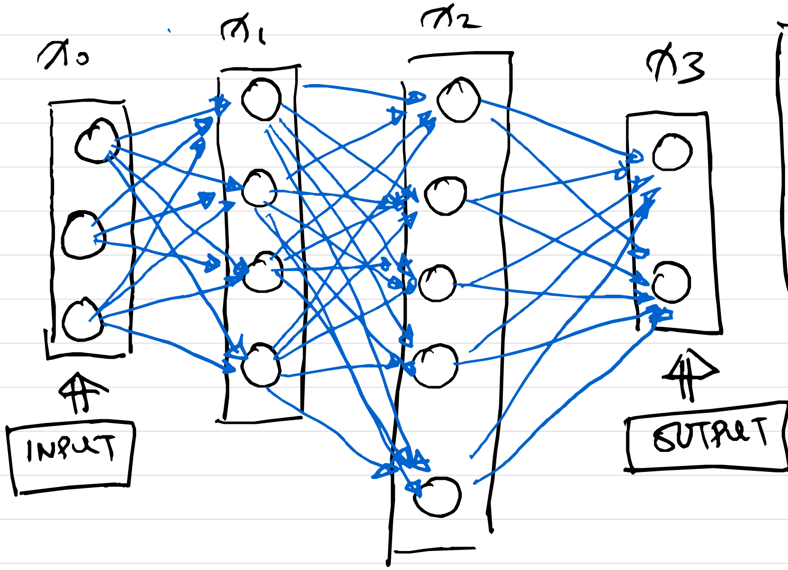


EXAMPLE OF FEEDFORWARD NETWORK

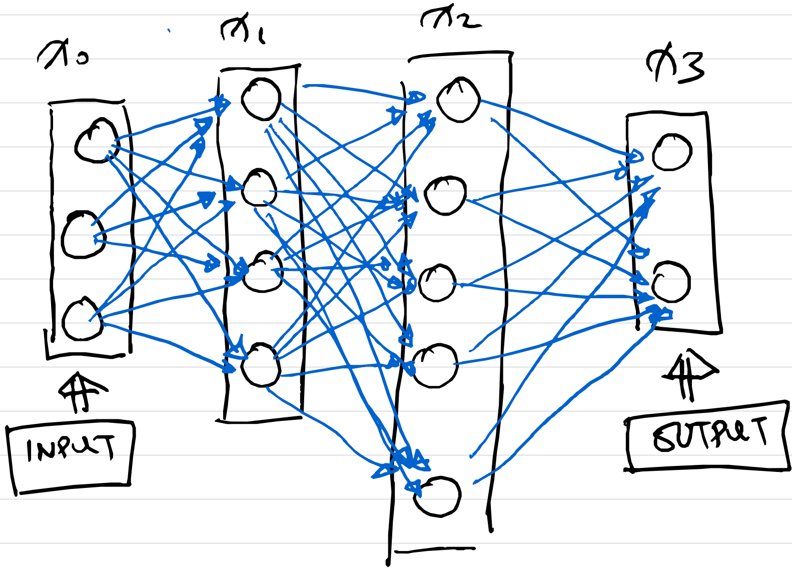
MULTI-LAYER PERCEPTRON (MLP)

$$x_i \in \mathbb{R}^{H \times 1}, \quad x_i = \begin{array}{cc} \begin{bmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \vdots \\ x_i^{(H)} \end{bmatrix} & \updownarrow H \end{array}$$

$$W_i \in \mathbb{R}^{D \times H}, \quad W_i = \begin{array}{cc} D \updownarrow & \begin{bmatrix} w_i^{(1,1)} & \cdots & w_i^{(1,H)} \\ \vdots & & \\ w_i^{(D,1)} & \cdots & w_i^{(D,H)} \end{bmatrix} \\ & \longleftarrow H \longrightarrow \end{array}$$

$$b_i \in \mathbb{R}^{D \times 1}, \quad b_i = \begin{array}{cc} \begin{bmatrix} b_i^{(1)} \\ \vdots \\ b_i^{(D)} \end{bmatrix} & \updownarrow D \end{array}$$

$$\boxed{X_{i+1} = f\left(w_i x_i + b_i\right)}$$



$$x_0 = \begin{bmatrix} x_0^{(1)} \\ x_0^{(2)} \\ x_0^{(3)} \end{bmatrix} \qquad x_1 = \begin{bmatrix} x_1^{(1)} \\ x_1^{(2)} \\ x_1^{(3)} \\ x_1^{(4)} \end{bmatrix} \qquad X_2 = \begin{bmatrix} x_2^{(1)} \\ x_2^{(2)} \\ x_2^{(3)} \\ x_2^{(4)} \\ x_2^{(5)} \end{bmatrix} \qquad x_3 = \begin{bmatrix} x_3^{(1)} \\ x_3^{(2)} \end{bmatrix}$$

