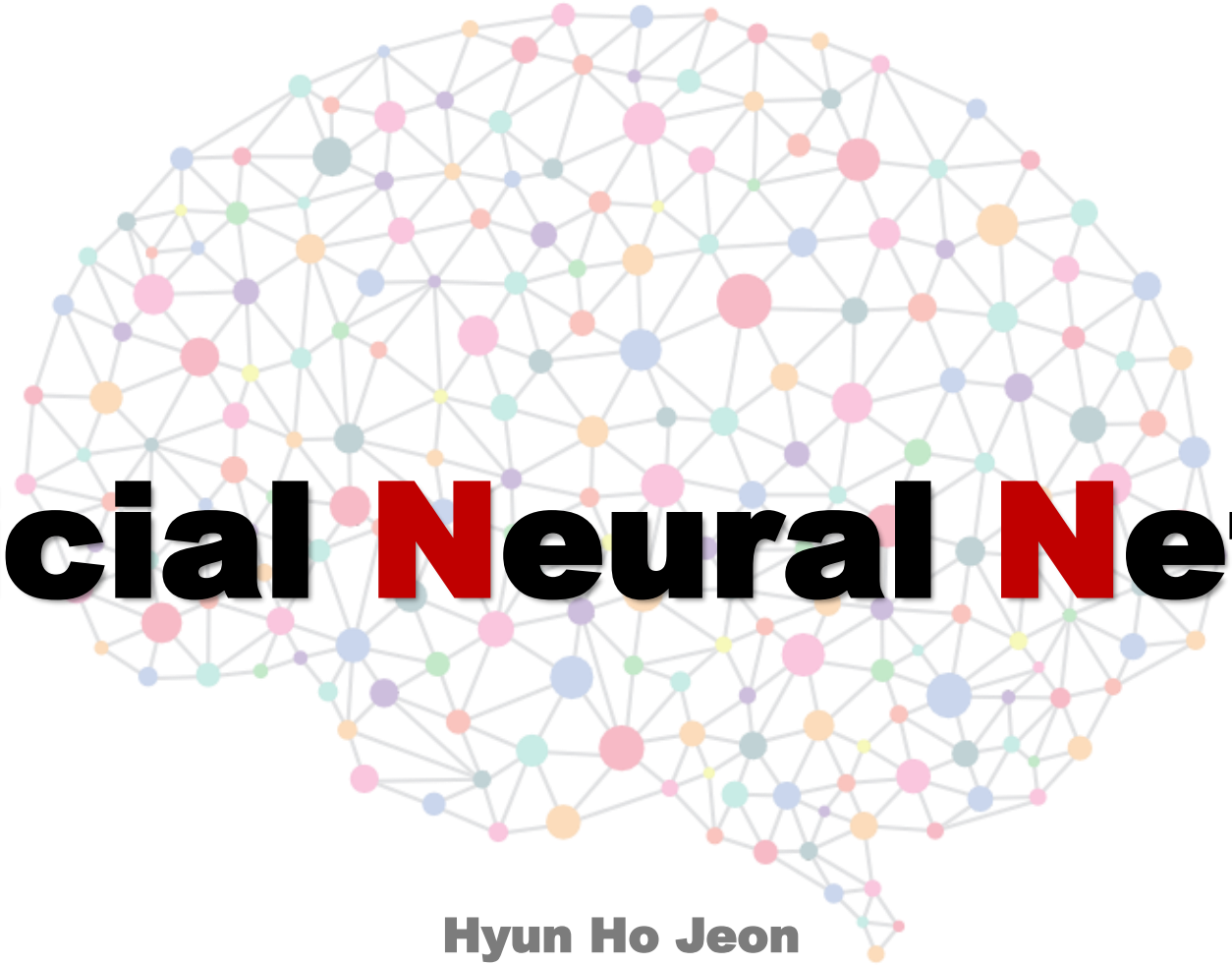
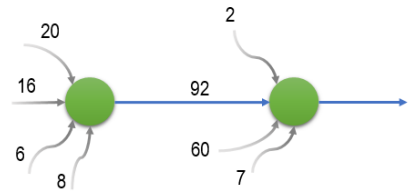


