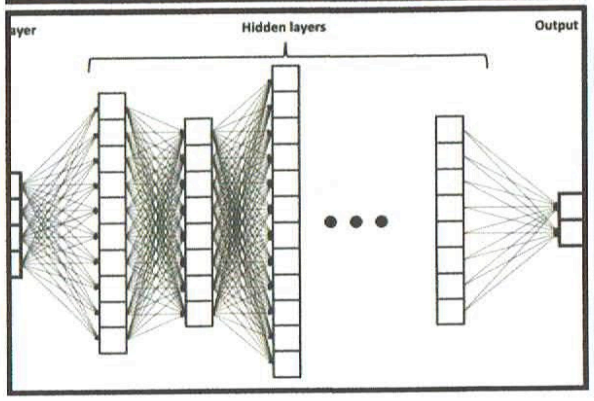
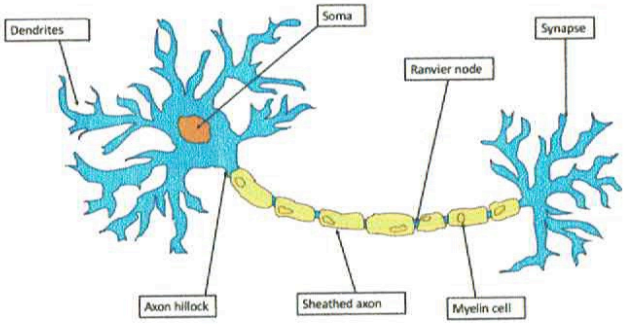
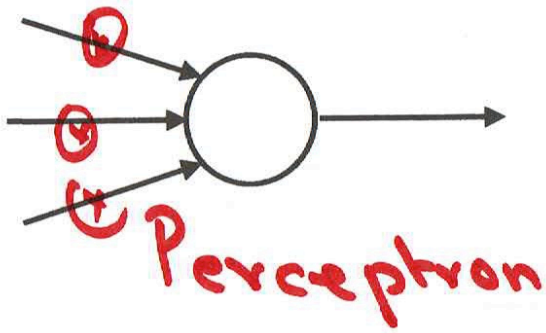
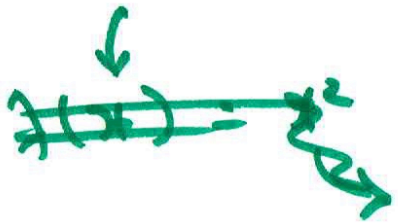
Building towards a complex task!

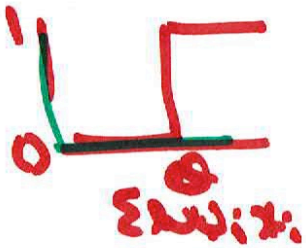






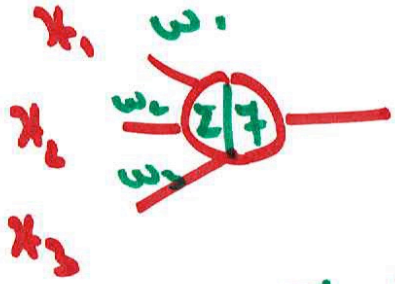
Perceptron!



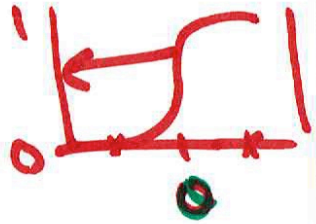


- What does it do?

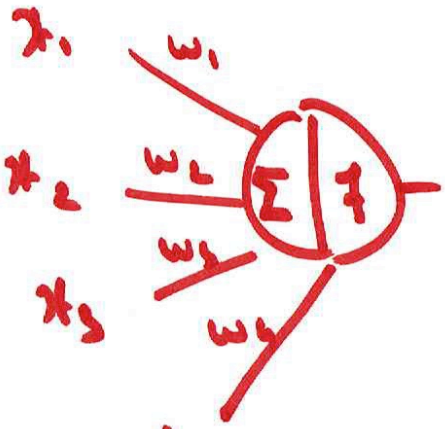
1958



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



- Bias



$$f(\Sigma w_i x_i + w_4)$$

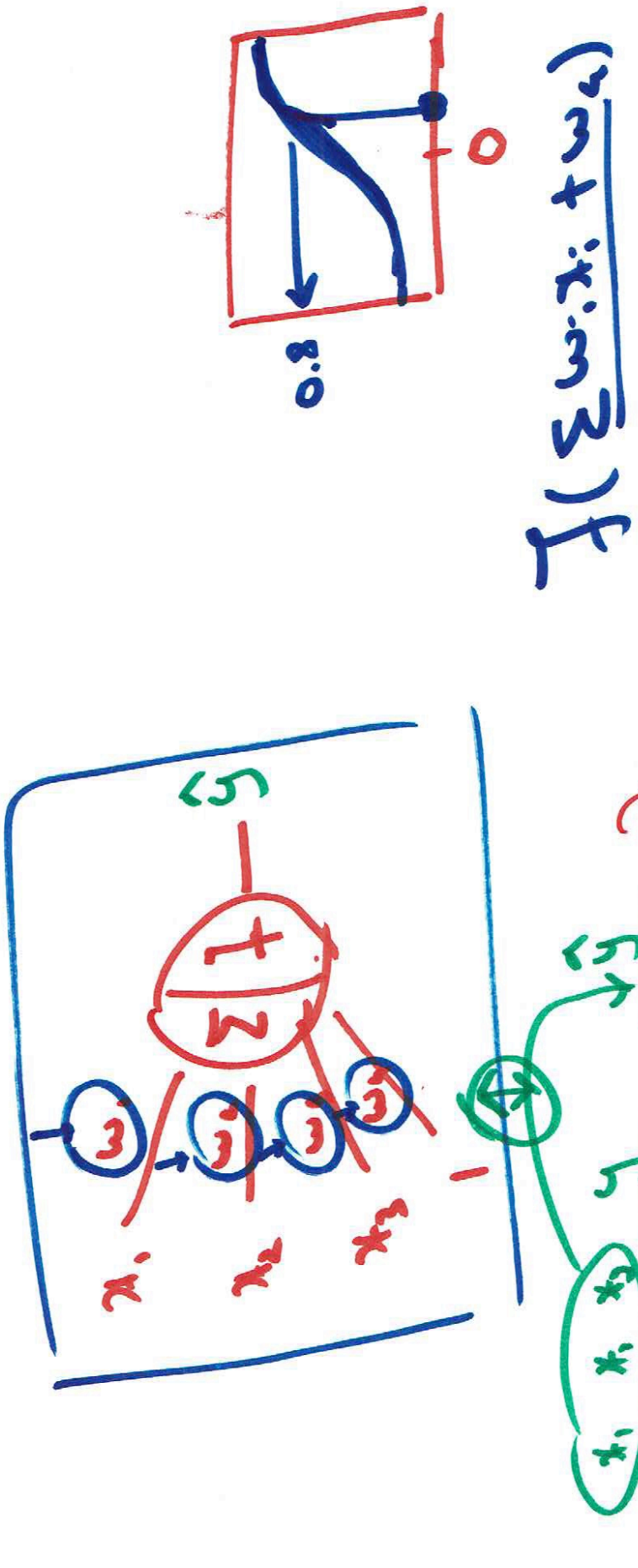
Perceptron

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

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

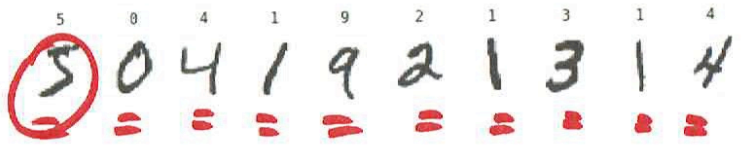
$\theta = 1 \Rightarrow$ "OR"

