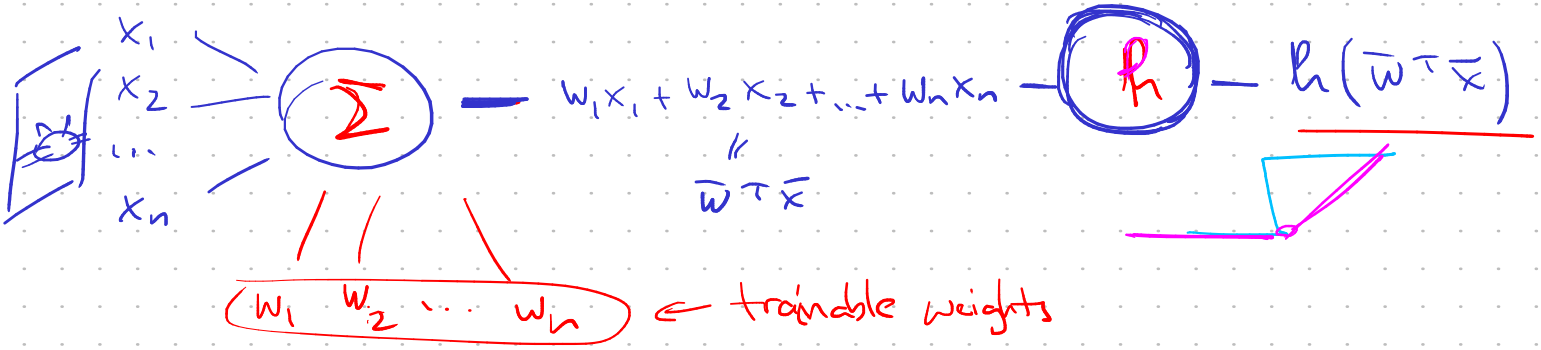


① Neuron & neural networks

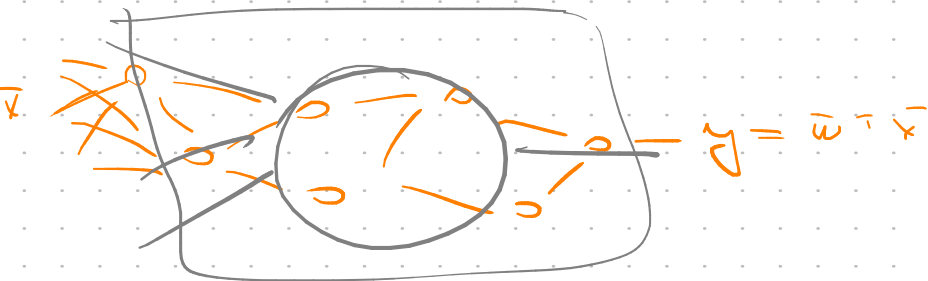
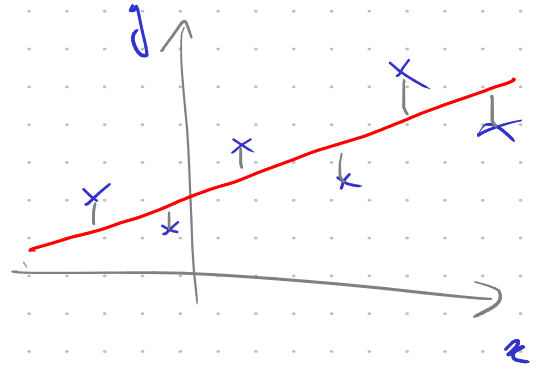
activation function



