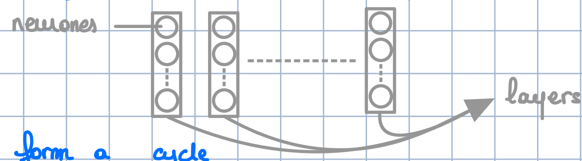
Lecture 1 : Feed - Forward Neural Network (FFNN)

- FFNN consists of layers of computers units (ie : neurons) usually connected in a feed-forward way



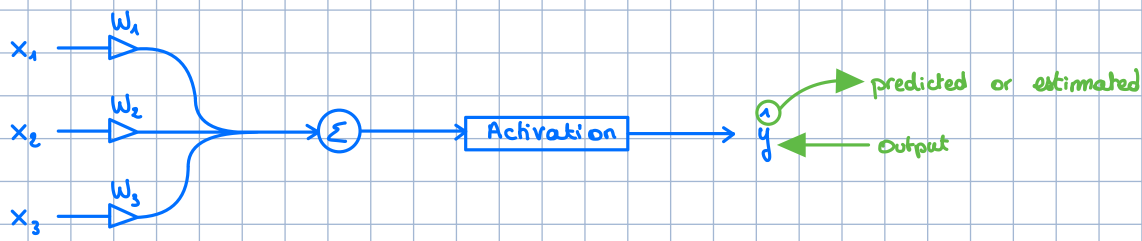
- FFNN is an ANN wherein connections between the nodes do not form a cycle

ANN → Artificial Neural Network ≠ Biological Neural Network

Types of ANN

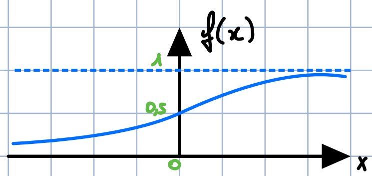
1) Feed - forward Neural Network (FFNN)
2) Feedback Neural Network (FBNN) / Recurrent Neural Network (RNN) → Memory !

A model for the artificial neuron



Basic types of activation fonctions

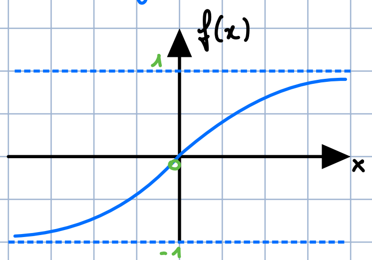
1) Sigmoid or logistic



$$f(x) = \frac{1}{1 + \exp(-x)}$$

$$f'(x) = f(x)(1 - f(x))$$

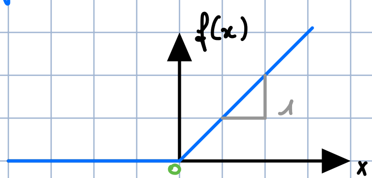
2) Hyperbolic tangent



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

$$f'(x) = 1 - f^2(x)$$

3) Rectified Linear Unit (ReLU)

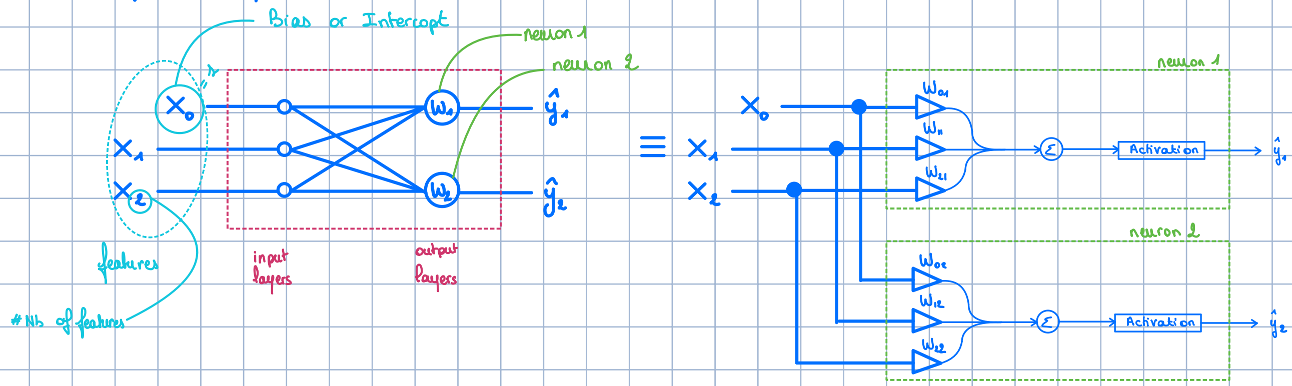


$$f(x) = \max\{0 ; x\} = x \, \mathbb{I}\{x > 0\}$$

$$= \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases}$$

$$f'(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

A FFNN of two layers with two features :