$\theta = 3 \Rightarrow$ "AND"

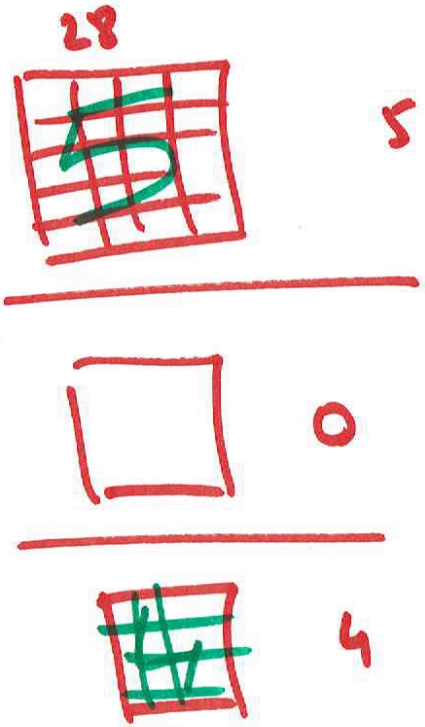


An Example

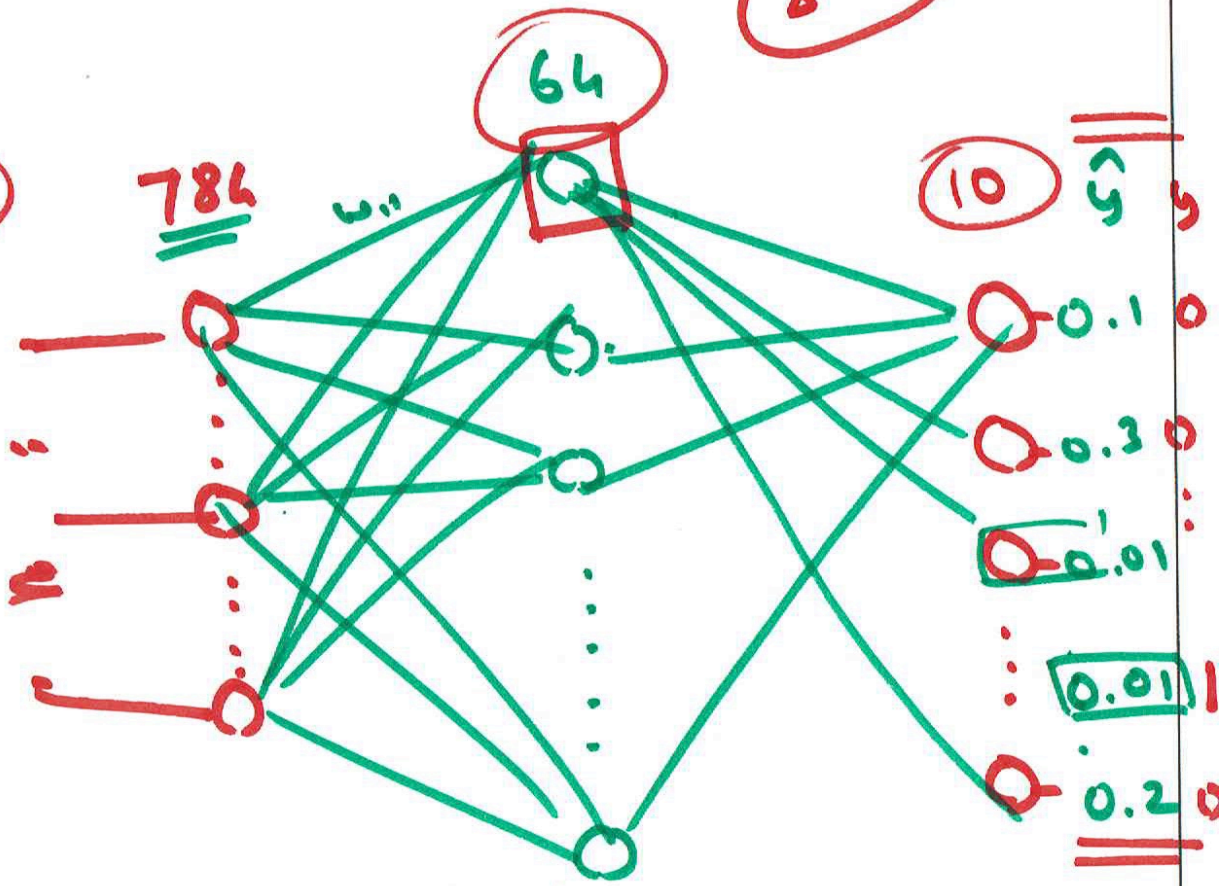
MNIST



60000 train



