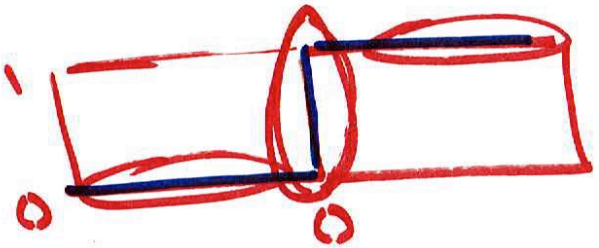
$$z_1 = f(\omega_{11} x_1 + \omega_{21} x_2 + \omega_{31} x_3 + b)$$

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


W^1 =
$h^1 \times d$ matrix
$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$
$\sum w_{ij} x_i + b$
$\hat{y} = f^n(\ldots f^2(W^2 f^1(W^1 X + b^1) + b^2) + b^3) \ldots)$
$h^n \times 1$
$h^1 \times d$ $d \times 1$
$h^1 \times 1$
$h^1 \times 1$
$h^2 \times h^1$
$h^2 \times 1$
$h^2 \times 1$



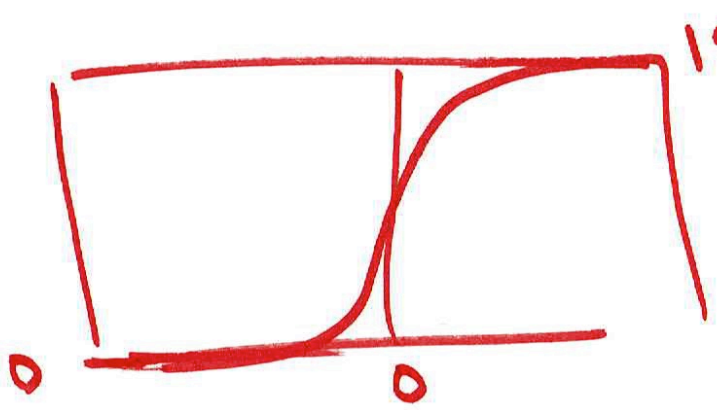
$$\text{Step}\left(w_1 x_1 + w_2 x_2 + w_3 x_3 + b\right)$$

