

# 8강 신경망모형 I

통계·데이터과학과 장영재 교수



- 01 신경망모형이란?
- 02 신경망의 등장
- 03 신경망의 구성 및 종류
- 04 신경망을 이용한 훈련
- 05 신경망의 장단점
- 06 딥러닝



# 1. 신경망모형이란?



## 01 신경망모형이란

- 신경망모형은 데이터마이닝의 분류 및 예측분야에서 주로 활용되는 계량적 학습방법
  - 인간이 뇌를 통해 문제를 처리하는 방법과 유사
    - 뇌의 기본 구조 조직인 뉴런(neuron)과 뉴런이 연결되어 일을 처리하게 되는데, 이와 유사하게 수학적 모형으로서의 뉴런이 서로 연결되어 네트워크를 형성하도록 구성된 것
  - 신경망은 각 독립적인 과업을 수행하는 뉴런이 연결되어 있고 정보가 많은 연결망을 통해 분산되어 있으므로 일부 뉴런의 문제가 발행하여도 신경망 전체에 큰 영향을 주게 될 가능성이 낮다는 특징
    - 뛰어난 병렬성(parallelism)과 결함 허용(fault tolerance) 능력

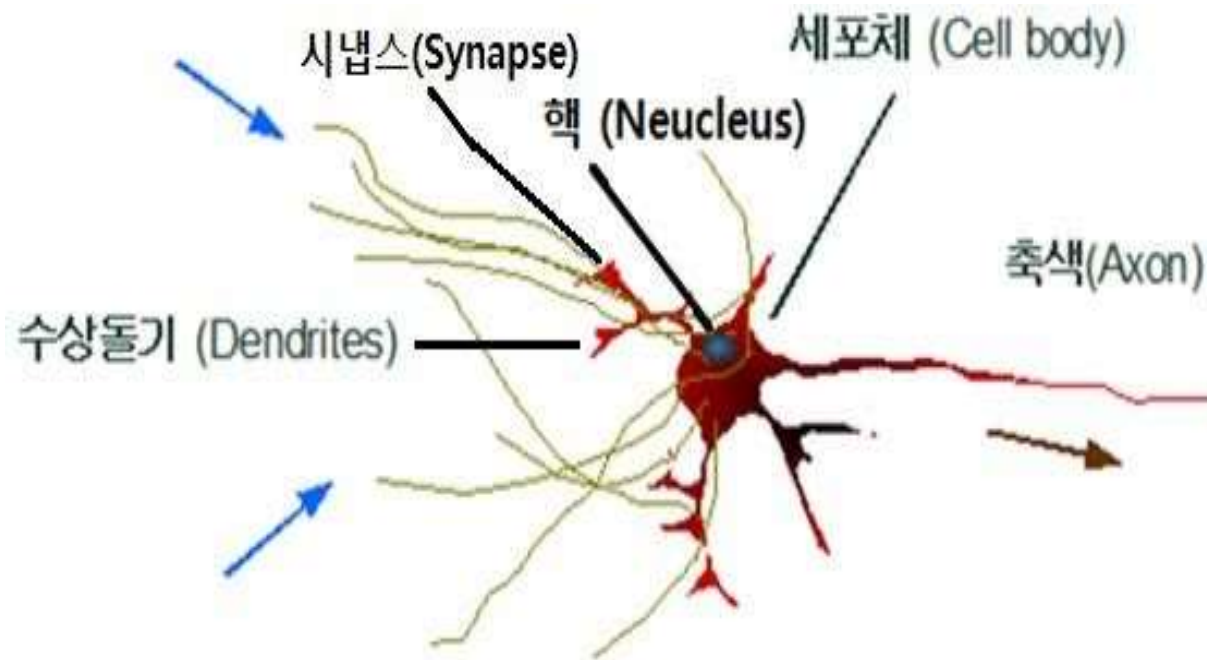


## 01 신경망모형이란

- 인간의 두뇌는 약  $10^{11}$  (1천억) 개의 뉴런(neuron)이라고 불리는 신경세포로 이루어져 있음
  - 수상돌기(dendrite)는 수많은 가지로 이루어져 있으며 신경세포가 신호를 받아들이는 부분
  - 축삭 또는 축삭(axon)은 세포체(cell body)로부터 길게 뻗어나가 있는 가지와 같은 모습이며 수상돌기와 세포체를 통해 전달된 정보를 다른 신경세포나 세포에 전달하는 부분
  - 신경세포들 사이의 신호전달을 위해 맞닿아 있는 부위를 시냅스 (synapse, 연접(連接))라고 하며 한 신경세포가 만들어내는 시냅스는 약 1,000여개 이상으로 알려져 있음



## 01 신경망모형이란



〈그림1〉 인간의 뉴런(neuron)