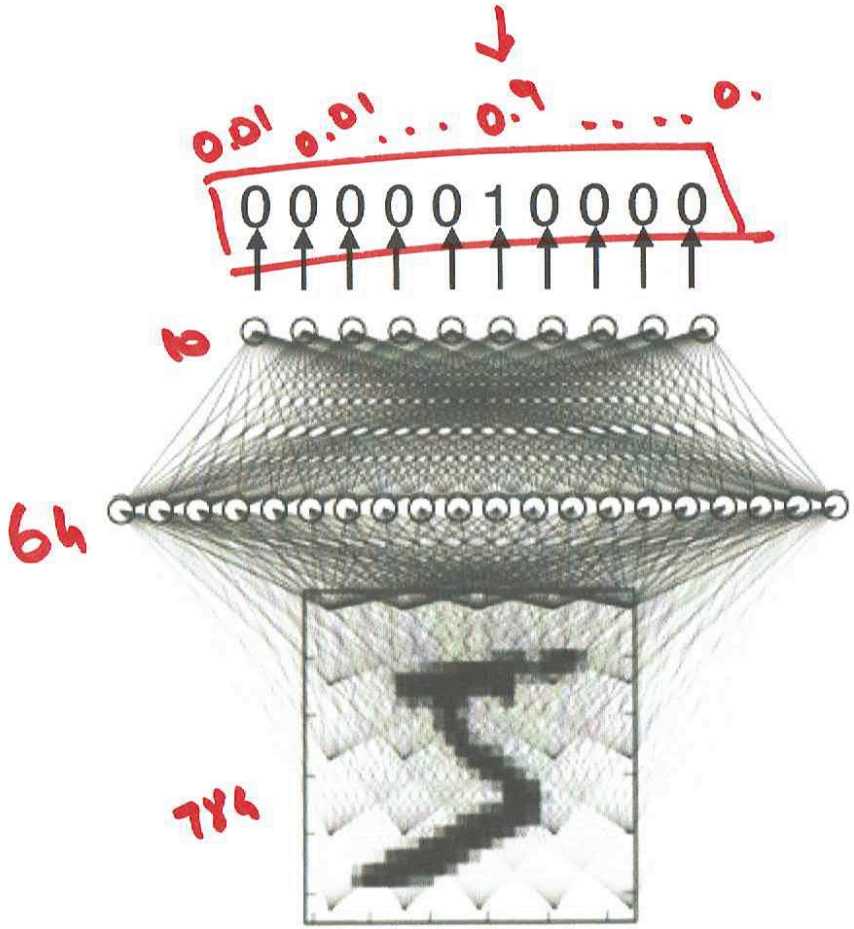
9

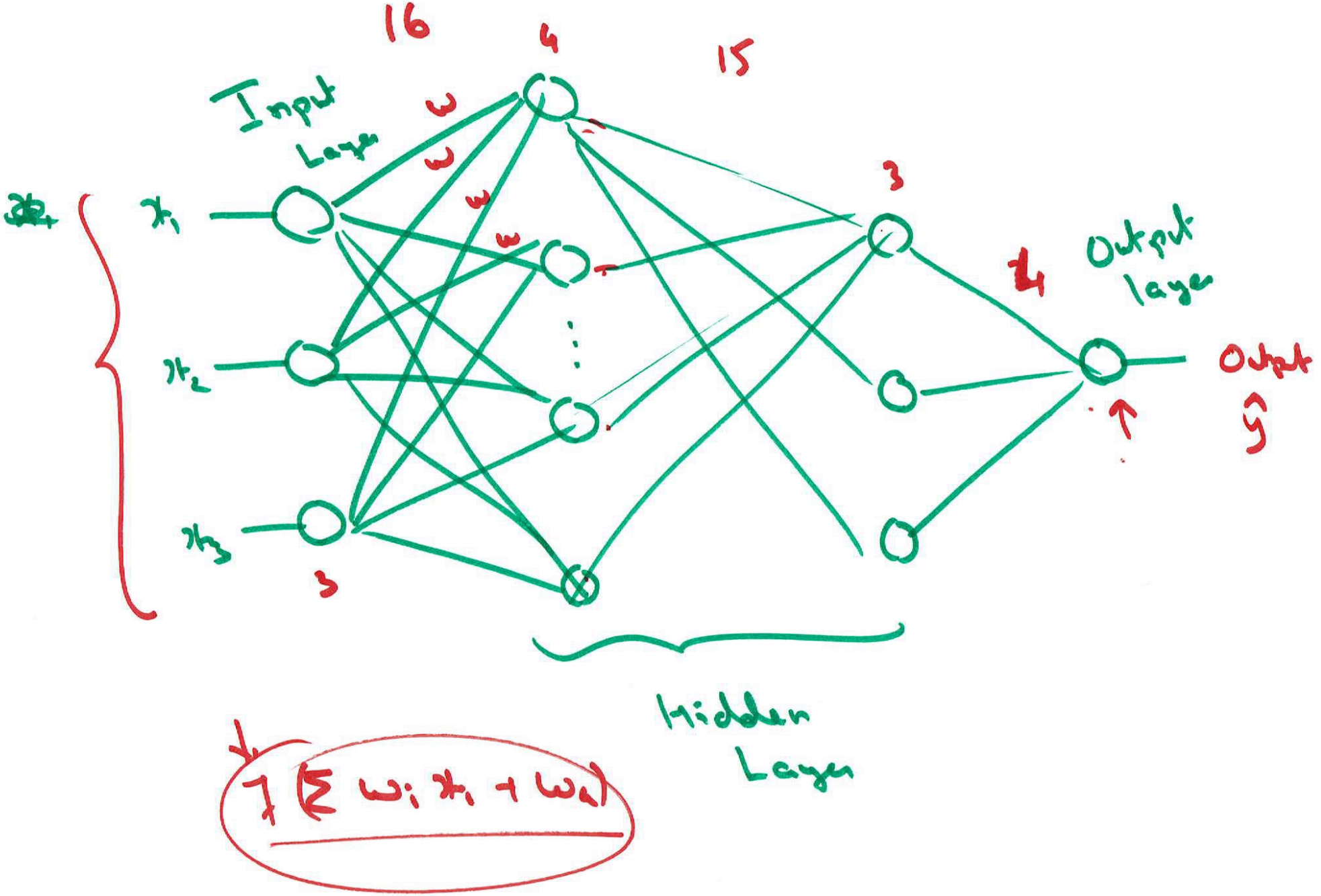


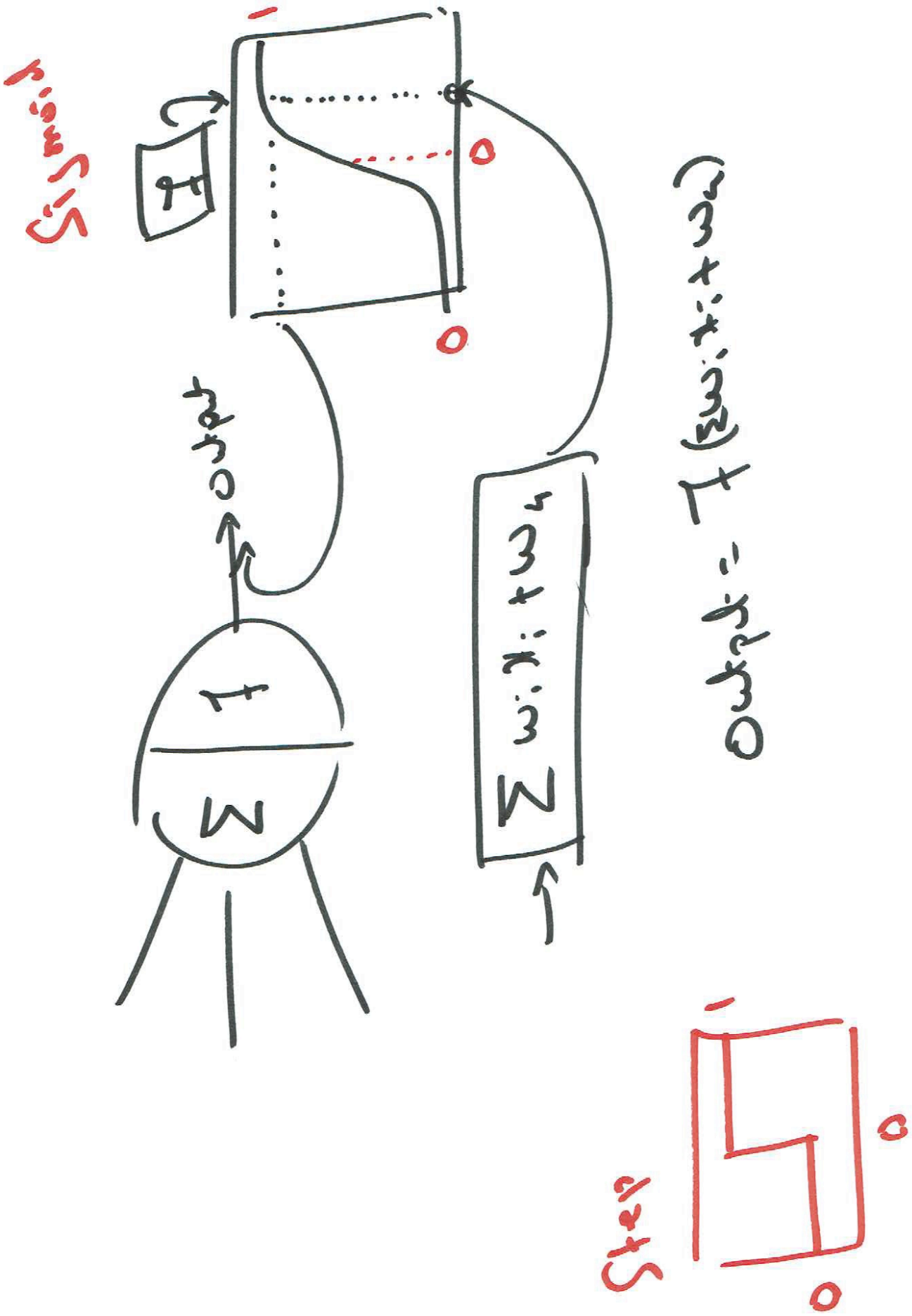
(784 × 64 + 64) + (64 × 10 + 10)