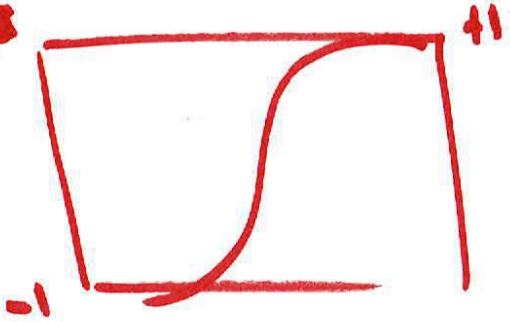


Sigmoid

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

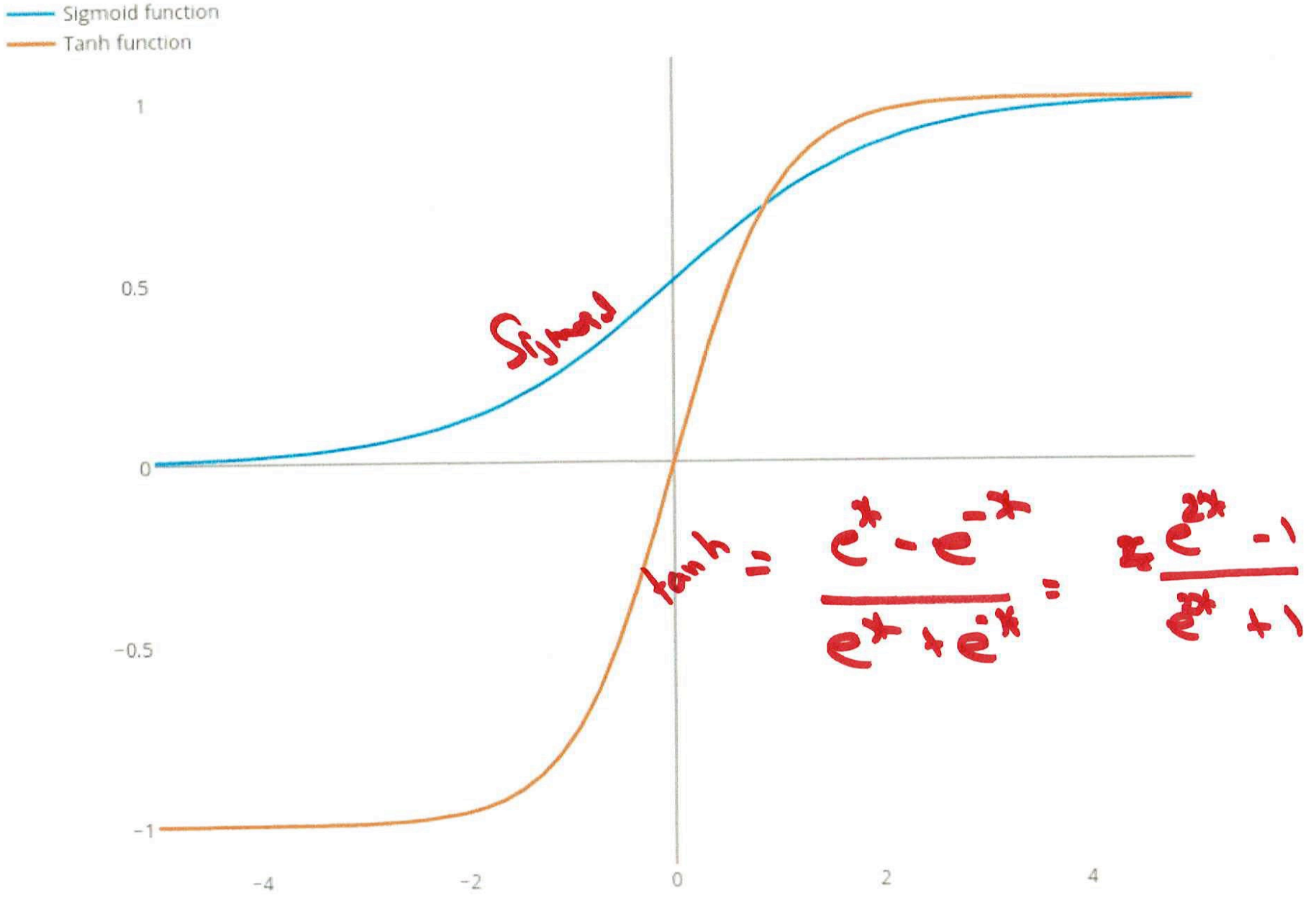
$$= \frac{e^{x}}{e^{x}+1}$$



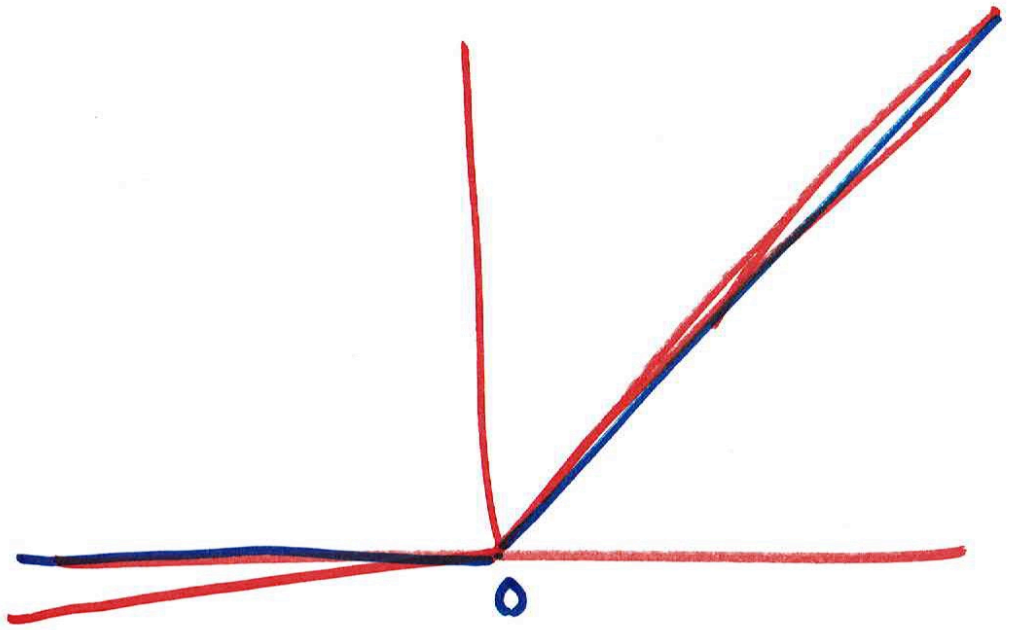


$$2\sigma(x) - 1$$

$$\tanh = 2\sigma(2x) - 1$$


Sigmoid function
Tanh function
1
0.5
0
−0.5
−1
−4
−2
0
2
4
Sigmoid
tanh = $\frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}$

