

1. 데이터 학습 개요



가위바위보 게임을 통한 학습 이해하기



- 우리의 뇌는 가위바위보에서 누가 이겼는지 어떻게 판단할까?
- 판단을 어떻게 학습했고, 결정을 내렸는지 설명하기 어려움

가위바위보 게임을 통한 학습 이해하기

- 0:가위, 1:바위, 2:보
- $My=1$, $computer = 2$



- If $my == 0$ and $computer == 0$ then 비김
- If $my == 0$ and $computer == 1$ then 짐
- If $my == 0$ and $computer == 2$ then 이김
- If $my == 1$ and $computer == 0$ then 이김
- If $my == 1$ and $computer == 1$ then 비김
- If $my == 1$ and $computer == 2$ then 짐
- If $my == 2$ and $computer == 0$ then 짐
- If $my == 2$ and $computer == 1$ then 이김
- If $my == 2$ and $computer == 2$ then 비김

- 프로그래밍 언어가 지원하는 if 문장을 사용해 구현 가능

가위바위보 게임을 통한 학습 이해하기



- 모든 경우의 수를 기억하고 있으면 상황에 맞는 **지식을 쉽게 추론할 수 있음**
- **학습은 모든 경우의 수를 머릿속에 저장하는 작업**

가위바위보 게임을 통한 학습 이해하기

● 0: 가위, 1:바위, 2:보

- my : 0 computer : 0 $\rightarrow 0 - 0 = 0$ \rightarrow 비김
- my : 0 computer : 1 $\rightarrow 0 - 1 = -1$ \rightarrow 짐
- my : 0 computer : 2 $\rightarrow 0 - 2 = -2$ \rightarrow 이김

- my : 1 computer : 0 $\rightarrow 1 - 0 = 1$ \rightarrow 이김
- my : 1 computer : 1 $\rightarrow 1 - 1 = 0$ \rightarrow 비김
- my : 1 computer : 2 $\rightarrow 1 - 2 = -1$ \rightarrow 짐

- my : 2 computer : 0 $\rightarrow 2 - 0 = 2$ \rightarrow 짐
- my : 2 computer : 1 $\rightarrow 2 - 1 = 1$ \rightarrow 이김
- my : 2 computer : 2 $\rightarrow 2 - 2 = 0$ \rightarrow 비김

비긴 경우 : 0

진 경우 : -1, 2

이긴 경우 : 2, 1

가위바위보 게임을 통한 학습 이해하기

- 0: 가위, 1:바위, 2:보
- 판단 공식 = $(my - computer + 3) \% 3$
- 0 : 비김, 1 : 이김, 2 : 짐

		computer		
		0:가위	1:바위	2:보
my	0:가위	비김 $(0-0+3)\%3=0$	짐 $(0-1+3)\%3=2$	이김 $(0-2+3)\%3=1$
	1:바위	이김 $(1-0+3)\%3=1$	비김 $(1-1+3)\%3=0$	짐 $(1-2+3)\%3=2$
	2:보	짐 $(2-0+3)\%3=2$	이김 $(2-1+3)\%3=1$	비김 $(2-2+3)\%3=0$

- 학습은 계산 가능한 수학 공식을 만드는 것
- 공식을 만들 수 있다고 해도 변수들의 범위 등이 조금만 변해도 다른 결과가 나옴

데이터 기반 학습

- (my, computer) → label
- (0,0) → 0(비김)
- (0,1) → 1(짐)
- (0,2) → 2(이김)

학습 데이터

0(가위)

1(바위)

1(짐)



예측값

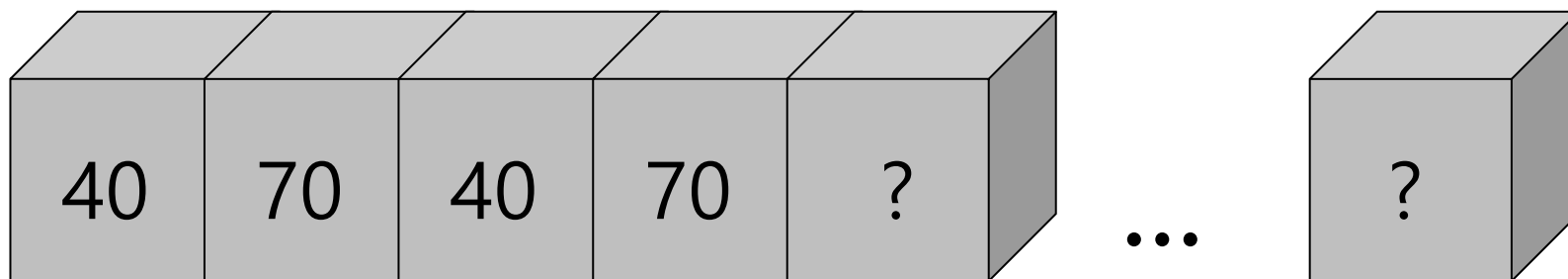
2(이김)



판단할 데이터 (0, 2)

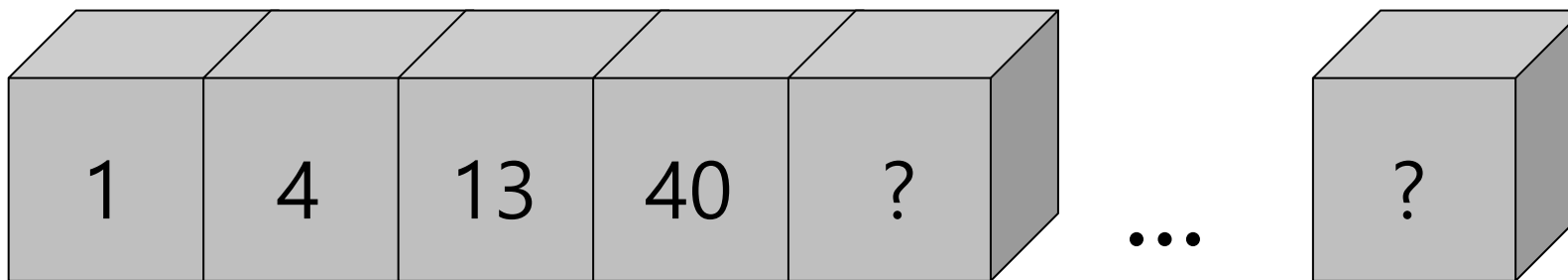
- 기계학습 및 딥러닝에서 학습은 학습 데이터를 가장 잘 설명할 수 있는 모델을 만들어가는 과정

Data 학습 #1



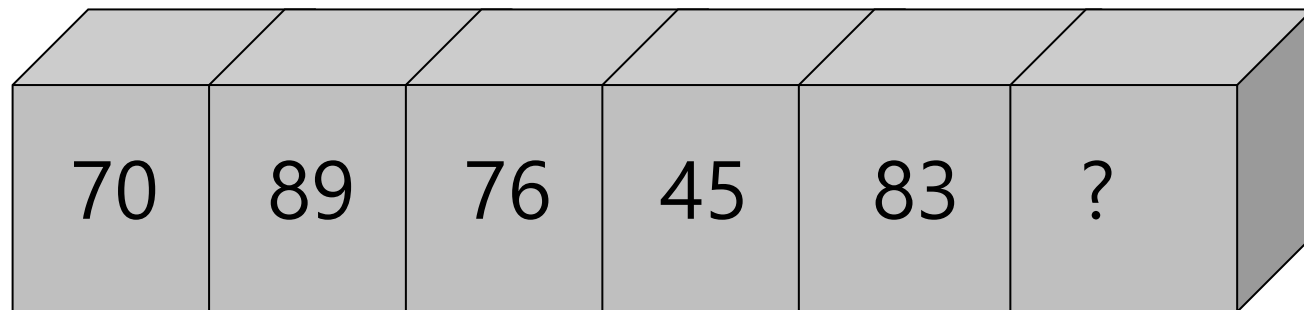
- 다음 숫자가 90일까요?
- 그럼 다음 숫자는 무엇일까요?
- 20번째 숫자는?
- 수학적으로 표현하면?