= 50890

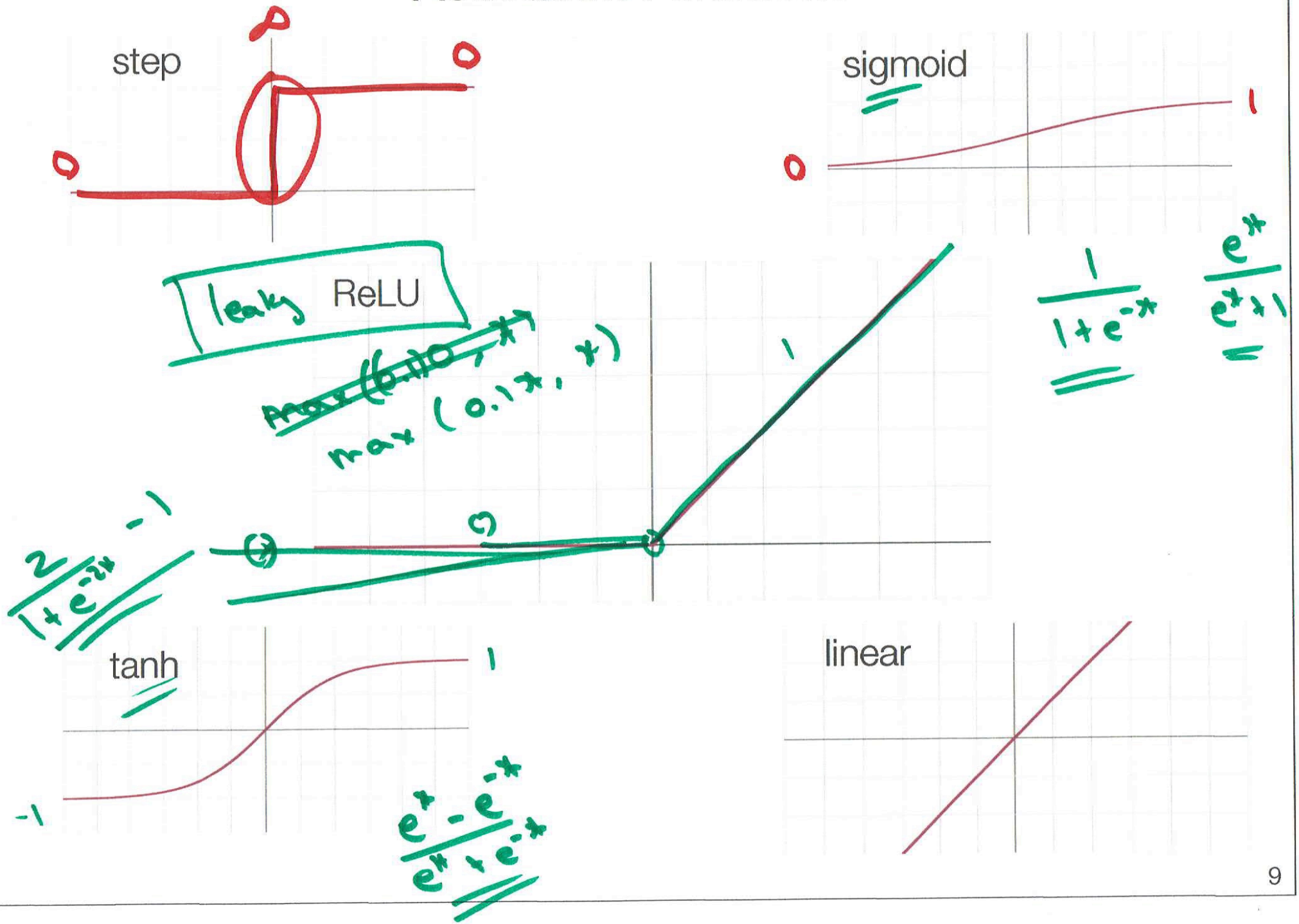
An Example

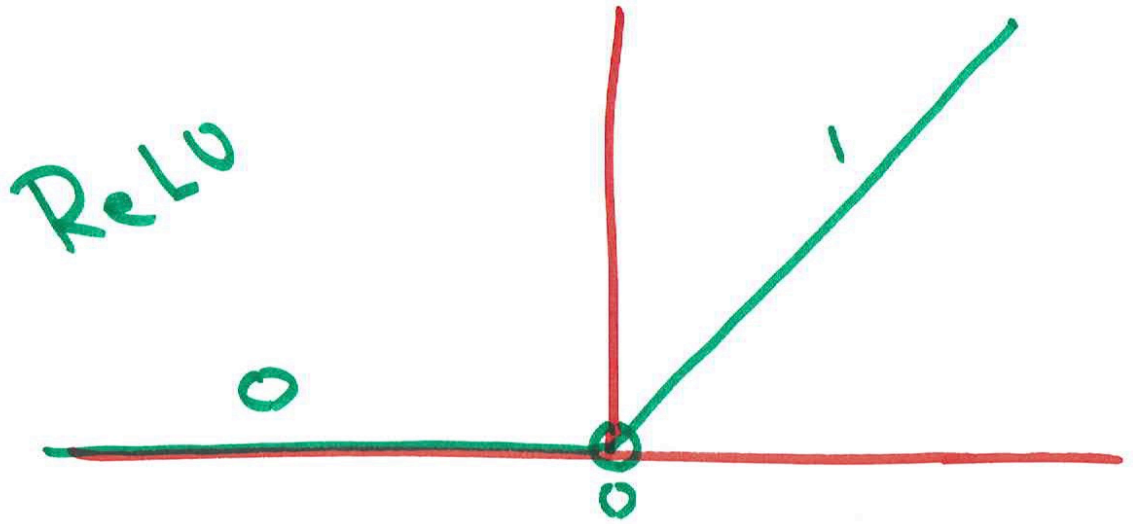






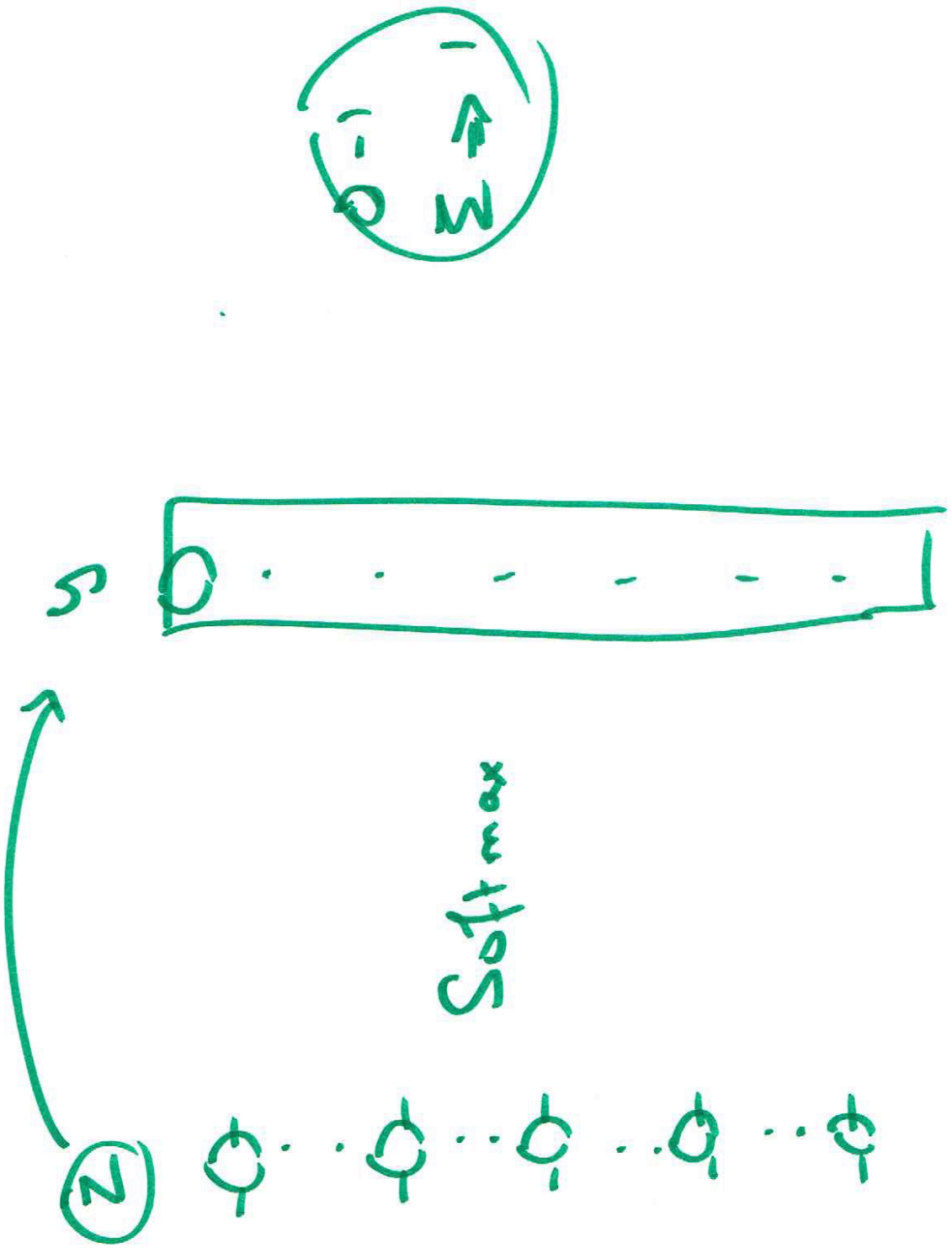
Activation Functions





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