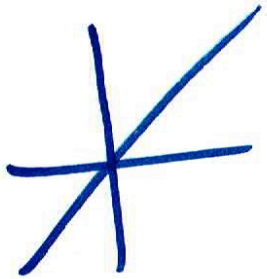
ReLU



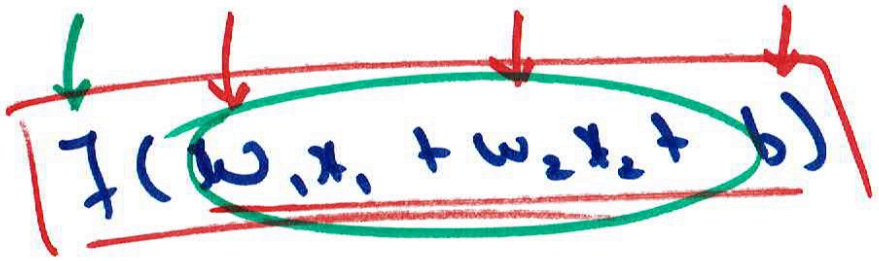
$$= \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

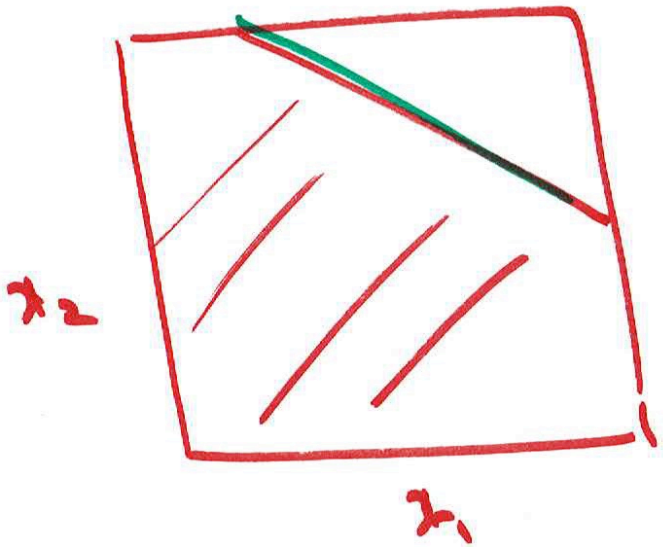
$$= \max(0, x)$$

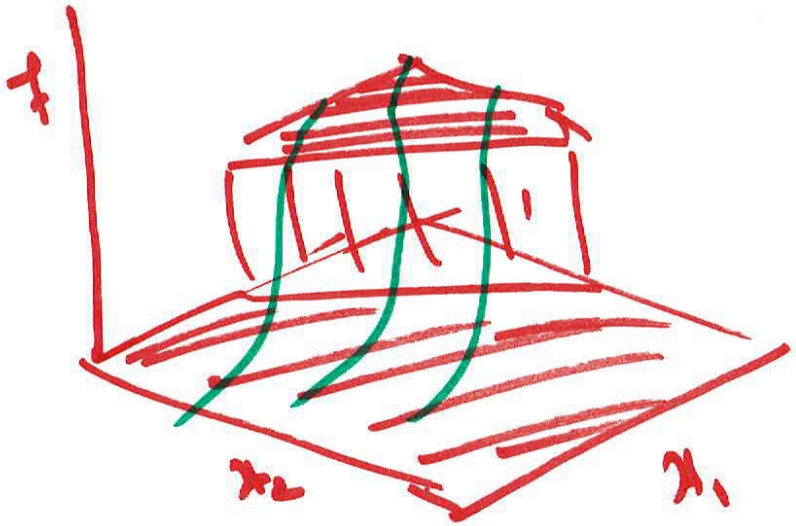
$$\max(0.01x, x)$$

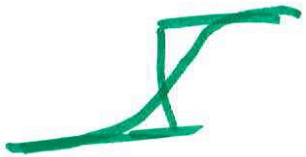
linear = 

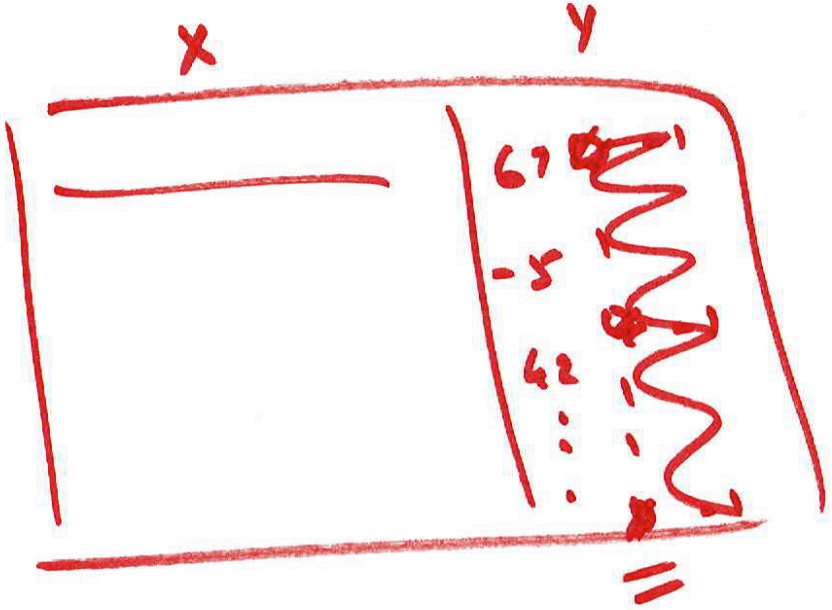
$$\text{Step} = f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$