Data 학습 #2



- 다음 숫자는 무엇일까요?
- 8번째 숫자는?
- 수학적으로 표현하면?

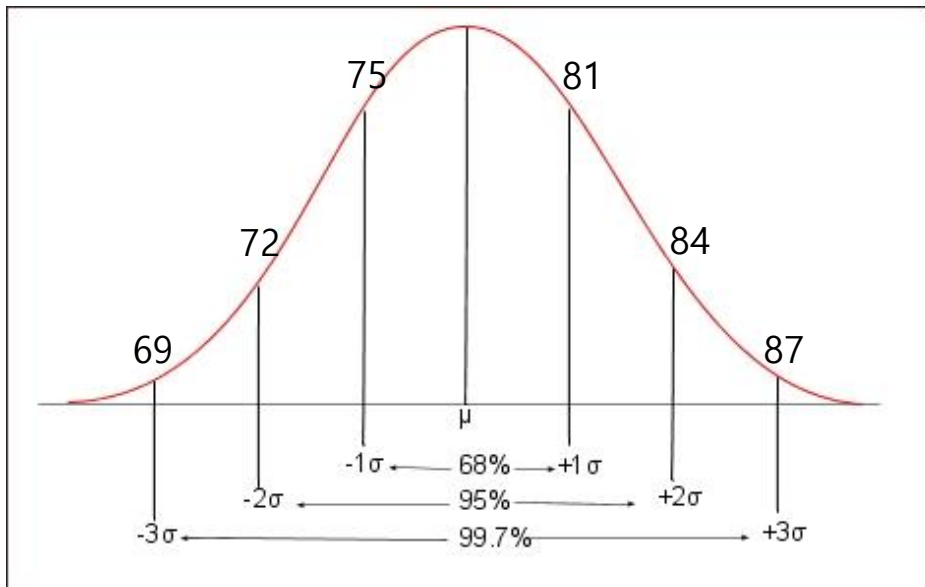
Data 학습 #3



- 70 89 76 45 83
- 다음 숫자는?
- 다음 숫자는 90일까요?
- 수학적으로 표현하면?

평균, 표준편차를 이용한 데이터 분석

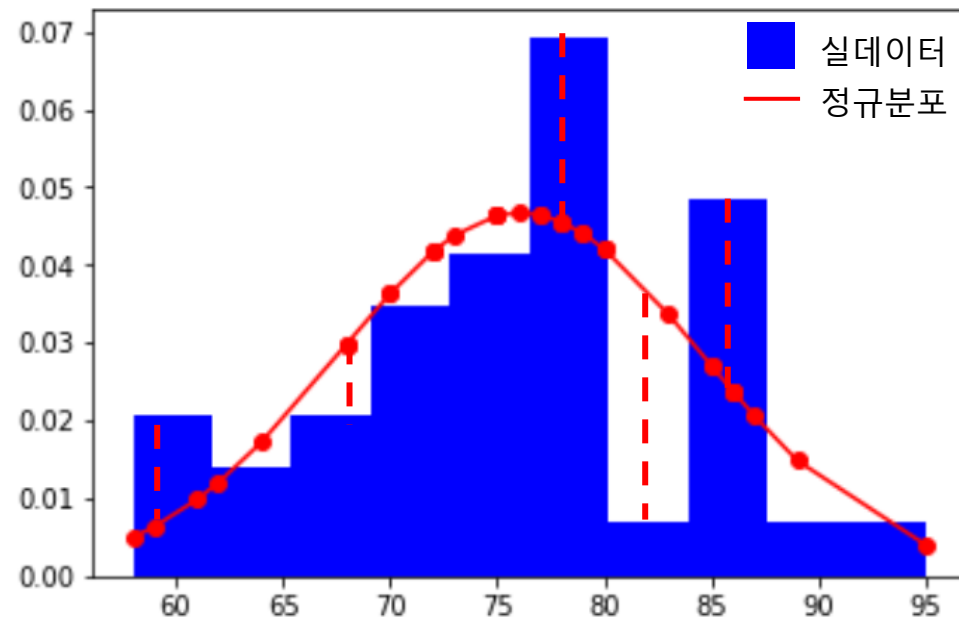
평균(μ)=78, 표준편차(σ)=3



- 다음 시험에서 87점을 맞을 확률은?
- 시험 구간 95%에서 점수 범위는 어떻게 되는가?
- 확률적으로 0.3%이하이면 비정상 점수로 판단한다면 몇 점이 비정상 점수인가?

학습모델 검증

- 학습모델은 Data를 얼마나 잘 설명하고 있나?
- 학습모델에 대한 검증이 필요함
- 그래프를 통한 학습 모델 검증
 - 시각적으로 가설의 적합성을 확인할 수 있음
 - 그러나 고차원 데이터는 시각화하기 어려움



통계 데이터 기반 학습

- 학습 데이터

- $X_1, X_2, X_3, X_4, X_5, \dots, X_n$

- 학습

- 평균과 표준편차 계산

- 학습 후 할 수 있는 것

- 평균을 중심으로 값 예측

- 분포 간 비교

- 표준편차를 가지고 신뢰성 판단

- 정상/비정상 데이터 판단

데이터 기반 지도학습

x	y
1	3
2	5
3	7
4	9
5	11
10	21
20	41
30	61

● x는 40일 때 y는?