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



z

y

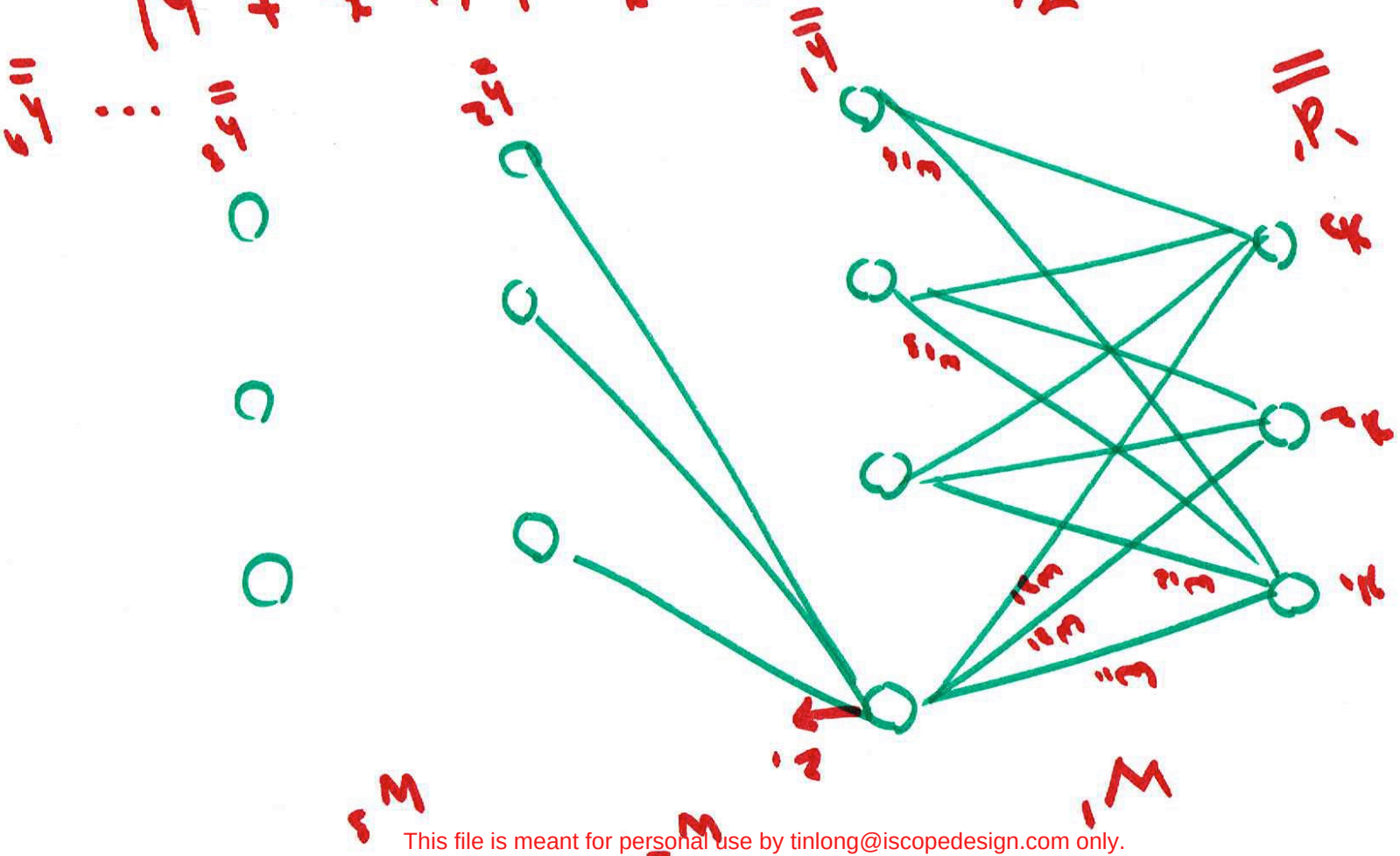
$$y_i = \frac{e^{z_i}}{\sum e^{z_i}}$$

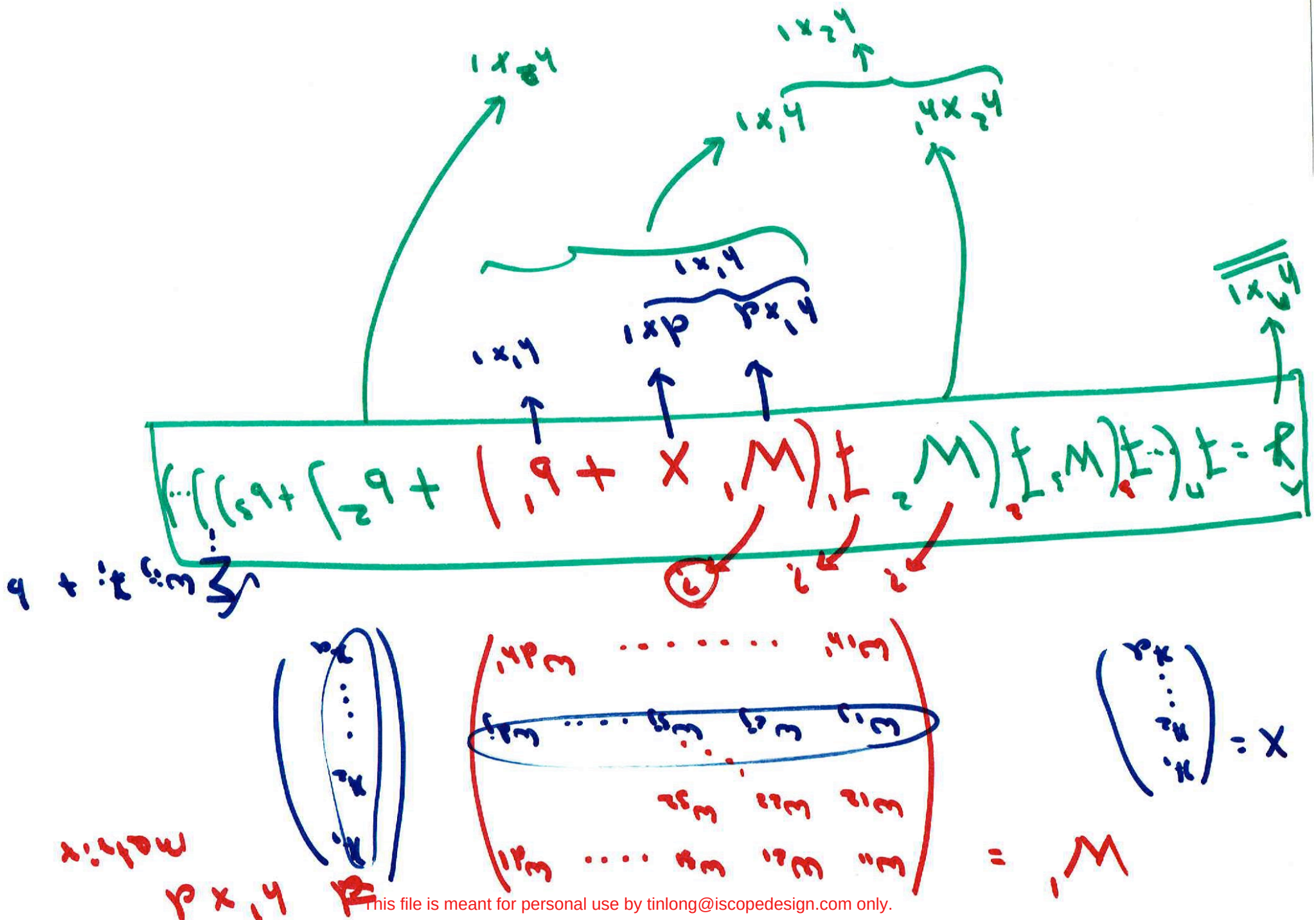
$$w^{new} = w^{old} - \eta \nabla_w L(w)$$