Output nodes

hidden layer

Classification → Sigmoid, tanh, Softmax

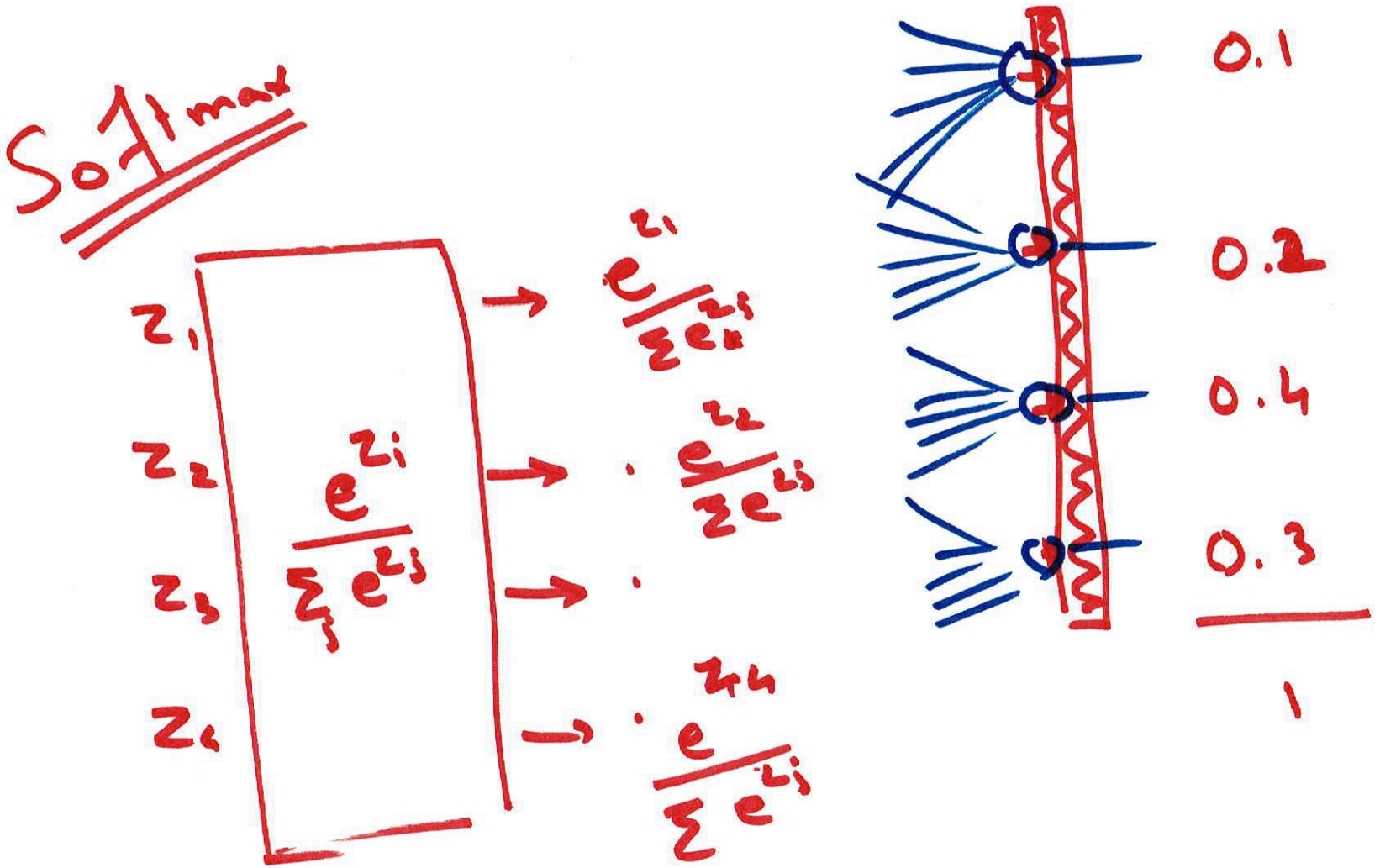
Reg → linear

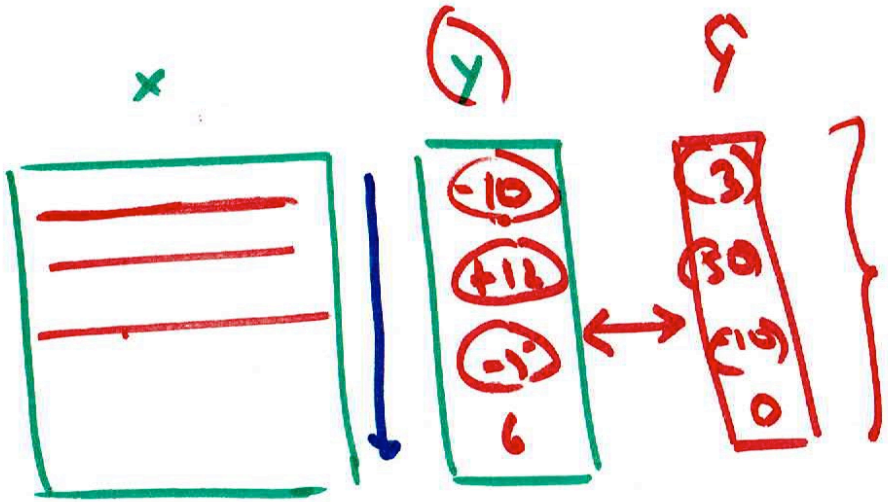
Sigmoid ✓

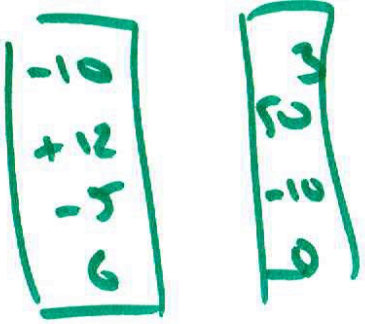
tanh ✓

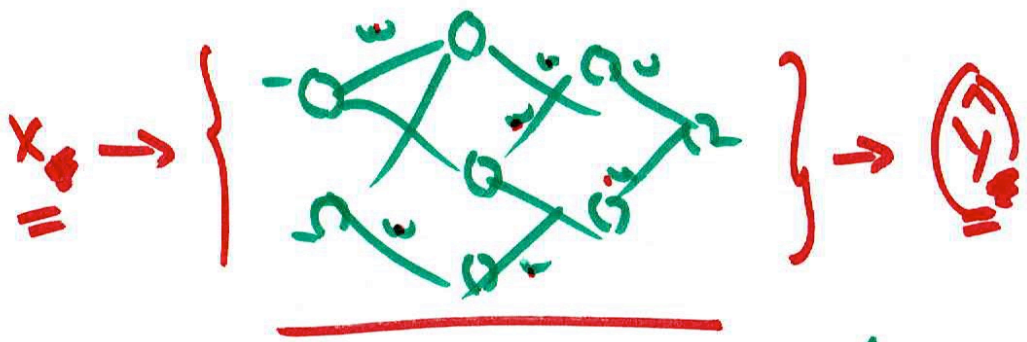
ReLU ✓

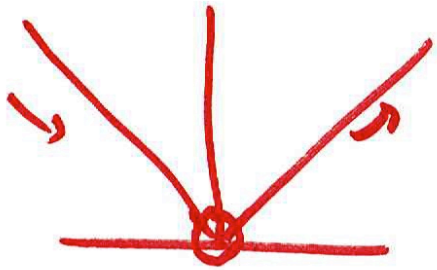
$$\hat{a} + \hat{b}(a + bx)$$