Artificial Neural Network

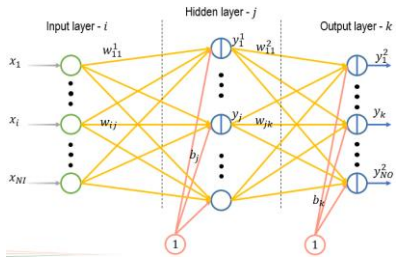
Hyun Ho Jeon
ISL Lab Seminar

Contents

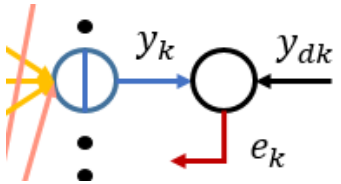
Introduction



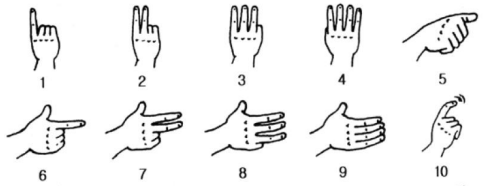
Neural Network



Back Propagation



Implementation



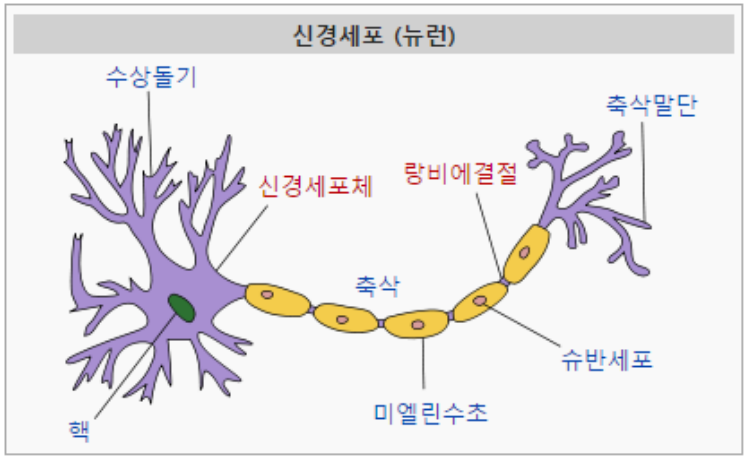
Introduction

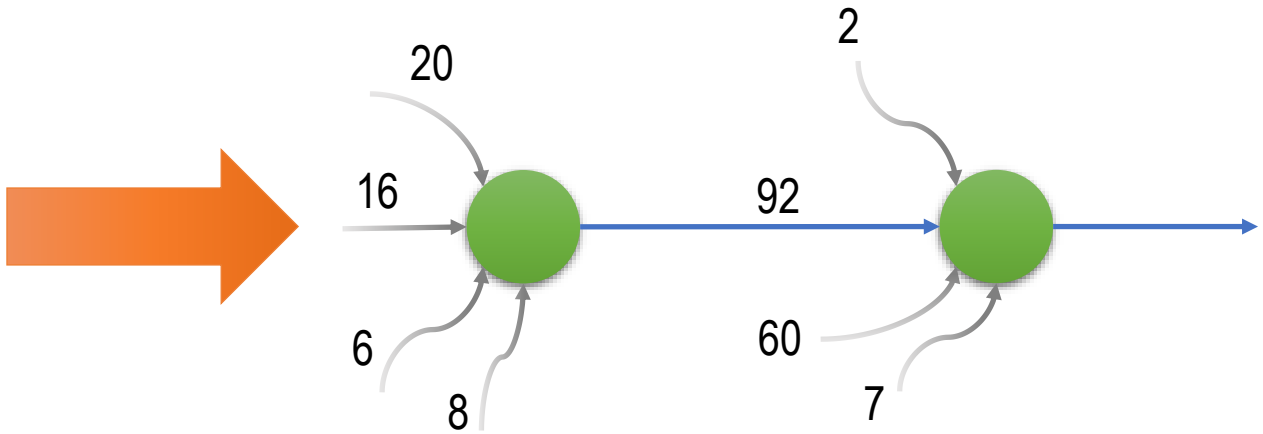
- **History**

Introduction

1943 MP → 1957 P → 1974 BP → 2006 DNN

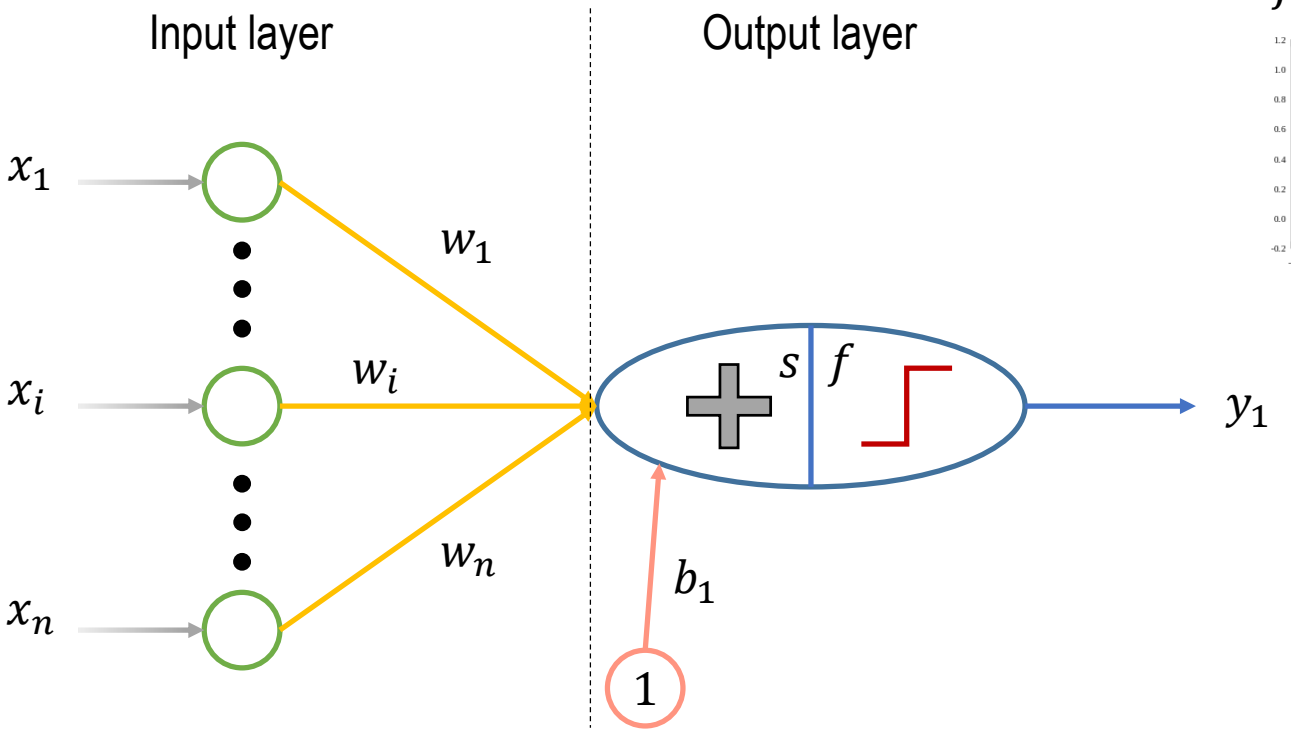
- Neuron



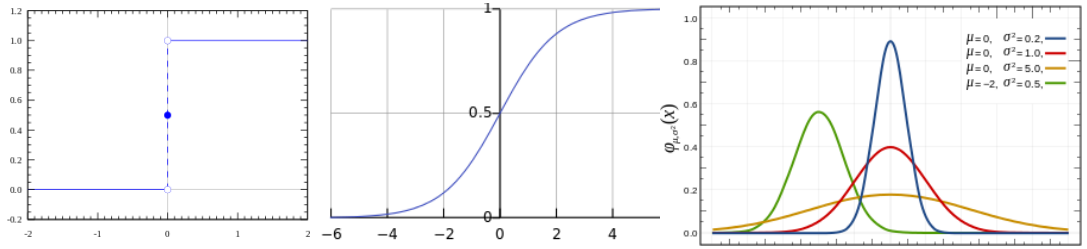


Neural Network

- **Single Layer Perceptron**



$f(t)$:

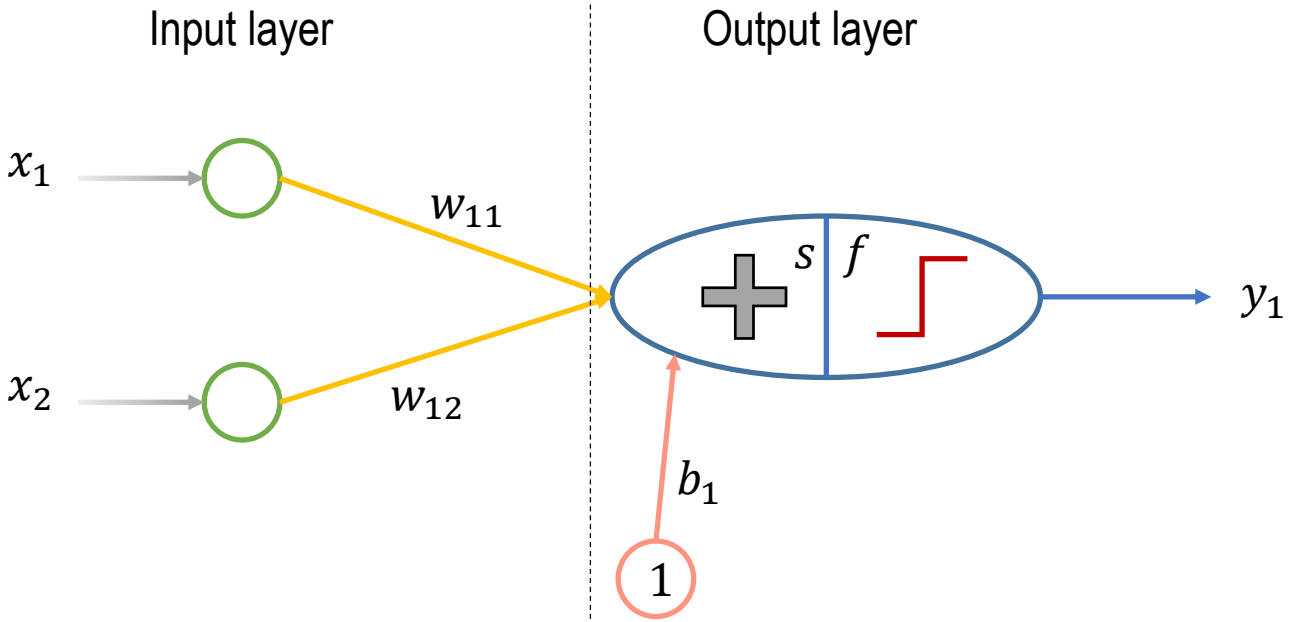


$$s = \sum_{i=1}^{n} x_i w_i + b_j$$

$$y = f(s)$$

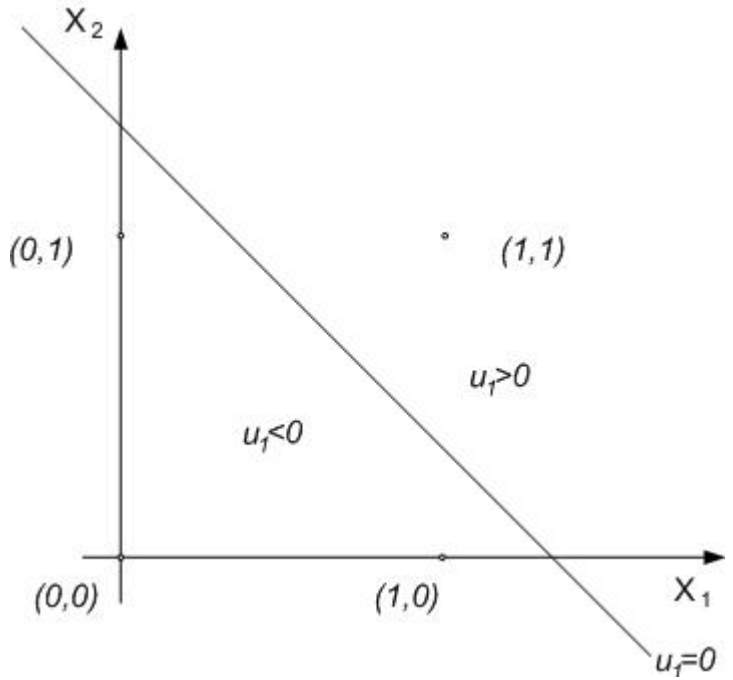
Neural Network

- **Single Layer Perceptron**
  - Example - AND



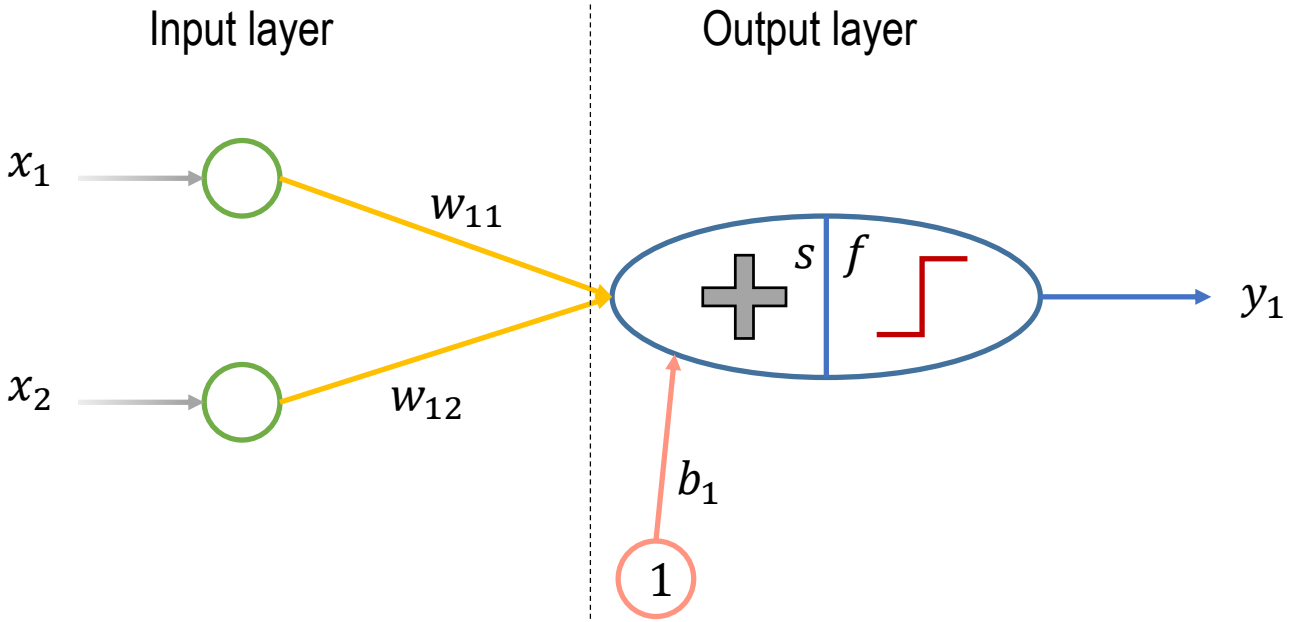
$$y_1 = f(w_{11}x_1 + w_{12}x_2 + b_1) = u_1$$

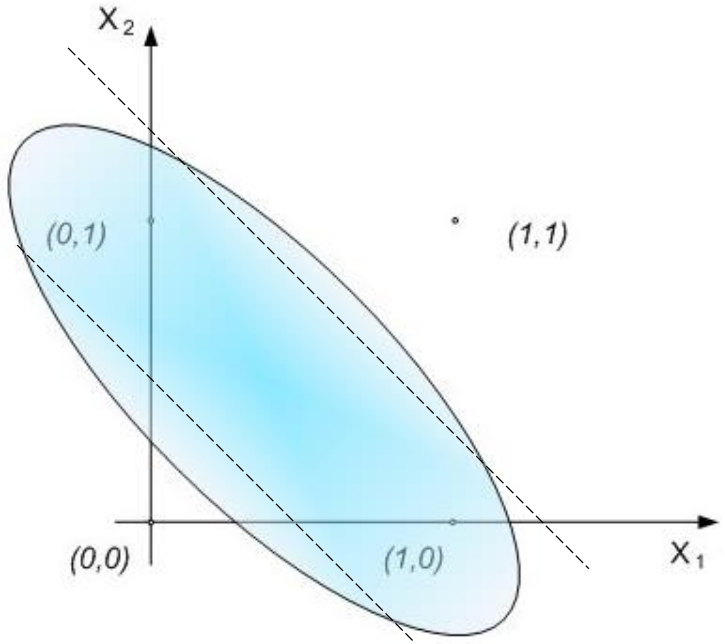
$$x_2 = -\frac{w_{11}}{w_{12}}x_1 - \frac{b_1}{w_{12}} \Rightarrow -x_1 + 1.5$$