$$X = \begin{array}{c|c} & 1 \;\cdots\cdots\; N \\ \begin{matrix}1\\2\\ \vdots \\ i \\ \vdots \\ I\end{matrix} & \begin{bmatrix} (x^{(1)})^T \\ \vdots \\ (x^{(I)})^T \end{bmatrix} \end{array} \qquad X^{(i)} = \begin{bmatrix} (x_1^{(i)})^T \\ (x_2^{(i)})^T \\ \vdots \\ (x_N^{(i)})^T \end{bmatrix}$$

$$\bar{X} = \begin{bmatrix} x_0^{(1)} = 1 \\ x_0^{(2)} = 1 \\ \vdots \\ x_0^{(I)} = 1 \end{bmatrix} \begin{bmatrix} (x^{(1)})^T \\ (x^{(2)})^T \\ \vdots \\ (x^{(I)})^T \end{bmatrix} \quad I \times (N+1)$$

Bias Column

bias

$$W = \begin{array}{c|c} \begin{matrix}0\\1\\ \vdots \\ N\end{matrix} & \begin{bmatrix} W_1 & \cdots\cdots & W_J \end{bmatrix} \end{array} \quad \hat{=} \text{ dimension of output variables}$$

$(N+1) \times J$

$$W = \begin{array}{c|c} \begin{matrix}0\\1\\ \vdots \\ N\end{matrix} & \begin{bmatrix} W_1 & W_2 \end{bmatrix} \end{array} = \begin{bmatrix} W_{01} & W_{02} \\ W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \quad (2+1) \times 2$$

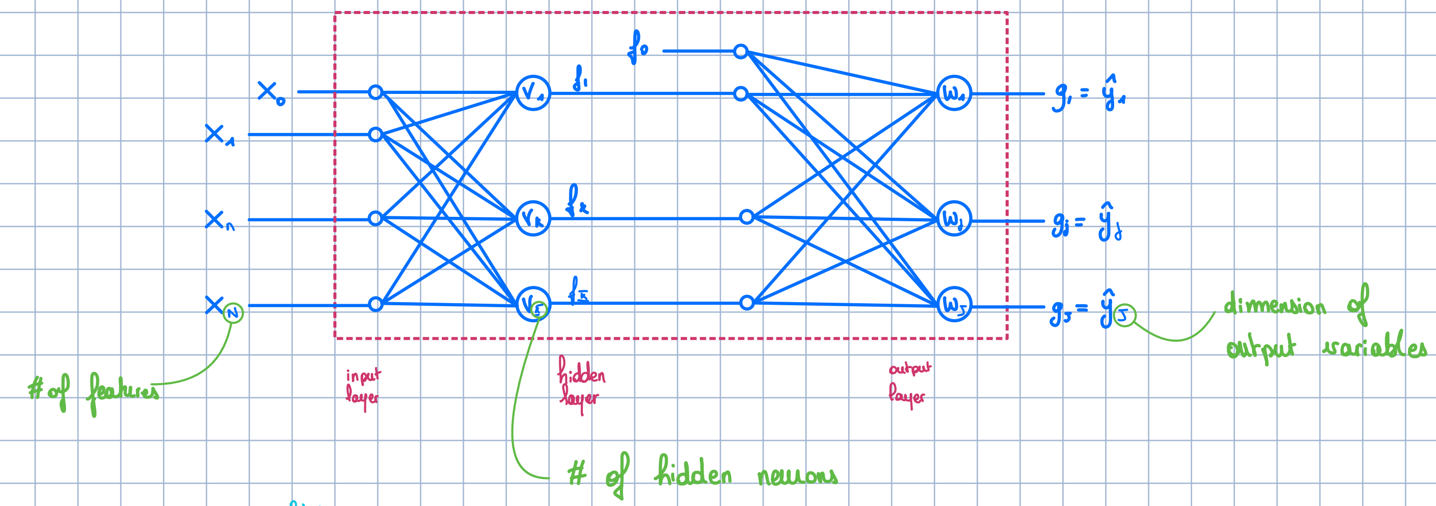
$$\bar{\bar{X}}_{I \times J} = \bar{X}_{I \times \cancel{(N+1)}} \cdot W_{\cancel{(N+1)} \times J} = \begin{array}{c|c} & 1 \;\; 2 \cdots\cdots J \\ \begin{matrix}1\\2\\ \vdots \\ I\end{matrix} & \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \end{array} \qquad \hat{y} = \begin{array}{c|c} & 1 \;\; 2 \cdots j \cdots J \\ \begin{matrix}1\\2\\ \vdots \\ i \\ \vdots \\ I\end{matrix} & \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \end{array}$$