데이터 기반 지도학습

x	y
1	3
2	5
3	7
4	9
5	11
10	21
20	41
30	61

데이터

가설

$$Y=2x+1$$



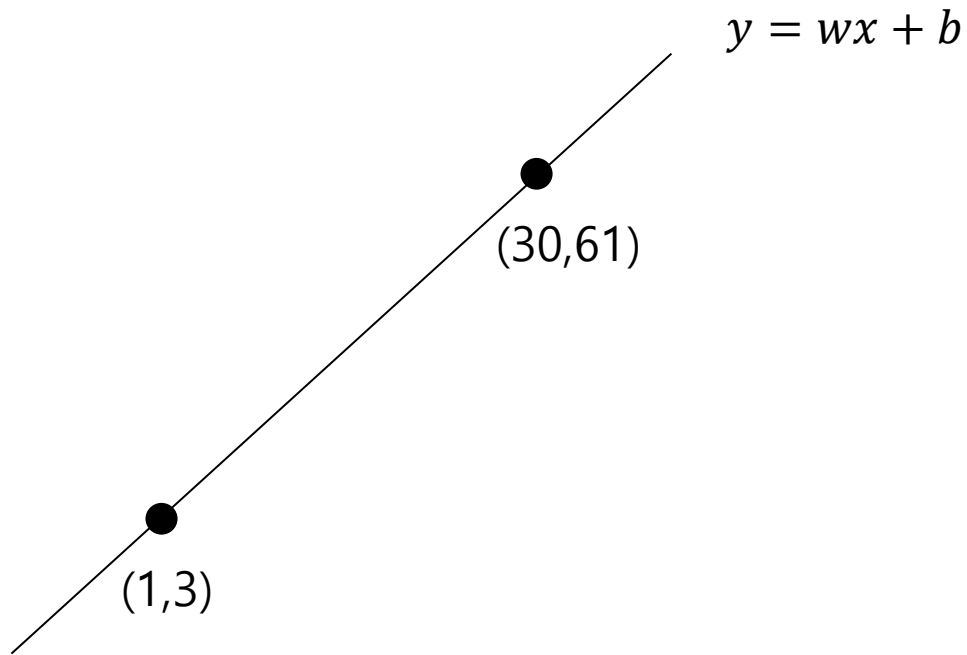
가설 검증

h	y - h 오차
3	0
5	0
7	0
9	0
11	0
21	0
41	0
61	0



x = 40 일 때 y = 81로 예측
x = 50 일 때 y = 101로 예측

직선의 방정식을 이용한 데이터 학습



x	y
1	3
2	5
3	7
4	9
5	11
10	21
20	41
30	61

- 기울기 $w = \frac{3-61}{1-30} = \frac{-58}{-29} = 2$

- y절편 $b = y - wx = 3 - w = 3 - 2 = 1$



$$y = 2x + 1$$

연립방정식을 이용한 데이터 학습

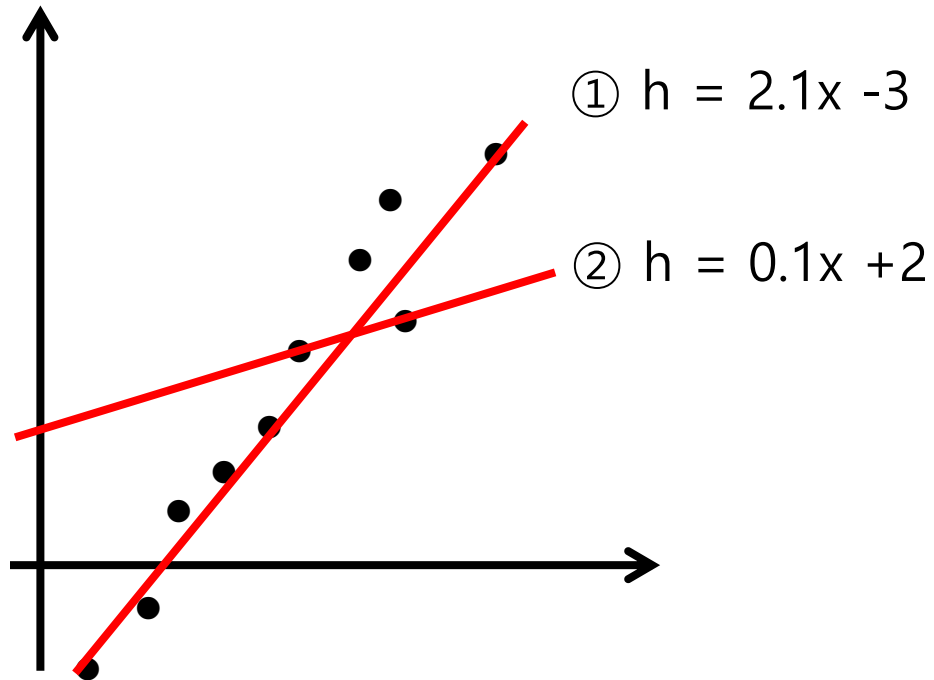
x	y
1	3
2	5
3	7
4	9
5	11
10	21
20	41
30	61



$$y = 2x + 1$$

- $y = 2x + 1$
- $3 = w \times 1 + b \quad \dots \textcircled{1}$
- $5 = w \times 2 + b \quad \dots \textcircled{2}$
- $\textcircled{1} - \textcircled{2} = -2 = -w$
 $w = 2$
- $(1, 3)$ 대입
 $3 = 2 + b$
 $b = 1$