Neural Network

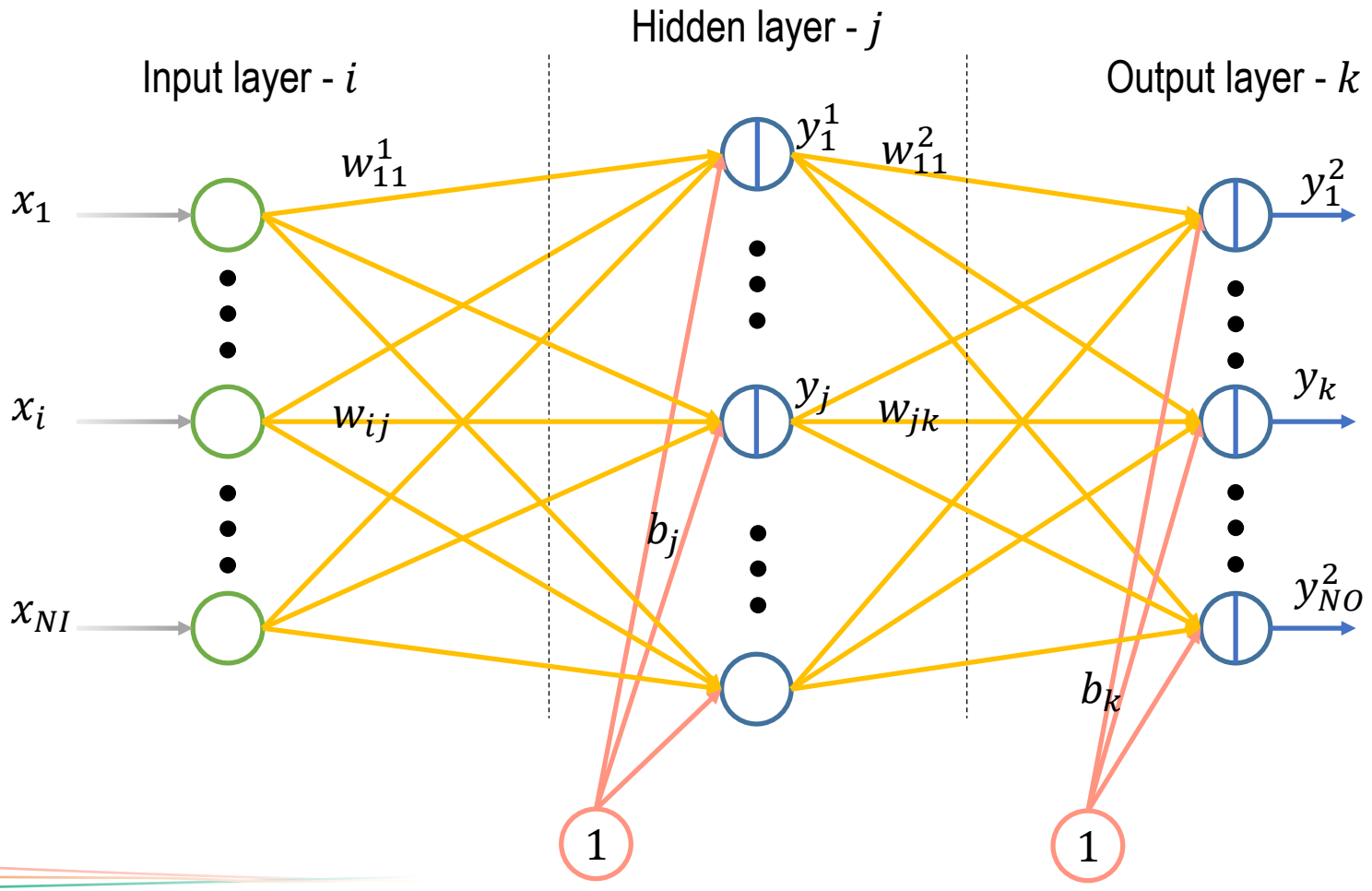
- **Single Layer Perceptron**
  - Example – XOR Problem (*Minsky, M.; S. Papert (1969).《An Introduction to Computational Geometry》. MIT Press.*)





Neural Network

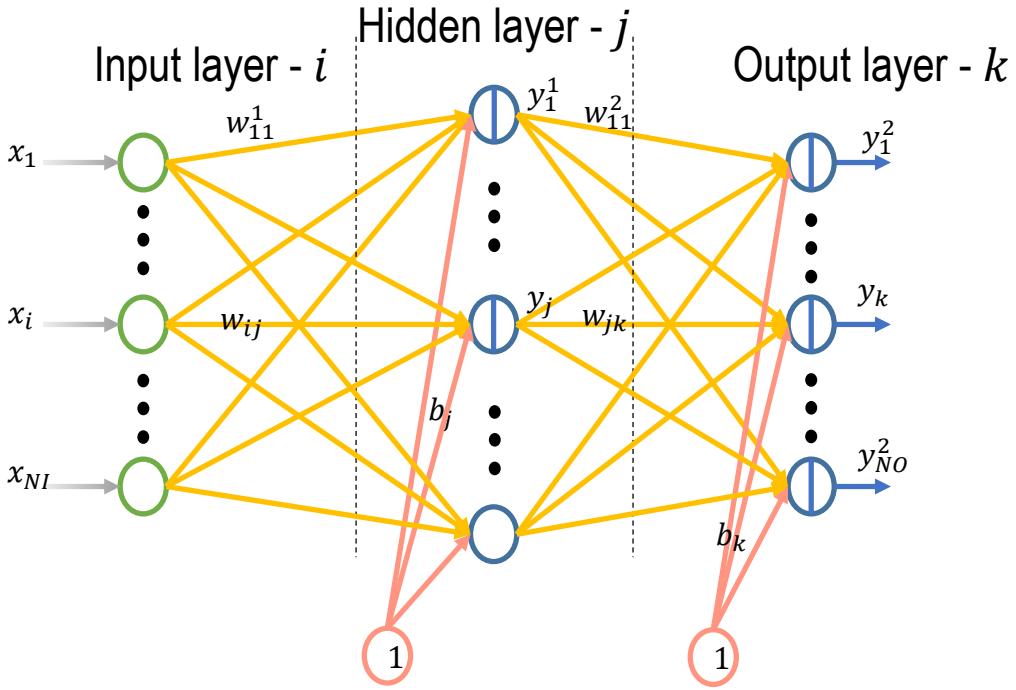
- **Multi Layer Perceptron (MLP)**



Neural Network

1943 MP → 1957 P → 1974 BP → 2006 DNN

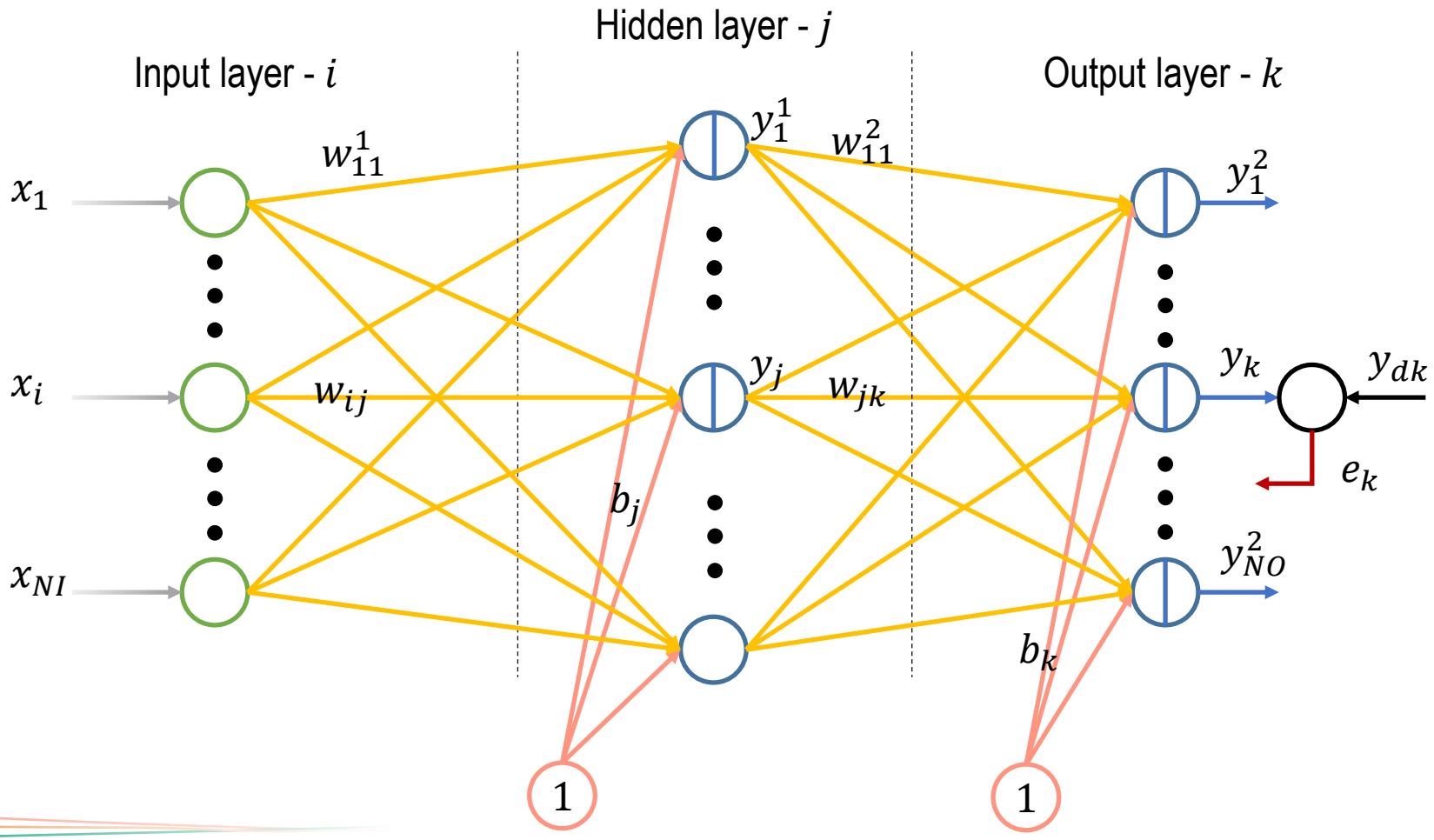
- **Multi Layer Perceptron (MLP)**



Structure	Types of Decision Regions	Exclusive-OR Problem	Classes with Meshed regions	Most General Region Shape
Single-Layer	Half Plane Bounded By Hyperplane	A B B A	B	
Two-Layer	Convex Open Or Closed Regions	A B B A	B	
Three-Layer	Arbitrary (Complexity Limited by No. of Nodes)	A B B A	B	

