

Part.02  
회귀분석

# | 회귀분석이란

FASTCAMPUS  
ONLINE

머신러닝과 데이터분석 A-Z

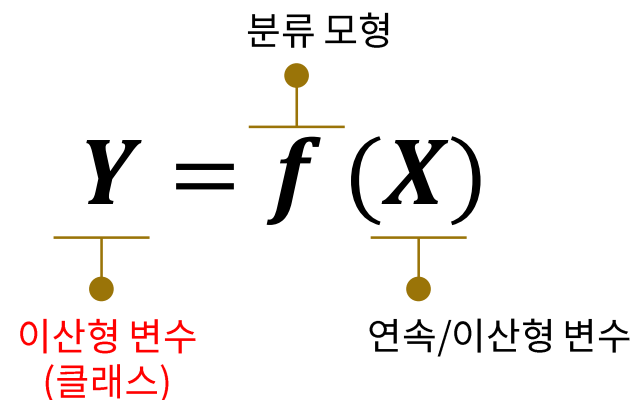
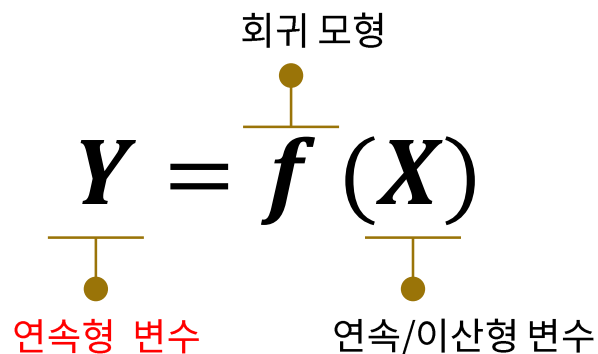
강사. 이경택

# I 회귀분석이란

## ■ 지도 학습(supervised learning)

$Y = f(X)$  에 대하여 입력 변수 ( $X$ )와 출력 변수 ( $Y$ ) 의 관계에 대하여 모델링하는것  
( $Y$ 에 대하여 예측 또는 분류하는 문제)

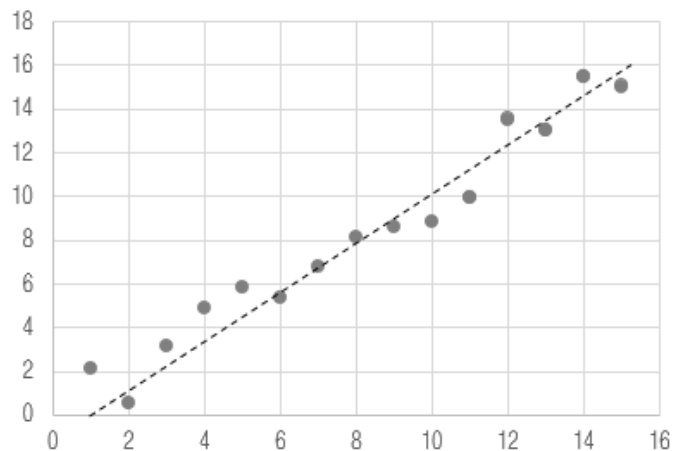
- 회귀 (regression): 입력 변수  $X$ 에 대해서 연속형 출력 변수  $Y$ 를 예측
- 분류 (classification): 입력 변수  $X$ 에 대해서 이산형 출력 변수  $Y$ (class)를 예측



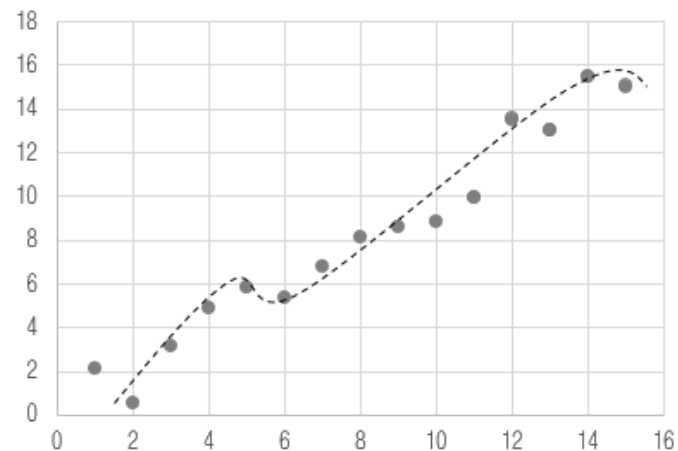
# I 회귀분석이란

## ■ 회귀분석

- 입력 변수인  $X$ 의 정보를 활용하여 출력 변수인  $Y$ 를 예측하는 방법
- 회귀분석 중 간단한 방법으로는 선형회귀분석(좌측 그림)이 있으며, 이를 바탕으로 더 복잡한 회귀분석(우측 그림)이 개발



선형회귀



비선형회귀

대부분의 분류모델(SVM, Decision Tree 등)로도 회귀가 가능함.