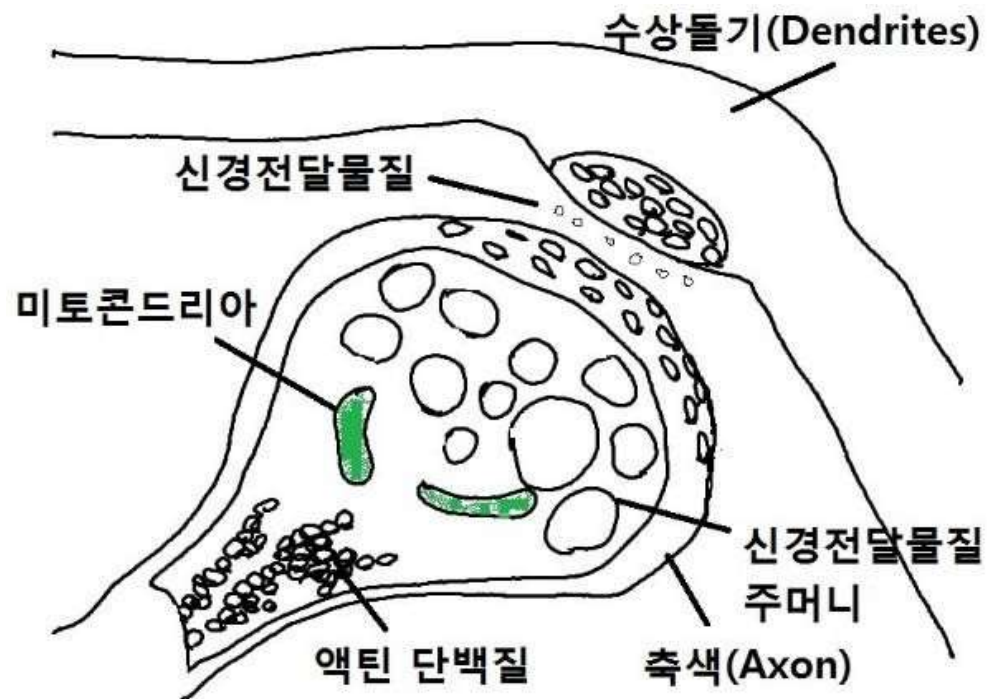


## 01 신경망모형이란

- 시냅스에서는 각 수상돌기의 정보를 그대로 통합하는 것이 아니라 종류에 따라 그 가중치를 달리하여 합해서 전달
- 종합된 정보의 값이 일정 수준보다 작으면 다른 축색으로 전달되지 않으나 그보다 크면 다음 단계로 전달



## 01 신경망모형이란



〈그림2〉 시냅스(synapse)의 구조





# 01 신경망모형이란

## ■ 인간의 신경전달체계는 크게 세 단계로 구성

- 두뇌(신경망)에 이르기 전까지의 단계가 수용체에서의 정보변환 단계
  - 외부의 자극(아날로그 정보)이 있을 때, 수용체(receptor)에서 이 자극을 전기적 신호(디지털 정보)로 변환하여 두뇌로 전달하게 되는 과정
- ‘두뇌(신경망)’이라고 표시된 것이 뉴런들의 복잡하고 유기적인 결합체
- 효용체(effector)에서의 신호 전달 단계는 두뇌에서 결정한 전기적 신호를 다시 아날로그 신호로 바꾸는 과정



## 01 신경망모형이란



〈그림3〉 인간 두뇌의 학습과정



## 2. 신경망의 등장



## 02 신경망의 등장

### 1 신경망의 등장배경

#### ■ 신경망은 정보처리 이론이 가지고 있던 한계를 극복하고자 하는 노력에서 비롯

- 정보처리 이론은 인간의 지능을 구성하는 중요한 요소인 상식의 추론과 패턴인식 기능을 설명하는데 한계
  - 정보처리 이론은 인간이 정보를 기호에 의해 처리하는 것으로 전제하고 있지만, 인간은 기호처리 이외에 다른 형태로도 정보를 처리하고 있기 때문
  - 신경망에서는 병렬처리 방식을 기본으로 하고 있으며 불완전한 자료를 가지고도 상황에 따라 최적의 의사결정을 하게 됨



## 02 신경망의 등장

### 1 신경망의 등장배경

■ 신경망에 대한 연구는 1940년대부터 이루어져 최근 급속도로 발전

연도	연구자	주요 개발내용
1943	매컬릭과 피츠	뉴런을 모형화
1949	헵	뉴런의 연결강도 조정하는 학습규칙 제안
1957	로젠블랫	실질적 신경망, 퍼셉트론 개발(문자인식)
1959	위드로우	Adaline(Adaptive linear) 개발
1969	민스키와 페이퍼트	다층신경망 개발(비선형성 해결)
1982	홉필드	역전파 알고리즘 개발(성능 개선)
1986	러멜하트와 맥클랜드	병렬분산처리 제안



## 02 신경망의 등장

### 1 신경망의 등장배경

- 다층신경망을 통해 대표적인 비선형적 문제로 꼽을 수 있는 XOR(Exclusive OR) 문제와 같은 문제를 해결하면서 한 단계 도약

- XOR은 배타적 논리합이라고도 하는데, 수리 논리학에서 주어진 2개의 명제 가운데 1개만 참일 경우를 판단하는 논리 연산

명제 P	명제 Q	XOR값
1	1	0
1	0	1
0	1	1
0	0	0

- 컴퓨터의 발달을 기반으로 1980년대에 홉필드(Hopfield)에 의해 역전파(backpropagation) 알고리즘이 제안되고 러멜하트(Rumelhart)와 맥클랜드(Mc Clelland)의 병렬분산처리에 관한 책이 발표되면서 다시 각광을 받기 시작