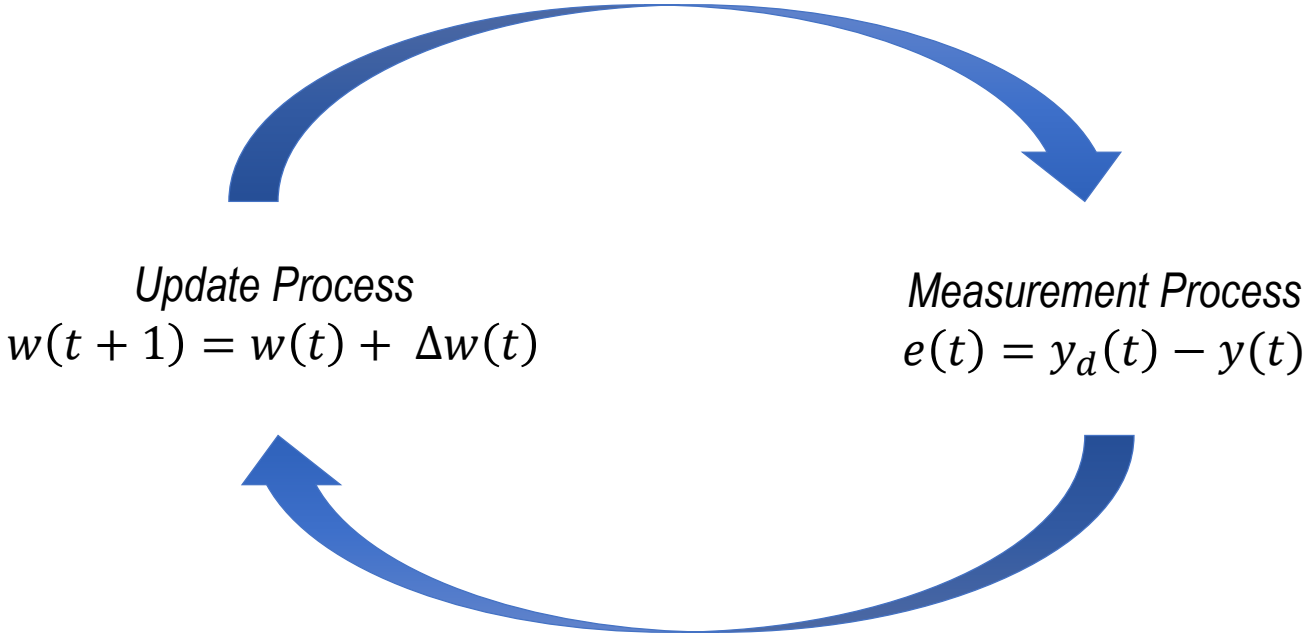
Back Propagation

- **Back Propagation** (Werbos : 1974, Parker : 1982)



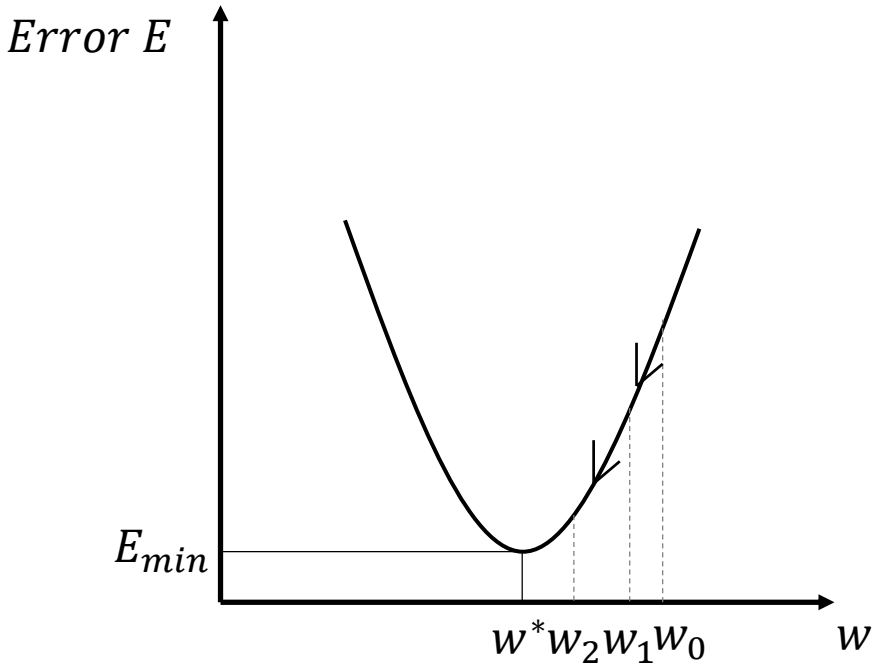
Back Propagation

- **Back Propagation** (Werbos : 1974, Parker : 1982)
  - Concept of gradient descent algorithm



Back Propagation

- **Back Propagation** (Werbos : 1974, Parker : 1982)
  - Gradient descent algorithm



Weight update

$$w(t+1) = w(t) - \Delta w(t)$$

Objective function

$$\mathrm{E} = \frac{1}{2}\sum_{k=1}^{NO} e_j^2$$

The gradient

$$\Delta w(t) = -\eta \frac{\partial E}{\partial w}$$

Back Propagation

- **Back Propagation** (Werbos : 1974, Parker : 1982)
  - Example

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial e_k}\frac{\partial e_k}{\partial w_{jk}} \Rightarrow \boxed{\Delta w_{jk}(t) = -\eta\frac{\partial E}{\partial w_{jk}} = \eta e_k f'(s_k) y_j}$$

$$= e_k \frac{\partial e_k}{\partial w_{jk}} \quad \Leftarrow \quad \frac{\partial E}{\partial e_k} = \frac{1}{2}\frac{\partial e_k^2}{\partial e_k} = e_k$$

$$= e_k \frac{\partial e_k}{\partial y_k}\frac{\partial y_k}{\partial w_{jk}}$$

$$= -e_k \frac{\partial y_k}{\partial w_{jk}} \quad \Leftarrow \quad \frac{\partial e_k}{\partial y_k} = \frac{\partial (yd_k - y_k)}{\partial y_k} = -1$$

