

# Introduction to Intelligent Systems / Lecture 3

## Feedforward Neural Network

Jae Yun JUN KIM\*

July 2, 2019

A **feedforward neural network (FFNN)** can be used to solve a classification problem. A FFNN is a network of neurons inspired from the **biological neural network**.

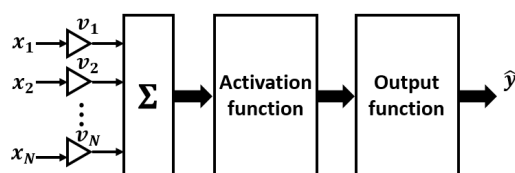


Figure 1: Model for an artificial neuron

### 1 FFNN of two layers

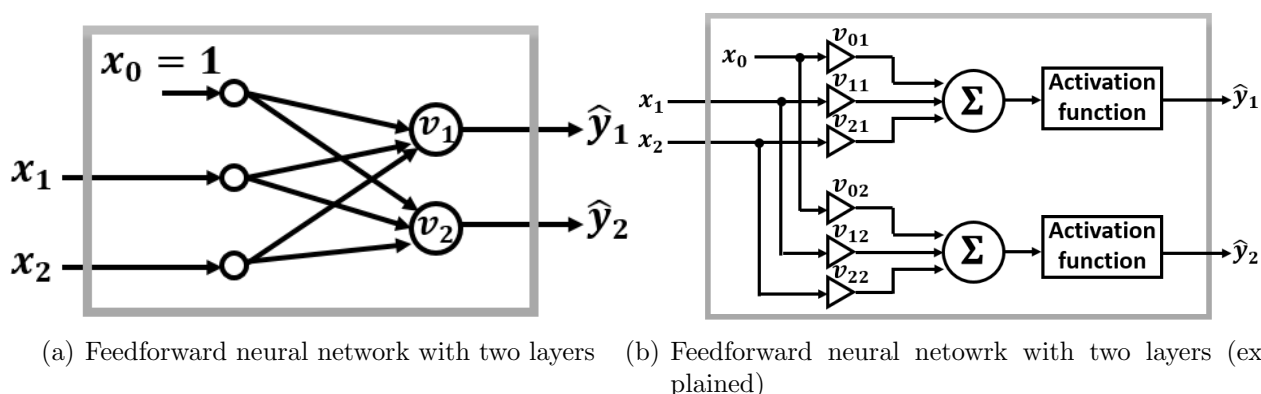


Figure 2: Feedforward neural network with two layers

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

## 2 FFNN of three layers

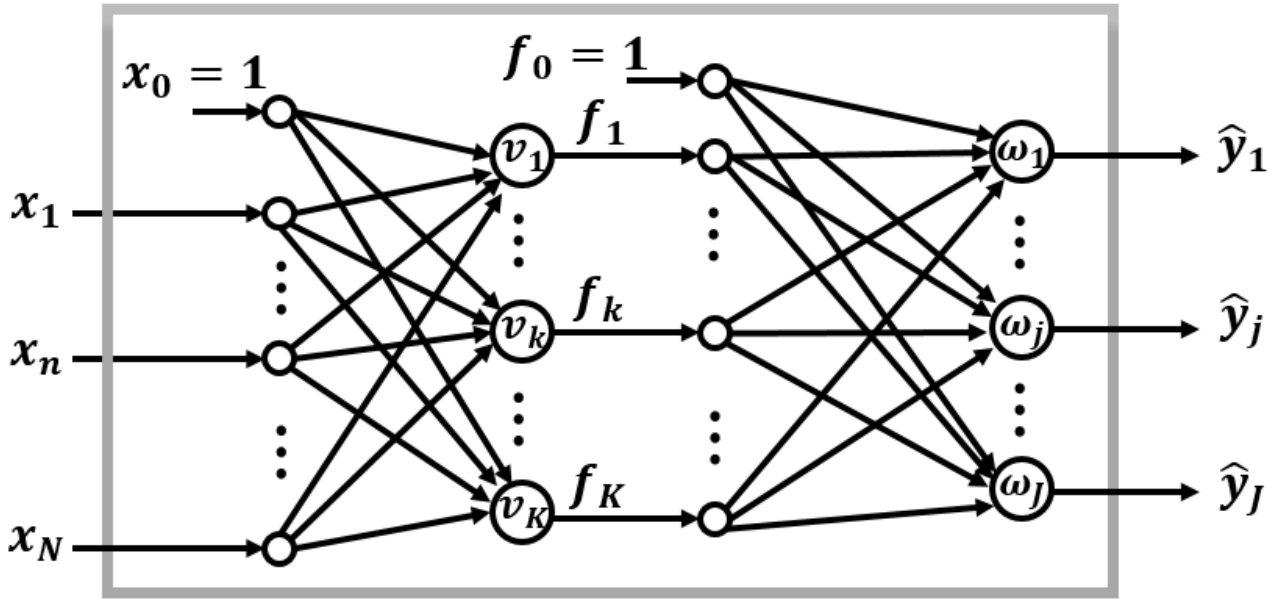


Figure 3: Feedforward neural network with three layers

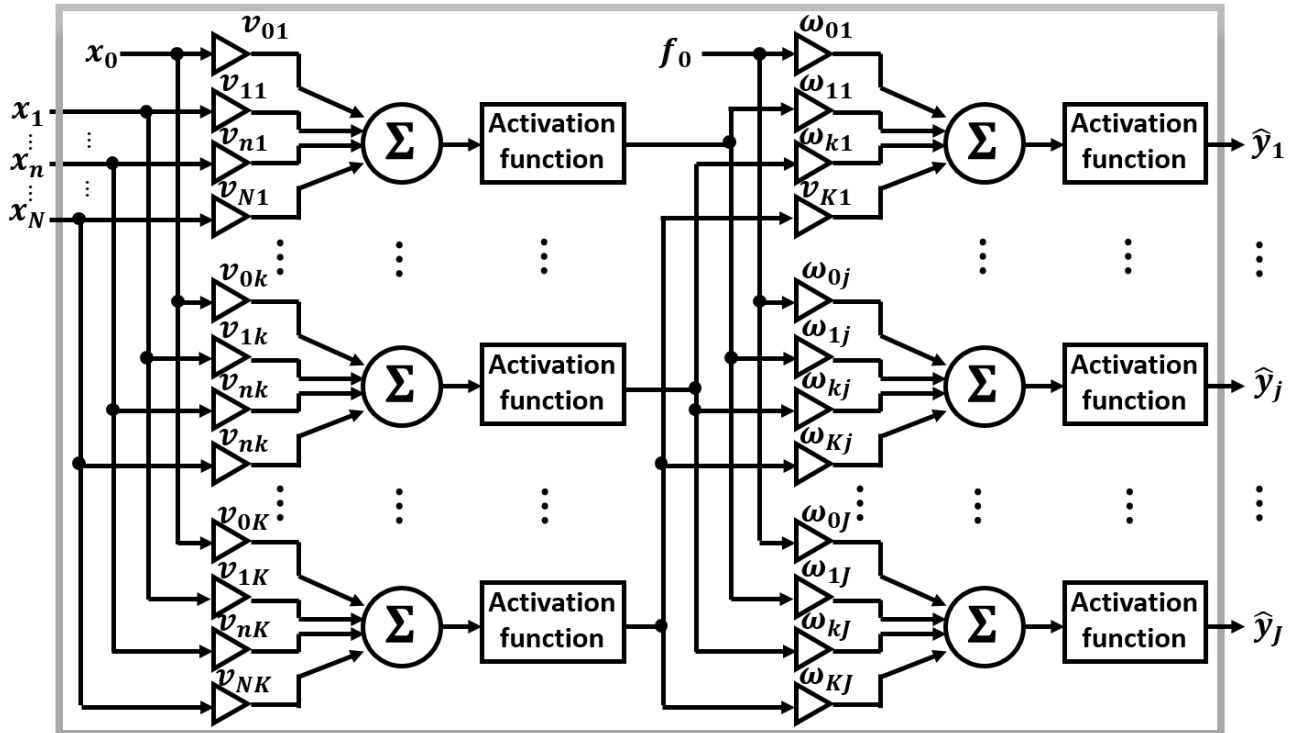


Figure 4: Feedforward neural network with three layers (explained)

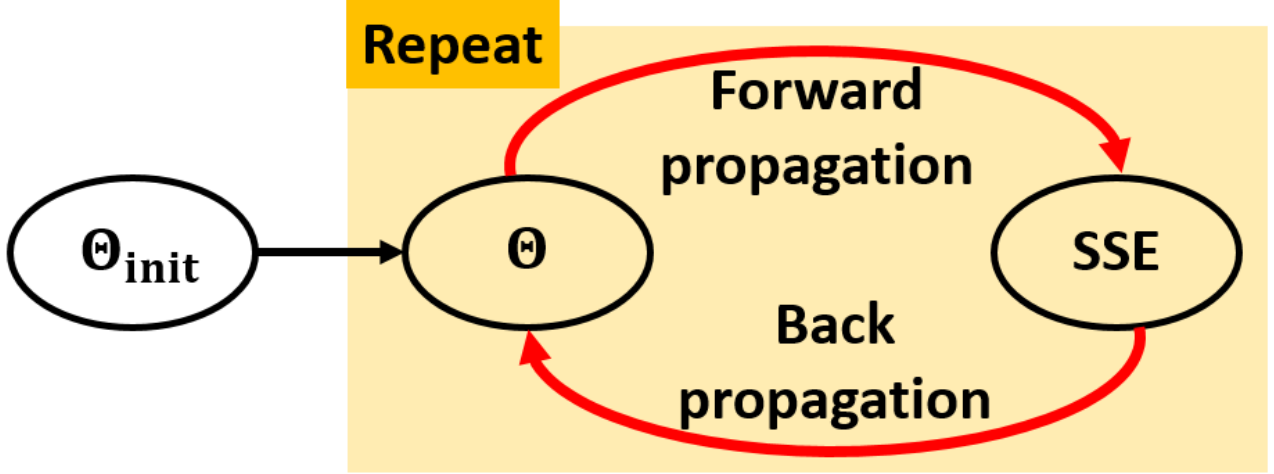


Figure 5: Forward propagation and backpropagation

### 3 Forward propagation and backpropagation

#### 3.1 Forward propagation

Input data matrix

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

Extended input data matrix

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

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

Define

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

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

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

$$\bar{\bar{F}} = \bar{F} \cdot \omega = \begin{bmatrix} (\bar{f}^{(1)})^T \cdot \omega \\ \vdots \\ (\bar{f}^{(I)})^T \cdot \omega \end{bmatrix} \quad (6)$$

Define

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

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

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

---

**Algorithm 1:** find labels

---

```

1 Initialize label;
2 for  $i = 1 : I$  do
3    $\text{label}(i) \leftarrow \arg \max_j G(i, j) - 1$ ;
4 return label;
```

---

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

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

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

### 3.2 Backpropagation

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

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

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

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

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

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

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

Hence,

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

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

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

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

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

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  $\frac{\partial g_j}{\partial v_{nk}}$ .

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

On the other hand,

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

Now,

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

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