

다층퍼셉트론 신경회로망

Multilayer Perceptron Neural Network

모델 Power

# 선형 회귀분석 vs 비선형 회귀분석

- 선형 회귀분석

$$\hat{y} = \Theta + \sum_{i=1}^p w_i x_i$$

- “선형”: 모델 파라미터인 beta 또는 w 들간의 관계
- x 대신  $x^2$ ,  $x^3$ ,  $e^x$ ,  $\log x$ ,  $\sin x$  가 대체 되어도 모두 선형 회귀분석

# 선형 회귀분석 vs 비선형 회귀분석

- 비선형 회귀분석
  - 모델 파라미터 간의 관계가 비 선형
  - 다양한 모델 존재
  - 그 가운데 가장 인기 좋은 신경회로망
  - 신경 회로망 가운데 가장 많이 사용되는 multi-layer perceptron 다층 퍼셉트론

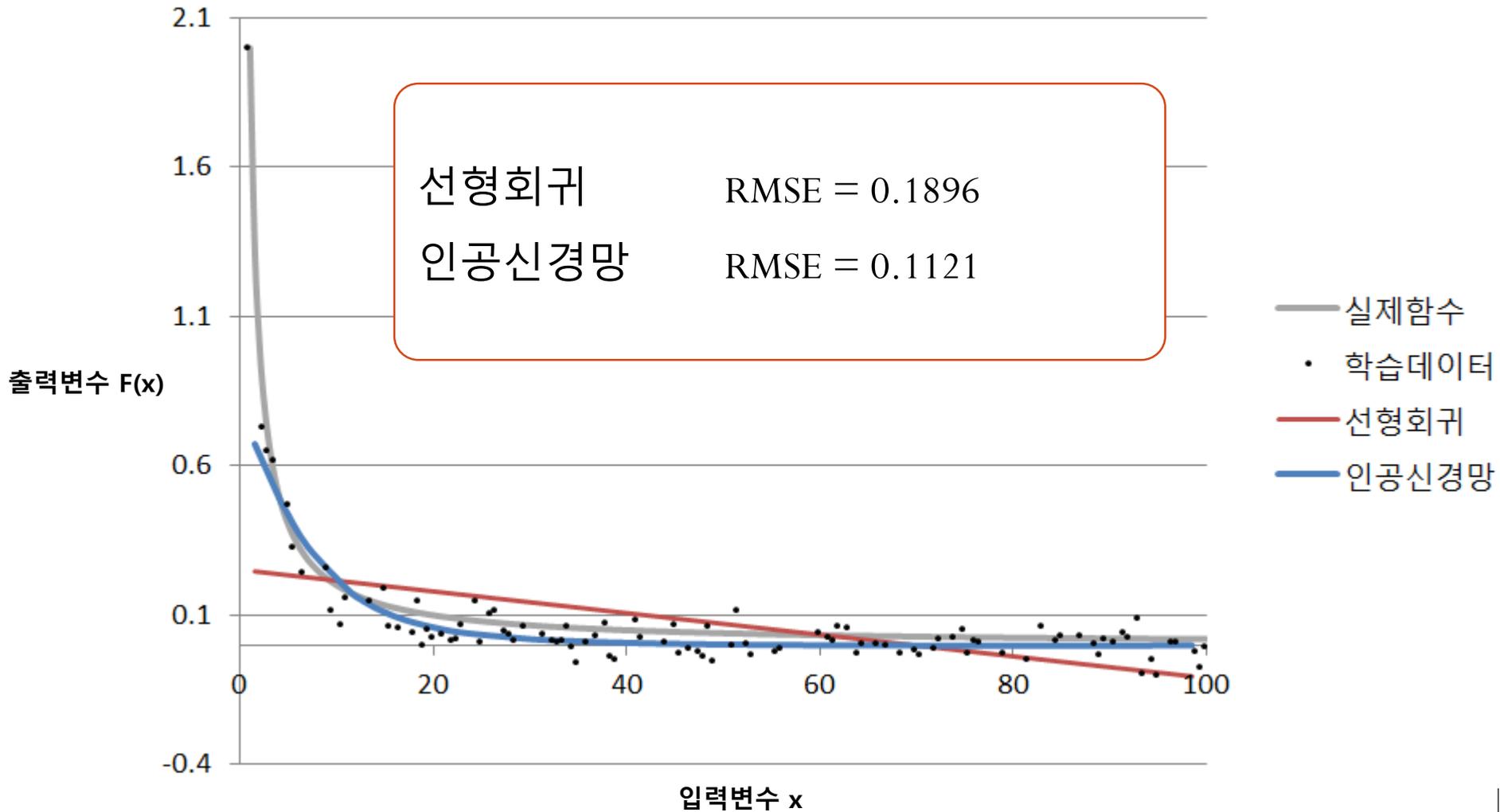
# 선형 회귀분석 vs 비선형 회귀분석

- 비선형 회귀분석
  - 다양한 모델 존재
  - 그 가운데 가장 인기 좋은 신경회로망
  - 신경 회로망 가운데 가장 많이 사용되는 multi-layer perceptron 다층 퍼셉트론
- 선형과 비선형의 비교
  - 선형은 직선 fit ( $x^2$  이나  $e^x$  없는 경우)
  - 비선형은 곡선 fit

