Artificial Neural Network (ANN): Feed-Forward Neural Network (FFNN)

Jae Yun JUN KIM[*]

February 3, 2022

An **artificial neural network (ANN)** can be used for instance to solve a classification problem.
An ANN is a network of neurons inspired from the **biological neural network**. A Feed-Forward Neural Network (FFNN) is a type of ANN, which is charaterized by not having feedback in its topology.

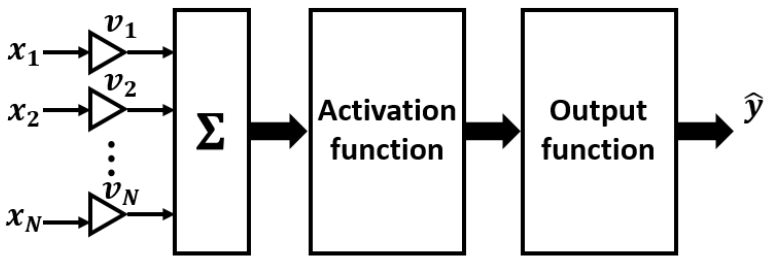
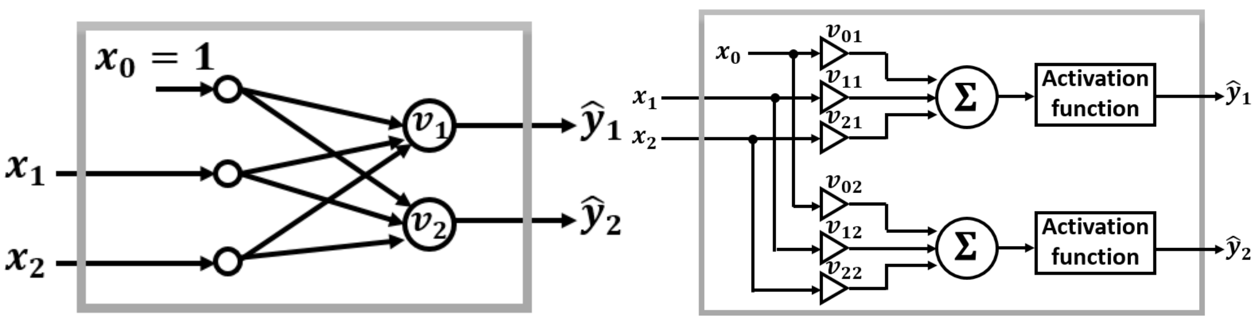


Figure 1: Model for an artificial neuron

1 FFNN of two layers



(a) Feed-forward neural network with two layers (b) Feed-forward neural netowrk with two layers (explained)

Figure 2: Feed-forward neural network with two layers

*ECE Paris Graduate School of Engineering, 37 quai de Grenelle 75015 Paris, France; jae-yun.jun-kim@ece.fr