$\uparrow$ input 3×1 4×1 5×1 $\uparrow$ output 2×1

So: $x_1 = f\left(W_0 \cdot X_0 + b_0\right)$

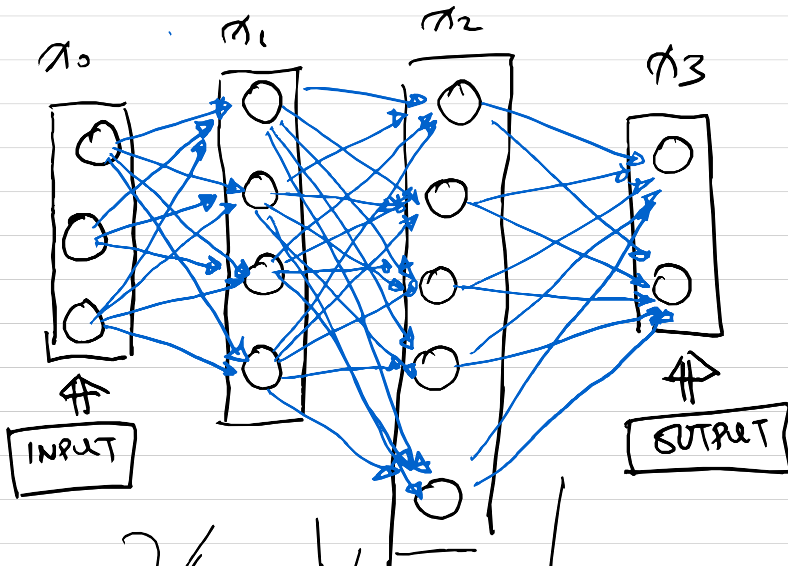
$\downarrow$

$X_2 = f\left(w_1 \cdot x_1 + b_1\right)$

$\downarrow$

$X_3 = f\left(w_2 x_2 + b_2\right)$

flow $\downarrow$



$W_0 : 4\times 3 \qquad W_1 : 5\times 4 \qquad W_2 : 2\times 5$

$$W_0 = \begin{array}{cc} 4 & \begin{bmatrix} w_0^{(1,1)} & \cdots & w_0^{(1,3)} \\ \vdots & & \\ w_0^{(4,1)} & \cdots & w_0^{(4,3)} \end{bmatrix} \\ & \longleftarrow 3 \longrightarrow \end{array}, \qquad b_0 = \begin{bmatrix} b_0^{(1)} \\ \vdots \\ b_0^{(4)} \end{bmatrix}$$

$$x_1 = f(W_0 \cdot X_0 + b_0) = f\left(\underbrace{\begin{bmatrix} W_0 \end{bmatrix}}_{3\times 4} \cdot \underbrace{\begin{bmatrix} x_0 \end{bmatrix}}_{3\times 1} + \underbrace{\begin{bmatrix} b_0 \end{bmatrix}}_{4\times 1} \right)$$

$$= f\left(\begin{bmatrix} w_0^{(1,1)} \cdot x_0^{(1)} + w_0^{(1,2)} \cdot x_0^{(2)} + w_0^{(1,3)} x_0^{(3)} + b_0^{(1)} \\ w_0^{(2,1)} \cdot x_0^{(1)} + w_0^{(2,2)} \cdot x_0^{(2)} + w_0^{(2,3)} x_0^{(3)} + b_0^{(2)} \\ w_0^{(3,1)} x_0^{(1)} + w_0^{(3,2)} x_0^{(2)} + w_0^{(3,3)} x_0^{(3)} + b_0^{(3)} \\ w_0^{(1,1)} x_0^{(1)} + w_0^{(4,2)} x_0^{(2)} + w_0^{(4,3)} x_0^{(3)} + b_0^{(4)} \end{bmatrix} \right) \Rightarrow$$

$$x_1 = f(W_0 \cdot X_0 + b_0) = f\left(\begin{bmatrix} W_0 \end{bmatrix} \cdot \begin{bmatrix} x_0 \end{bmatrix} + \begin{bmatrix} b_0 \end{bmatrix} \right)$$

$$= \begin{bmatrix} f\left(W_0^{(1,1)} \cdot x_0^{(1)} + W_0^{(1,2)} \cdot x_0^{(2)} + W_0^{(1,3)} x_0^{(3)} + b_0^{(1)}\right) \\ f\left(W_0^{(2,1)} \cdot x_0^{(1)} + W_0^{(2,2)} \cdot x_0^{(2)} + W_0^{(2,3)} x_0^{(3)} + b_0^{(2)}\right) \\ f\left(W_0^{(3,1)} x_0^{(1)} + W_0^{(3,2)} x_0^{(2)} + W_0^{(3,3)} x_0^{(3)} + b_0^{(3)}\right) \\ f\left(W_0^{(4,1)} x_0^{(1)} + W_0^{(4,2)} x_0^{(2)} + W_0^{(4,3)} x_0^{(3)} + b_0^{(4)}\right) \end{bmatrix}$$

$\uparrow$

$\swarrow$ This is the 2nd layer

Similarly

$$X_2 = f(W_1 \cdot X_1 + b_1)$$

$$X_3 = f(W_2 X_2 + b_2)$$

INPUT

$$X_1 = f(W_0 \cdot x_0 + b_0)$$

$$X_2 = f(W_1 \cdot x_1 + b_1)$$

$$X_3 = f(W_2 x_2 + b_2)$$

OUTPUT

$\boxed{x_0 : 3\times1}$ $\quad x_1 : 4\times1 \quad x_2 : 5\times1 \quad x_3 : 2\times1$

INPUT

COMPUTED FROM x_0

COMPUTED FORM x_1

COMPUTED FROM x_2

FEED-FORWARD NETWORK

TRAINABLE PARAMETERS

$W_0 : 4\times3$ matrix $\qquad b_0 : 4\times1$ vector

$W_1 : 5\times4$ matrix $\qquad b_1 : 5\times1$ vector

$W_2 : 2\times5$ matrix $\qquad b_2 : 2\times1$ vector

TOTAL # of parameters of network $= 4\times3+4+5\times4+5+2\times5+2 = 53$