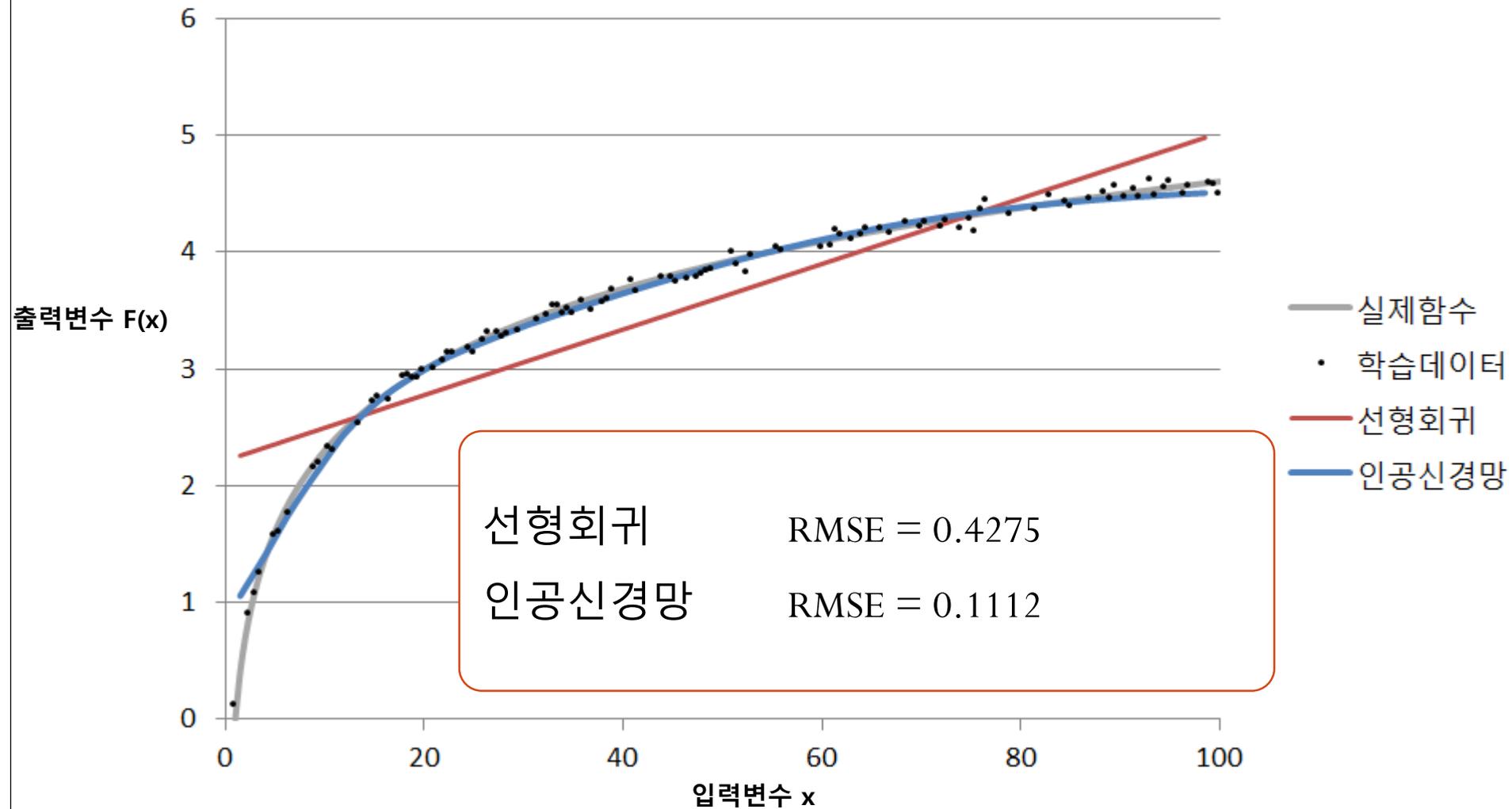
# 함수 fit

- 주어진 함수  $f$  로부터 데이터  $D$  생성
- 생성된  $D$  를 바탕으로  $y = f'(X)$  구축
- 모델  $f'$  와  $f$  비교 ( $f'$  이  $f$  와 비슷한가?)
  
- 주어진 함수  $f$ 
  - $y = 2/x$
  - $y = \log_2 x$
  - $y = \exp(-0.2 * x)$
  - $y = \sin(x)$

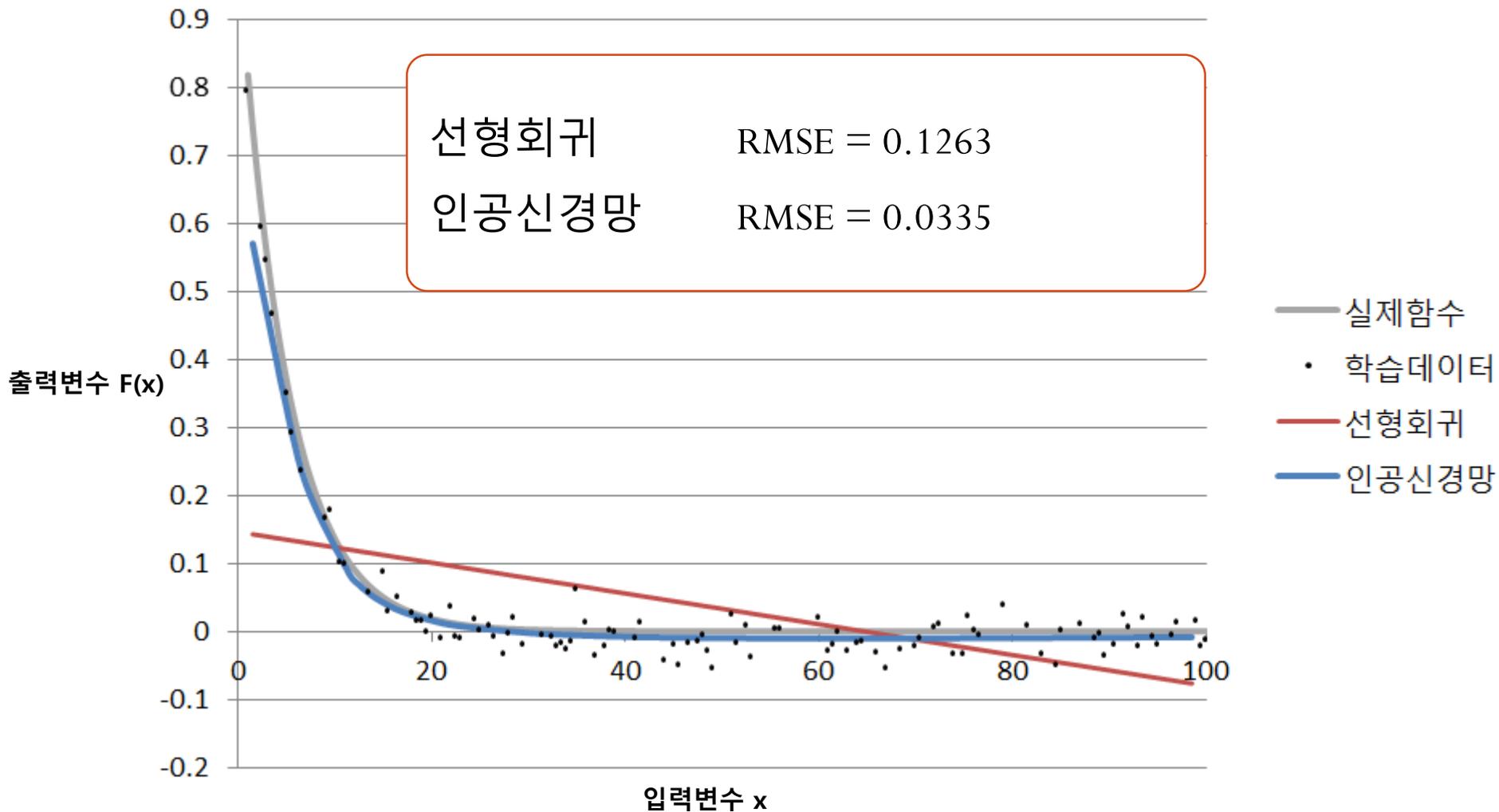
$$F(x) = 2/x \quad (1 \leq x \leq 100)$$