데이터 기반 학습 최적화

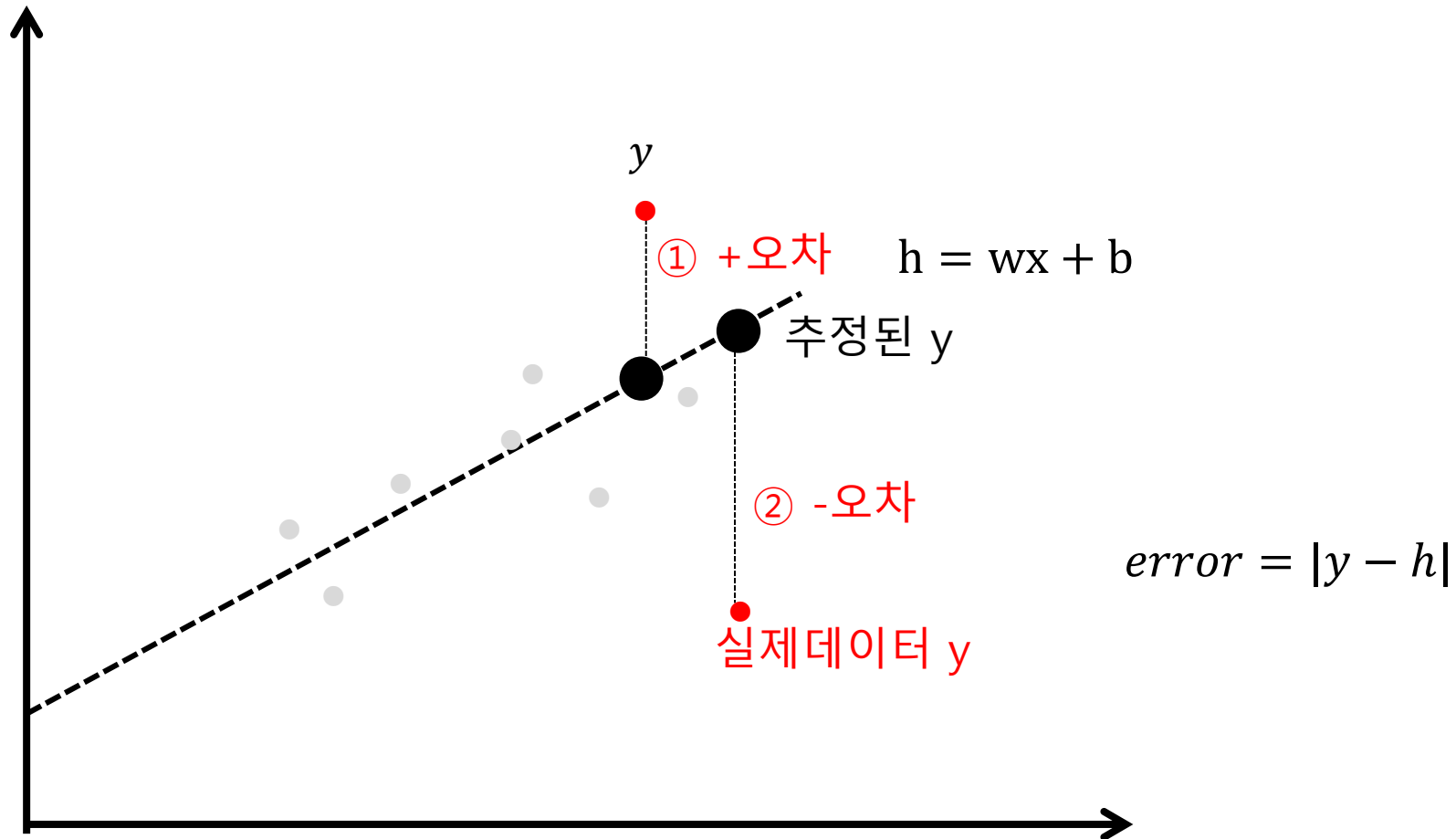


오차가 적은 직선의 방정식 추정

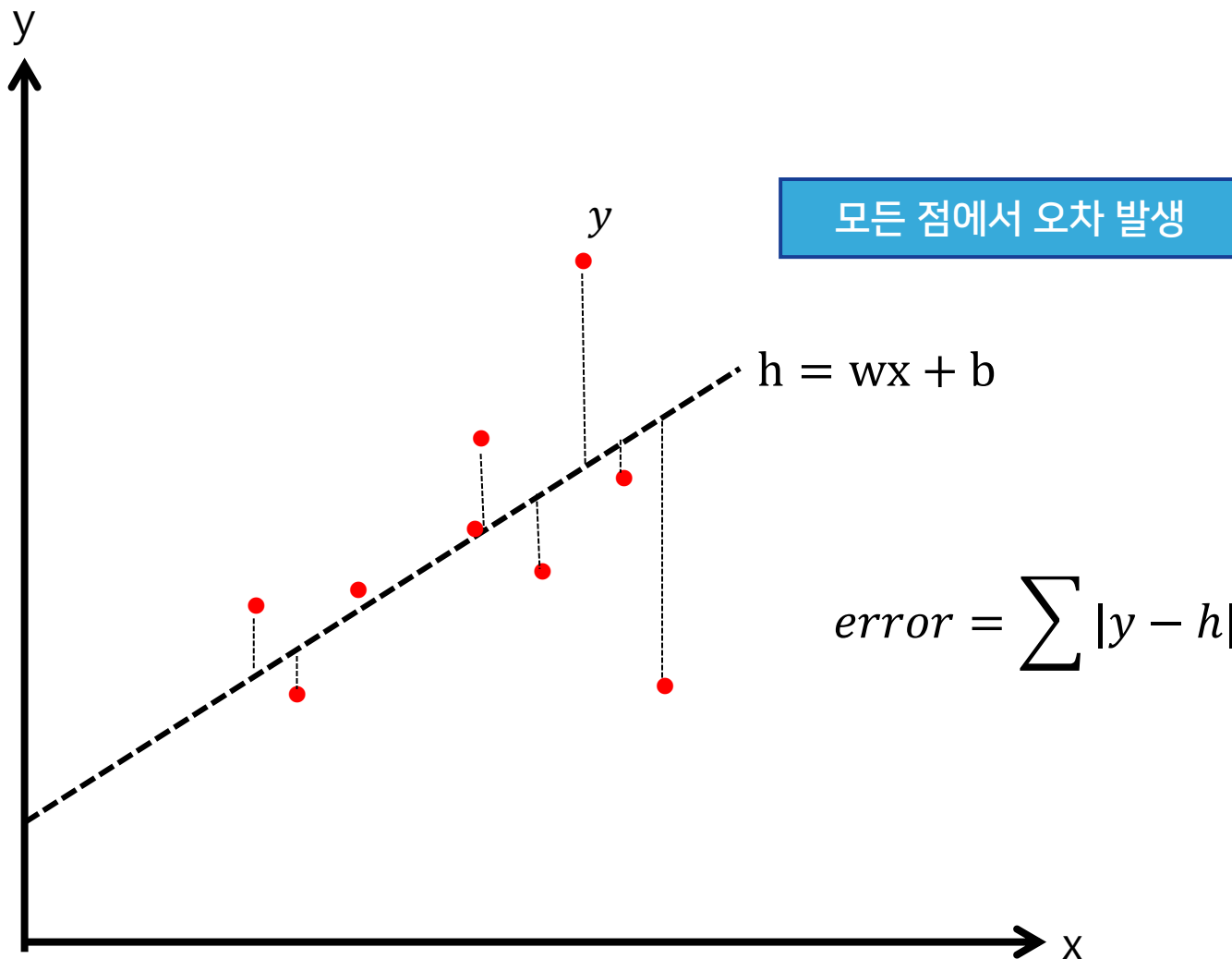
$$y = 2.1x - 3$$

- 임의로 두 개의 데이터 사용
- 가장 작은 데이터와 가장 큰 데이터 사용

오차의 발생



오차의 정의



오차에 대한 수학적 정의

- 평균 절대값 오차 MAE(Mean Absolute Error)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - h_i|$$

- 평균 제곱 오차 MSE(Mean Square Error)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - h_i)^2$$

- 평균 제곱근 오차 RMSE(Root Mean Square Error)

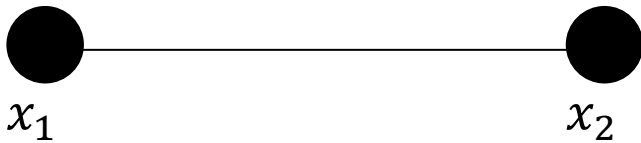
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - h_i)^2}$$