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









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

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

L_2 loss
MSE
SSE

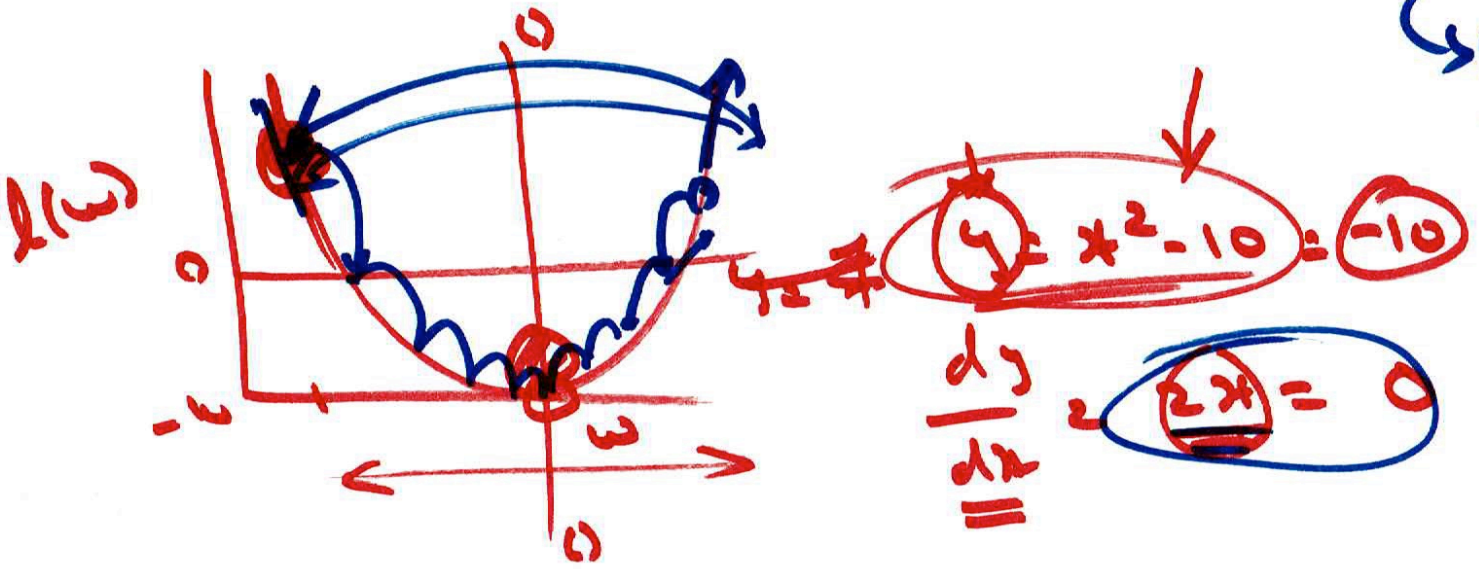
Reg

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

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 ~~Changing~~ Changing my $W^1, W^2, \ldots W^n$