2 FFNN of three layers

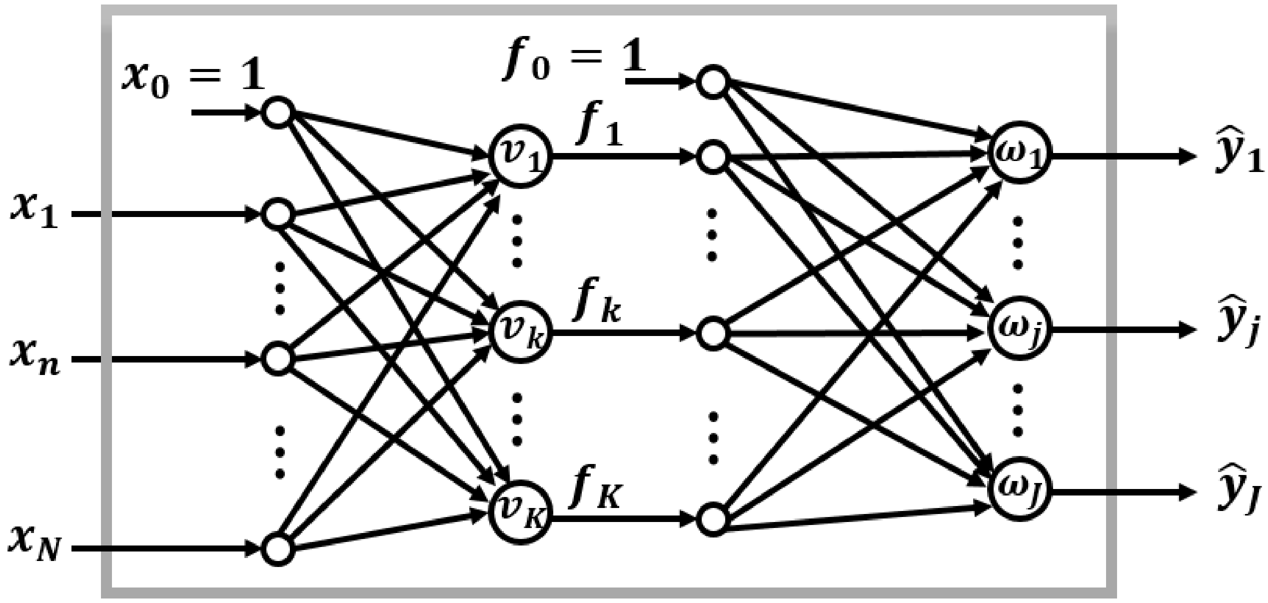


Figure 3: Feed-forward neural network with three layers

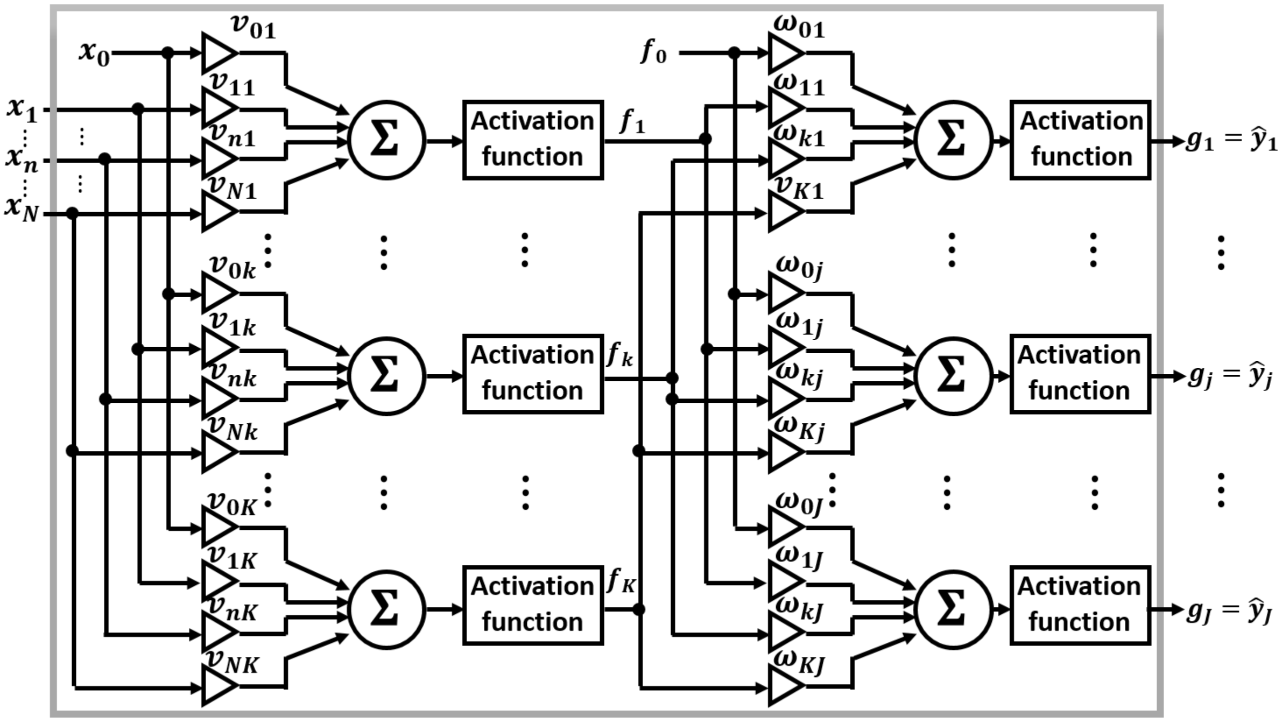


Figure 4: Feed-forward neural network with three layers (explained)