가장 보편적으로 사용

절대값 오차와 제곱근 오차의 기하학적 해석

- 1차원 공간에서는 동일하나, 다차원 공간에서 두 방법은 의미가 달라짐

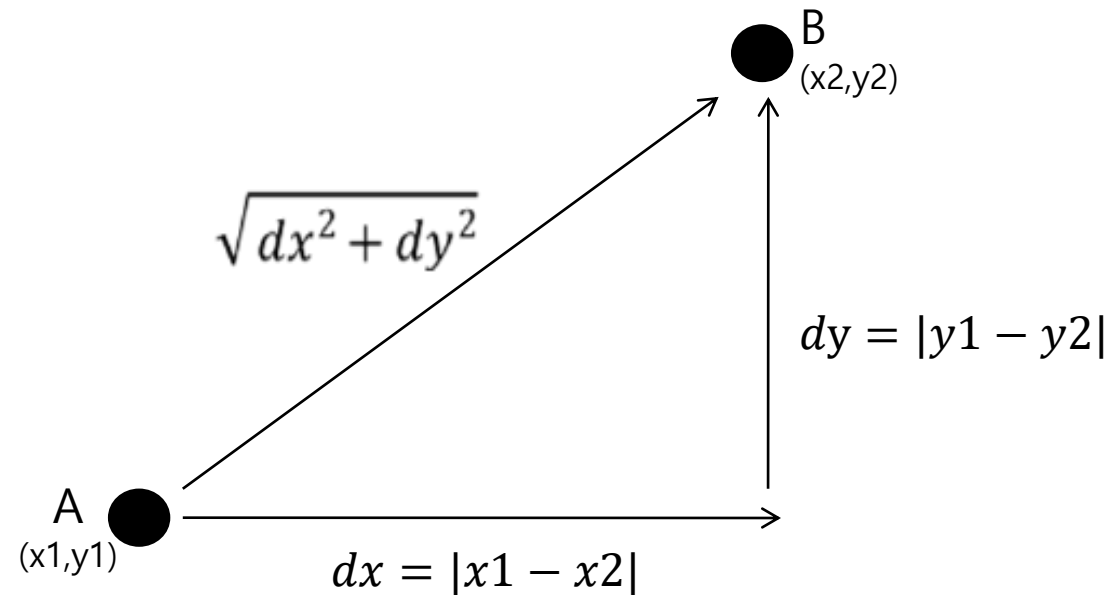
거리 : 1차원 공간의 직선의 길이



$$|x_1 - x_2| = \sqrt{\|x_1 - x_2\|^2}$$

절대값 오차와 제곱근 오차의 기하학적 해석

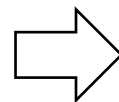
- 평균 제곱 오차가 더 정확하여 일반적으로 사용



1차 연립방정식을 이용한 데이터 학습

$$h = wx + b$$

x	y
1	2
2	4.4
3	6.4



$$w \times 1 + b = 2$$

$$w \times 2 + b = 4.4$$

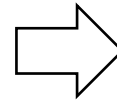
$$w \times 3 + b = 6.4$$

- 평균 제곱근 오차 RMSE(Root Mean Square Error)
 - 정확한 해가 존재하지 않으므로 가장 근사한 해를 구해야 함
 - 최소자승법, 고유값 분해, 역행렬

행렬을 이용한 학습해 구하기

$$h = wx + b$$

x	y
1	2
2	4.4



$$w \times 1 + b = 2$$

$$w \times 2 + b = 4.4$$

$$w \times 1 + b \times 1 = 2$$

$$w \times 2 + b \times 1 = 4.4$$

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 4.4 \end{bmatrix}$$

역행렬을 이용한 최적해 구하기

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 4.4 \end{bmatrix}$$

$$\begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 4.4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}^{-1} = \frac{1}{1-2} \begin{bmatrix} 1 & -1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$$

$$\begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 4.4 \end{bmatrix} = \begin{bmatrix} 2.4 \\ -0.4 \end{bmatrix}$$

역행렬 공식

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$h = wx + b$$

x	y
1	2
2	4.4



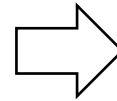
$$h = 2.4x - 0.4$$

역행렬을 이용한 최적해 구하기

- 정방행렬이 아닐 경우 의사역행렬 공식 적용

$$h = wx + b$$

x	y
1	2
2	4.4
3	6.4



$$w \times 1 + b = 2$$

$$w \times 2 + b = 4.4$$

$$w \times 3 + b = 6.4$$

$$w \times 1 + b \times 1 = 2$$

$$w \times 2 + b \times 1 = 4.4$$

$$w \times 3 + b \times 1 = 6.4$$

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 4.4 \\ 6.4 \end{bmatrix}$$

역행렬을 이용한 최적해 구하기

의사 역행렬

$$A^+ = (A^T A)^{-1} A^T$$

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 4.4 \\ 6.4 \end{bmatrix}$$

$$\begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}^+ \begin{bmatrix} 2 \\ 4.4 \\ 6.4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}^+ = \left(\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5 & -1 \\ -1 & 2.33 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -0.5 & 0 & 0.5 \\ 1.33 & 0.33 & -0.66 \end{bmatrix}$$

역행렬을 이용한 최적해 구하기

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 4.4 \\ 6.4 \end{bmatrix}$$

$$\begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}^+ \begin{bmatrix} 2 \\ 4.4 \\ 6.4 \end{bmatrix}$$

$$\begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} -0.5 & 0 & 0.5 \\ 1.33 & 0.33 & -0.86 \end{bmatrix} \begin{bmatrix} 2 \\ 4.4 \\ 6.4 \end{bmatrix} = \begin{bmatrix} 2.2 \\ -0.13 \end{bmatrix}$$