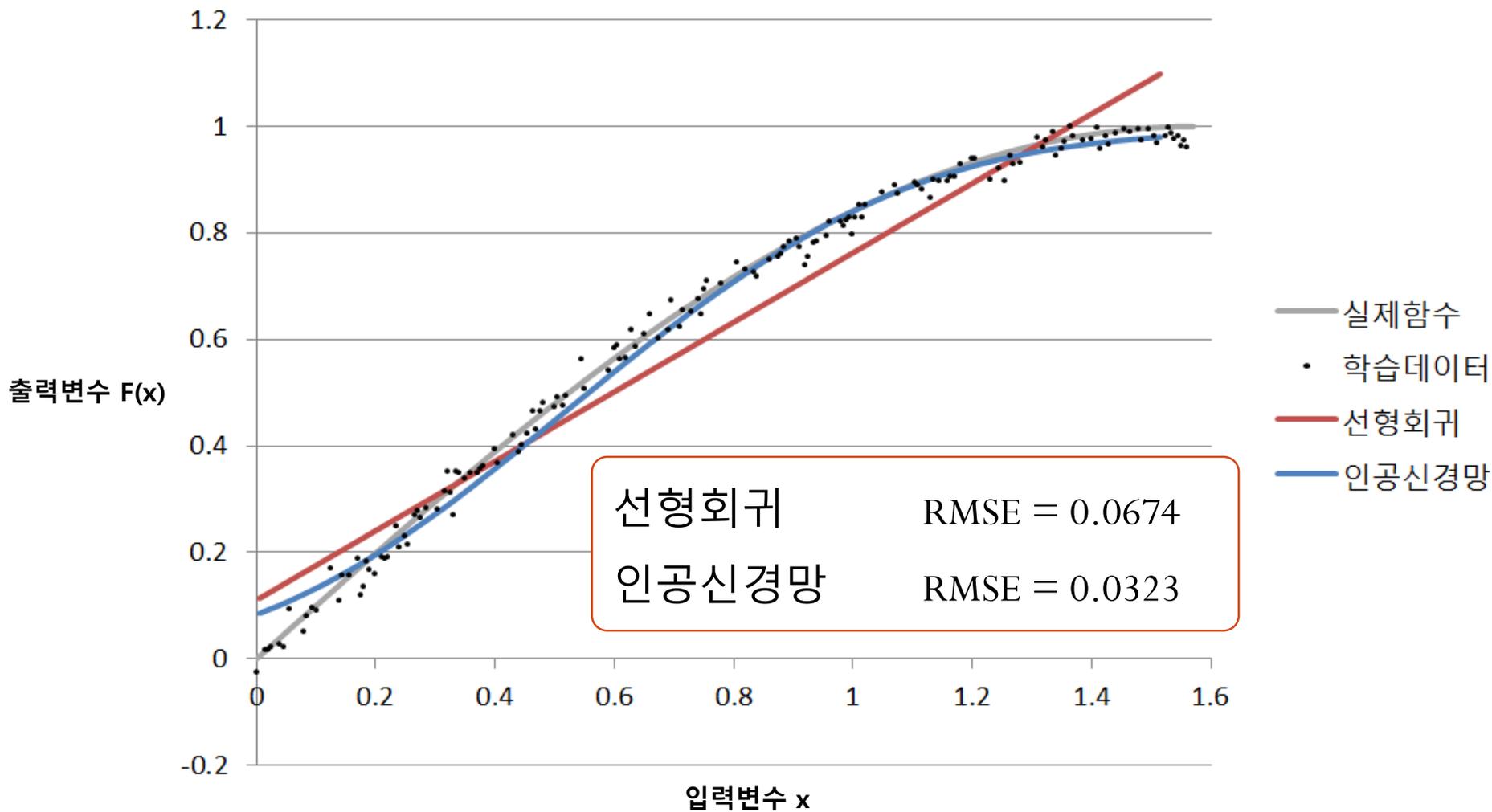
$$F(x) = \log(x) \quad (1 \leq x \leq 100)$$



$$F(x) = \exp(-0.2 \cdot x) \quad (1 \leq x \leq 100)$$



$$F(x) = \sin(x) \quad (0 \leq x \leq \pi/2)$$



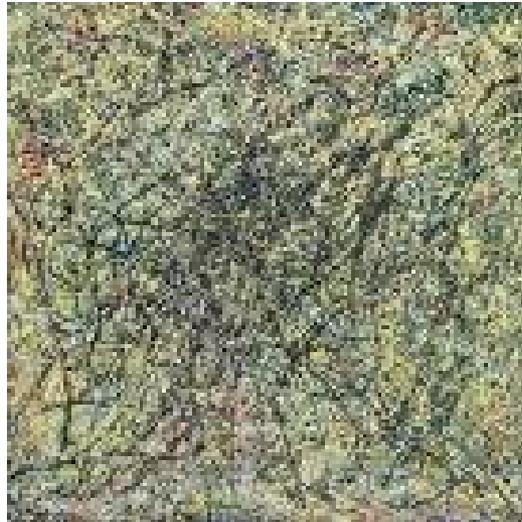
# 선형으로도 fit 가능하지만...

- 선형 회귀분석으로도  $2/x$ ,  $\log x$ ,  $e(-0.2x)$ ,  $\sin x$  항을 넣으면 위 함수들을 정확히 fit 할 수 있음
- 그러나 현실적으로 데이터 세트  $D$  만 주어졌을 때에, 어떤 “비선형 항”을 넣어야 하는지 판단 불가
- 따라서 신경망과 같은 general nonlinear model 이 사용성 측면에서 뛰어남

# 모델 구조

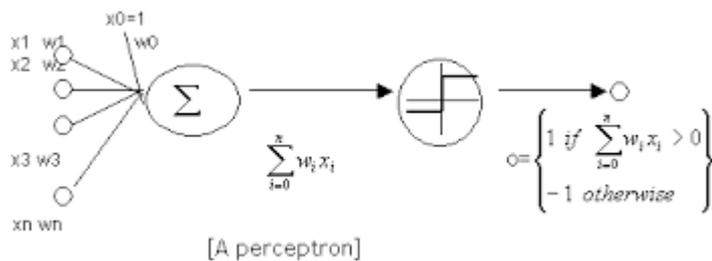
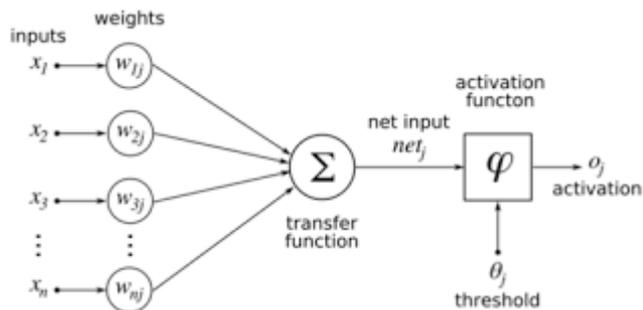
# Neural networks

