Chapter. 02

선형대수학 응용

최소자승법과 선형회귀

FASTCAMPUS
ONLINE
금융공학/퀀트

강사. 장순용

I 키포인트

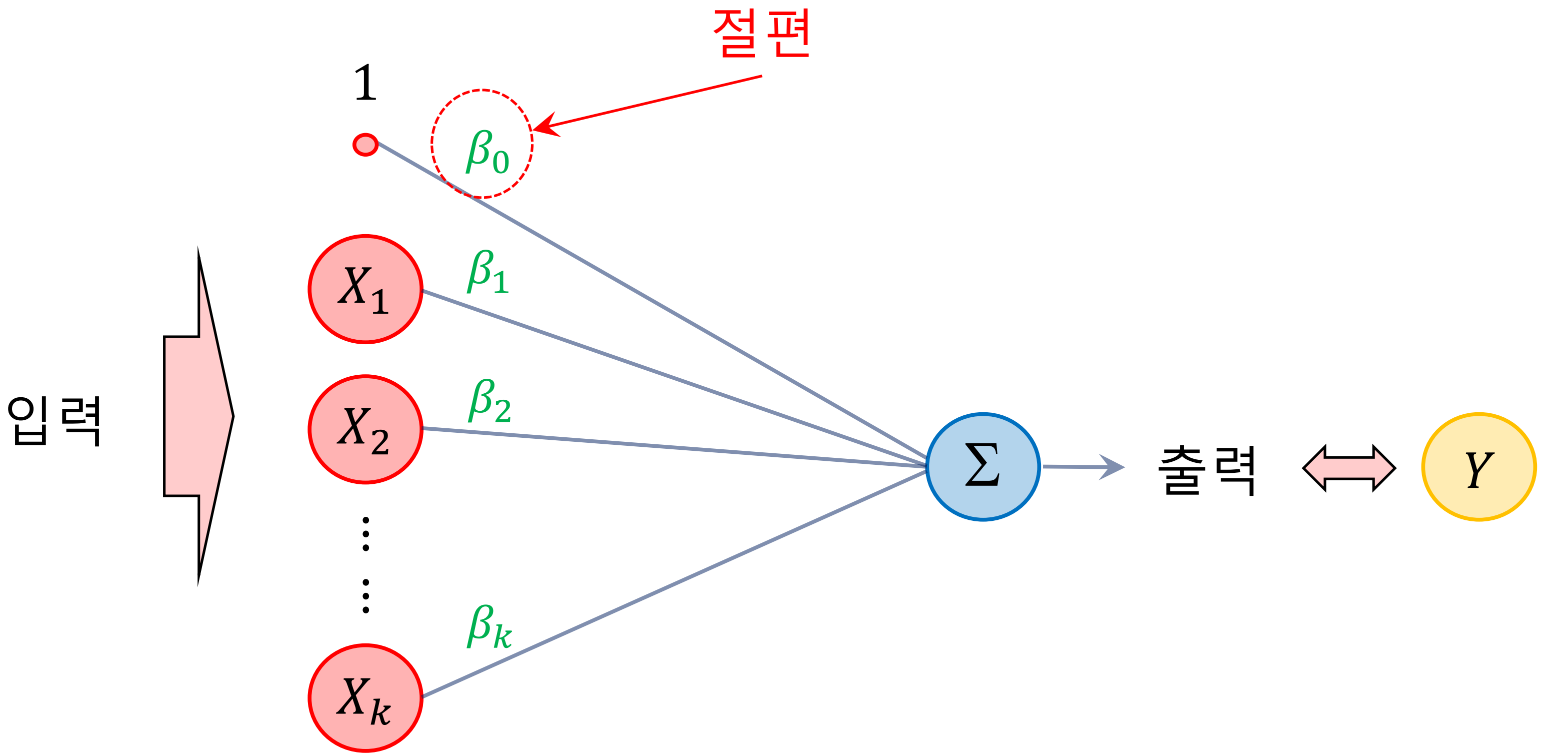
- 선형회귀.
- 최소자승법 (Ordinary Least Squares, OLS).
- 가역행렬의 도출.