$$h = wx + b$$

x	y	h
1	2	2.07
2	4	4.27
3	6.4	6.47

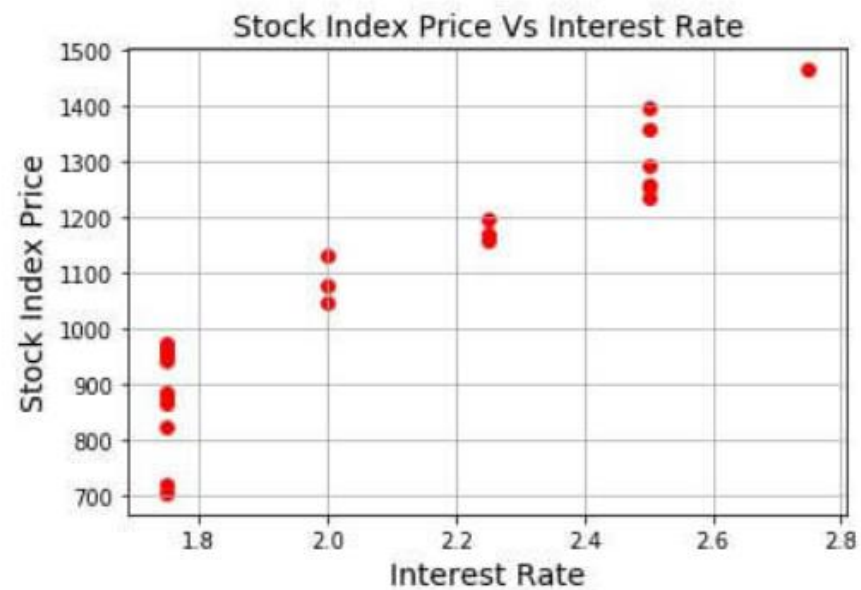


$$h = 2.2x - 0.1$$

데이터 기반 학습

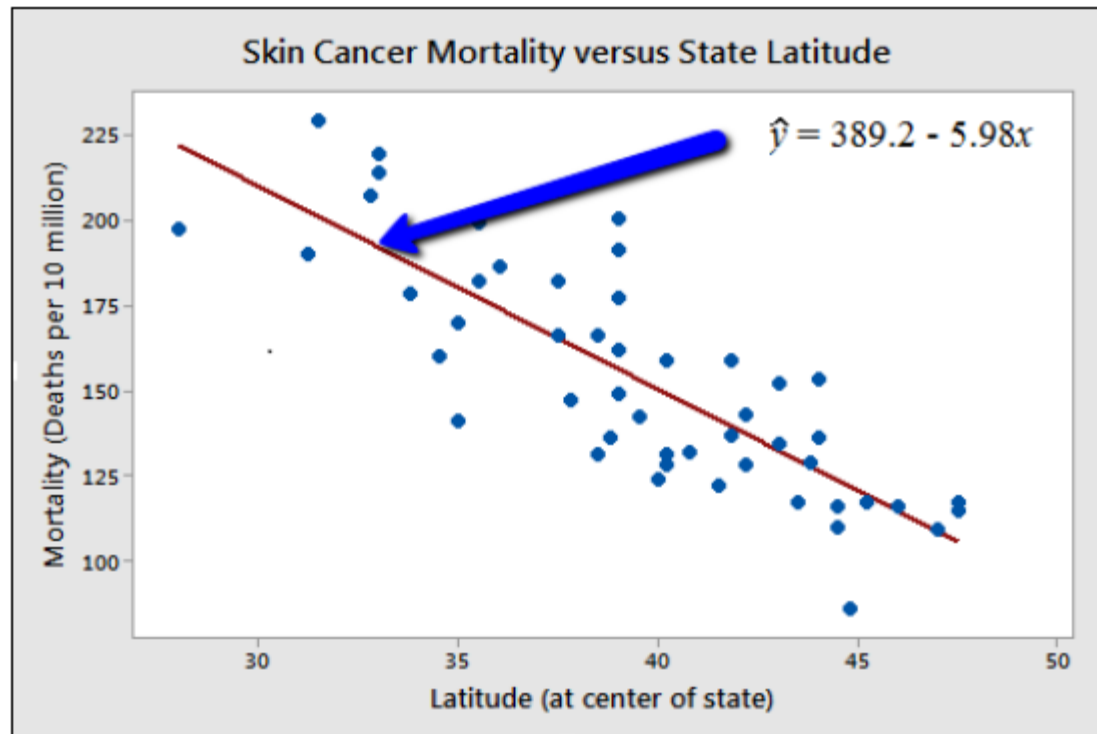
- 학습 데이터
 - (x, y) 데이터
- 학습의 정의
 - 가설 : $h=wx+b$
 - 기울기 x 와 y 절편 b
- 학습 방법
 - 선형 연립방정식과 역행렬을 이용한 미지수 w, b 추정
- 예측
 - w 와 b 로 이루어진 가설 함수를 사용해 y 값 예측

주가 예측



<https://datatofish.com/multiple-linear-regression-python/>

피부암 사망률 예측



<https://newonlinecourses.science.psu.edu/stat501/node/251/>

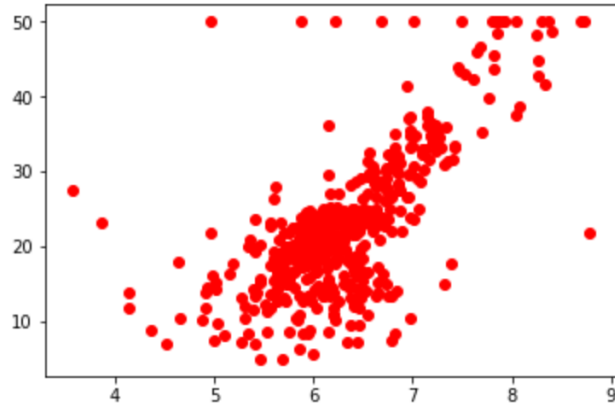
베네통 매출/광고 예측

Year	Sales (Million Euro)	Advertising (Million Euro)
1	651	23
2	762	26
3	856	30
4	1,063	34
5	1,190	43
6	1,298	48
7	1,421	52
8	1,440	57
9	1,518	58

<https://www.displayr.com/what-is-linear-regression/>

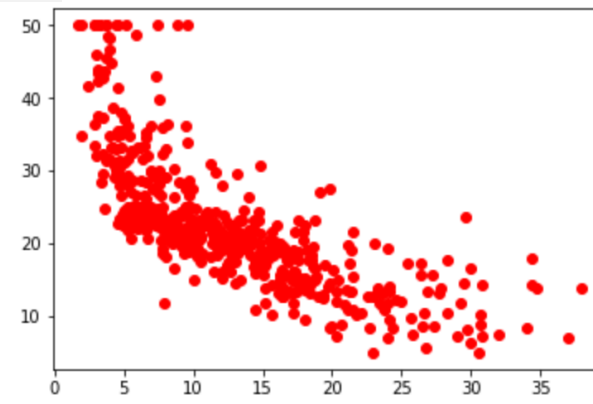
보스턴 집값 예측

집값



RM(주택당 평균방수)

집값

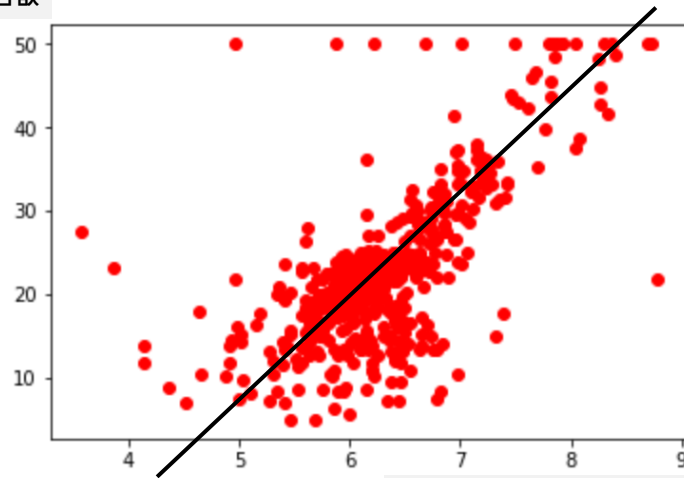


LSTAT(인구 감소 비율)

CRIM 타운별 1인당 범죄율
ZN 25,000 평방피트를 초과하는 거주지역의 비율
INDUS 비소매상업지역이 점유하고 있는 토지의 비율
CHAS 찰스강에 대한 더미변수
NOX 10ppm 당 농축 일산화질소
RM 주택 1가구당 평균 방의 개수
AGE 1940년 이전에 건축된 소유주택의 비율
DIS 5개의 보스턴 직업센터까지의 접근성 지수
RAD 방사형 도로까지의 접근성 지수
TAX 10,000 달러 당 재산세율
PTRATIO 타운별 학생/교사 비율
B 타운별 흑인 비율
LSTAT 모집단의 하위계층 비율
MEDV 본인 소유의 주택가격(단위: \$1,000)

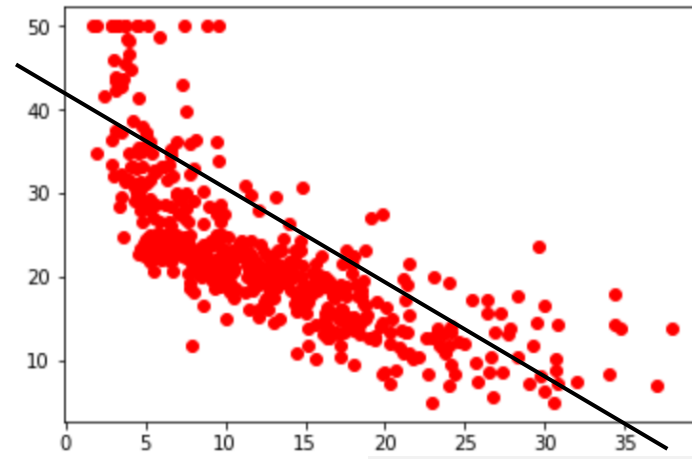
데이터 경향성 분석

집값



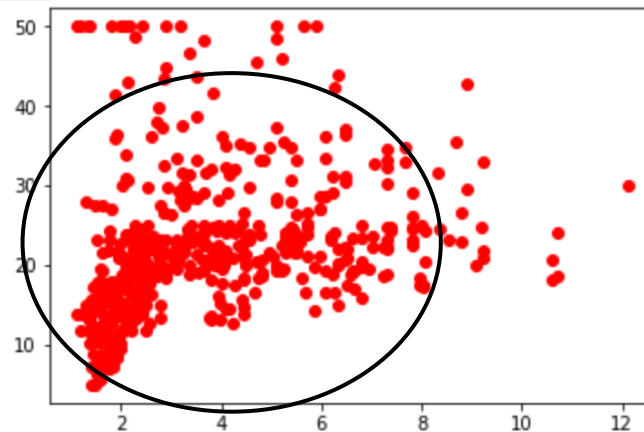
RM(주택당평균방수)