# I 회귀분석이란

## ■ 단순 선형 회귀분석

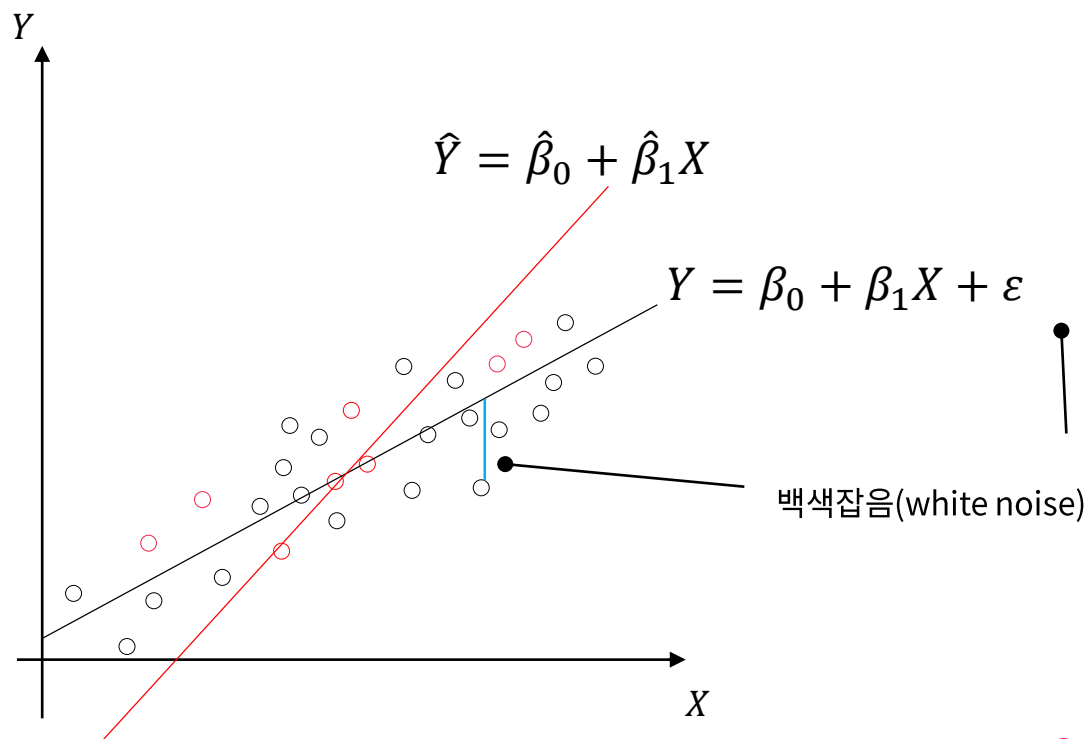
$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- 입력 변수가  $X$ , 출력 변수가  $Y$ 일 때, 단순선형회귀의 회귀식은 검은 선으로 나타낼 수 있음
- $\beta_0$ 는 절편(intercept),  $\beta_1$ 은 기울기(slope)이며 합쳐서 회귀계수(coefficients)로도 불림

- 검은 점: 모집단의 모든 데이터
- 빨간 점: 학습집합의 데이터
- 실제  $\beta_0$ 와  $\beta_1$ 은 구할 수 없는 계수로 데이터(학습집합)를 통해 이 둘을 추정해서 사용

## ■ 가정 :

$\varepsilon_i \sim i.i.d N(0, \sigma^2)$ ,  $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$ ,  $X$ 와  $Y$ 는 선형관계



# I 회귀분석이란

- 단순 선형 회귀분석
  - 우리가 알고 싶은 식

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

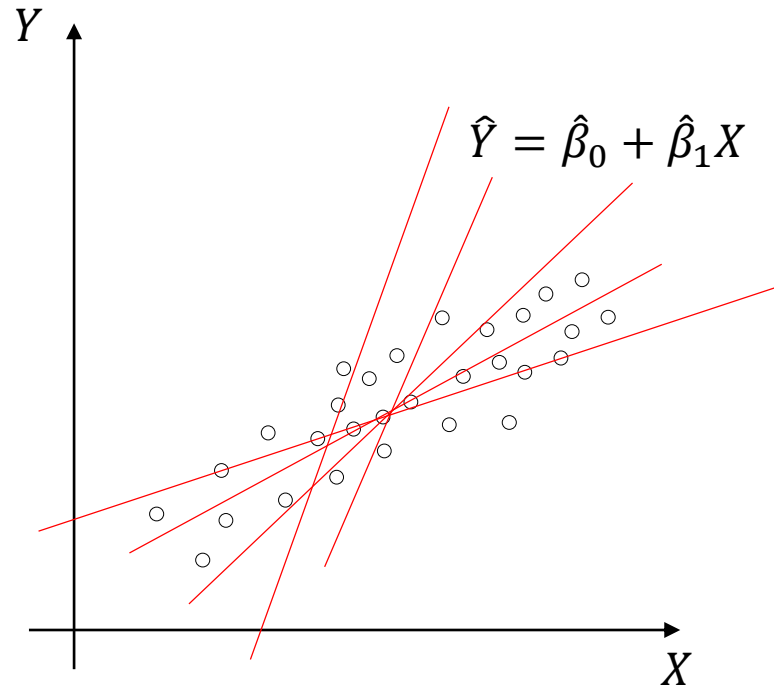
- 우리가 추정 해야하는 식

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

- 어떻게 추정 할까?

# I 회귀분석이란

- 단순 선형 회귀분석
  - 어떻게 추정 할까?
    - 여러 개의 직선 중 가장 좋은 직선은?



➤ 직선과 데이터의 차이가 평균적으로 가장 작아지는 직선

Part.02  
회귀분석

# | 회귀계수추정

FASTCAMPUS  
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택