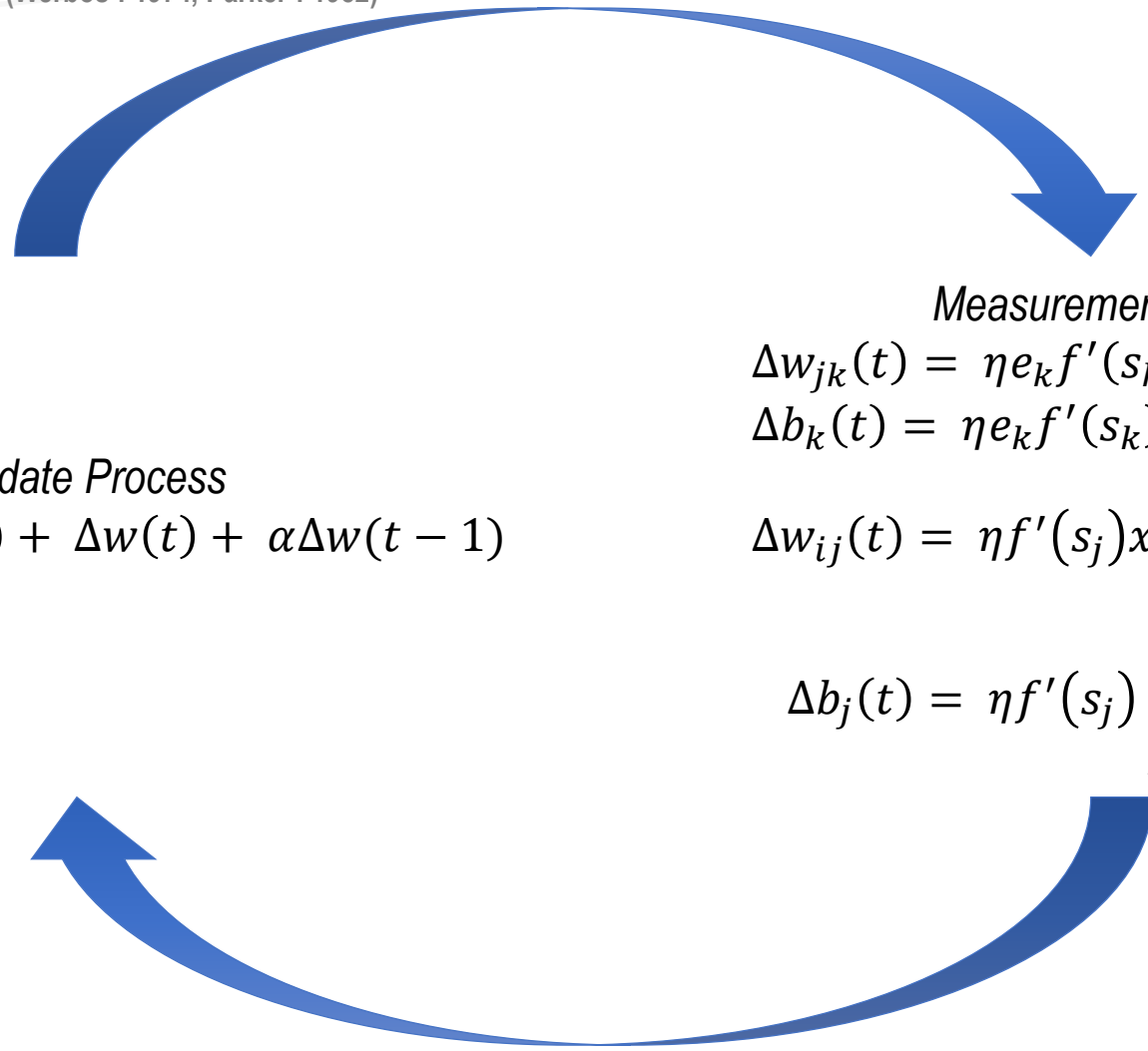
$$= -e_k \frac{\partial y_k}{\partial s_k}\frac{\partial s_k}{\partial w_{jk}}$$

$$= -e_k f'(s_k)\frac{\partial s_k}{\partial w_{jk}} \quad \Leftarrow \quad \frac{\partial y_k}{\partial s_k} = \frac{\partial f(s_k)}{\partial s_k} = f'(s_k)$$

$$= -e_k f'(s_k) y_j \quad \Leftarrow \quad \frac{\partial s_k}{\partial w_{jk}} = \frac{\partial}{\partial w_{jk}}\sum_{j=1}^{NH} w_{jk} y_j + b_k = y_j$$

Back Propagation

- **Back Propagation** (Werbos : 1974, Parker : 1982)



Measurement Process

$$\Delta w_{jk}(t) = \eta e_k f'(s_k) y_j$$

$$\Delta b_k(t) = \eta e_k f'(s_k)$$

$$\Delta w_{ij}(t) = \eta f'(s_j) x_i \sum_{k=1}^{NO} e_k f'(s_k) w_{jk}$$

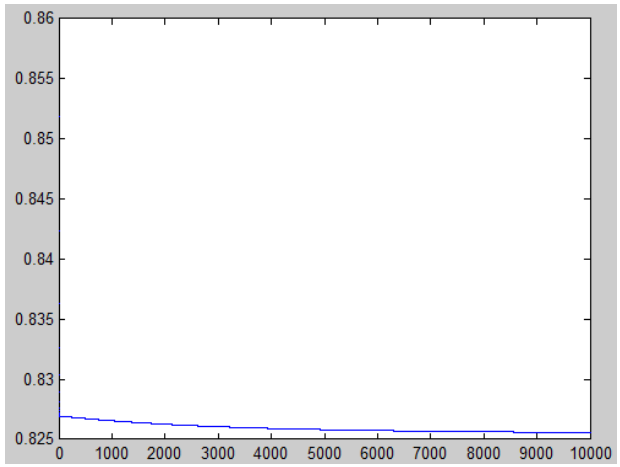
$$\Delta b_j(t) = \eta f'(s_j) \sum_{k=1}^{NO} e_k f'(s_k) w_{jk}$$

Update Process

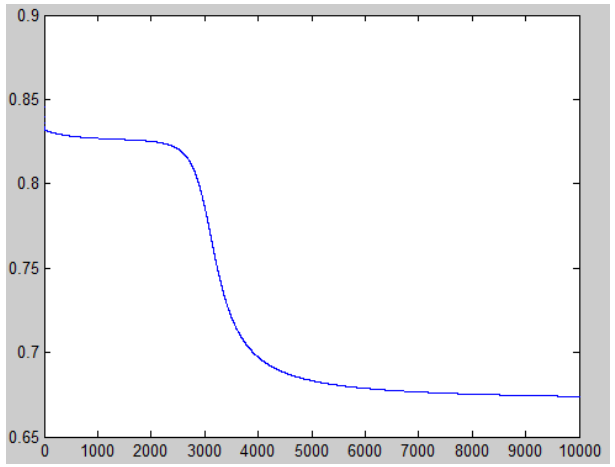
$$w(t+1) = w(t) + \Delta w(t) + \alpha \Delta w(t-1)$$