- 신경회로망

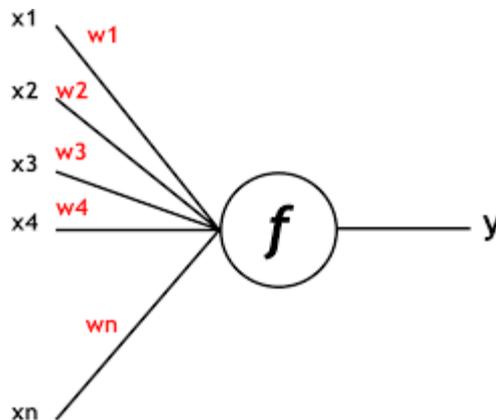
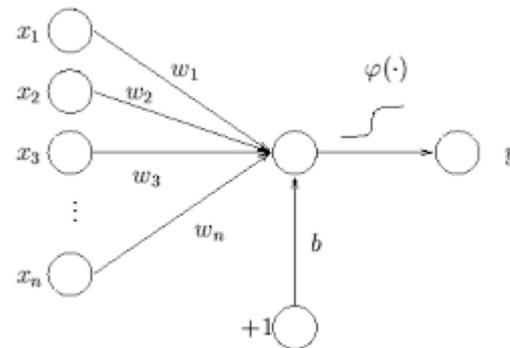


- 인간: ~1천억 개 뉴론 들이 10조 개의 시냅스를 통해 연결됨

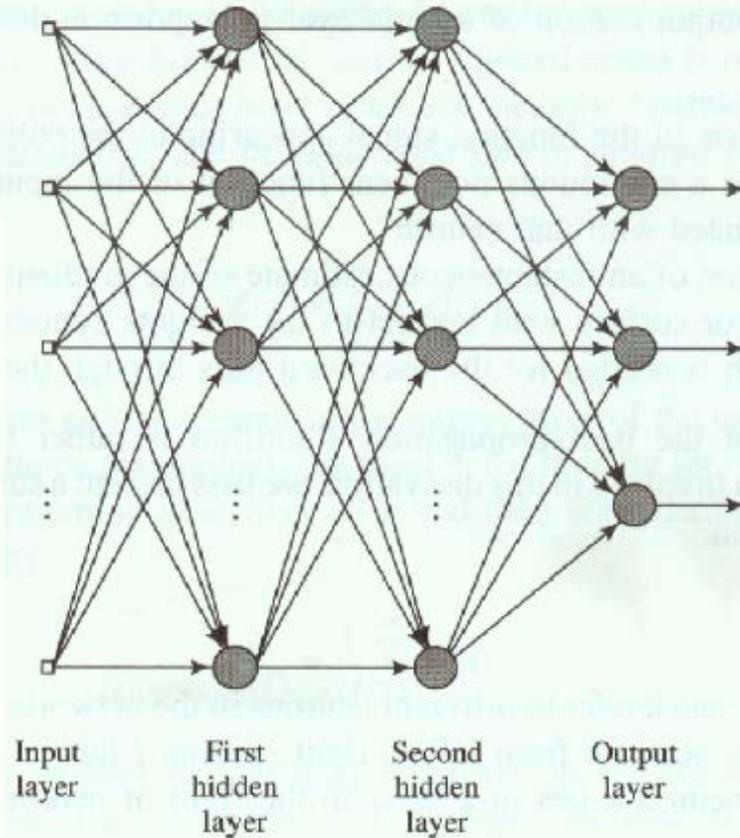
# 단층 퍼셉트론 single-layer perceptron



[A perceptron]



# 다층 퍼셉트론 multi-layer perceptron

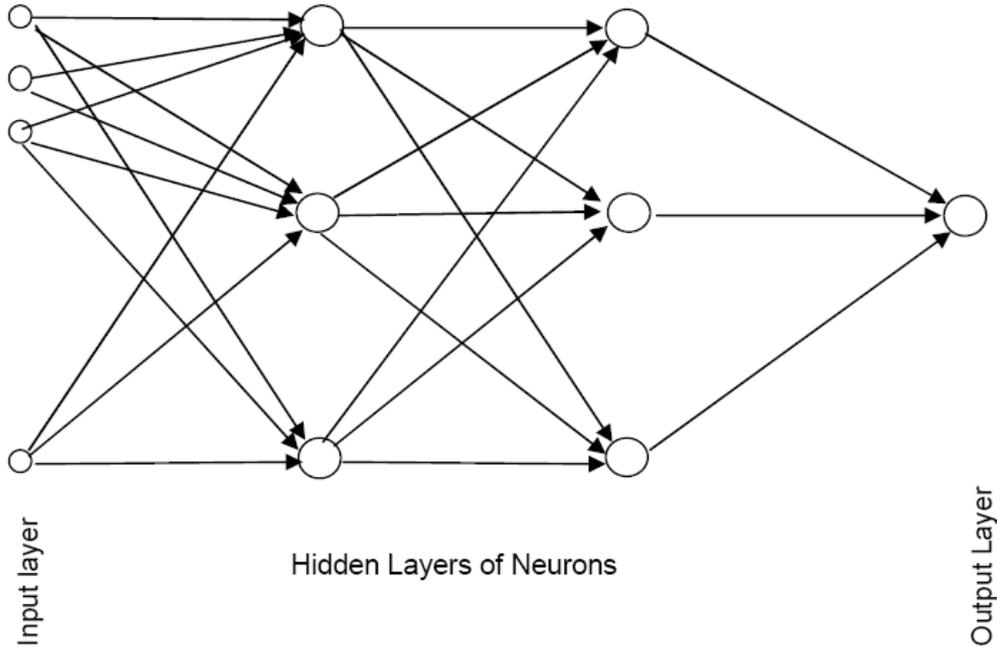


**FIGURE 6.1** Architectural graph of a multilayer perceptron with two hidden layers.

# 망 구조

- 노드, 뉴런 (회귀식 변수)
- 노드 층
  - 입력층 input layer
  - 은닉층 hidden layer
  - 출력층 output layer
- 에지, 시냅스 (회귀식 계수)