3 Forward propagation and backpropagation

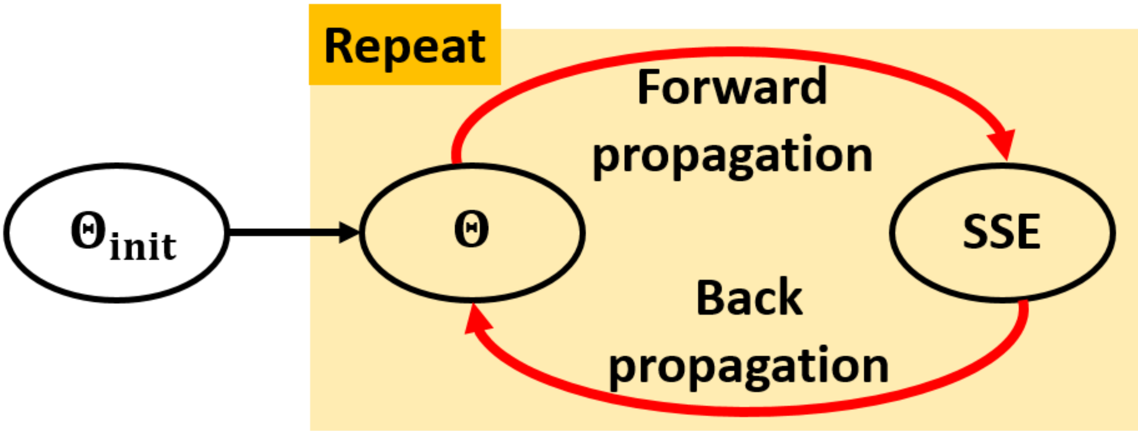


Figure 5: Forward propagation and backpropagation

3.1 Forward propagation

Input data matrix

$$X = \begin{bmatrix} \left(x^{(1)}\right)^T \\ \vdots \\ \left(x^{(I)}\right)^T \end{bmatrix}, \qquad \text{where} \quad x^{(i)} \in \mathbb{R}^{N\times 1} \tag{1}$$

Extended input data matrix

$$\overline{X} = \begin{bmatrix} \mathcal{I} & X \end{bmatrix} = \begin{bmatrix} \left(\overline{x}^{(1)}\right)^T \\ \vdots \\ \left(\overline{x}^{(I)}\right)^T \end{bmatrix}, \qquad \text{where} \quad \mathcal{I} = [1, \cdots, 1]^T \in \mathbb{R}^{I\times 1} \tag{2}$$

$$\overline{\overline{X}} = \overline{X} \cdot v = \begin{bmatrix} \left(\overline{x}^{(1)}\right)^T \cdot v \\ \vdots \\ \left(\overline{x}^{(I)}\right)^T \cdot v \end{bmatrix} \tag{3}$$

Define

$$F = \left(1 + \exp\left(-\overline{\overline{X}}\right)\right)^{-1} = \begin{bmatrix} \left(f^{(1)}\right)^T \\ \vdots \\ \left(f^{(I)}\right)^T \end{bmatrix} \tag{4}$$

where the operator $^{-1}$ represents the elementwise inversion.

$$\overline{F} = \begin{bmatrix} \mathcal{I} & F \end{bmatrix} = \begin{bmatrix} \left(\overline{f}^{(1)}\right)^T \\ \vdots \\ \left(\overline{f}^{(I)}\right)^T \end{bmatrix}, \qquad \text{where} \quad \mathcal{I} = [1, \cdots, 1]^T \in \mathbb{R}^{I\times 1} \tag{5}$$

$$\overline{\overline{F}} = \overline{F} \cdot \omega = \begin{bmatrix} \left(\overline{f}^{(1)}\right)^T \cdot \omega \\ \vdots \\ \left(\overline{f}^{(I)}\right)^T \cdot \omega \end{bmatrix} \tag{6}$$