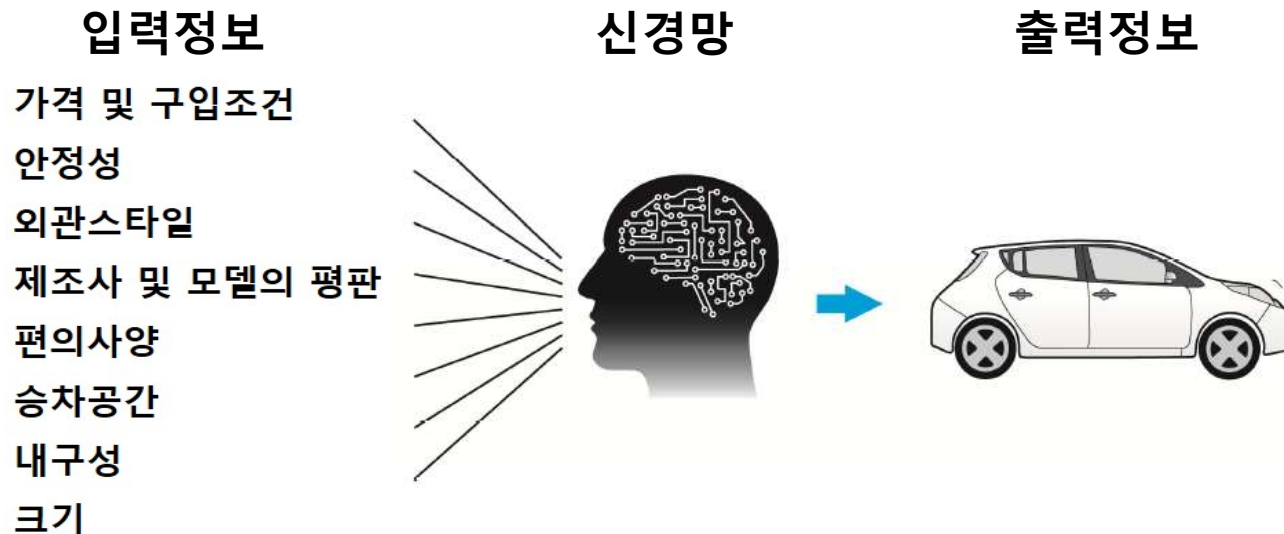


## 02 신경망의 등장

### 2 신경망의 응용

#### ■ 신경망모형이 개발되기 이전 기존의 예측모형은 엄격한 통계적 가정 하에 구성

- 우리가 당면하는 문제는 단순하거나 선형적이지 않은 문제들, 즉 매우 복잡하고 비선형적인 문제들인 경우가 많음



〈그림4〉 자동차 구입에 관한 의사결정

### 3. 신경망의 구성 및 종류



## 03 신경망의 구성 및 종류

### 1 신경망의 구성요소

- 입력층 : 입력변수에 대응되는 노드로 구성
- 은닉층 : 입력층으로부터 전달되는 변수값들의 선형결합을 비선형함수로 처리하여 출력층 또는 은닉층에 전달
- 출력층 : 목표변수에 대응되는 노드
  - 단층신경망 : 입력층과 출력층으로만 구성
  - 다층신경망 : 입력층, 은닉층, 출력층으로 구성



## 03 신경망의 구성 및 종류

### 2 신경망의 종류

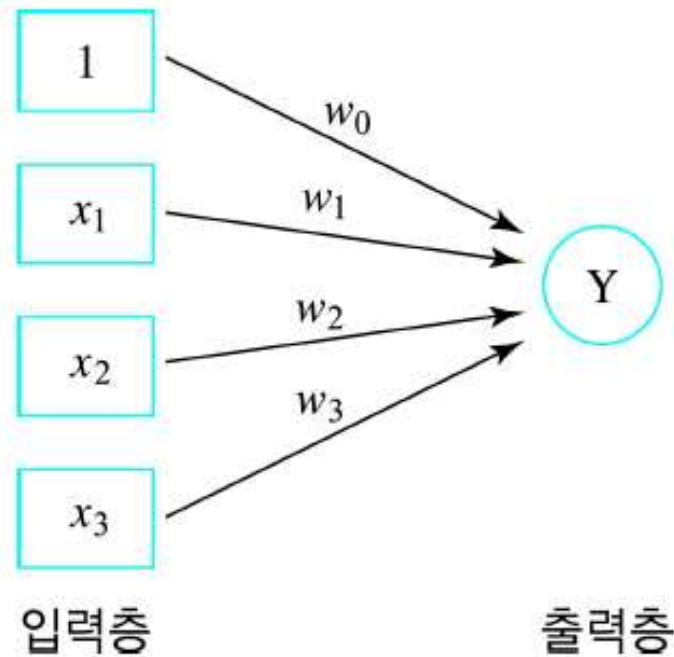
#### (1) 단층신경망(single-layer perception)

- 입력층은 3개의 변수( $x_1, x_2, x_3$ )로 구성
- '1'로 표시된 것은 모형의 상수항을 의미
- 출력층은 특별한 변환 없이 입력변수들의 가중평균으로 표현  
(가중치는 입력노드 옆에 표기된  $w_0, w_1, w_2, w_3$  등)
- 합성함수 : 입력변수를 결합하는 함수
- 활성화함수 : 결합된 입력정보를 변형