I 선형회귀

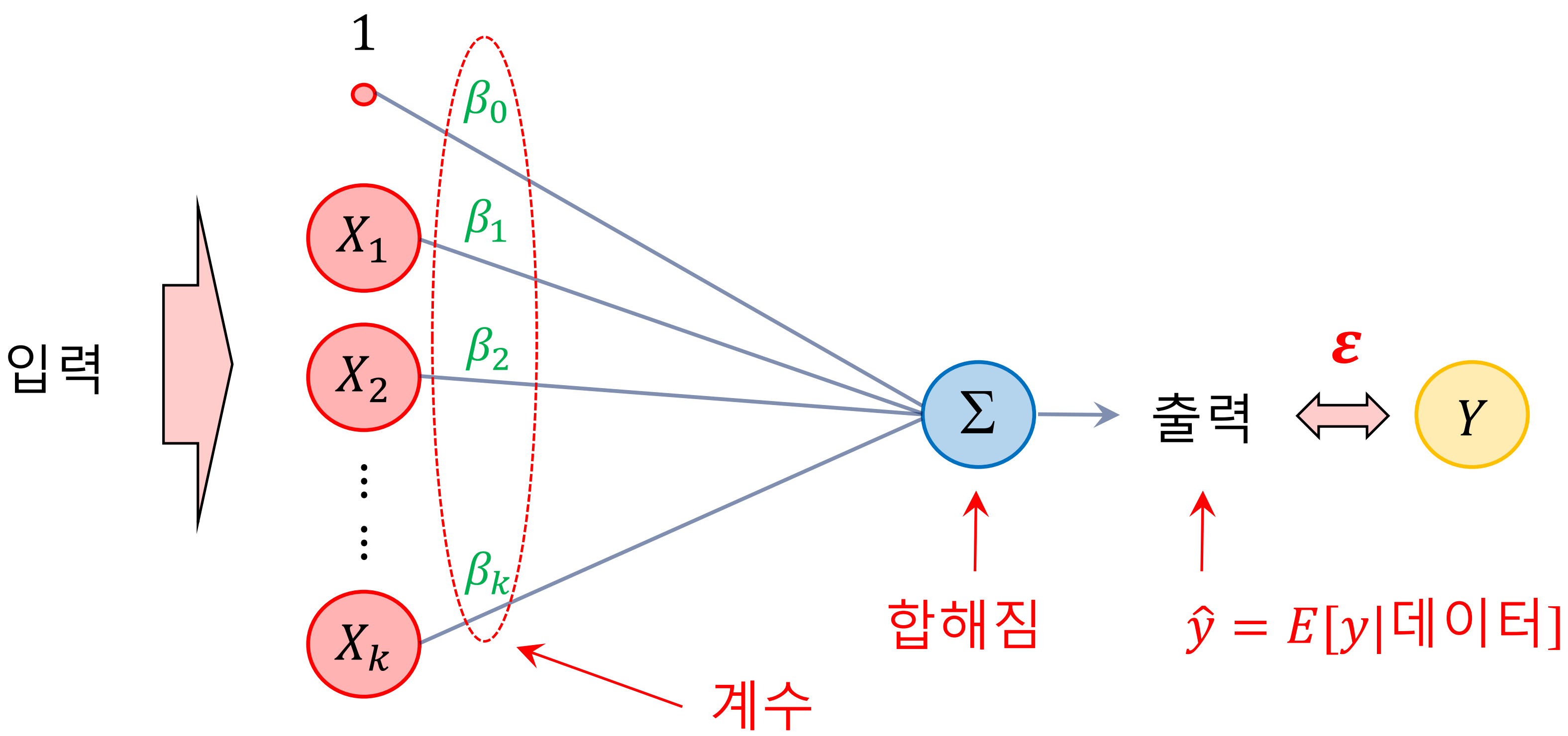
- 회귀모형:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + \varepsilon$$