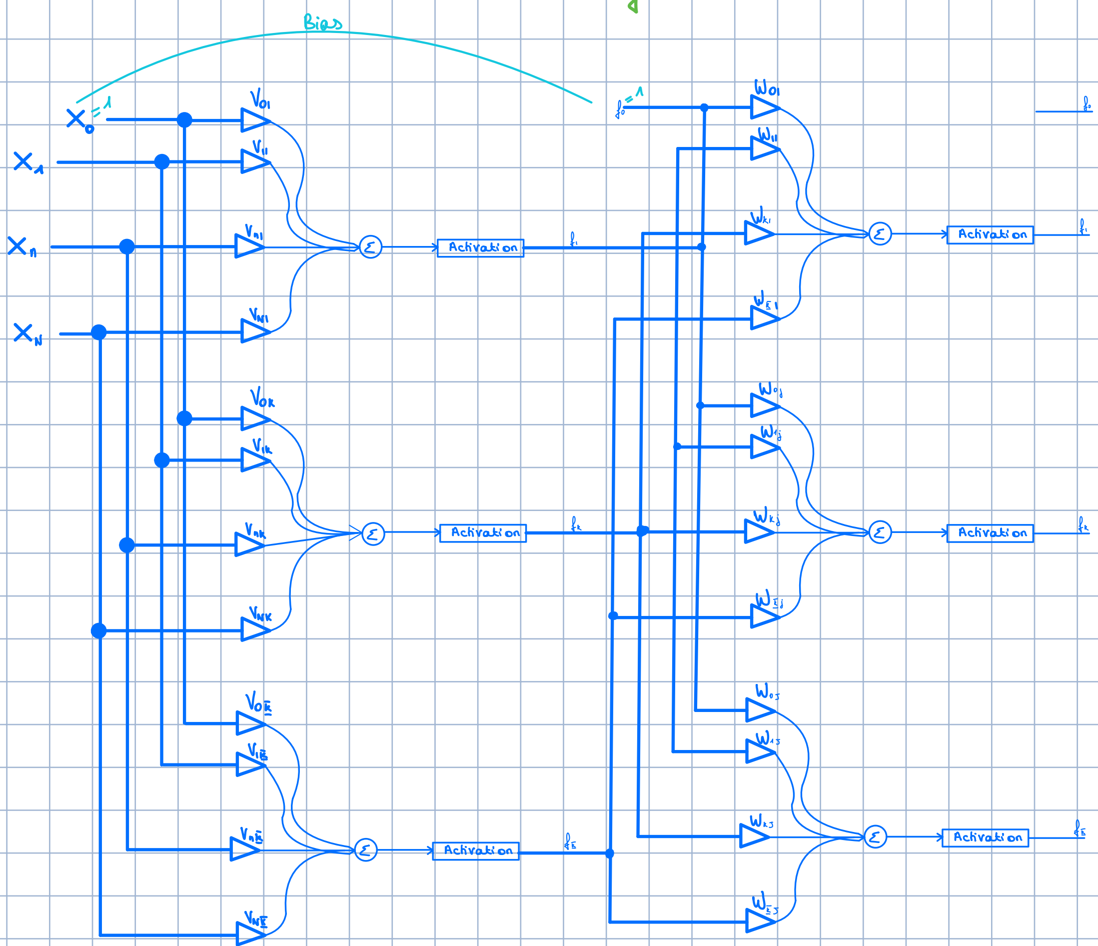
exemple :

$$\hat{y} = \begin{array}{c|c} & 1 \qquad 2 \\ \begin{matrix}1\\2\\ \vdots \\ i \\ \vdots \\ I\end{matrix} & \begin{bmatrix} \hat{y}_1^{(i)} & \hat{y}_2^{(i)} \end{bmatrix} \end{array} = (\hat{y}^{(i)})^T \qquad \hat{y}^{(i)} = \begin{bmatrix} \hat{y}_1^{(i)} \\ \hat{y}_2^{(i)} \end{bmatrix} = \begin{bmatrix} h(x^{(i)}; w_1) \\ h(x^{(i)}; w_2) \end{bmatrix}$$