## 03 신경망의 구성 및 종류

### 2 신경망의 종류



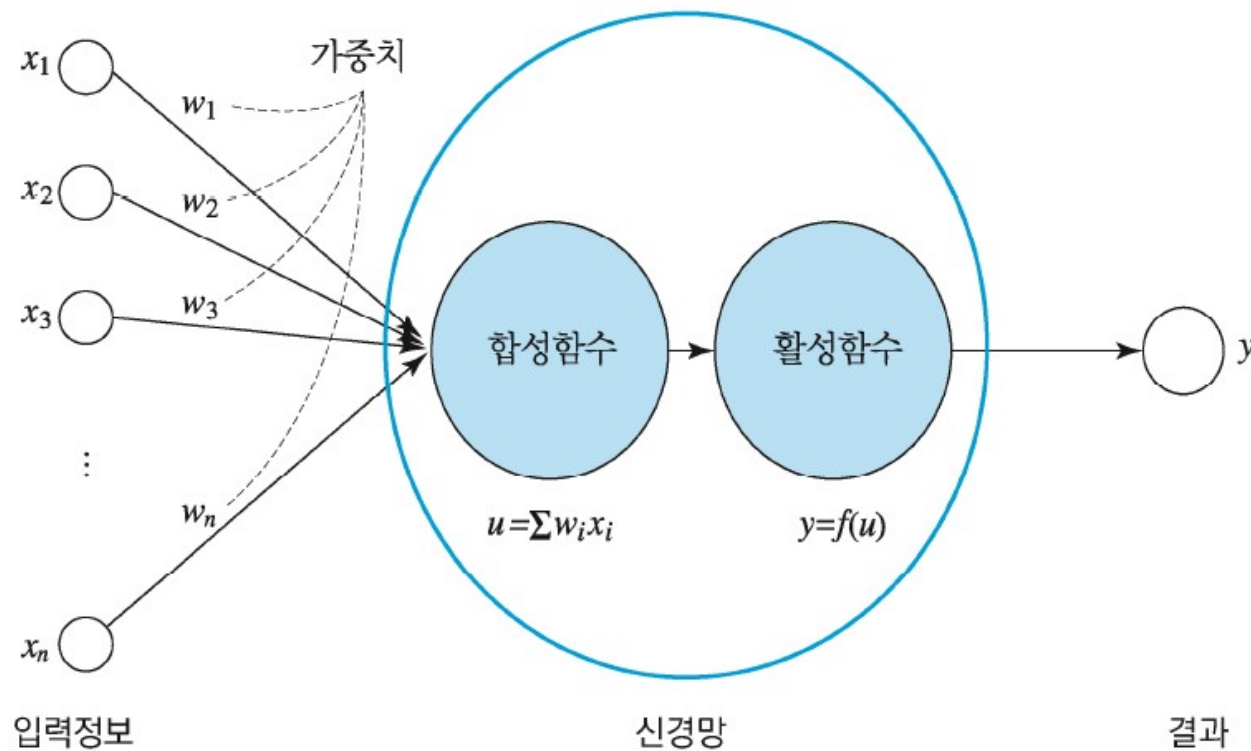
$$y_1 = w_0 + w_1x_1 + w_2x_2 + w_3x_3$$

〈그림5〉 간단한 단층신경망 (single-layer perception) 구조



## 03 신경망의 구성 및 종류

### 2 신경망의 종류



〈그림6〉 신경망의 구조





## 03 신경망의 구성 및 종류

### 2 신경망의 종류

- 활성화함수(또는 전이함수(transfer function))는 일반적으로 시그모이드(sigmoid)함수가 많이 사용됨
  - 시그모이드 함수 : S자 형태의 비선형함수로서 출력값은 0과 1 사이 값을 지니게 되는 미분 가능한 함수
  - 역전파 학습 알고리즘의 특성에 잘 맞으며 복잡한 유형의 의사결정 문제에도 효과적으로 적용
  - 큰 입력값에 대해서도 출력값이 급격히 변화하지 않으며 작은 입력값도 놓치지 않는다는 장점



## 03 신경망의 구성 및 종류

### 2 신경망의 종류

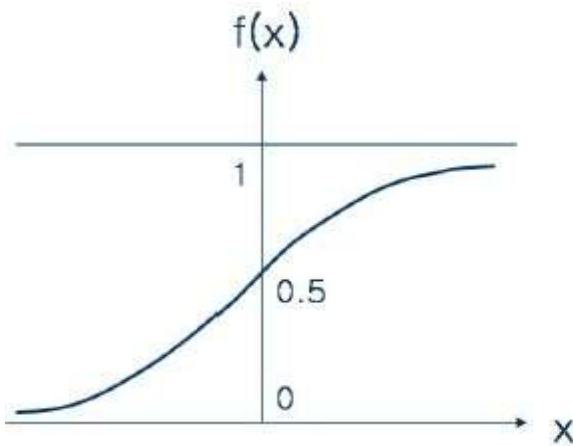
#### 〈활성함수의 종류〉

함수	함수형태	범위
logistic(sigmoid)	$\frac{1}{1+e^{-x}}$	$(0, 1)$
tanh	$\frac{e^x - e^{-x}}{e^x + e^{-x}} = 2 \times \text{logistic}(2x) - 1$	$(-1, 1)$
linear(identity)	$x$	$(-\infty, \infty)$
Gauss	$e^{-x^2/2}$	$(0, 1)$
threshold	$\begin{matrix} 0 & \text{if } x < \theta \\ 1 & \text{if } x \geq \theta \end{matrix}$	0 또는 1



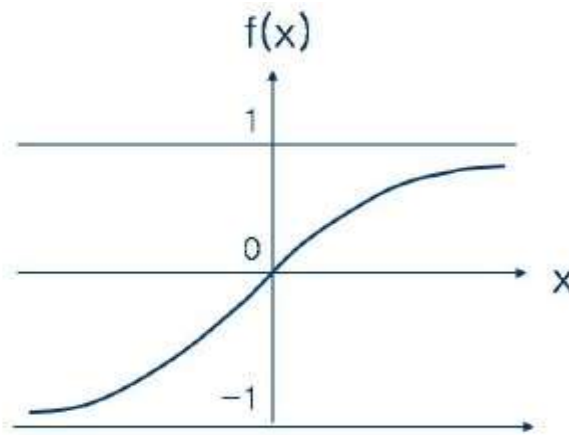
## 03 신경망의 구성 및 종류

### 2 신경망의 종류



단극성 시그모이드 함수

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



양극성 시그모이드 함수 (tanh과 유사)

