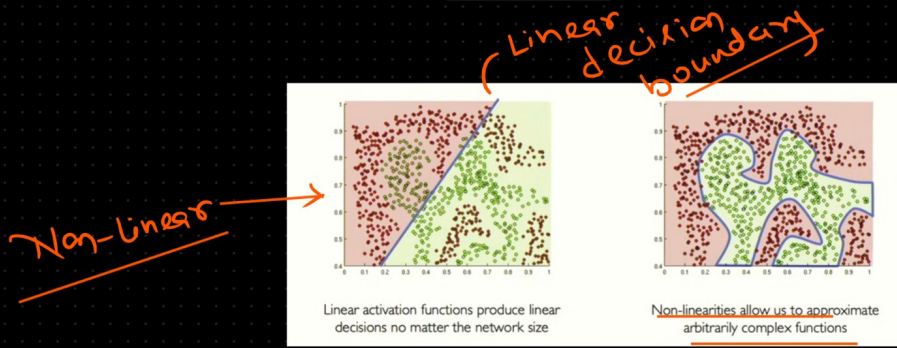


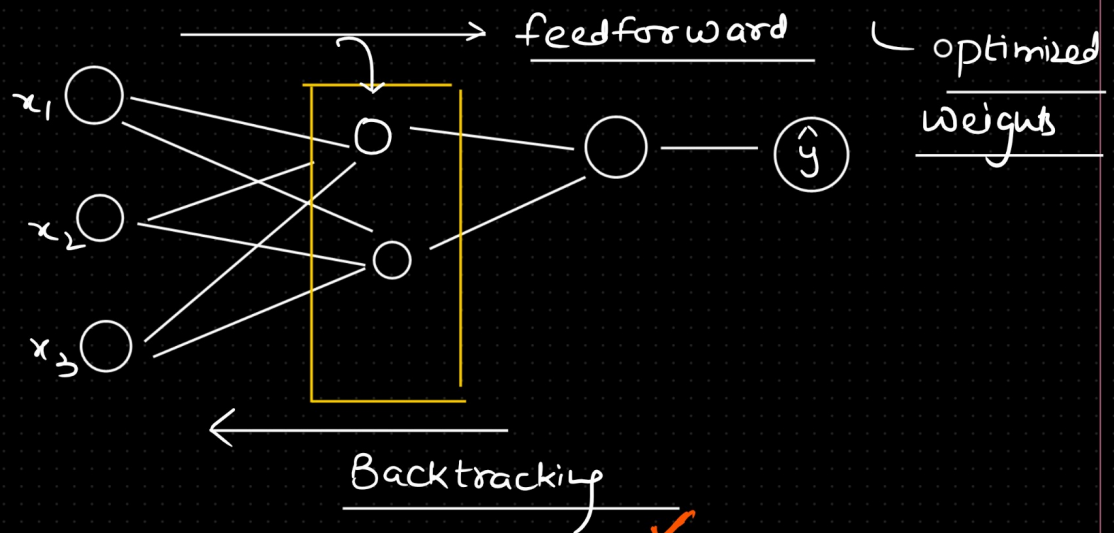
Activation function



Properties

① Non-linear function

② Differentiable → Training



$$\underline{w_{new}} = \underline{w_{old}} - \eta \frac{\delta L}{\delta w}$$

③ Computationally inexpensive
(Training → fast)

④ zero-centered → tanh



-1, 1

Linear Activation Function

↳ Linear decision boundary

$$\left\{ \begin{array}{l} \underline{f(x) = x} \end{array} \right.$$

(Rarely used)

Logistic Regression

binary \propto Sigmoid \rightarrow output layer

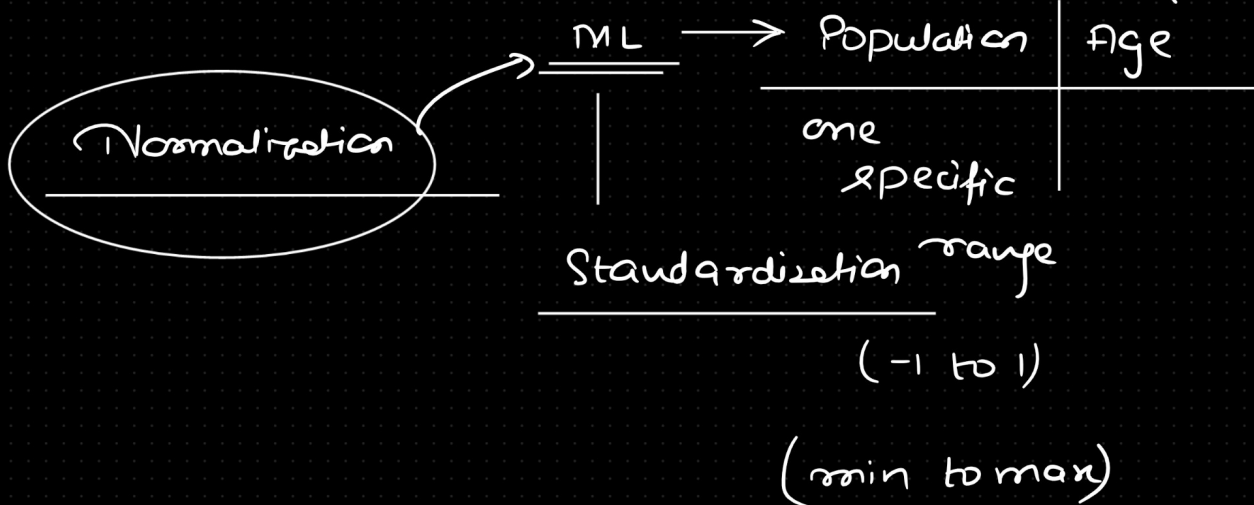
Classification $f(x) = \frac{1}{1 + e^{-x}}$

↳ Non-Linear activation function

Positive \rightarrow Range = (0, 1)

Value

✓
Millions
↗ Trillions ↖ ✓
100



\rightarrow Not zero-centered

Tanh

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

$$\text{Range} = (-1, 1)$$

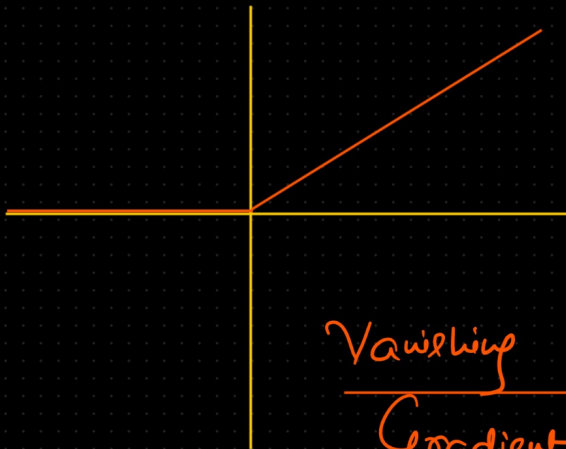
→ Symmetrical distribution

→ zero-centered

ReLU — Non-Linear
Activation
function → Hidden
layer

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

$$\text{Range} = [0, \infty)$$



Vanishing
Gradient
Descent

Softmax function \rightarrow Output layer

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^N e^{x_j}}$$

\downarrow
multiclass
Classification

Range = (0,1)