$$\bar{x} \rightarrow \Sigma \rightarrow \bar{w}^T \bar{x} = y$$

$\bar{w}$

linear regression

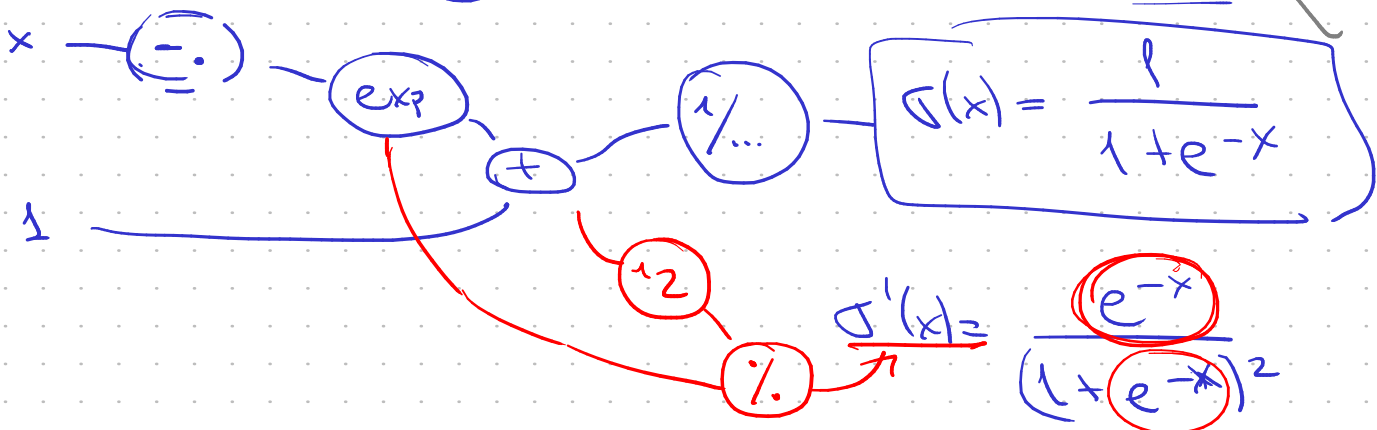
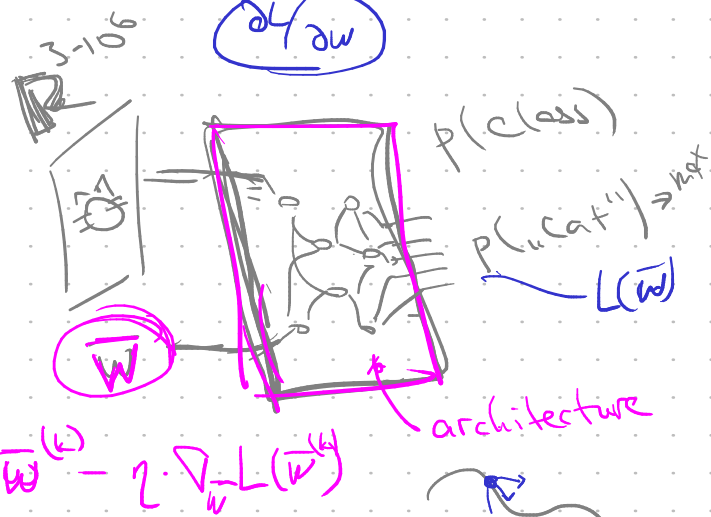
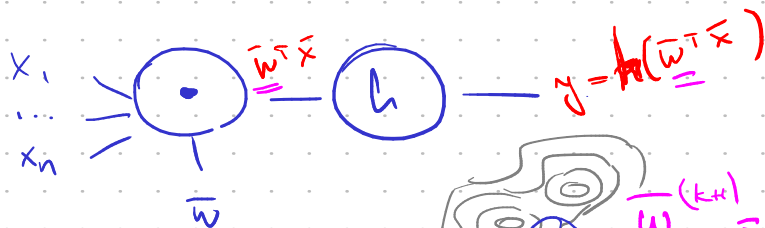