Activation function (sigmoid)

A FFNN of three layers for the general case :





Fwd

$$X = \begin{array}{c c} & \begin{matrix} 1 & \cdots & N \end{matrix} \\ \begin{matrix} 1 \\ \vdots \\ I \end{matrix} & \begin{bmatrix} (x^{(1)})^T \\ \vdots \\ (x^{(I)})^T \end{bmatrix} \end{array}$$

$$\bar{X} = \begin{bmatrix} 1 & \\ \vdots & X \\ 1 & \end{bmatrix} = \begin{bmatrix} (\bar{x}^{(1)})^T \\ \vdots \\ (\bar{x}^{(I)})^T \end{bmatrix}$$

Bias column

$$\bar{\bar{X}} = \bar{X}_{I \times (N+1)} \cdot V_{(N+1) \times K} = \begin{bmatrix} (\bar{x}^{(1)})^T V \\ \vdots \\ (\bar{x}^{(I)})^T V \end{bmatrix}$$

Model parameters

elementwise inversion

$$F = (1 + \exp(-\bar{\bar{X}}))^{-1} = \begin{bmatrix} (f^{(1)})^T \\ \vdots \\ (f^{(I)})^T \end{bmatrix}_{I \times \underline{K}}$$

$$\bar{F} = \begin{bmatrix} 1 & \\ \vdots & F \\ 1 & \end{bmatrix} = \begin{bmatrix} (\bar{f}^{(1)})^T \\ \vdots \\ (\bar{f}^{(I)})^T \end{bmatrix}_{I \times \underline{K}}$$

$$\bar{\bar{F}} = \bar{F}_{I \times (\underline{K}+1)} \cdot V_{(\underline{K}+1) \times J} = \begin{bmatrix} (\bar{f}^{(1)})^T W \\ \vdots \\ (\bar{f}^{(I)})^T W \end{bmatrix}_{I \times J}$$

$$G = (1 + \exp(-\bar{\bar{F}}))^{-1} = \begin{bmatrix} (g^{(1)})^T \\ \vdots \\ (g^{(I)})^T \end{bmatrix}_{I \times J}$$

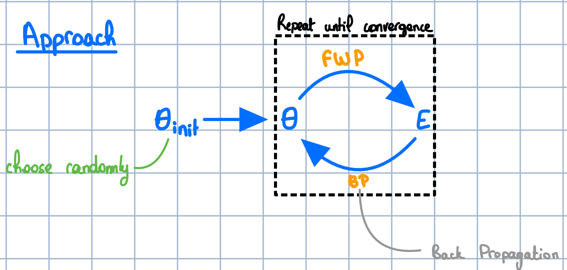
E = Sum of squared Errors (SSE)

$$= \frac{1}{2} \sum_{i=1}^{I} \| \hat{y}^{(i)} - y^{(i)} \|^2 = \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} (\hat{y}_j^{(i)} - y_j^{(i)})^2$$

Goal of FFNN : $\min_{\boxed{V, W}} E$

model parameters : $\theta = \{v, w\}$

<u>Approach</u>



Back propagation (BP)

itera = 1, 2, ... IterMax

Gradient descent

$$W_{kj}^{(itera)} \longleftarrow W_{kj}^{(itera-1)} - \alpha_1 \frac{\partial E}{\partial W_{kj}}$$

$$V_{nk}^{(itera)} \longleftarrow V_{nk}^{(itera-1)} - \alpha_1 \frac{\partial E}{\partial V_{nk}}$$