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

〈그림7〉 시그모이드 함수의 종류

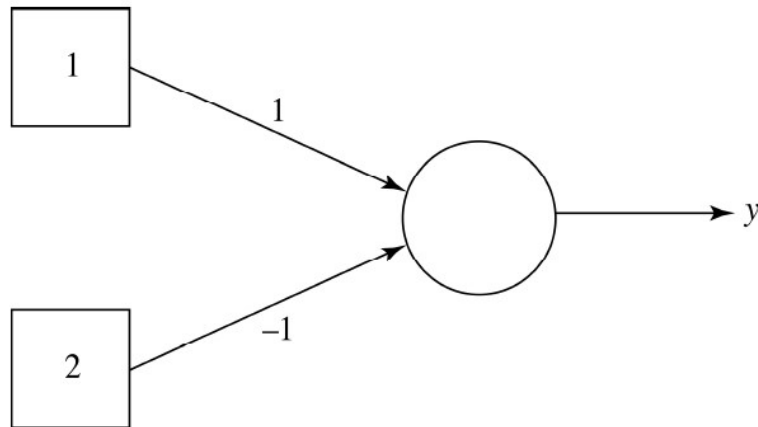


## 03 신경망의 구성 및 종류

### 2 신경망의 종류

- 예제 5-2: 다음과 같은 신경망 모형에서 합성함수로는 가중합을 이용하고, 활성화 함수는 하이퍼볼릭 탄젠트 함수를 이용하였을 때, 출력값은 어떻게 산출되는가?

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

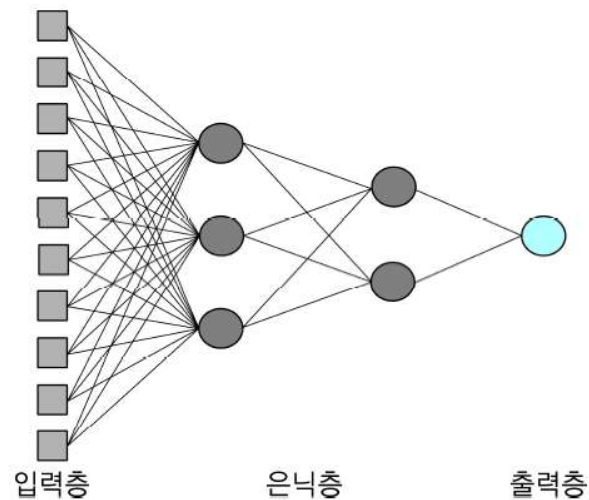


## 03 신경망의 구성 및 종류

### 2 신경망의 종류

#### (2) 다층신경망(multi-layer perception)

- 비선형적인 신경망 작동원리를 구현하기 위해서는 다층신경망(multi-layer perceptron)이 적절



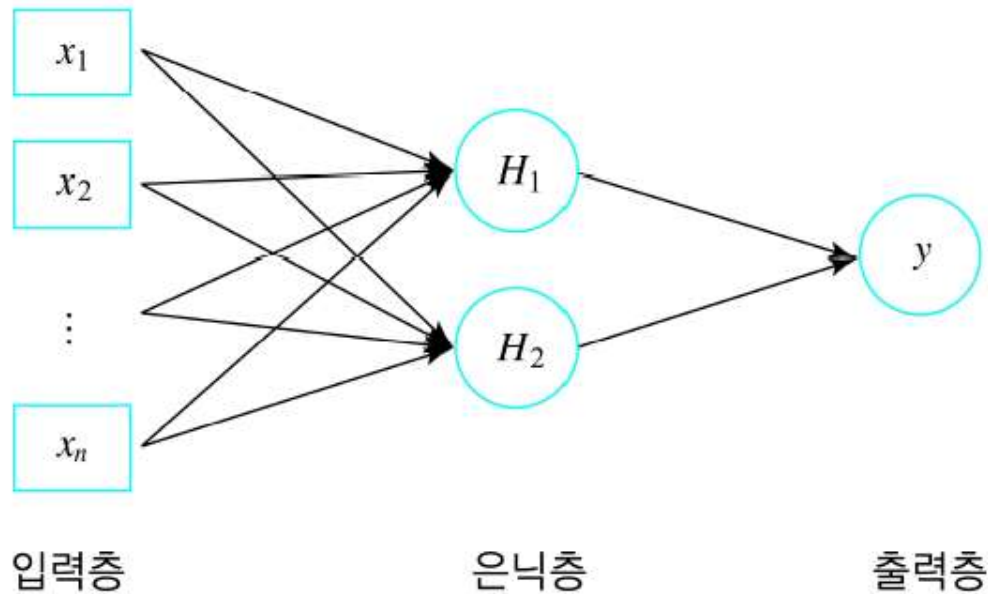
〈그림8〉 다층신경망(multi-layer perception) 구조



## 03 신경망의 구성 및 종류

### 2 신경망의 종류

- 다음은 은닉층이 하나인 다층신경망의 사례로서 입력노드는  $n$ 개이며 은닉층은 두 개의 노드로 구성되어 있음



〈그림9〉 다층신경망의 예





## 03 신경망의 구성 및 종류

### 2 신경망의 종류

$$H_1 = f_1(w_{01} + w_{11}x_1 + \cdots + w_{n1}x_n)$$

$$H_2 = f_2(w_{02} + w_{12}x_1 + \cdots + w_{n2}x_n)$$

$$y = g(w_{00} + w_{10}H_1 + w_{20}H_2)$$

- $f_1$ 과  $f_2$ 는 활성화함수이고,  $H_1$ 과  $H_2$ 는 은닉층을 의미하는 변수
- $g$ 는 은닉층  $H_1$ 과  $H_2$ 를 선형결합하여 조정하는 활성화함수
- $x_i (i = 1, \dots, n)$ 는 입력정보,  $y$ 는 출력정보