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

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

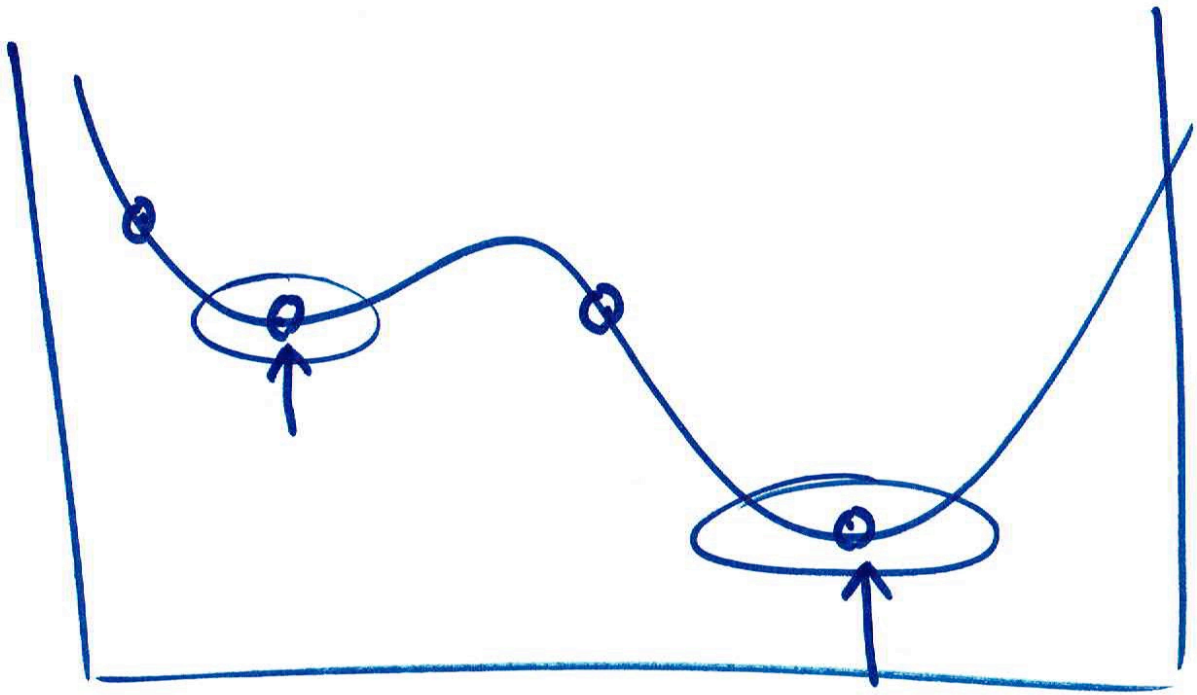
$x = 0$

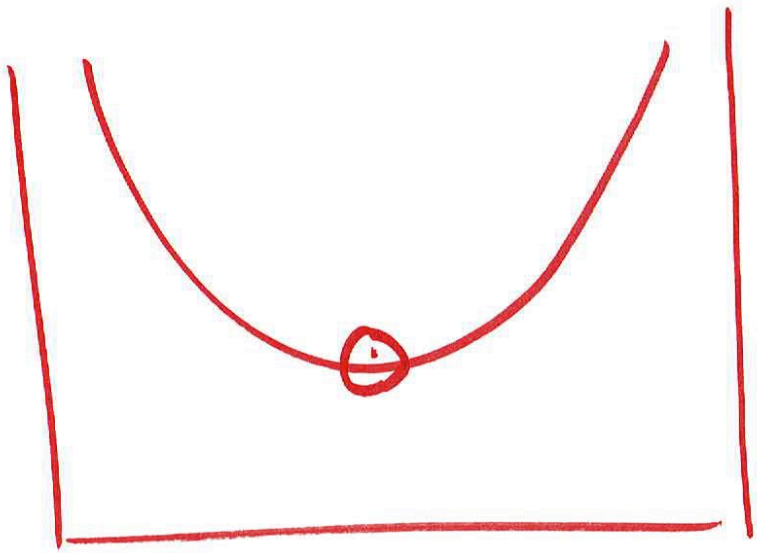
$\frac{dL}{dw} = \square = 0$

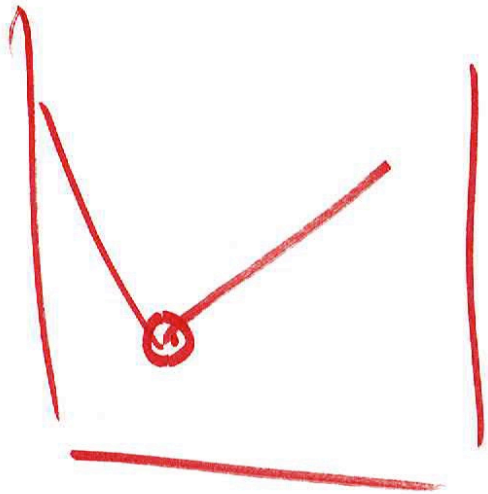
$\frac{dl}{dw}$

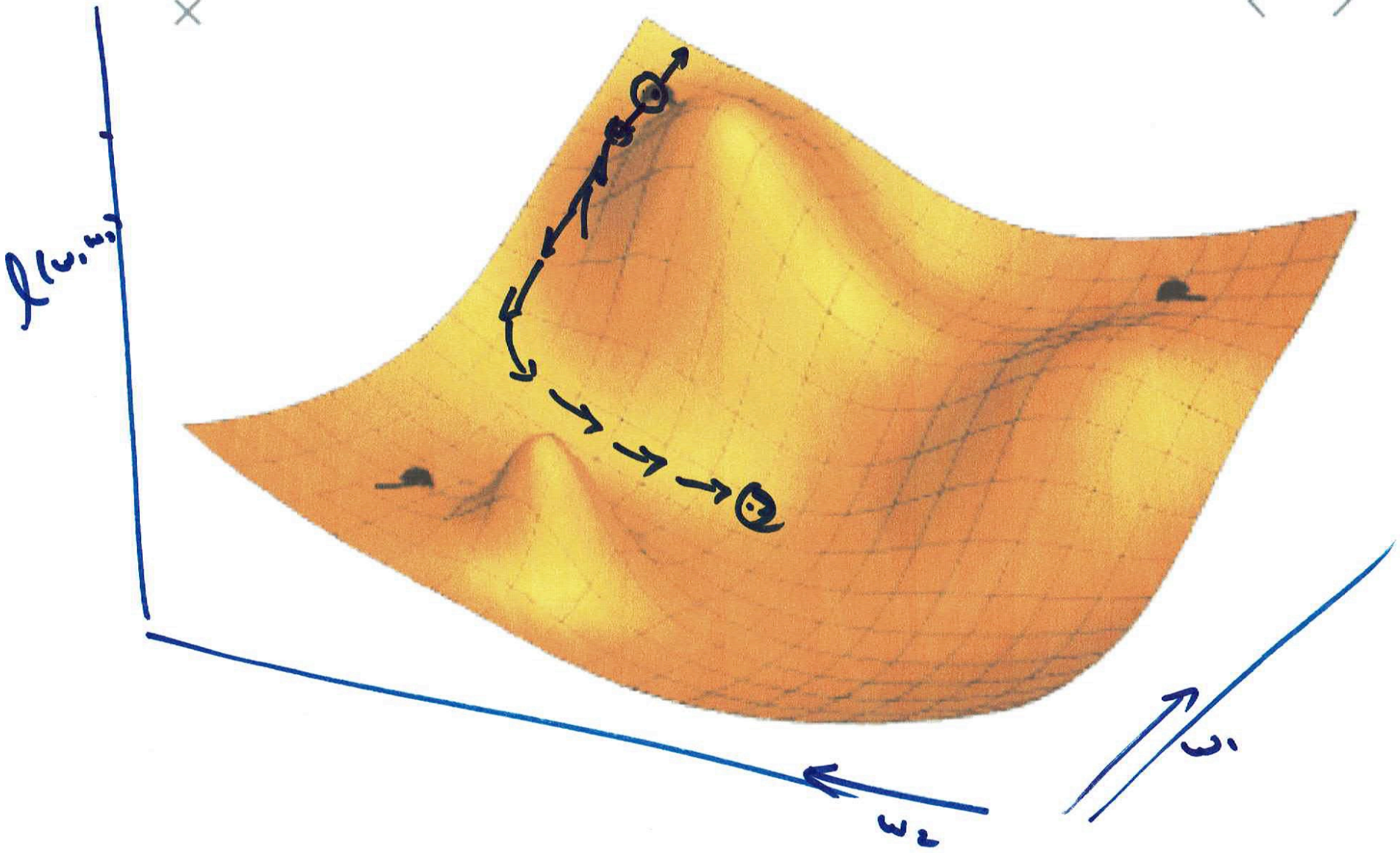
$w^{new} = w - \eta \nabla_w l$

η ↑ learning rate








$h(w_1, w_2)$
w_2
w_1

$$\omega^{new} = \omega^{old} - \eta \nabla_\omega l(\omega)$$

$$= \omega^{old} - \frac{1}{N} \eta \sum \nabla_\omega l_i(\omega) \leftarrow$$

SGD

$$\omega^{new} = \omega^{old} - \eta \nabla_\omega l_i(\omega) \leftarrow$$

$$\omega^{new} = \omega^{old} - \frac{1}{N} \eta \sum_{\text{over a mini batch}} \nabla_\omega l_i(\omega)$$

Function of w $(\ell(w))$

Loss

$$L: \frac{1}{N}\sum_{i} \left(y_i = f^n(\ldots f(W^2 f^1(W^1 x + b^1) + b^2)) \right)^2$$

Chain Rule

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

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

Back propagation