Universal approximation theorems



Connectionism

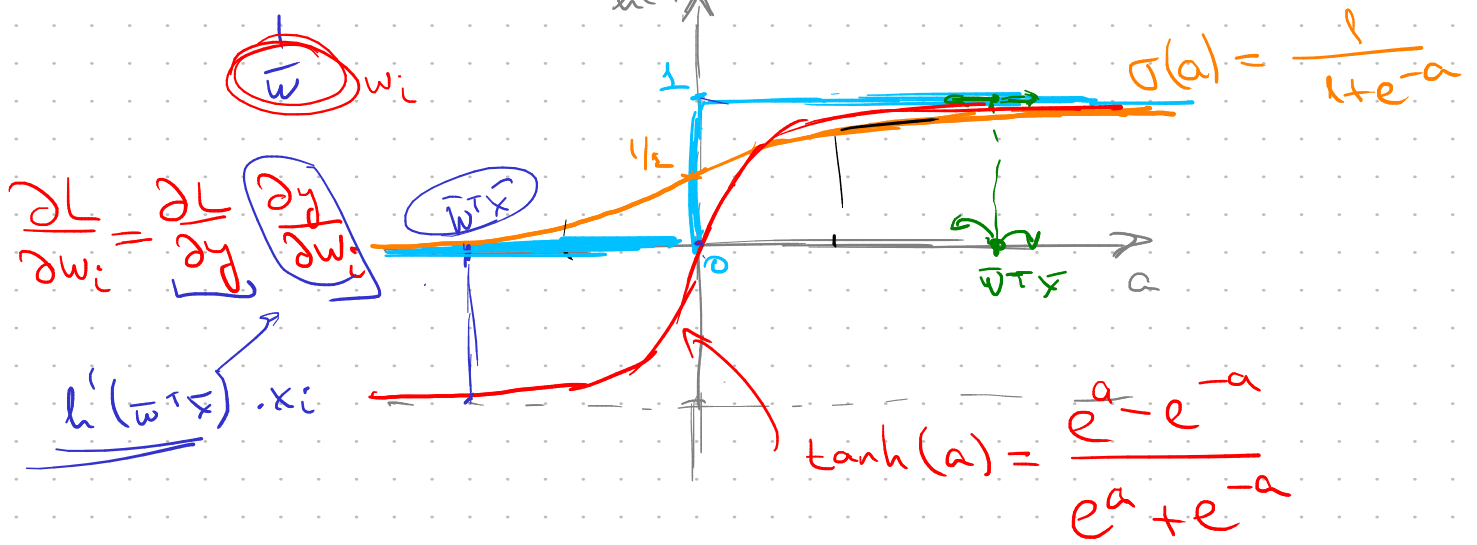
Computational graph



② Activation functions

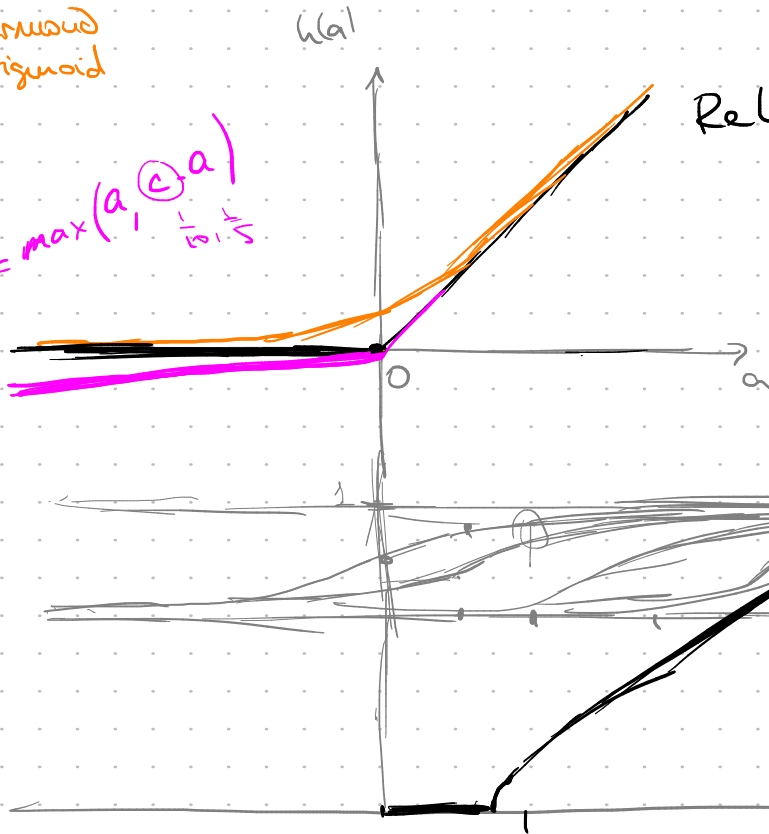


1943 / 1958



curved Sigmoid

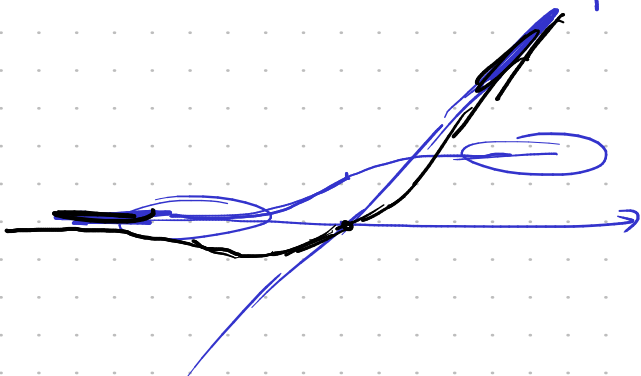
Leaky ReLU = $\max(a, \frac{1}{\alpha} a)$



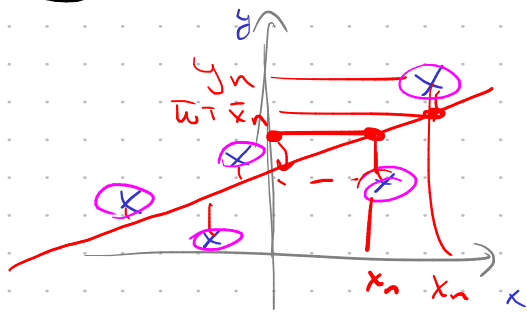
$\text{ReLU}(a) = \max(0, a) =$
if $a \geq 0$ then a
else 0

$\text{Swish}(a) = a \cdot \sigma(p - a)$