## 03 신경망의 구성 및 종류

### 2 신경망의 종류

#### (3) 기타 신경망

- RBF(Radial Basis Function) 신경망  
: 이 신경망은 은닉층이 하나이며  
(은닉층의 노드가 하나라는 말이 아님에 주의)  
입력층과 은닉층의 연결 시에는 가중치 없이 입력값이 그대로 전달.  
활성함수로서는 정규분포 형태의 함수가 사용
- EBF(Elliptical Basis Function) 신경망  
: RBF 신경망을 보완하기 위해서 제안된 것으로 입력정보의 선형결합값  
을 RBF 신경망에 대입하여 작성



## 4. 신경망을 이용한 훈련



## 04 신경망을 이용한 훈련

### 1 신경망의 목적함수

#### ■ 신경망 모델을 구축하는 과정을 두 단계로 요약

- ① 입력변수 또는 노드의 개수, 은닉층의 수와 노드의 개수, 활성화함수의 종류 등을 결정하는 단계
- ② 가중치( $w_{ij}$ )를 추정하는 단계로 구분 - 신경망 학습(learning) 또는 훈련(training)으로 통계학의 관점에서 보면 모수추정(parameter estimation)
  - 신경망의 학습은 지도학습과 자율학습 두 가지 경우 모두에 해당



## 04 신경망을 이용한 훈련

### 1 신경망의 목적함수

#### 〈신경망의 학습과정〉

1. 연결강도에 대한 초기값을 설정  
→ 출력정보를 계산
2. 학습을 통해 구한 출력값의 추정치와 학습자료에서 기준이 되는 출력값과 비교
3. 이 과정을 출력값과 학습값의 차이가 일정수준이 될 때까지 반복



## 04 신경망을 이용한 훈련

### 1 신경망의 목적함수

■ 연결강도인 가중치를 적절히 추정하기 위해서는 목적함수(objective function)을 적절하게 정의해야함

- 일반적으로 많이 쓰이는 목적함수로는 선형모형에서 사용되는 오차제곱합

$$\sum (y_i - p_i)^2$$

$$\sum_{i=1}^n (y_i - p_i)^2 = \sum_{i=1}^n [y_i - g(w_{00} + w_{10}H_{1i} + w_{20}H_{2i})]^2$$

- 분류 문제에 관해서는 통상 로그우도함수(log likelihood function)을 사용

$$-\sum_{i=1}^n (y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i))$$





## 04 신경망을 이용한 훈련

### 2 역전파 알고리즘(Back propagation)

#### (1) 역전파 알고리즘의 개요

- 신경망의 목적함수는 연결강도에 대하여 비선형 함수이기 때문에 이를 최적화하는 가중치를 찾는다는 것은 어려움  
→ 수치해석 방법을 이용하여 가중치를 산출



## 04 신경망을 이용한 훈련

### 2 역전파 알고리즘(Back propagation)

#### (1) 역전파 알고리즘의 개요

〈 홉필드(Hopfield)가 제안한 역전파(back propagation)알고리즘〉

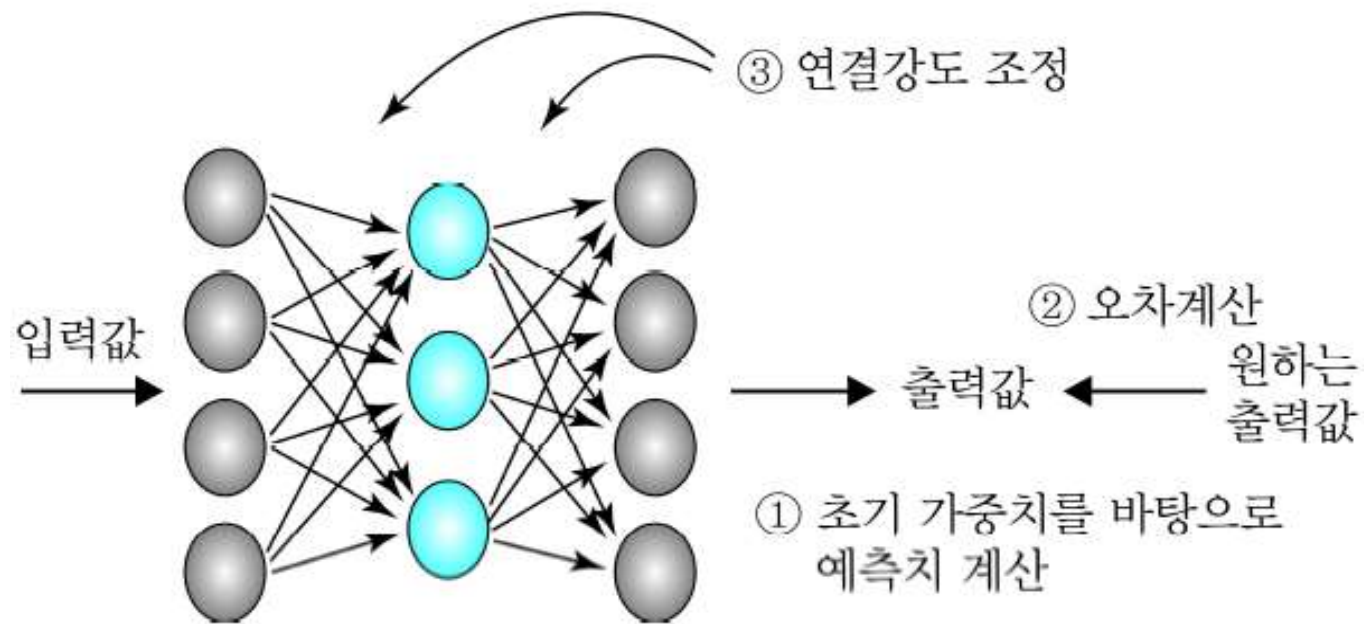
1. 첫 단계에서 초기 가중치와 목표함수를 최적화하는 기준을 정하고 초기 가중치를 바탕으로 예측치를 계산
2. 두 번째 단계에서는 실제 출력값과 예측치 사이의 오차를 계산
3. 마지막 단계에서는 두 번째 단계에서 구한 오차를 은닉층과 입력층으로 역전파 시켜서 가중치(연결강도)를 새로 조절