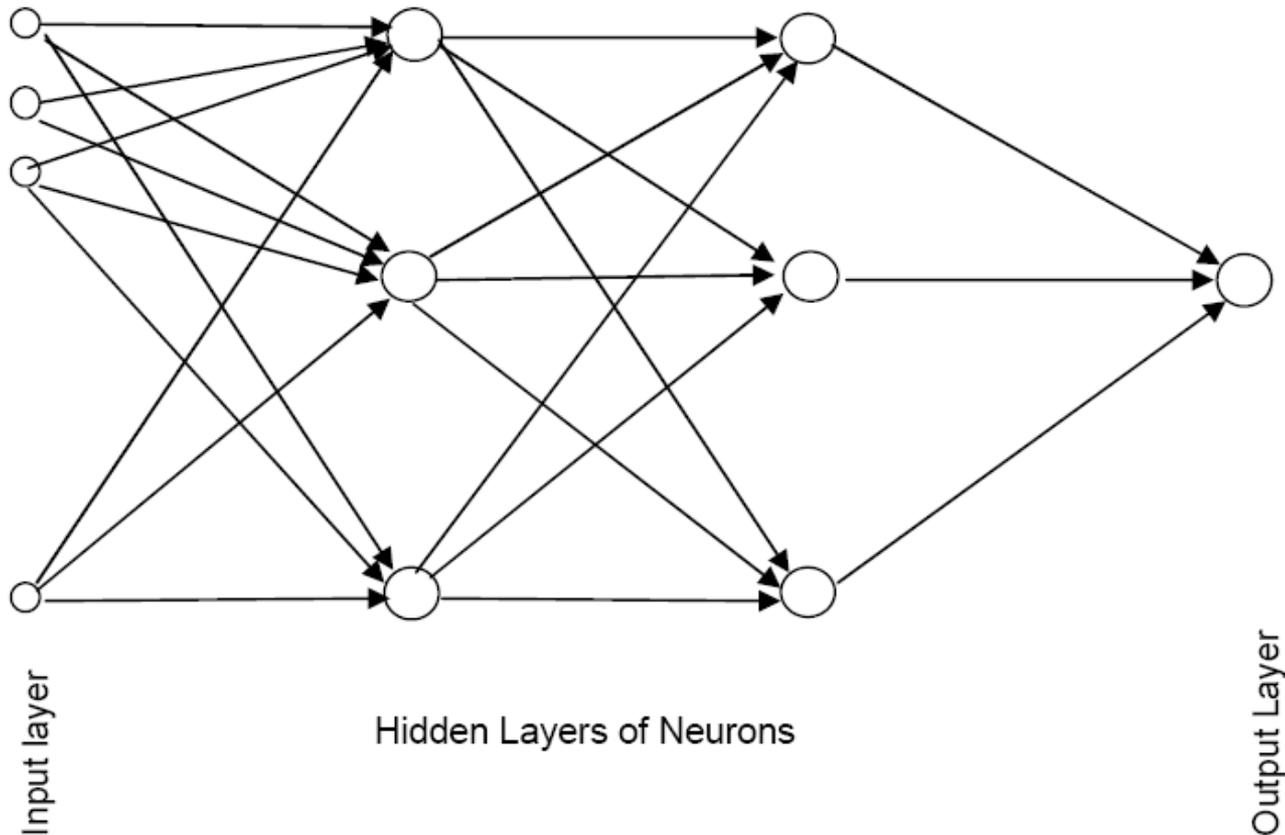
# 3층 퍼셉트론 구조



\* 영국식 층 계산 법

# 3층 퍼셉트론 구조

- input layer: input nodes = input or **independent variables  $x$**
- output layer: output node = output or **dependent variable  $y$**
- hidden layer: hidden nodes = ?  $h$



# 각 노드에서 하는 계산

- (bio) Action potential, nonlinearity, threshold, synapse, other neuron's

$$output_j = g(\theta_j + \sum_{i=1}^p w_{ij} x_i)$$

# 1층 퍼셉트론 구조는?

- P 개의 입력 노드와 1개의 출력노드를 가진...
- $P=3$
- 시냅스 수는?

# 선형 회귀 모델 Linear Regression!!

출력 노드가 하나이고 중간층이 없는 망은, 여기서  $g$ 는 항등함수, 선형 회귀분석 모델과 같은 형태를 취한다.

$$\hat{y} = \Theta + \sum_{i=1}^p w_i x_i$$

## 2층 퍼셉트론 구조는?

- P 개의 입력 노드, H 개의 은닉 노드, 1 개의 출력 노드
- $P=3, H=4$
- 시냅스 수는?

# 비선형 2층 퍼셉트론 모델

- Nonlinear regression 인 경우,
  - 히든 노드의  $g$  함수는 sigmoid 이고,
  - 출력 노드의  $g$  함수는 identity (or linear) 를 사용

- 수식으로 표현하면

$$y = \theta_0 + \sum_{j=1}^H w_j \left\{ g\left(\theta_j + \sum_{i=1}^p w_{ij} x_i\right) \right\}$$

- Logistic Regression 몇 개?

# Example – Using fat & salt content to predict consumer acceptance of cheese

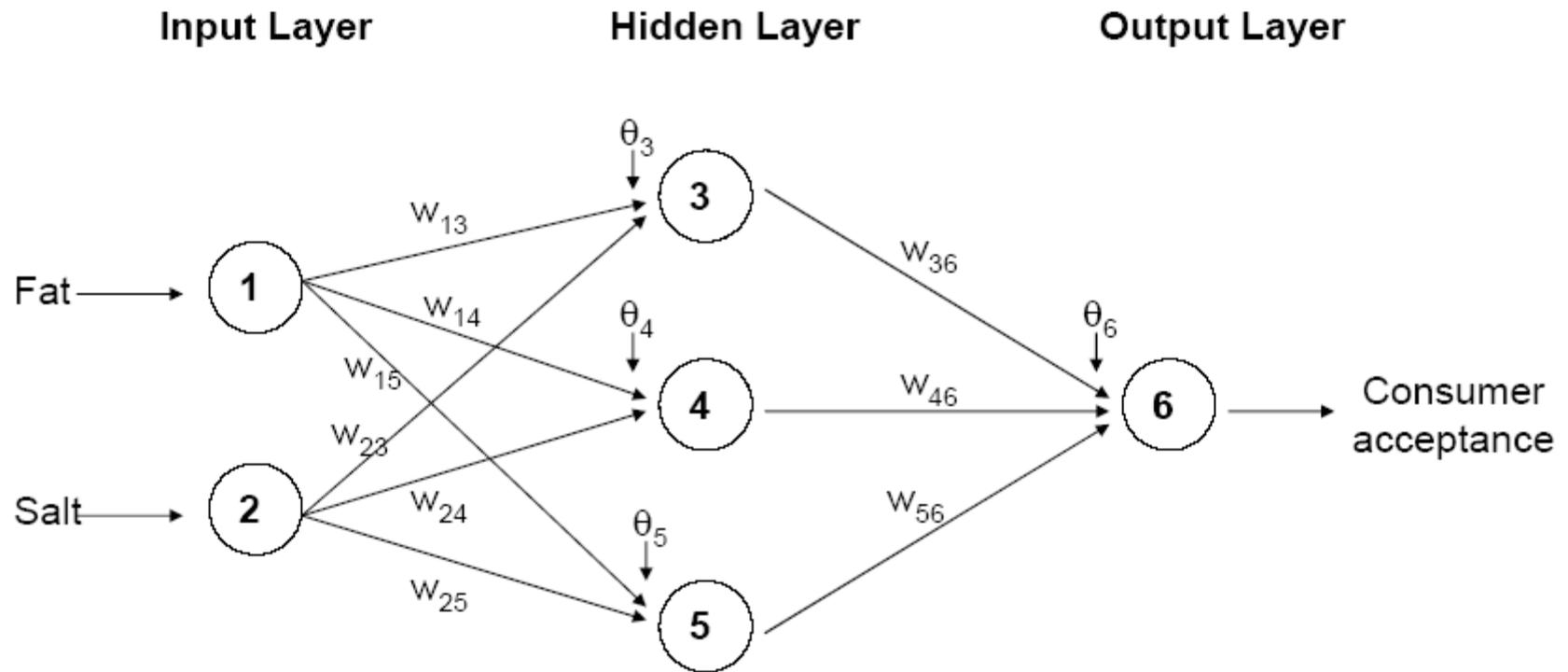


Figure 11.2: Neural network for the tiny example. Circles represent nodes,  $w_{i,j}$  on arrows are weights, and  $\theta_j$  are node bias values.