집값



LSTAT(인구 감소 비율)

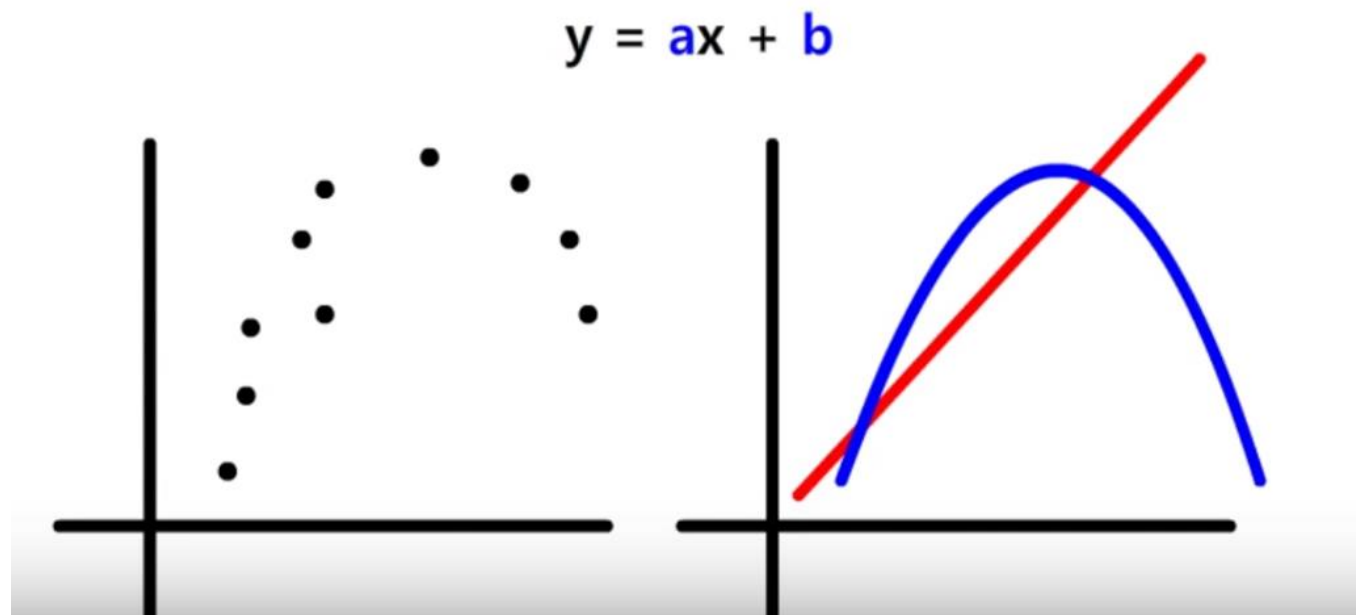
집값



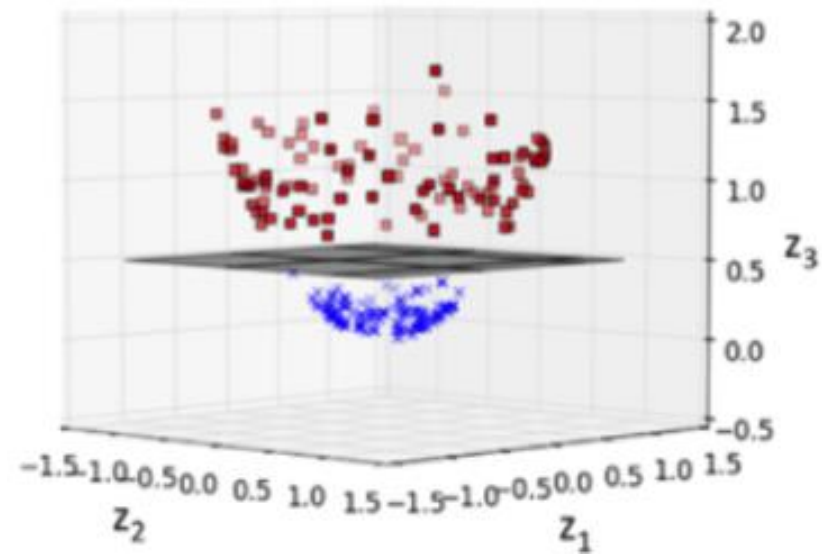
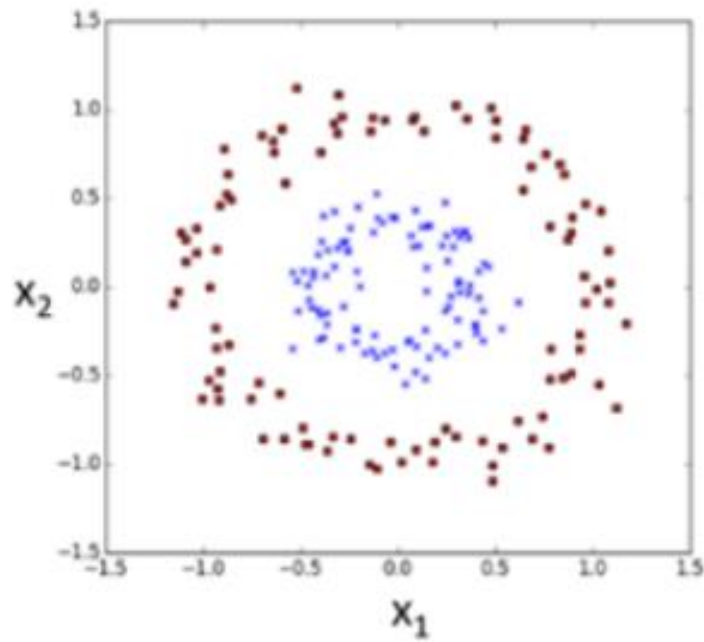
NOX(농축일산화질소의 양)

비선형 모델이면?