Define

$$G = \left(1 + \exp\left(-\overline{\overline{F}}\right)\right)^{-1} = \begin{bmatrix} \left(g^{(1)}\right)^T \\ \vdots \\ \left(g^{(I)}\right)^T \end{bmatrix} \tag{7}$$

where $^{-1}$ represents once again the elementwise inversion.

Then, one can find the labels using the following algorithm:

Algorithm 1: find labels

```
Initialize label;
for i = 0 : I do
    label(i) ← arg max_j G(i, j);
return label;
```

Let us define the **sum of squared errors (SSE)** as

$$E = \frac{1}{2}\sum_{i=1}^{I}\left(\hat{y}^{(i)} - y^{(i)}\right)^2 = \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\hat{y}_j^{(i)} - y_j^{(i)}\right)^2 = \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(g_j^{(i)} - y_j^{(i)}\right)^2 \tag{8}$$

where $g_j^{(i)} = g\left(\overline{f}\left(\overline{x}^{(i)}\right)\right) = g \circ \overline{f}\left(\overline{x}^{(i)}\right)$.

3.2 Backpropagation

In this course, the backpropagation is performed using the batch gradient descent algorithm. The update rules are

$$\begin{aligned} v_{nk} &\leftarrow v_{nk} - \alpha_v \frac{\partial E}{\partial v_{nk}} \\ w_{kj} &\leftarrow w_{kj} - \alpha_w \frac{\partial E}{\partial w_{kj}} \end{aligned} \tag{9}$$

But, what are $\frac{\partial E}{\partial v_{nk}}$ and $\frac{\partial E}{\partial w_{kj}}$?

Let us compute first $\frac{\partial E}{\partial w_{kj}}$.

$$\frac{\partial E}{\partial w_{kj}} = \frac{\partial}{\partial w_{kj}}\left[\frac{1}{2}\sum_{i=1}^{I}\sum_{j'=1}^{J}\left(g_{j'}^{(i)} - y_{j'}^{(i)}\right)^2\right] = \sum_{i=1}^{I}\frac{\partial}{\partial w_{kj}}\left[\frac{1}{2}\sum_{j'=1}^{J}\left(g_{j'}^{(i)} - y_{j'}^{(i)}\right)^2\right] = \sum_{i=1}^{I}\left(g_j^{(i)} - y_j^{(i)}\right)\frac{\partial g_j^{(i)}}{\partial w_{kj}} \tag{10}$$

For simplicity, let us drop the notation $^{(i)}$ and compute $\frac{\partial g_j}{\partial w_{kj}}$.

$$
\begin{aligned}
\frac{\partial g_j}{\partial w_{kj}} &= \frac{\partial}{\partial w_{kj}} \left(\frac{1}{1 + e^{\left(-(\overline{f})^T w_j\right)}} \right) \\
&= \frac{-e^{\left(-(\overline{f})^T w_j\right)}}{\left(1 + e^{\left(-(\overline{f})^T w_j\right)}\right)^2} (-\overline{f}_k) \\
&= \frac{1}{1 + e^{\left(-(\overline{f})^T w_j\right)}} \left(1 - \frac{1}{1 + e^{\left(-(\overline{f})^T w_j\right)}}\right) \overline{f}_k \\
&= g_j(1 - g_j)\overline{f}_k
\end{aligned} \tag{11}
$$

Hence,

$$
\frac{\partial E}{\partial w_{jk}} = \sum_{i=1}^{I} \left(g_j^{(i)} - y_j^{(i)}\right) g_j^{(i)} \left(1 - g_j^{(i)}\right) \overline{f}_k^{(i)} \tag{12}
$$

$$
w_{kj} \leftarrow w_{kj} - \alpha_1 \sum_{i=1}^{I} \left(g_j^{(i)} - y_j^{(i)}\right) g_j^{(i)} \left(1 - g_j^{(i)}\right) \overline{f}_k^{(i)} \tag{13}
$$

where $k = \{0, 1, \cdots, K\}$ and $j = \{1, 2, \cdots, J\}$.