위의 3단계를 반복 적용하여 가중치 값이 거의 변하지 않거나 일정해지면 반복을 멈추고 그 값으로 가중치의 값을 결정



## 04 신경망을 이용한 훈련

### 2 역전파 알고리즘(Back propagation)



〈그림9〉 역전파 알고리즘 요약

## 04 신경망을 이용한 훈련

### 2 역전파 알고리즘(Back propagation)

#### (2) 초기 가중치 및 학습률

- 가중치의 초기값을 정하는 방법으로는 각 모수에 난수를 이용하여 임의로 값을 지정하고 목적함수 관점에서 성능이 가장 좋은 초기값을 선택(일정 영역의 균일분포(Uniform Distribution)로부터 산출된 난수를 이용)
- 역전파 알고리즘 3단계에서 연결가중치의 조절 정도를 학습률(Learning Rate)이라고 한다. 학습률이 높으면 가중치 값이 빠르게 변화



## 04 신경망을 이용한 훈련

### 2 역전파 알고리즘(Back propagation)

#### (2) 초기 가중치 및 학습률

- 학습률이  $\alpha$ 라고 하고 추정할 가중치의 개수가  $N$ 일 경우 학습패턴의 수  $p$ 는 다음과 같이 계산

$$p = \frac{N}{1 - \alpha}$$

- 처음에는 크게, 그리고 반복수가 증가하면서 점점 작아지도록 설정



## 04 신경망을 이용한 훈련

### 3 신경망 모형의 작성 6단계

■ Berry and Linoff(1997)의 신경망모형의작성 6단계

- 1) 입력 및 출력변수의식별
- 2) 입력, 출력 변수값을적절한 범위의 값으로 변환
- 3) 신경망의구조를 설정
- 4) 자료를 이용하여 신경망 학습
- 5) 학습에 이용된 자료가 아닌 실제 자료에 대하여 4)에서 작성된 모형이 적절한지 검정하여 신경망모형을 완성
- 6) 5)에서 작성된 모형을 이용하여 새로운 현상을 예측 또는 분류



## 04 신경망을 이용한 훈련

### 3 신경망 모형의 작성 6단계

#### I 주요 단계별 특징

##### ① 변수 선택 및 변환 (1,2단계)

- 범주형 입력변수의 경우에는 모든 범주에서 일정 빈도 이상의 값을 가져야 함
- 연속형 입력변수인 경우 입력값들의 범위가 변수 간에 많은 차이가 없어야 함
- 연속형 자료일 경우 일반적으로 변수의 표준화 과정을 거침

$$\text{조정변수값} = \frac{\text{실제값} - \text{최솟값}}{\text{최댓값} - \text{최솟값}}$$

(최대값은 예상보다 조금 크게, 최소값은 예상보다 조금 작게 설정)





## 04 신경망을 이용한 훈련

### 3 신경망 모형의 작성 6단계

#### ① 변수 선택 및 변환 (1,2단계)

- 필요할 경우 연속형 변수를 범주형 자료로 전환  
ex) 소득 변수를 매우 낮음, 낮음, 중간, 높음, 매우 높음 등으로 변환
- 순서가 있는 자료의 경우에는 연속형 자료처럼 표준화하여  
0과 1사이의값으로 표현  
ex) 0,1,2,3,4,5를 0,0.2,0.4,0.6,0.8,1 등으로 표현



## 04 신경망을 이용한 훈련

### 3 신경망 모형의 작성 6단계

#### ① 변수 선택 및 변환 (1, 2단계)

- 온도계식 방법: 예를 들어 기업평가의 값이 A, B, C, D, F라면 다음과 같이 숫자값으로 변환

$F \rightarrow 00000 \rightarrow 0$	$D \rightarrow 10000 \rightarrow 0.5$
$C \rightarrow 11000 \rightarrow 0.5 + 0.25 = 0.75$	$B \rightarrow 11100 \rightarrow 0.5 + 0.25 + 0.125 = 0.875$
$A \rightarrow 11110 \rightarrow 0.5 + 0.25 + 0.125 + 0.0625 = 0.9375$	

- 순서가 의미 없는 성별과 같은 범주형 자료의 경우 범주별로 다른 입력변수로 지정
- 범주형 자료의 경우에는 모든 범주에 일정 빈도이상의 값을 가지도록 해야 하기 때문에, 지나치게 많은 범주로 나누지 말고 범주가 지나치게 많은 경우 범주를 적절하게 병합하여 사용