# Example - Data

<i>Obs.</i>	<i>Fat Score</i>	<i>Salt Score</i>	<i>Acceptance</i>
1	0.2	0.9	1
2	0.1	0.1	0
3	0.2	0.4	0
4	0.2	0.5	0
5	0.4	0.5	1
6	0.3	0.8	1

# 모델 작동

# 입력층

입력층에서, 입력 = 출력

- E.g., record #1에서:

지방 입력 = 출력 = 0.2

염분 입력 = 출력 = 0.9

입력층의 출력 = 은닉층으로 입력

# 은닉층

이 예에서, 은닉층은 3개의 노드를 가짐

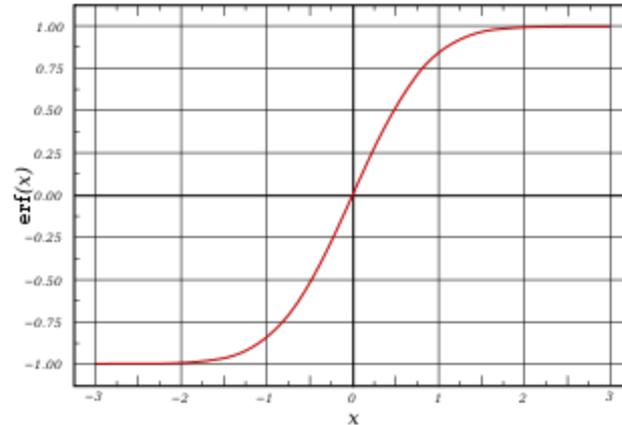
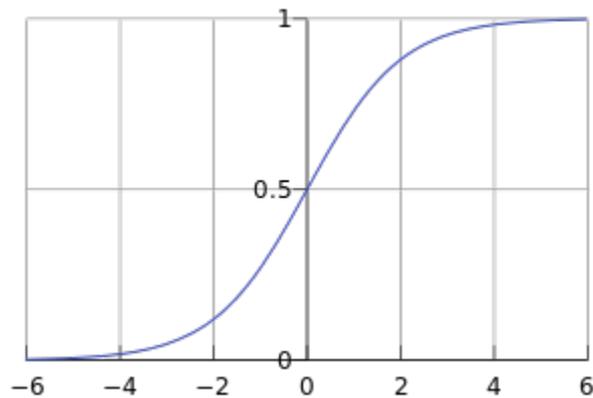
각 노드는 전체 입력 노드의 출력을 입력함

각 은닉층의 출력은 입력 가중치 합에 대한 함수

$$output_j = g(\theta_j + \sum_{i=1}^p w_{ij} x_i)$$

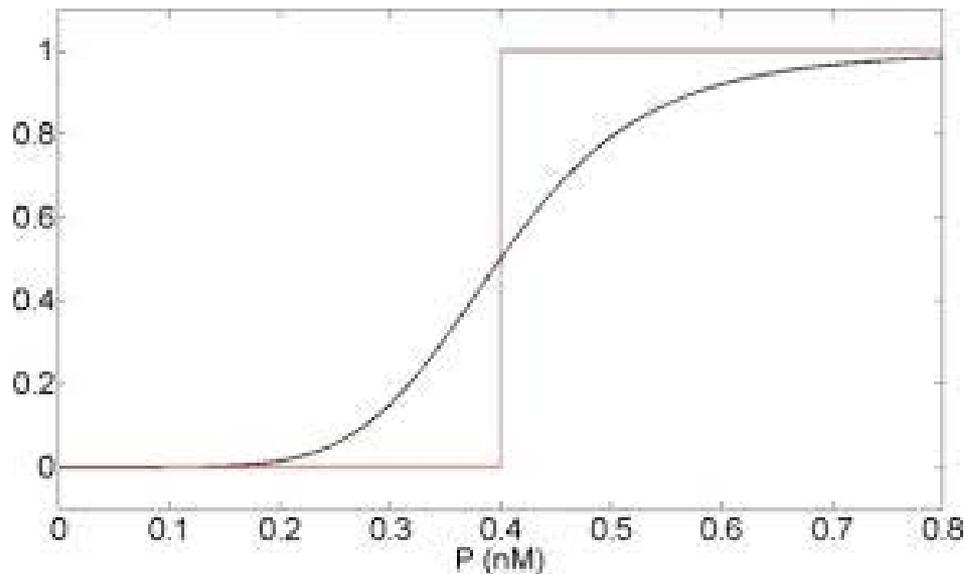
# Function g?

- $g(x) = 1/(1+\exp(-x))$
- 시그모이드, 로지스틱
- 뉴론의 활성화 함수 또는 학습 함수



# Function g?

- $g(x) = 1/(1+\exp(-k*x))$
- $k$  값이 아주 크면, 시그모이드, 로지스틱 함수는 어떤 모양이 되는가?



