

청년을 응원합니다!

K-Digital Training

AI 비즈니스 컨설팅: 데이터 기반 분석 및 문제 해결 과정



기초통계 및 경영통계

강사 김경하



고용노동부

삼성 **KPMG**



현대경제연구원



i-Growth



기초 통계 with 데이터분석, 머신러닝

강의 내용

- 통계 이해
 - 기초통계
 - 가설검정
- 데이터분석, 머신러닝을 위한 수학
 - 미분 기본 원리 이해
 - 확률과 조건부 확률

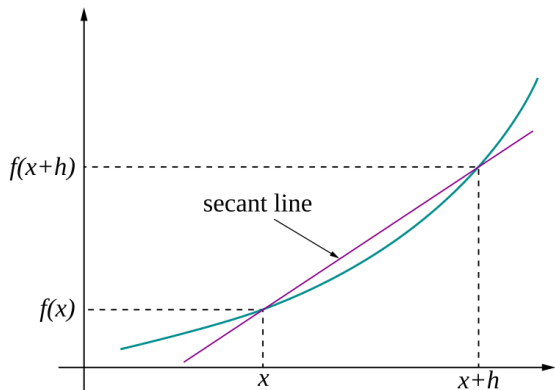
학습 목표

- 미분계수의 의미를 안다.
- 머신러닝에서 미분이 사용되는 기본 원리를 이해한다.
- 확률과 조건부 확률을 이해한다.
- 머신러닝에서의 조건부 확률이 사용되는 기본 원리를 이해한다.

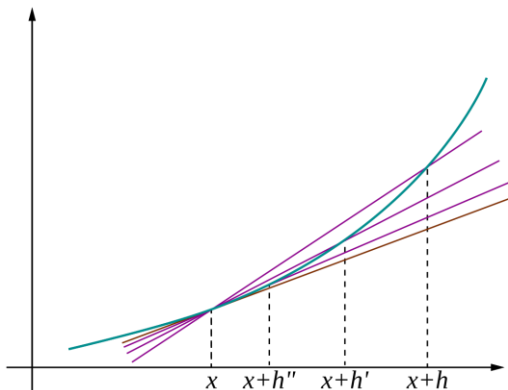
머신러닝에서 미분 활용

<https://ko.wikipedia.org/wiki/%EB%AF%B8%EB%B6%84>

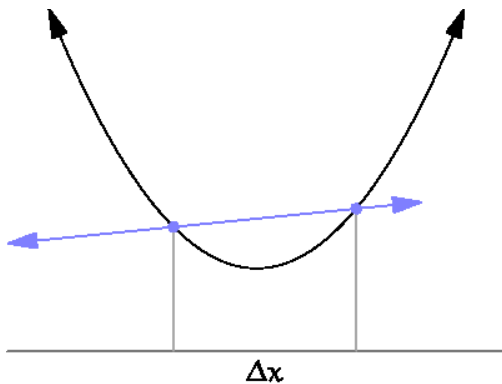
- 미분(differentiation) = 도함수
 - 어떤 함수의 **순간 변화율(미분계수)**를 구하는 것
- 기하학적 관점에서 미분
 - 주어진 곡선의 접선을 구하는 문제와 동치됨



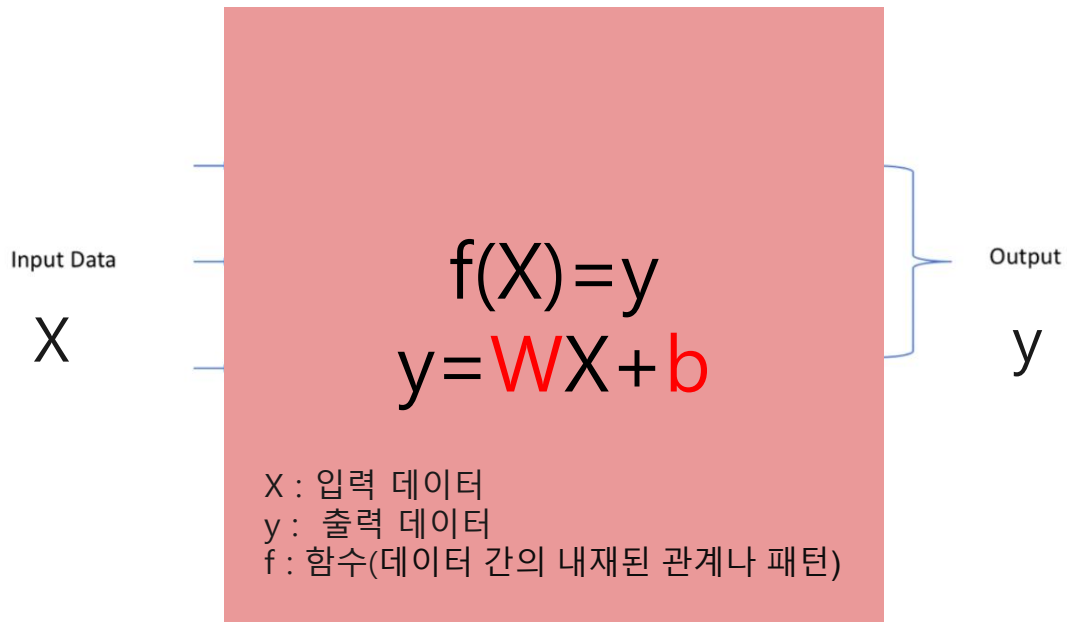
$$\bar{k} = \frac{\Delta y}{\Delta x} = \frac{f(x+h) - f(x)}{(x+h) - x}$$