I 선형회귀 구조



I 선형회귀 예측



1
β_0
β_1
β_2
β_k
입력
X_1
X_2
X_k
Σ
출력
ε
Y
합해짐
$\hat{y} = E[y|$데이터$]$
계수

I 선형회귀 연립방정식

- 개개 방정식:

$$y_j = \beta_0 + \underbrace{\beta_1 x_{j,1} + \beta_2 x_{j,2} + \cdots + \beta_K x_{j,K}}_{n \;>>\; K} + \varepsilon_j$$

$j \in [1, n]$

$n >> K$

Over-determined!

I 선형회귀 연립방정식

- 선형대수학적 표기:

$$\boldsymbol{Y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

I 선형회귀 연립방정식

- 선형대수학적 표기:

$$\boldsymbol{Y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

I 선형회귀 연립방정식

- 선형대수학적 표기:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,K} \\ 1 & x_{2,1} & \cdots & x_{2,K} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,K} \end{bmatrix}$$

I 선형회귀 연립방정식

- 선형대수학적 표기:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}$$

I 선형회귀 연립방정식

- 선형대수학적 표기:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

I 선형회귀의 OLS 해 도출

- $\|\boldsymbol{\varepsilon}\|^2$ 를 최소화 하는 가장 "근접한 해" $\boldsymbol{\beta}$를 도출해 내고자 한다.
- $\|\boldsymbol{\varepsilon}\|^2$ 의 최소값 조건은 다음과 같이 미분으로 나타낼 수 있다.

$$\frac{d\|\boldsymbol{\varepsilon}\|^2}{d\beta} = 0$$

I 선형회귀의 OLS 해 도출

- 선형 방정식을 대입하여 전개해 본다:

$$\frac{d\|\boldsymbol{\varepsilon}\|^2}{d\beta} = \frac{d\|\boldsymbol{Y} - \boldsymbol{X\beta}\|^2}{d\beta}$$

$$= \frac{d\{(\boldsymbol{Y} - \boldsymbol{X\beta})^t(\boldsymbol{Y} - \boldsymbol{X\beta})\}}{d\beta}$$

$$= \frac{d\{\boldsymbol{Y^tY} - \boldsymbol{\beta^tX^tY} - \boldsymbol{Y^tX\beta} + \boldsymbol{\beta^tX^tX\beta}\}}{d\beta}$$

$$= -2\boldsymbol{X^tY} + 2\boldsymbol{X^tX\beta} = 0$$

I 선형회귀의 OLS 해 도출

- 이전 슬라이드의 조건을 정리하면 다음 OLS 해를 얻는다:

$$\boldsymbol{\beta} = [(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t]\boldsymbol{Y}$$

$$\Uparrow$$

$\|\boldsymbol{\varepsilon}\|^2$를 최소화 하는 계수벡터이다.

I 선형회귀의 OLS 해 도출

- 이전 슬라이드의 조건을 정리하면 다음 OLS 해를 얻는다:

$$\boldsymbol{\beta} = [(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t]\boldsymbol{Y}$$

가역행렬 !!!

(Pseudoinverse)

$\Rightarrow$ 미지수보다 방정식 (조건)이 더 많은 상황이었다.

I 선형회귀의 OLS 해 도출

- 이전 슬라이드의 조건을 정리하면 다음 OLS 해를 얻는다:

$$\boldsymbol{\beta} = [(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t]\boldsymbol{Y}$$

⇒ 다음과 같이 단 1행의 코딩으로 구현할 수 있다 (Python).

```
beta = np.dot(np.linalg.inv(np.dot(X.T,X)),np.dot(X.T,Y))
```

I 끝.

감사합니다.

FAST CAMPUS
ONLINE
장순용 강사.