대부분의 문제는 비선형 분포

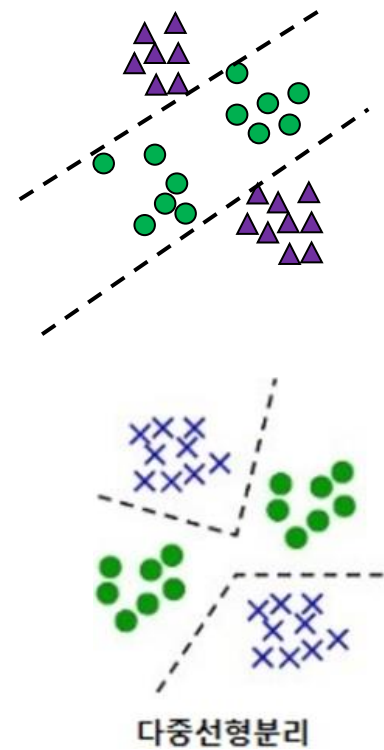
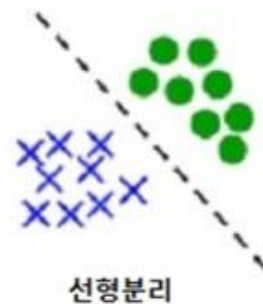
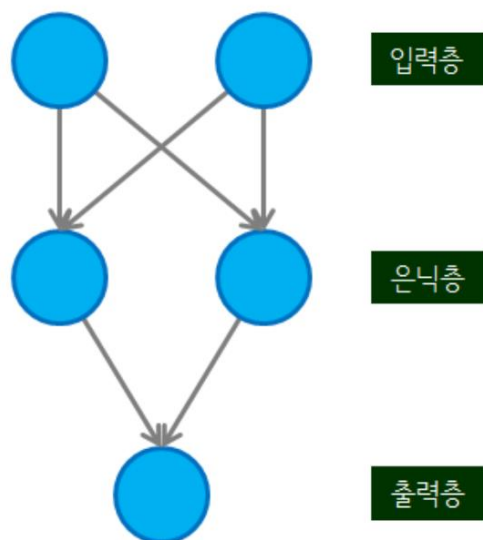


고차원 변환을 통해 선형 모델링

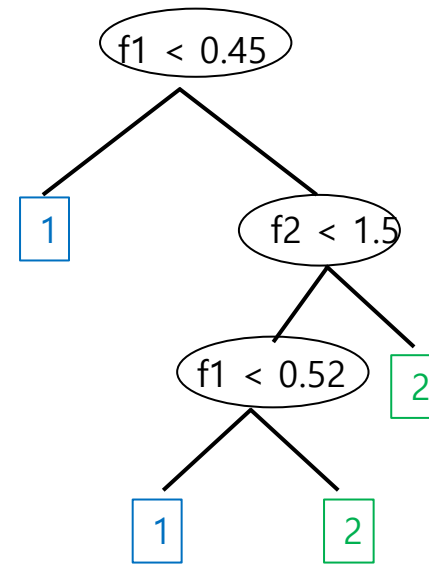
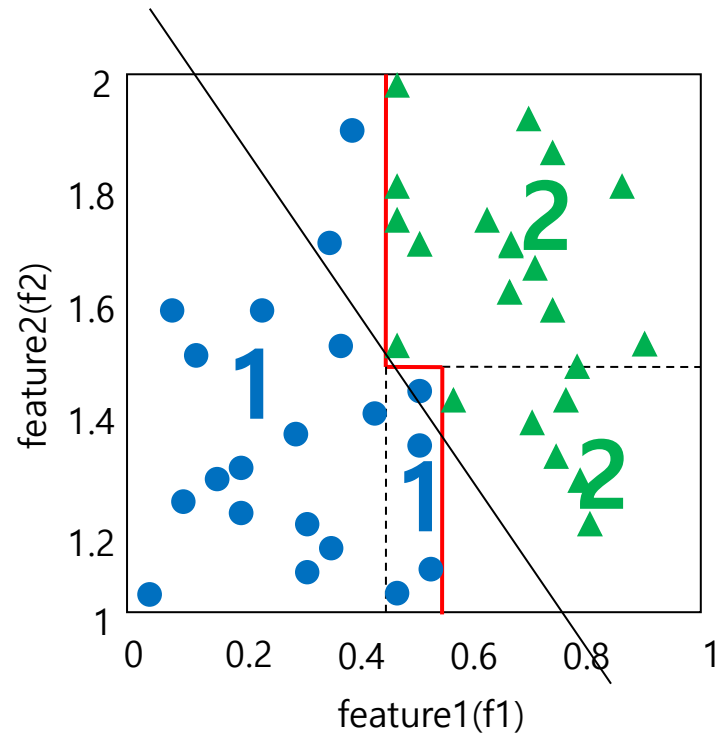


다층신경망(Multi-Layer Perception)

다층 신경망을 통한 비선형 문제 해결



다층 분류기를 이용한 문제 해결



딥러닝(Deep Learning)

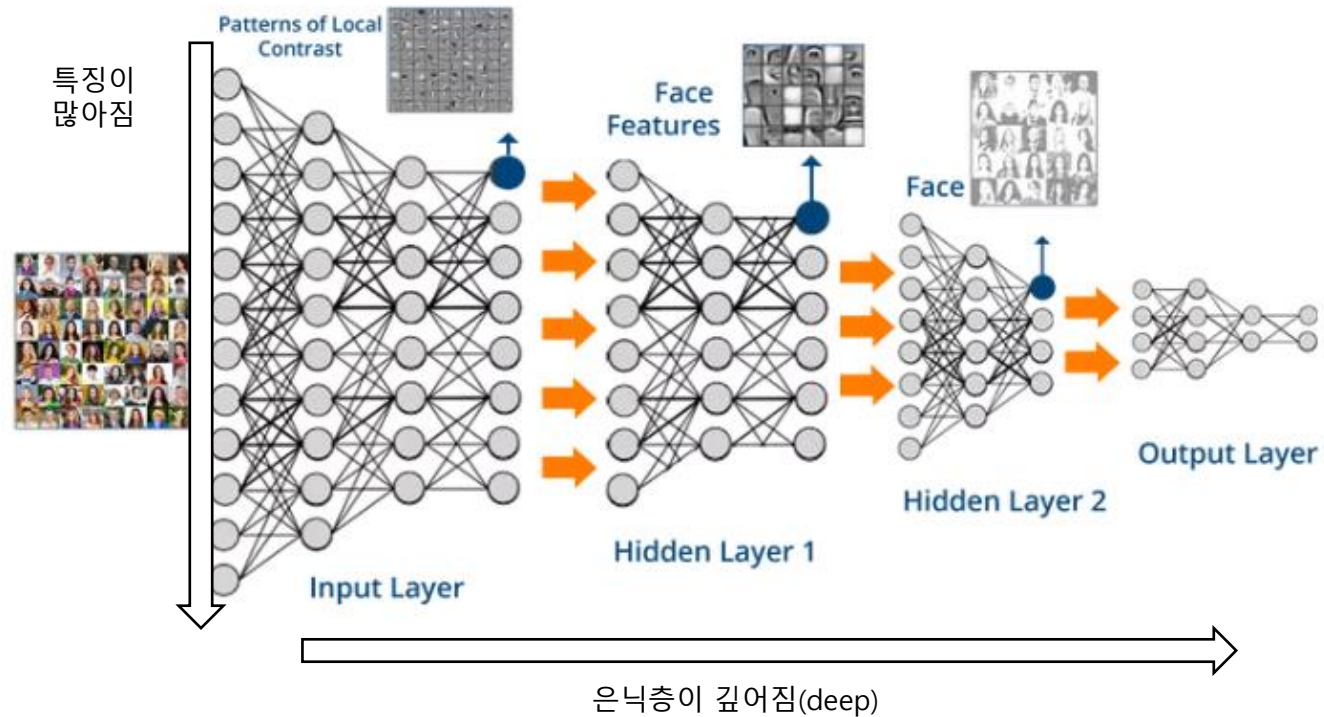


Image from: <https://cdn.edureka.co/blog/wp-content/uploads/2017/05/Deep-Neural-Network-What-is-Deep-Learning-Edureka.png>

딥러닝 (Deep Learning)

Revolution of Depth