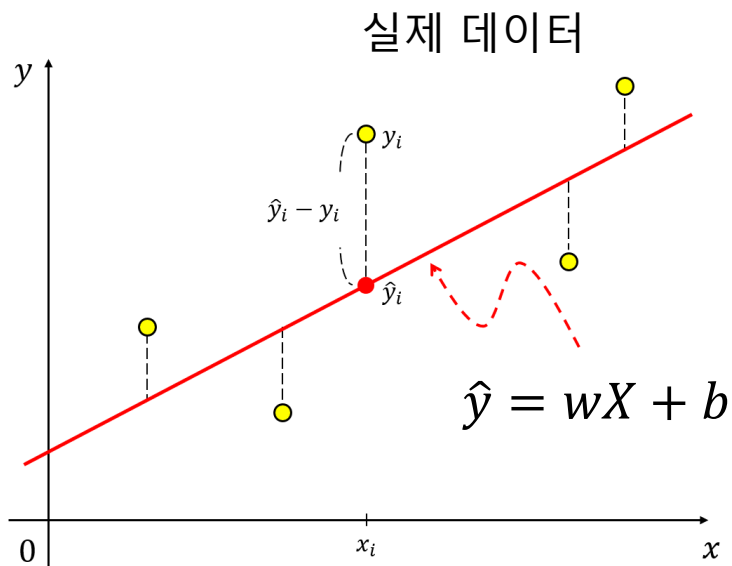
$$\lim_{x \text{ 증가량} \rightarrow 0} \frac{y \text{ 증가량}}{x \text{ 증가량}} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{(x+h) - x}$$