Implementation

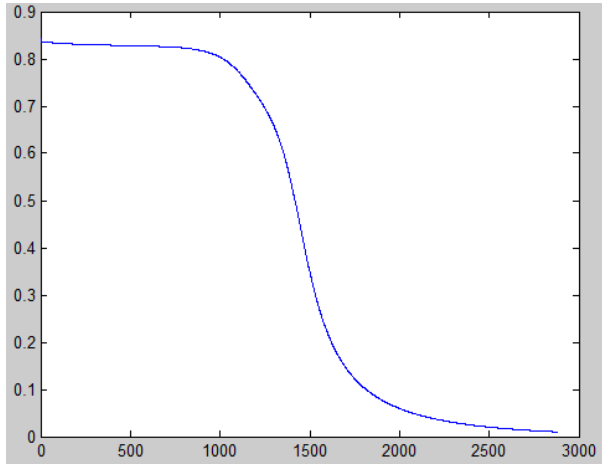
- XOR Classification



$\eta = 0.1$



$\eta = 0.5$



$\eta = 0.9$

Implementation

- **XOR Classification** $\eta = 0.9$

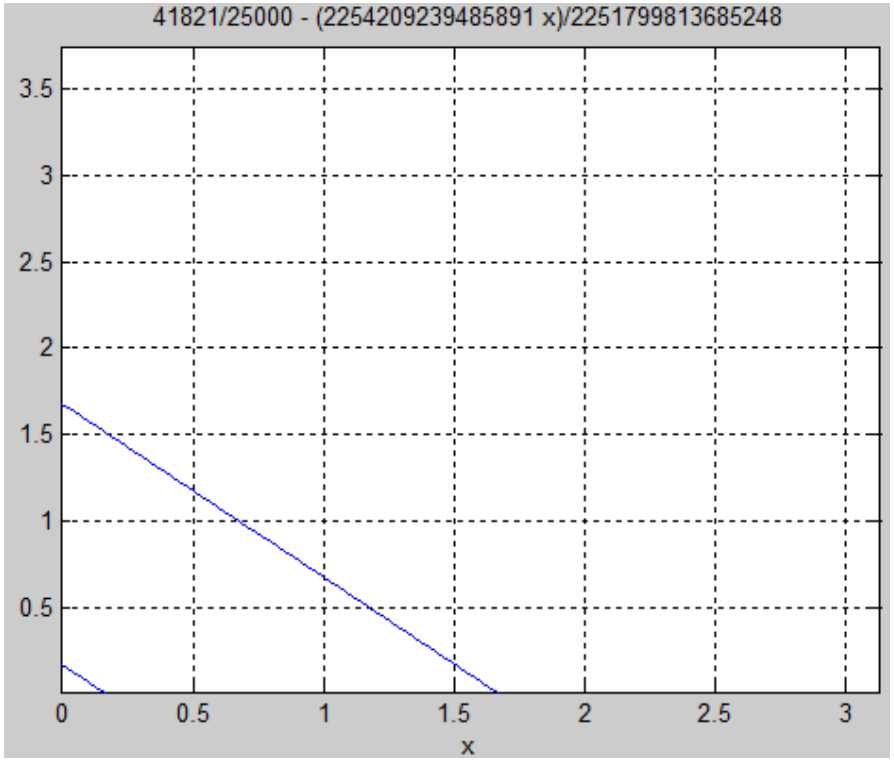
- Parameter

```
clc;
clear all
clf

syms x;
xi=[0.1 0.1; 0.1 0.9; 0.9 0.1; 0.9 0.9];
yd=[0.1; 0.9; 0.9; 0.1];
pattern=4;
eta=0.5;
beta=0.5;
alpha=0.9;
error_threshold=0.01;
in=2;
hidden=2;
out=1;
```

- Output

```
yk =

    0.1023


yk =

    0.8957


yk =

    0.8957


yk =

    0.1076


rms_error =

    0.0100
```

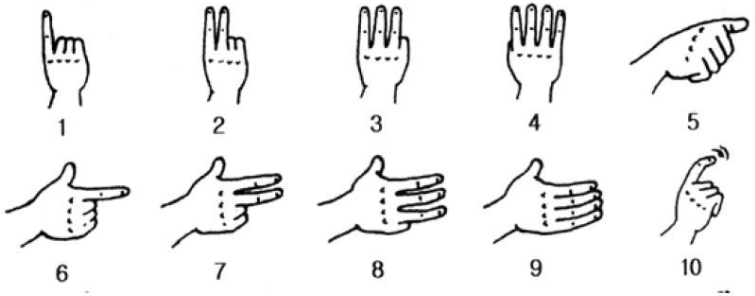
- Classification

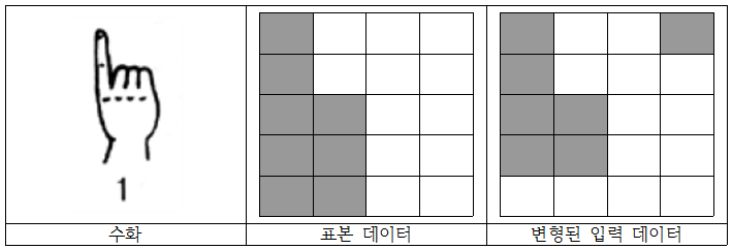


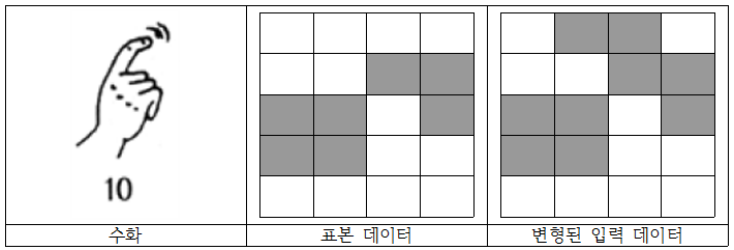
Implementation

- **Deafsign Classification**

- Model







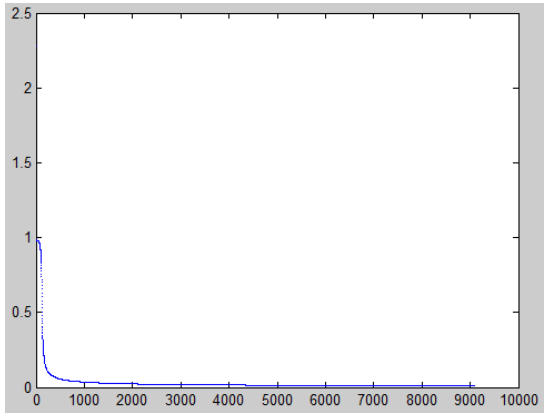
- x_i & y_d

```
xi=[ 1 0 0 0 1 0 0 0 1 1 0 0 1 1 0 0 1 1 0 0;   %1
     1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0;   %2
     1 1 1 0 1 1 1 0 1 1 0 0 1 1 0 0 1 1 0 0;   %3
     1 1 1 1 1 1 1 1 1 1 0 0 1 1 0 0 1 1 0 0;   %4
     0 0 1 1 0 0 0 0 1 1 0 0 1 1 0 0 0 0 0 0;   %5
     0 0 0 0 1 0 1 1 1 1 0 0 1 1 0 0 0 0 0 0;   %6
     0 0 0 0 1 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0;   %7
     0 0 0 0 1 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0;   %8
     0 0 0 0 1 0 1 1 1 1 1 1 1 1 1 1 0 0 1 1;   %9
     0 0 0 0 0 0 1 1 1 0 1 1 1 0 0 0 0 0 0 0]:  %10 인풋 데이터
    |
yd=[1 0 0 0 0 0 0 0 0 0;
    0 1 0 0 0 0 0 0 0 0;
    0 0 1 0 0 0 0 0 0 0;
    0 0 0 1 0 0 0 0 0 0;
    0 0 0 0 1 0 0 0 0 0;
    0 0 0 0 0 1 0 0 0 0;
    0 0 0 0 0 0 1 0 0 0;
    0 0 0 0 0 0 0 1 0 0;
    0 0 0 0 0 0 0 0 1 0;
    0 0 0 0 0 0 0 0 0 1];    %아웃풋 데이터
```