GELU
SwiGLU



③ Gradient descent & neural networks

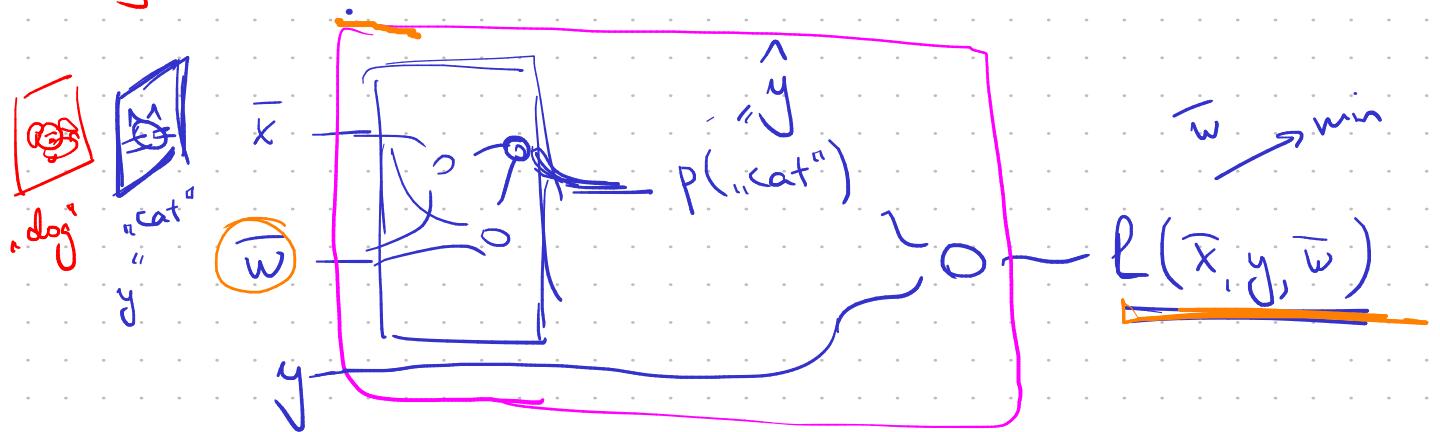


$$D = \{(x_n, y_n)\}_{n=1}^N$$

$$y \approx w_0 + w_1 x$$

$$L(\bar{w}) = \frac{1}{N} \sum_{n=1}^N (y_n - (w_0 + w_1 x_n))^2$$

Learning: $L(\bar{w}) \xrightarrow{\bar{w}} \min$



$$L(\bar{w}) \xrightarrow{\bar{w}} \min$$

$$\bar{w}^{(k+1)} = \bar{w}^{(k)} - \underbrace{\eta \cdot \nabla_{\bar{w}} L(\bar{w}^{(k)})}_{\text{gradient descent step}}$$