- 문제해결을 위해 입력 데이터(X)에 대응하는 정답(y)를 함께 입력받아, 서로에게 대응하는 함수식(f)를 찾는 것

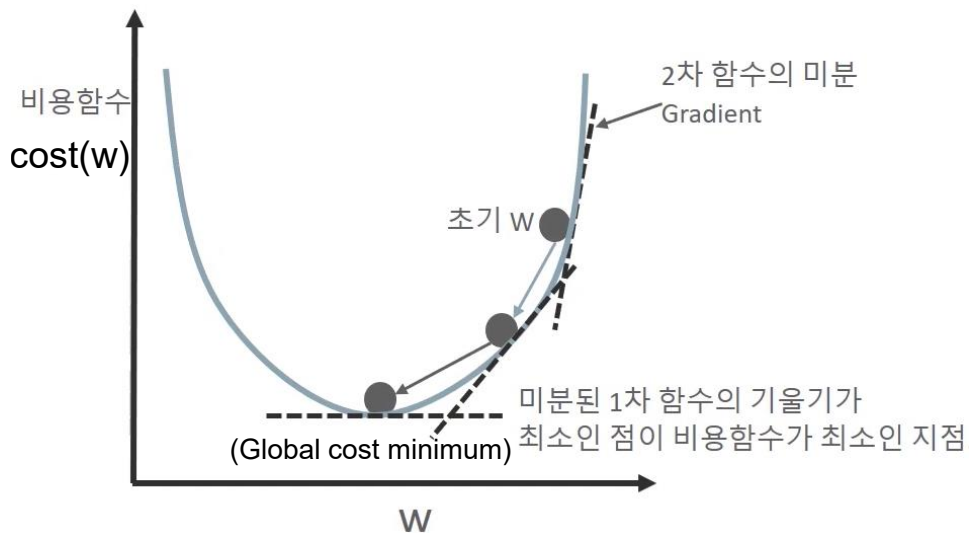


경사하강법의 미분계수 의미



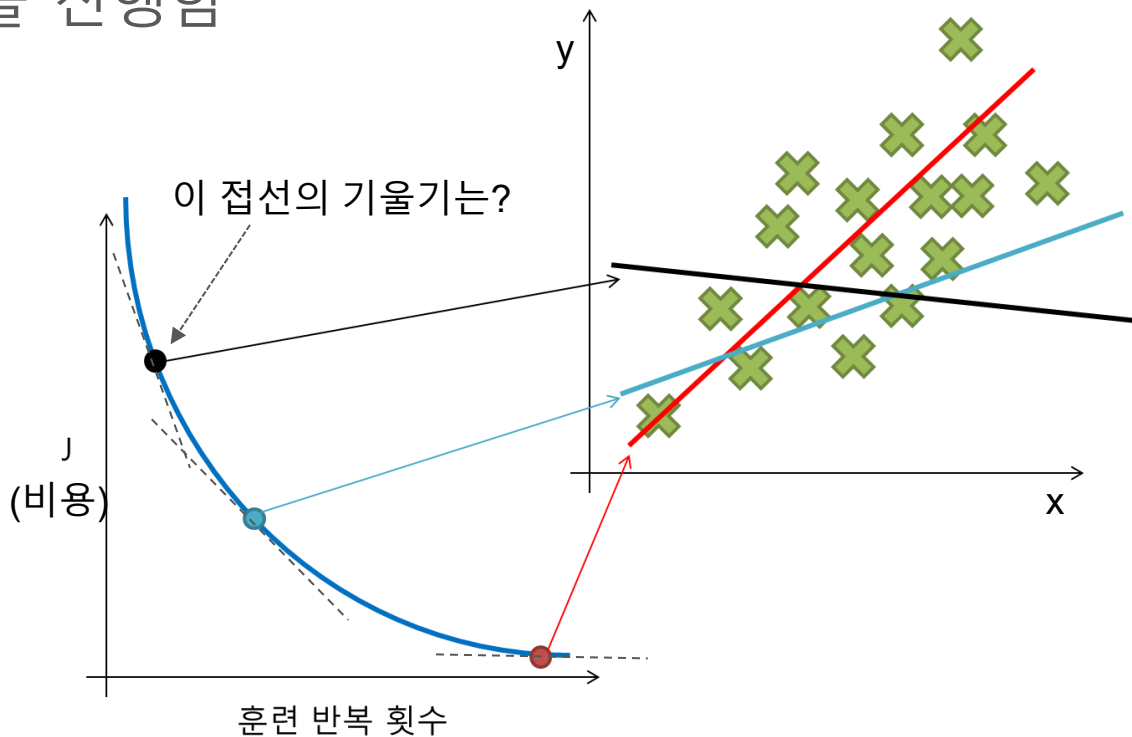
가설함수 $H(x) = \hat{y}$

w : 기울기
 b : 절편



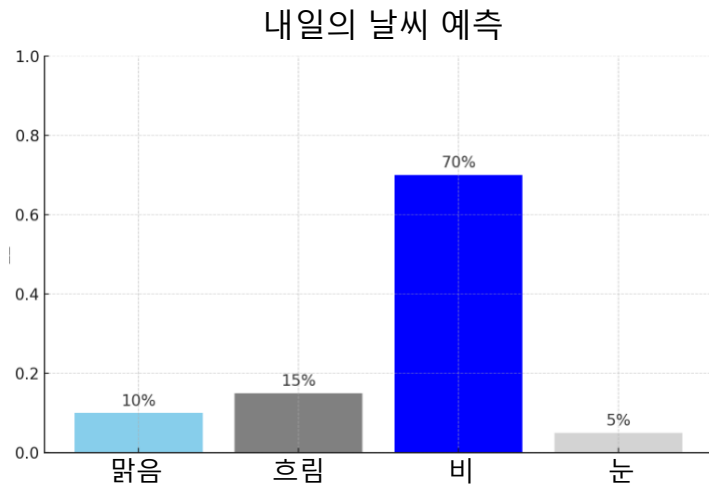
$$cost(w) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

- 접선의 기울기가 음수일 때, 기울기가 증가하는 방향으로 학습을 진행함



확률과 조건부 확률

- 통계의 목적
 - 통계는 아직 벌어지지 않은 일을 예측하기 위해 존재함
 - 우리는 미래를 알 수 없지만, 가능성을 숫자와 함수로 표현함으로써 준비할 수 있음
 - 지금은 보이지 않는 미래의 사건을, **데이터와 확률**을 통해 들여다보는 것이 바로 통계



어떤 일이 일어날 가능성을 숫자로 표현한 것

- 실험 : 주사위를 한번 던진다.
- 표본 공간 : 주사위에서 나올 수 있는 모든 결과
($\{1, 2, 3, 4, 5, 6\}$)
- 근원사건 : 1의 눈이 나오는 사건 $\{1\}$
- 복합사건 : 짝수가 나오는 사건 $\{2, 4, 6\}$



확률의 공리주의적 정의

- 모든 사건 e 에 대해 0과 1사이 값을 가진다.

$$0 \leq P(e) \leq 1$$

- 표본공간의 확률의 합은 1 이다

$$P(S) = 1$$

- 각 사건의 확률을 모두 더하면 값은 1이다.