노드 3의 출력: 문제가 예측이면  $g$ 가 identity 함수이고, 분류이면  $g$ 가 로지스틱

$$output_j = g(\theta_j + \sum_{i=1}^p w_{ij} x_i)$$

$$output_3 = \frac{1}{1 + e^{-[-0.3 + (0.05)(0.2) + (0.01)(0.9)]}} = 0.43$$

# 신경망의 초기 통과

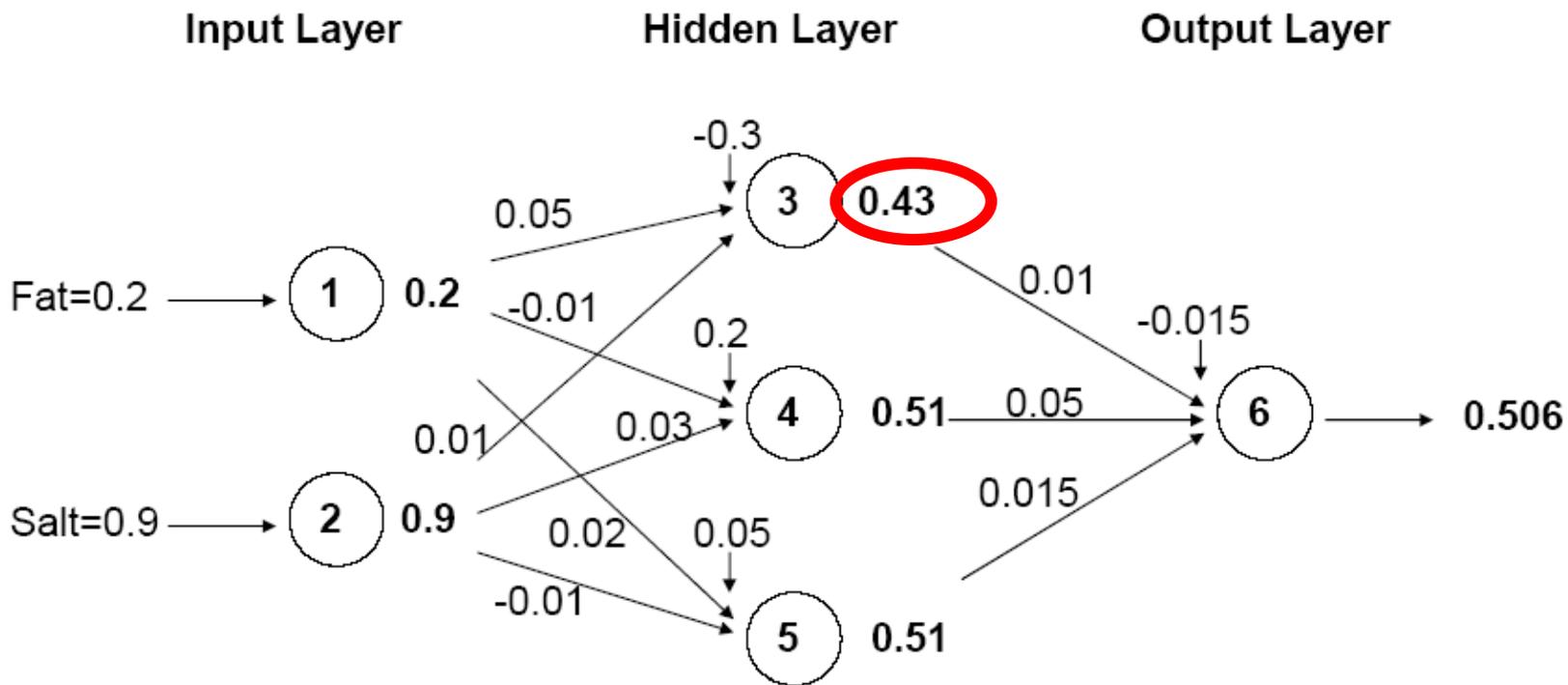


Figure 11.3: Computing node outputs (in boldface type) using the first observation in the tiny example and a logistic function.

# 출력층

마지막 중간층의 출력이 출력층의 입력이 됨

위와 같은 함수 사용, i.e. 가중평균의  $g$  함수

$$output_6 = \frac{1}{1 + e^{-[-0.015 + (0.01)(0.43) + (0.05)(0.507) + (0.015)(0.511)]}} = 0.506$$

# 출력 노드

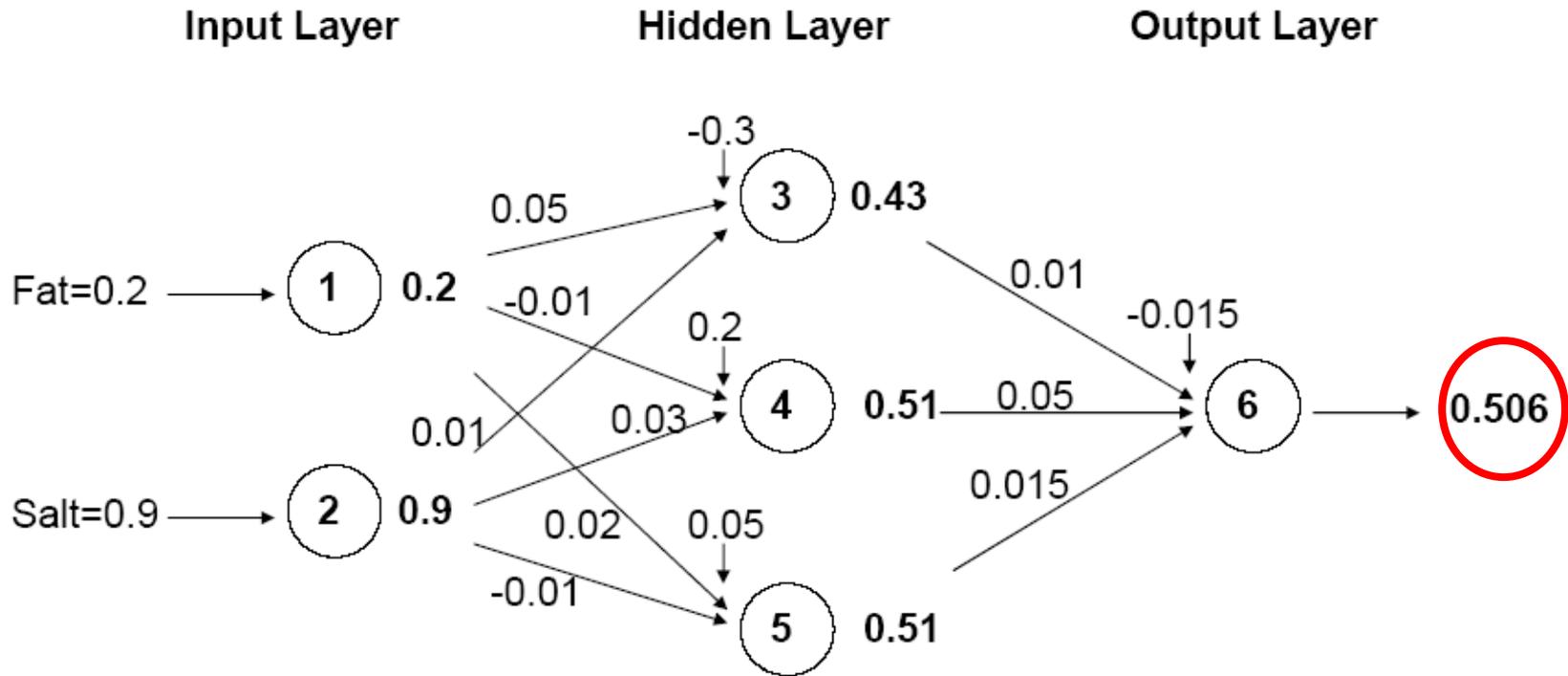
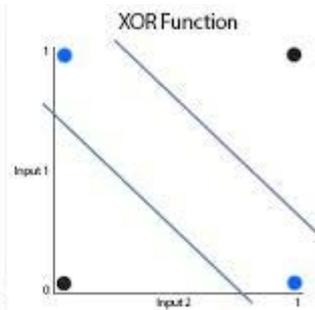
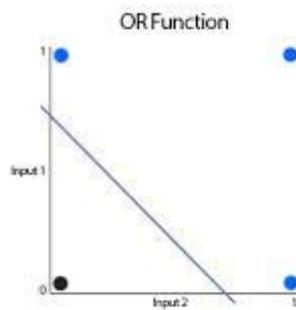
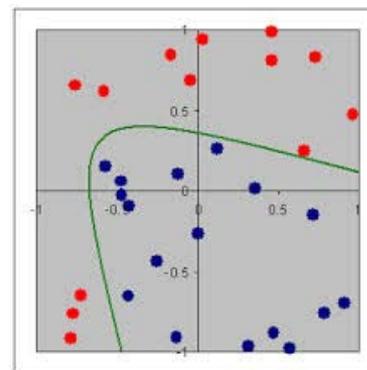
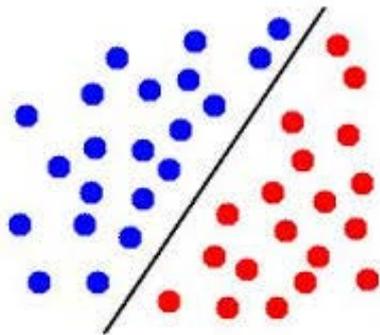