Descent — Learning rate — gradients

$$\begin{cases} k = 0, 1, 2, \dots, \underline{K} \\ j = 1, 2, \dots, J \end{cases}$$

$$\begin{cases} n = 0, 1, \dots, N \\ k = 1, 2, \dots, \underline{K} \end{cases}$$

$$\frac{\partial E}{\partial W_{kj}} = \frac{\partial}{\partial W_{kj}} \left[\frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} (\hat{y}_j^{(i)} - y_j^{(i)})^2 \right] = \frac{1}{2} \sum_{i=1}^{I} \frac{\partial}{\partial W_{kj}} \left[(g_1^{(i)} - y_1^{(i)})^2 + \dots + (g_j^{(i)} - y_j^{(i)})^2 + \dots + (g_J^{(i)} - y_J^{(i)})^2 \right]$$

$$= \frac{1}{2} \sum_{i=1}^{I} 2 (g_j^{(i)} - y_j^{(i)}) \frac{\partial g_j^{(i)}}{\partial W_{kj}} = \sum_{i=1}^{I} (g_j^{(i)} - y_j^{(i)}) \, g_j^{(i)} (1 - g_j^{(i)}) \, \bar{f}_k^{(i)}$$

$$\frac{\partial g_j^{(i)}}{\partial W_{kj}} = \frac{\partial}{\partial W_{kj}} \left[\frac{1}{1 + \exp(-(\bar{f}^{(i)})^T w_j)} \right] = \frac{-\exp(-(\bar{f}^{(i)})^T w_j)}{(1 + \exp(-(\bar{f}^{(i)})^T w_j))^2} (-\bar{f}_k^{(i)})$$

$$= \frac{1}{1+\exp(-(\bar{f}^{(i)})^T w_j)} \left(1 - \frac{1}{1+\exp(-(\bar{f}^{(i)})^T w_j)}\right) \bar{f}_k^{(i)} = g_j^{(i)}(1-g_j^{(i)})\, \bar{f}_k^{(i}$$

$$\frac{\partial E}{\partial V_{nk}} = \frac{\partial}{\partial V_{nk}} \left[\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}(\hat{g}_j^{(i)} - y_j^{(i)})^2\right] = \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\partial}{\partial V_{nk}}\left[(g_j^{(i)} - y_j^{(i)})^2\right] = \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J} 2\,(g_j^{(i)} - y_j^{(i)}) \frac{\partial g_j^{(i)}}{\partial V_{nk}}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J} (g_j^{(i)} - y_j^{(i)})\, g_j^{(i)}(1-g_j^{(i)})\, W_{kj}\, f_k^{(i)}(1-f_k^{(i)})\, \bar{x}_n^{(i)}$$

$$\frac{\partial g_j^{(i)}}{\partial V_{nk}} = \frac{\partial g_j^{(i)}}{\partial \bar{f}_k^{(i)}} \frac{\partial \bar{f}_k^{(i)}}{\partial V_{nk}}$$

$$\frac{\partial g_j^{(i)}}{\partial \bar{f}_k^{(i)}} = \frac{\partial}{\partial \bar{f}_k^{(i)}} \left[\frac{1}{1+\exp(-(\bar{f}^{(i)})^T w_j)}\right] = \frac{-\exp(-(\bar{f}^{(i)})^T w_j)}{(1+\exp(-(\bar{f}^{(i)})^T w_j))^2}\,(-W_{kj})$$

$$= \frac{1}{1+\exp(-(\bar{f}^{(i)})^T w_j)} \left(1 - \frac{1}{1+\exp(-(\bar{f}^{(i)})^T w_j)}\right) W_{kj} = g_j^{(i)}(1-g_j^{(i)})\, W_{kj}$$

$$\frac{\partial \bar{f}_k^{(i)}}{\partial V_{nk}} = \frac{\partial}{\partial V_{nk}} \left[\frac{1}{1+\exp(-(\bar{x}^{(i)})^T V_k)}\right] = \frac{-\exp(-(\bar{x}^{(i)})^T V_k)\,(-\bar{x}_n^{(i)})}{(1+\exp(-(\bar{x}^{(i)})^T V_k))^2} \quad \frac{\beta'}{\beta^2}$$

$$= \frac{1}{1+\exp(-(\bar{x}^{(i)})^T V_k)} \left(1 - \frac{1}{1+\exp(-(\bar{x}^{(i)})^T V_k)}\right) = f_k^{(i)}(1-f_k^{(i)})\, \bar{x}_n^{(i)}$$