## 04 신경망을 이용한 훈련

### 3 신경망 모형의 작성 6단계

#### ② 신경망 구조의 설정 (3단계)

- 입력정보와 출력정보가 정해져 있는 상태에서 은닉층의 마디수, 합성함수, 활성화함수 등을 변화시키면서 모형을 탐색
- 통상 신경망이 범용근사자(universal approximator)로서 역할을 하기 위해서는 은닉층을 1개 정도만 설정

\* 1개의 은닉층의 마디수가 지나치게 많아지면 은닉층의 수를 2개로 작성



## 04 신경망을 이용한 훈련

### 3 신경망 모형의 작성 6단계

#### 〈은닉층의 마디수를 정하는 방법〉

- 1) 은닉마디수를 여러가지로 변화시켜서 최적모형을 선택하는 방법 (Trial and Error)
- 2) 단순한 신경망모형에서 복잡한 모형으로 신경망모형을 키워가면서 은닉층의 마디의 수를 결정하는 방법 (Constructive Algorithm)
- 3) 복잡한 신경망모형에서 단순한 신경망모형으로 축소시키면서 모형을 선택하는 방법 (Pruning)

(통상 은닉층의 마디 수는 입력층의 마디의 수의 두 배를 넘지 않도록 권고)



## 04 신경망을 이용한 훈련

### 3 신경망 모형의 작성 6단계

- 은닉 마디의 수를 보다 객관적으로 정하려면 마디 수를 정하는 과정에서 특정한 모형선택기준을 이용

→ 적합도 함수의 최소화  $E + \lambda \sum W_{ij}^2$

( $E$ 는 적합도,  $\lambda$ 는 가중치 제곱합의 벌칙항,  $W_{ij}$ 는 가중치(연결강도))

- 신경망의 경우 목표변수의 출력값은 연속형으로 산출되므로 분류의 경우 출력값을 범주형 출력값으로 환산



## 04 신경망을 이용한 훈련

### 3 신경망 모형의 작성 6단계

#### ③ 신경망의 학습 (4단계)

- 역전파 알고리즘: 비선형적인 추정방법과 반복적인 추정방법을 통해 가중치를 추정
- 일반적으로  $n$ 개 노드의 입력변수와  $h$ 개 노드의 은닉층, 1개 노드의 출력층이 있을 경우 상수항을 포함하여  $(n+1) \times h + h + 1$ 개의 가중치가 필요
- 신경망이 과적합되는 것은 은닉층의 수 및 은닉마디가 지나치게 클 경우 발생
  - 기존 자료는 매우 잘 적합하지만, 새로운 자료에 대해서는 예측력이 저하됨



## 04 신경망을 이용한 훈련

### 4 민감도 분석

- 신경망은 분류 및 예측에 좋은 결과를 제공하지 못하지만 입력정보와 출력정보 간의 관계가 불명확한 블랙박스
  - 신경망의 불투명성을 완화하기 위해 민감도 분석(Sensitivity Analysis)이 필요
  - 민감도 분석이란 입력변수들의 상대적인 중요도를 간접적으로 파악하는 과정
    - 입력변수 각각의 평균을 찾고 입력변수들의 평균값에서의 목표변수값을 구한 뒤 마지막으로 각 입력변수값이 변할 때 마다 출력변수의 변화를 측정
      - 입력변수의 상대적인 중요도 측정이 가능





## 5. 신경망의 장단점



## 05 신경망 모형의 장단점

### ■ 신경망의 장점

- 예외적인 경우도 있으나 대체로 의사결정나무나 회귀분석보다 분류 및 예측력 측면에서 우수
- 입력변수들과 출력변수들 간의 관계가 복잡한 비선형 형태일 때 더 유용
- 견고성에 의해 자료의 잡음에 크게 영향을 받지 않으며 계량적인 변수뿐만 아니라 정성적인 변수도 한꺼번에 효과적으로 신속하게 처리할 수 있음
- 기존의 통계적 방법과는 달리 여러 가지 통계적 가정을 필요로 하지 않기 때문에 자료를 마음대로 활용할 수 있음



## 05 신경망 모형의 장단점

### ■ 신경망의 단점

- 학습은 했지만 그 과정이 투명하지 않고 복잡하다는 것
  - 신경망이 분류의 문제에 관해 출력값을 산출했다 하더라도 왜 그렇게 분류하였는지를 설명하지 못한다는 것
- 실무적인 측면에 있어서 신경망 가중치 설정 등에는 전문성이 필요하므로 비전문가가 쉽게 이용하기 어렵다는 점도 제약으로 작용
- 신경망은 잘못된 입력정보에 둔감하기 때문에 입력정보의 오류가 오랜 기간이 지난 후에야 출력결과로 나타난다는 점



## 6.딥러닝



## 06 딥러닝

- 딥러닝은 기계학습 기법 중 하나로 신경망모형으로부터 비롯
  - 신경망모형의 학습 과정의 개념과 동일하다고 할 수 있으며 은닉층이 많이 쌓여가면서 복잡하고 깊은 구조로 발전
- 딥러닝의 출현은 SNS 관계망이나 인터넷 등을 통해 생성되는 빅데이터와 클라우드 컴퓨팅 환경 등의 하드웨어 발전에 기반
- 더불어 이러한 방대한 데이터에 필요한 연산량을 CPU 대신 보다 저렴한 GPU 병렬 프로그래밍으로 해결할 수 있는 방법이 개발되면서 가격 제약도 약화



다음시간 안내

# 9강. 신경망 모형 II