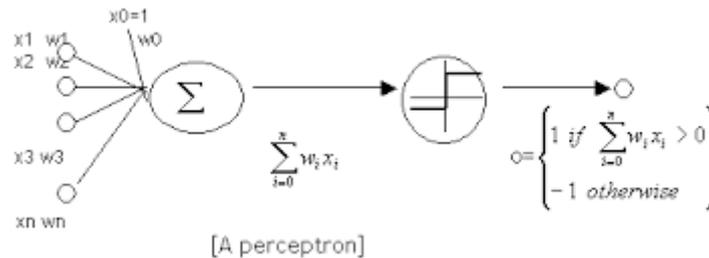
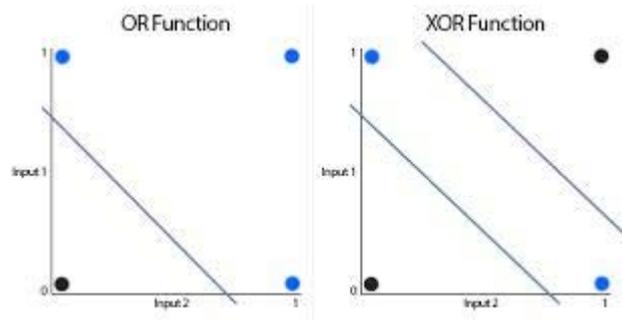
Figure 11.3: Computing node outputs (in boldface type) using the first observation in the tiny example and a logistic function.

# 비선형 분리가능성



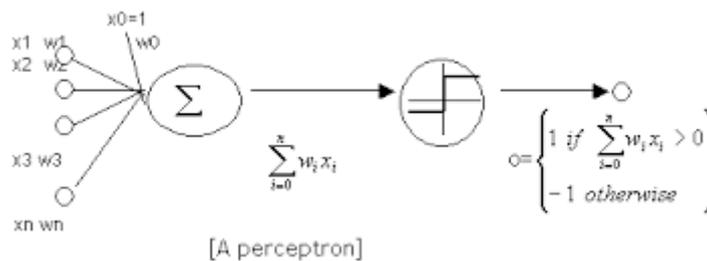
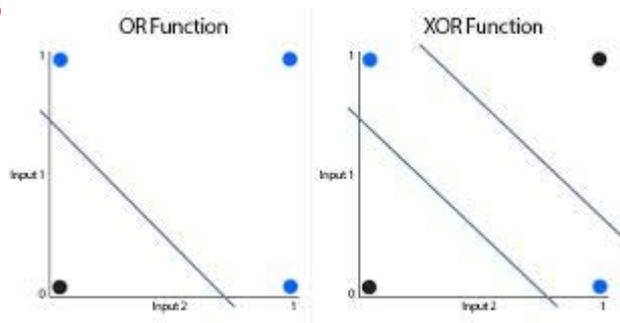
# 비선형 분리가능성 nonlinear separability

- OR 문제
  - “선형 Decision Boundary”
  - 1층 perceptron 으로 분리 가능



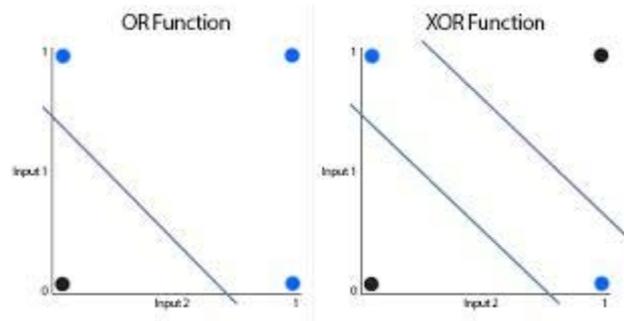
# 비선형 분리가능성 nonlinear separability

- OR 문제
  - “선형 Decision Boundary”
  - 1층 perceptron 으로 분리 가능
  - **How? Give w's**



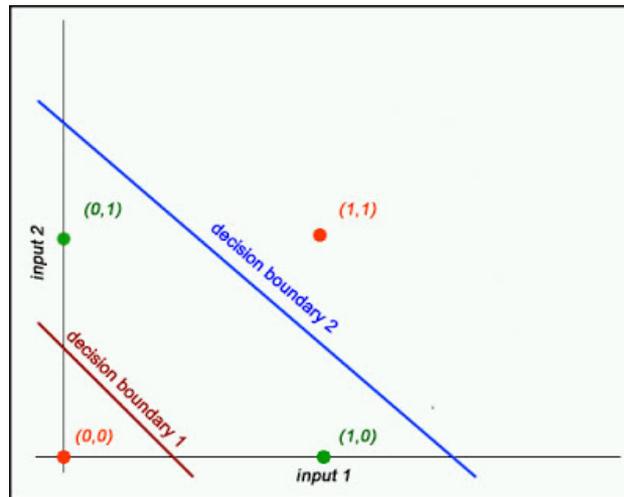
# 비선형 분리가능성 nonlinear separability

- XOR 문제
  - 1층 perceptron 으로 분리 불가능
  - 1969 “Perceptron” by Minsky
  - 여러 개의 1층 perceptron 으로는 분리 가능!



# 비선형 분리가능성 nonlinear separability

- XOR 문제
  - How? Stacking!



# Stacked “2-layer perceptron”