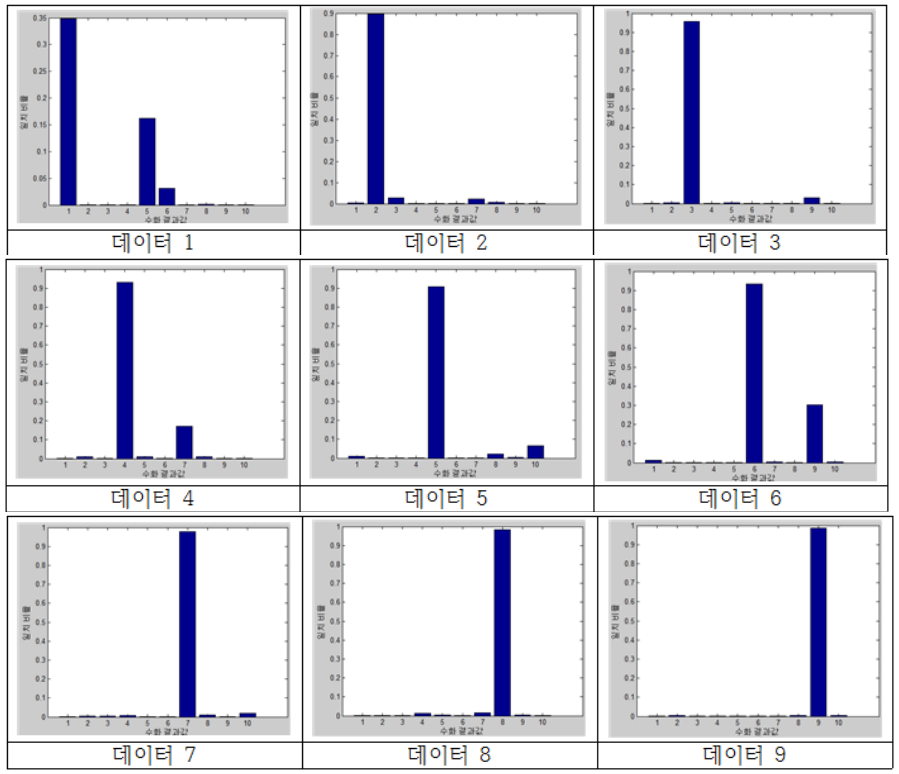
Implementation

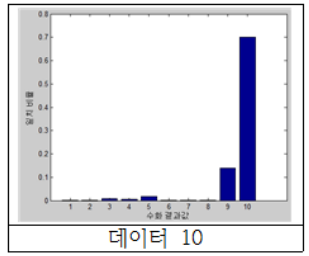
- **Deafsign Classification**

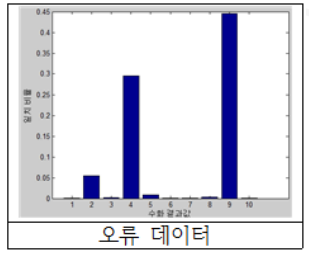
- Learning



- result



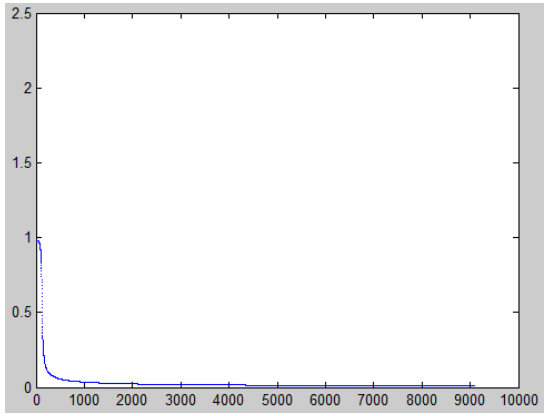




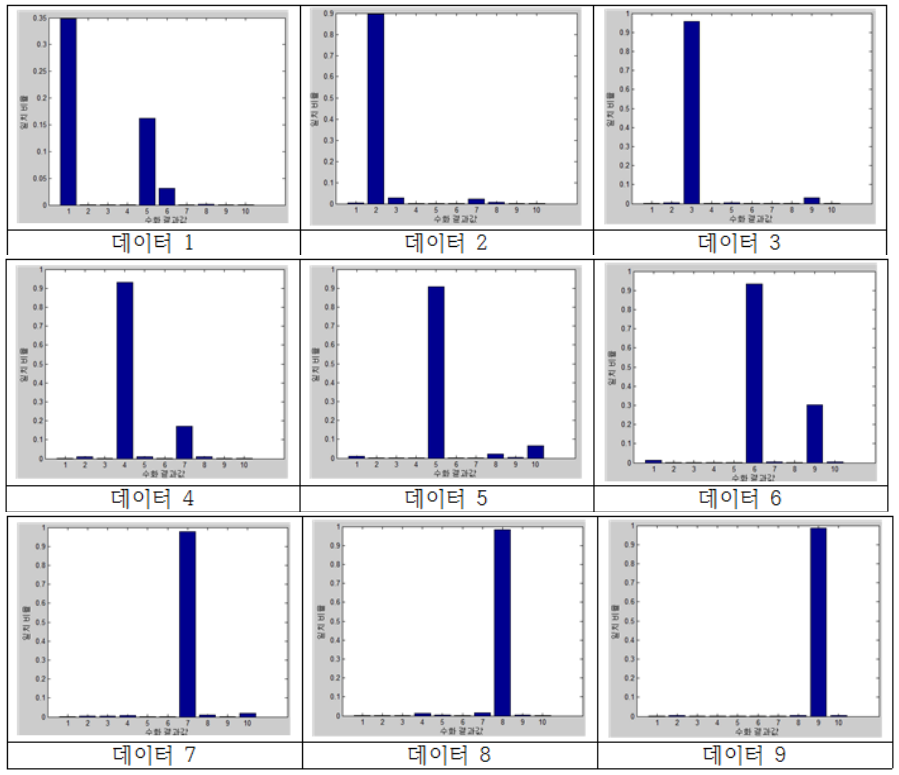
Implementation

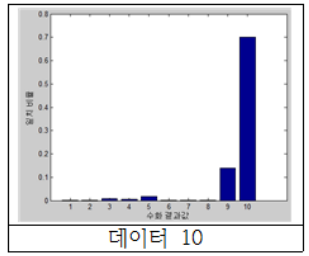
- **Deafsign Classification**

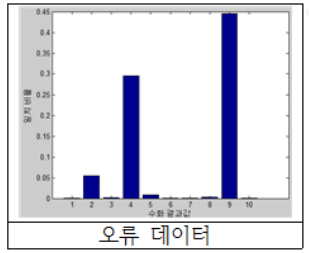
- Learning



- result



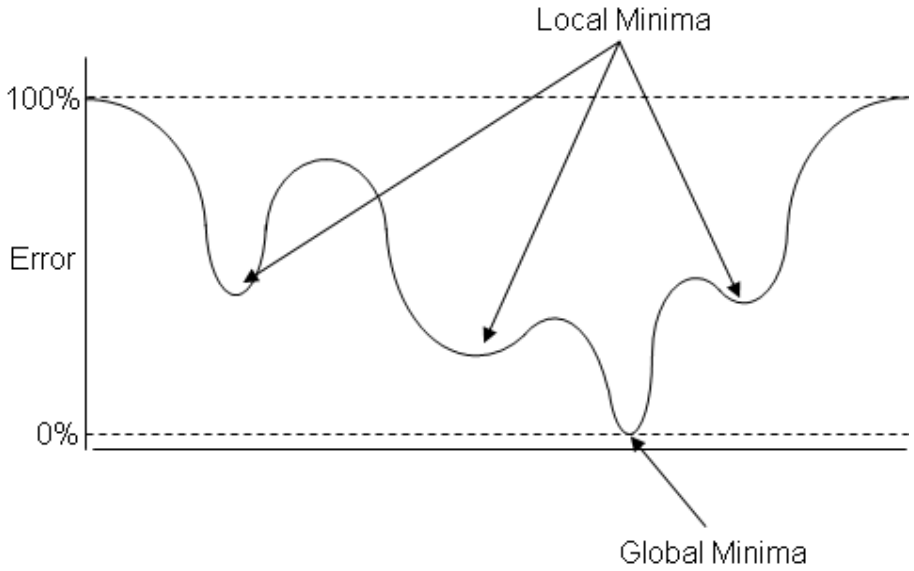




Deep Learning

- **Local minima problem**

- Local minima



- Unsupervised Learning => Pre-training

Depp Learning

1943 MP → 1957 P → 1974 BP → 2006 DNN

- Deep Neural Network

Leon A. Gatys, Alexander S. Ecker, Matthias Bethge. ” A Neural Algorithm of Artistic Style”.

Q&A