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

$$z_j = \sum_i w_{ij} x_i + b_j$$

$$z_1 = f(w_{11} x_1 + w_{21} x_2 + w_{31} x_3 + b_1)$$





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

Sigmoid

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

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

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

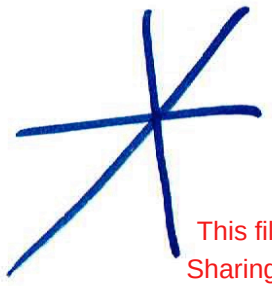
$$\tanh = 2\sigma(2x) - 1$$

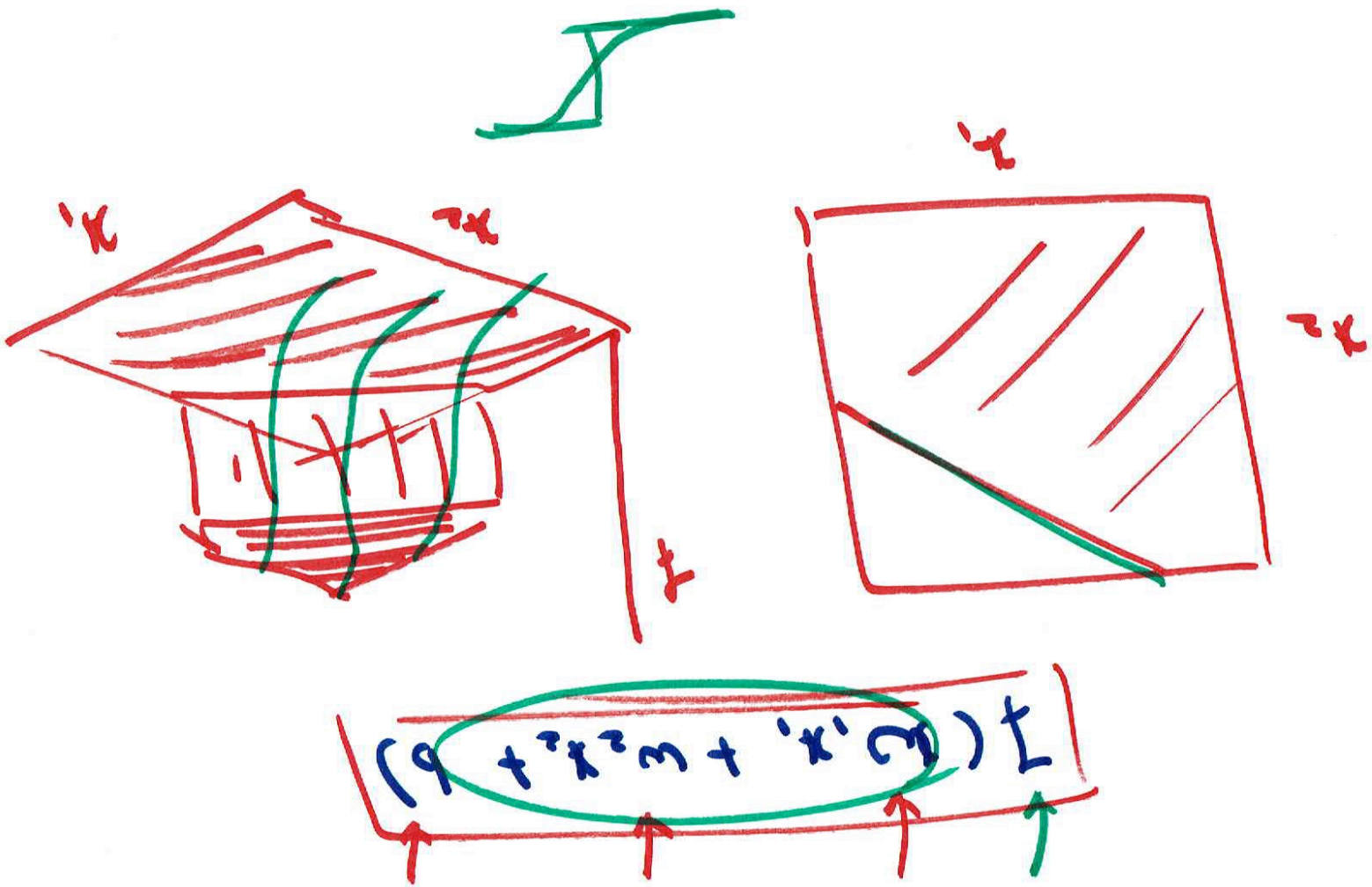
ReLU

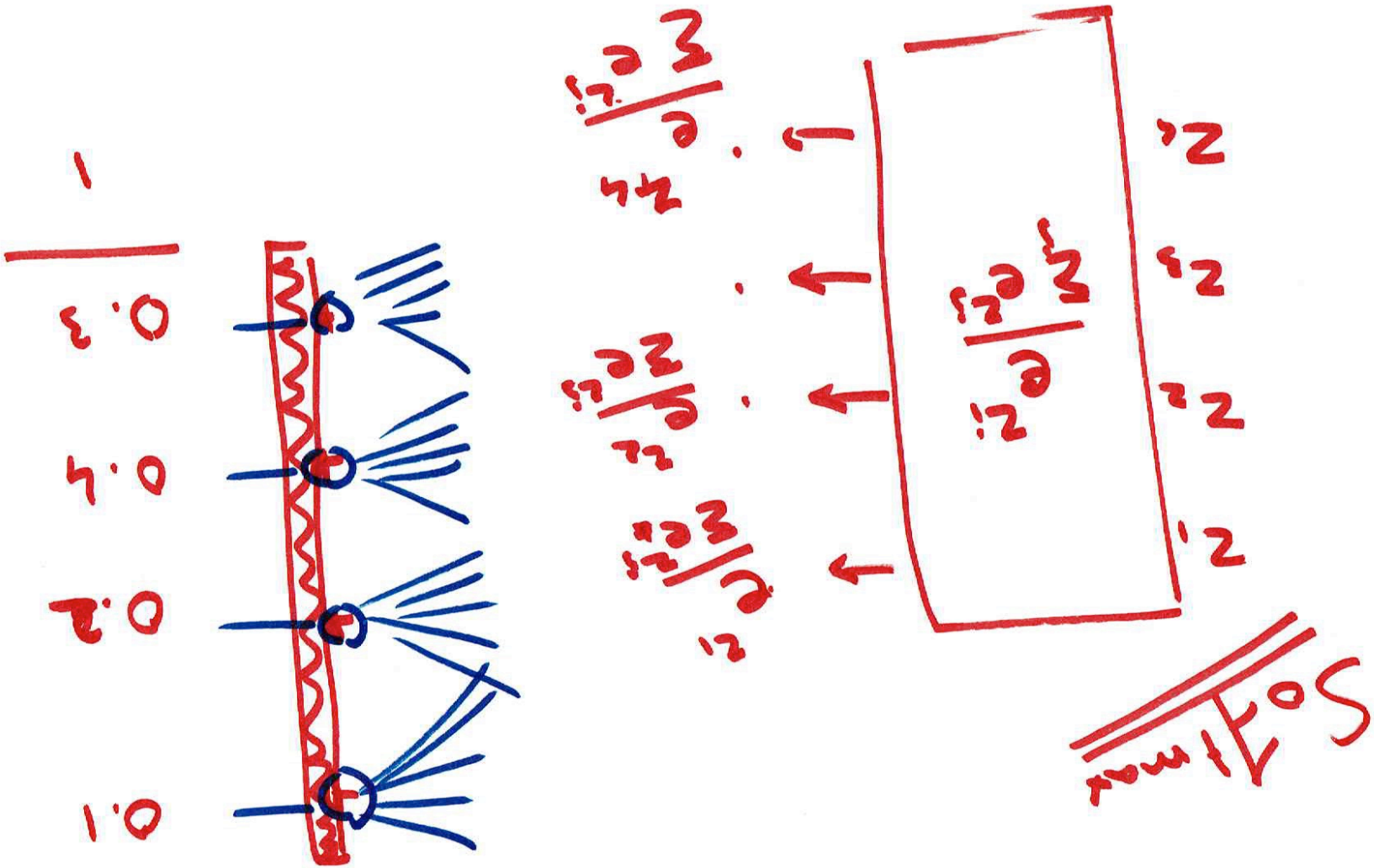
$$= \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