$$\sum_{i=1}^n P(e_i) = P(e_1) + P(e_2) + \cdots + P(e_n) = 1$$

사건 B가 발생했다는 조건에서 사건 A가 발생할 확률

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- 사건 A : 3의 배수의 눈이 나오는 사건 $A = \{3, 6\}$
- 사건 B : 5 이상의 눈이 나오는 사건 $B = \{5, 6\}$
- $P(A|B)$: 5이상의 눈이 나왔다는 조건에서 3의 배수가 나올 확률
 - 5이상의 눈인 것을 알고 있다면 그 준이 3의 배수일 확률은?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{6\})}{P(\{5, 6\})} = \frac{\frac{1}{6}}{\frac{2}{6}} = \frac{1}{2}$$

[팀 미션]

- 조건부 확률이 머신러닝에서 어떻게 활용되는지 조사 발표하기
- 발표 10 : 40 ~

조건부 확률 활용

- 머신러닝에서 조건부 확률
 - $P(Y|X)$: 입력값 X 가 주어졌을 때 정답 Y 일 확률
- 머신러닝에서 사용
 - 입력 X (예: 이미지, 텍스트, 숫자 등)가 주어졌을 때, 어떤 출력 Y (예: 고양이 인지, 긍정적인지, 판매될지 등) 확률적으로 추정하는 것

조건부 확률 활용

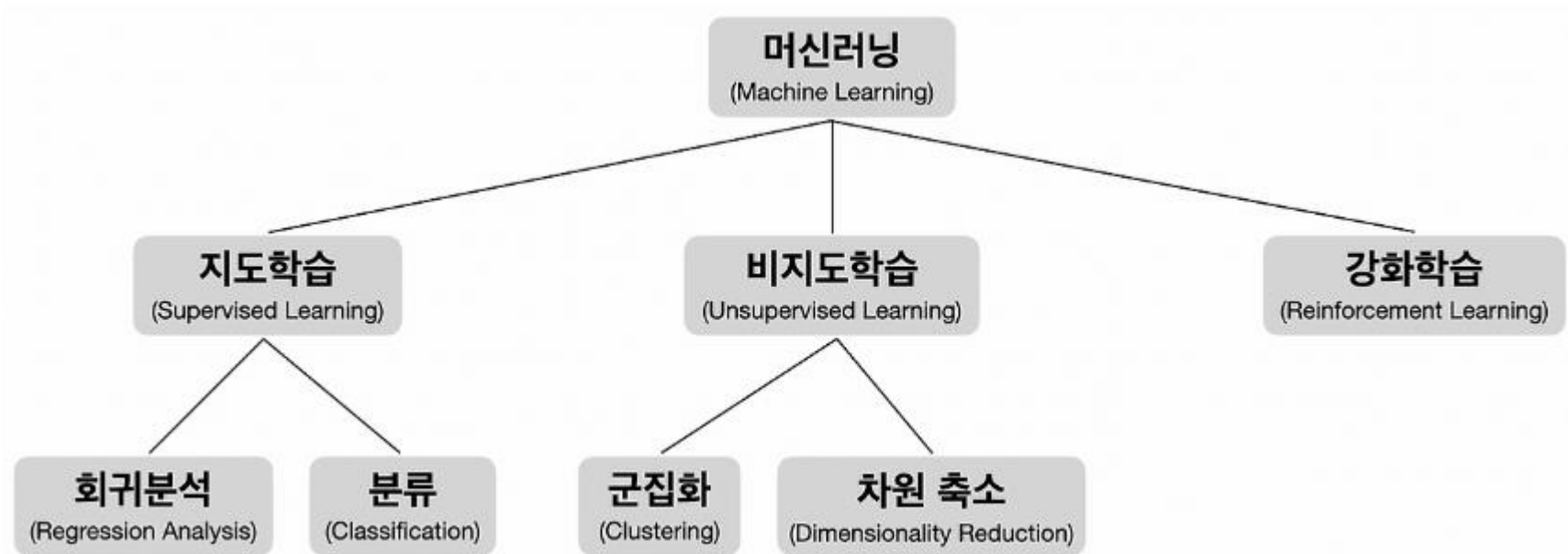
- 분류 문제

- $P(\text{환자가 특정 질병을 가짐} \mid \text{환자의 증상, 검사 결과, 병력})$
예: $P(\text{당뇨병} \mid \text{혈당 수치, 체중, 나이, 가족력})$

- 회귀 문제

- 집의 크기(X)가 주어졌을 때 가격(Y)을 예측한다고 하면
회귀 모델은 평균값(예: 기대값)을 추정함
 $P(\text{가격} \mid \text{집 크기, 위치, 연식, 방 개수})$

머신러닝 학습 방법



감사합니다