Now, let us compute $\frac{\partial E}{\partial v_{nk}}$.

$$
\frac{\partial E}{\partial v_{nk}} = \frac{\partial}{\partial w_{nk}} \left[\frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \left(g_j^{(i)} - y_j^{(i)}\right)^2 \right] = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\partial}{\partial v_{nk}} \left[\frac{1}{2} \left(g_j^{(i)} - y_j^{(i)}\right)^2 \right] \tag{14}
$$

Then,

$$
\frac{\partial}{\partial v_{nk}} \left[\frac{1}{2} \left(g_j^{(i)} - y_j^{(i)}\right)^2 \right] = \left(g_j^{(i)} - y_j^{(i)}\right) \frac{\partial g_j^{(i)}}{\partial v_{nk}} \tag{15}
$$

For simplicity, let us drop the notation $^{(i)}$ and compute $\frac{\partial g_j^{(i)}}{\partial v_{nk}}$.

$$
\frac{\partial g_j}{\partial v_{nk}} = \frac{\partial g_j}{\partial \overline{f}_k} \frac{\partial \overline{f}_k}{\partial v_{nk}} \tag{16}
$$

On the other hand,

$$
\begin{aligned}
\frac{\partial g_j}{\partial \overline{f}_k} &= \frac{\partial}{\partial \overline{f}_k} \left(\frac{1}{1 + e^{\left(-(\overline{f})^T w_j\right)}} \right) \\
&= \frac{-e^{\left(-(\overline{f})^T w_j\right)}}{\left(1 + e\left(-(\overline{f})^T w_j\right)\right)^2} (-w_{kj}) \\
&= \frac{1}{1 + e^{\left(-(\overline{f})^T w_j\right)}} \left(1 - \frac{1}{1 + e^{\left(-(\overline{f})^T w_j\right)}}\right) w_{kj} \\
&= g_j(1 - g_j) w_{kj}
\end{aligned} \tag{17}
$$

Now,

$$
\begin{aligned}
\frac{\partial \overline{f}_k}{\partial v_{nk}} &= \frac{\partial f_k}{\partial v_{nk}} \\
&= \frac{\partial}{\partial v_{nk}} \left(\frac{1}{1 + e^{(-(\overline{x})^T v_k)}} \right) \\
&= \frac{-e^{\left(-(\overline{x})^T v_k\right)}}{\left(1 + e^{(-(\overline{x})^T v_k)}\right)^2} (-\overline{x}_n) \\
&= \frac{1}{1 + e^{(-(\overline{x})^T v_k)}} \left(1 - \frac{1}{1 + e^{(-(\overline{x})^T v_k)}} \right) \overline{x}_n \\
&= f_k (1 - f_k) \overline{x}_n
\end{aligned} \tag{18}
$$

Hence,

$$
\frac{\partial E}{\partial v_{nk}} = \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \left(g_j^{(i)} - y_j^{(i)} \right) g_j^{(i)} \left(1 - g_j^{(i)} \right) w_{kj} f_k^{(i)} \left(1 - f_k^{(i)} \right) \bar{x}_n^{(i)} \right\} \tag{19}
$$

Now, by considering all training examples,

$$
v_{nk} \leftarrow v_{nk} - \alpha_2 \sum_{i=1}^{I} \sum_{j=1}^{J} \left\{ \left(g_j^{(i)} - y_j^{(i)} \right) g_j^{(i)} \left(1 - g_j^{(i)} \right) w_{kj} f_k^{(i)} \left(1 - f_k^{(i)} \right) \bar{x}_n^{(i)} \right\} \tag{20}
$$

where $n = \{0, 1, \cdots, N\}$ and $k = \{1, 2, \cdots, K\}$.