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

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







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

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

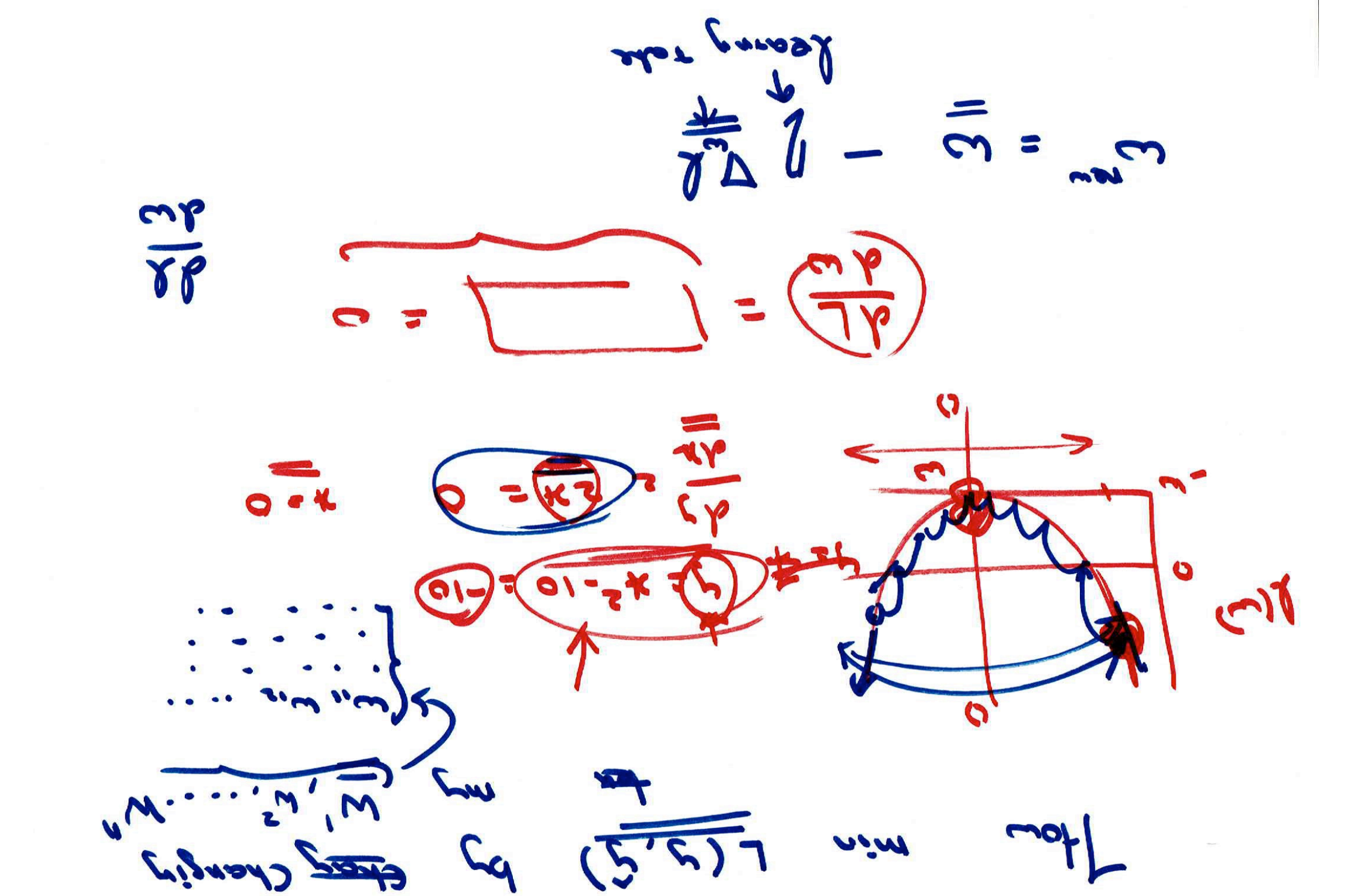
L_2 loss
MSE
SSE

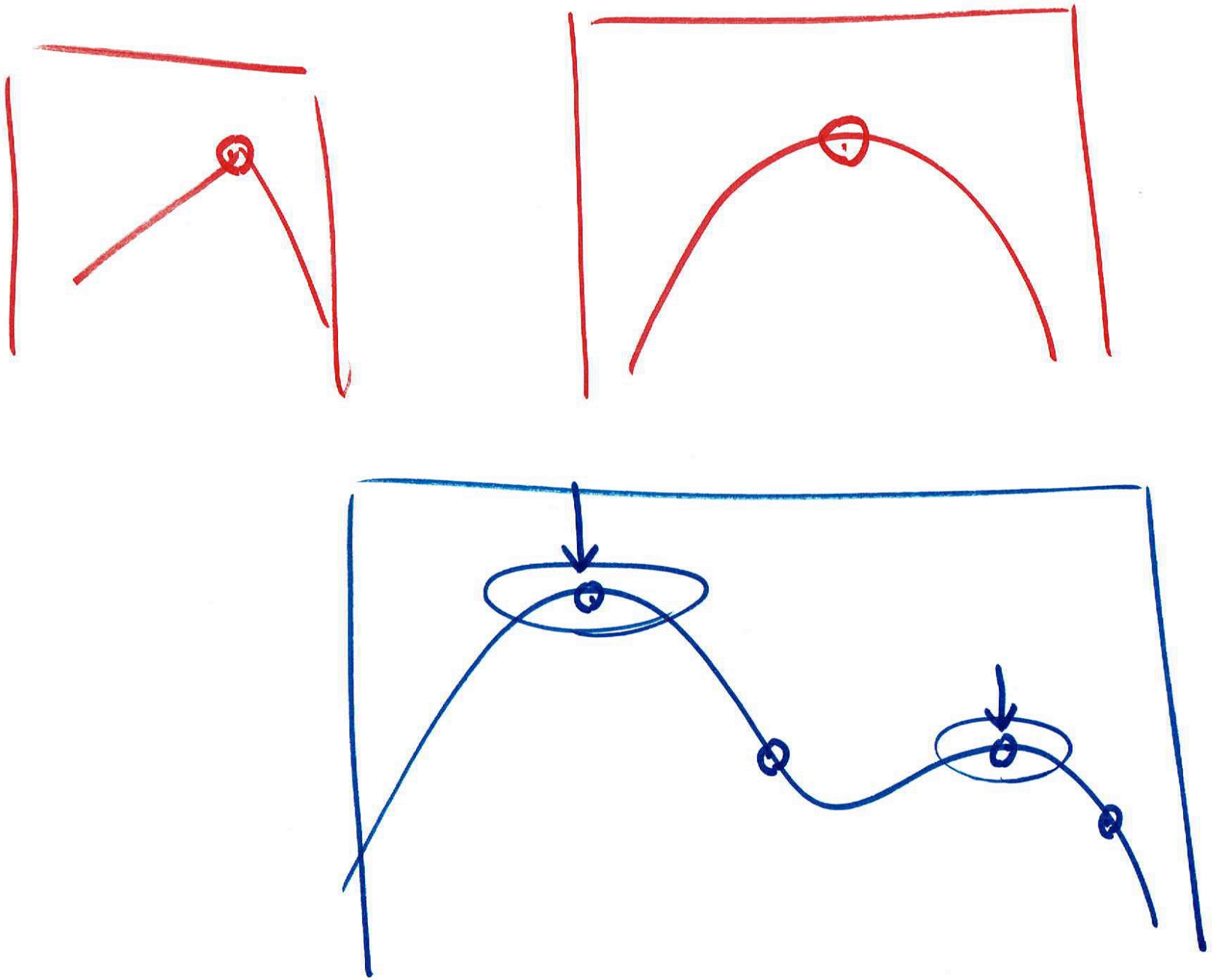
Reg

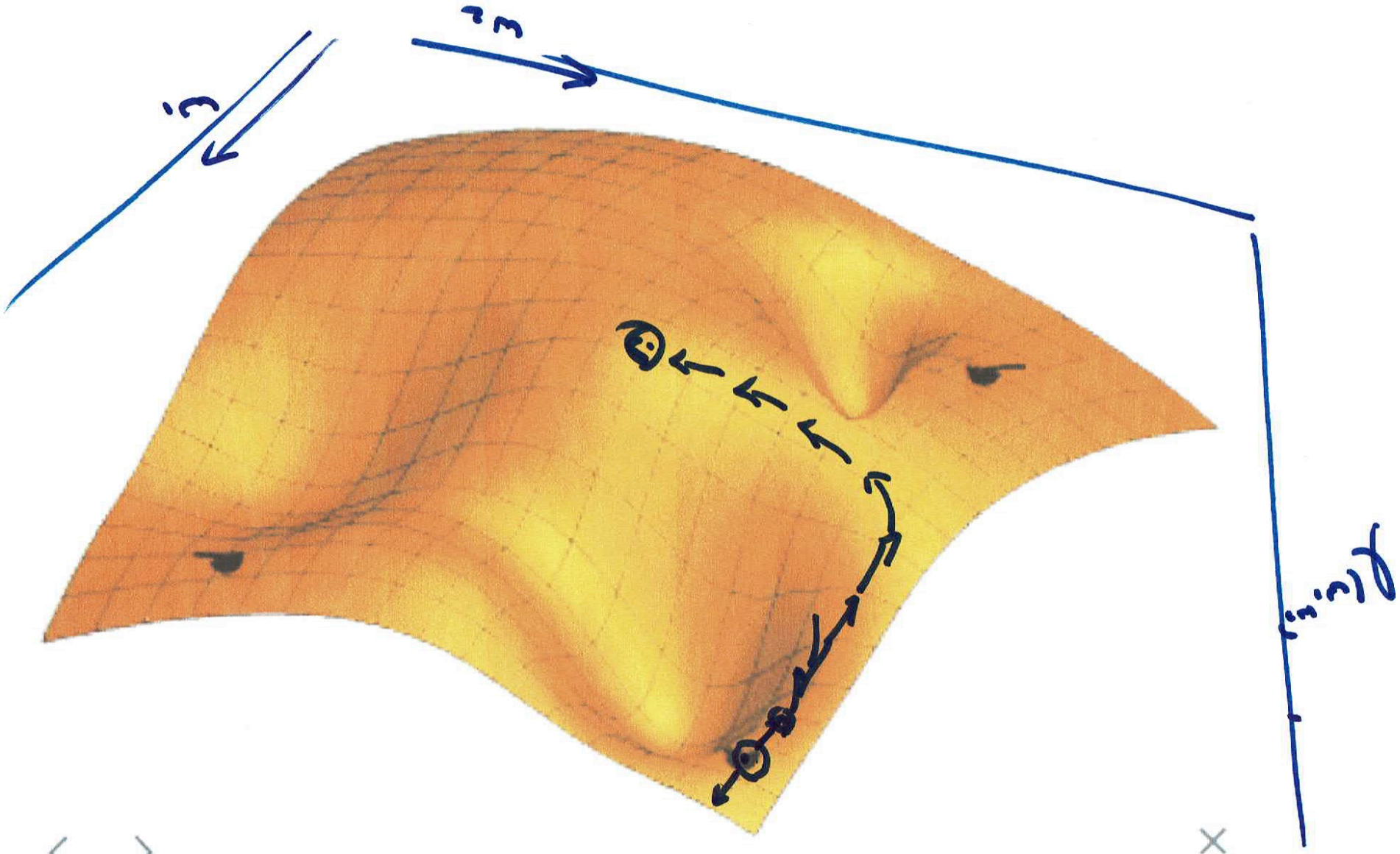
$$L(y, \hat{y}) = l(w)$$

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

Cross entropy loss







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

SGD

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

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

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

Loss

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

function of ω $(l(\omega))$

Chain Rule

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

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

Back propagation