Chapter06

선형회귀

선형회귀의 원리

FASTCAMPUS ONLINE

금융공학/퀀트

강사. 장순용

I 키포인트

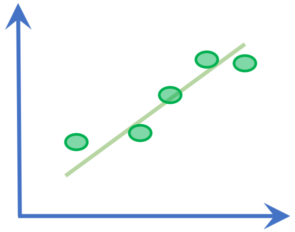
- 선형회귀의 원리.
- 선형회귀 모형의 해석.

I 통계 예측모형

통계 예측모형

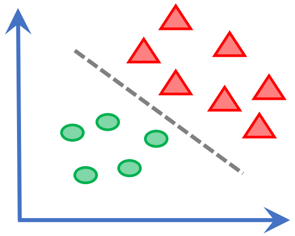
수치 예측

$Y = 13.45,\ \ 73,\ \ 9.5,\ \ .....$



분류 예측

$Y =$ red, green, blue, ...



I 통계 예측모형

통계 예측의 평가

수치 예측

MSE, MAE, RMSE, 등.

분류 예측

Accuracy (정확도), Precision (정밀도), Recall (재현율, 민감도), 등.

I 선형회귀 개요

- 선형회귀는 대표적인 수치 예측 방법이다.
- 한 개 이상의 독립변수 (설명변수)가 있다: $X_1, X_2, \ldots, X_K$
- 한개의 종속변수가 있다: Y
- 선형 관계를 전제한다: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + \varepsilon$
- 이외에도 여러 가지의 전제조건이 있다 ⇒ 잔차 분석 (later)

I 선형회귀 목적

1. 종속변수를 설명하는 독립변수를 밝혀냄.

예). 아파트의 가격은 면적, 위치, 방의 수 등으로 설명할 수 있다.

(?)

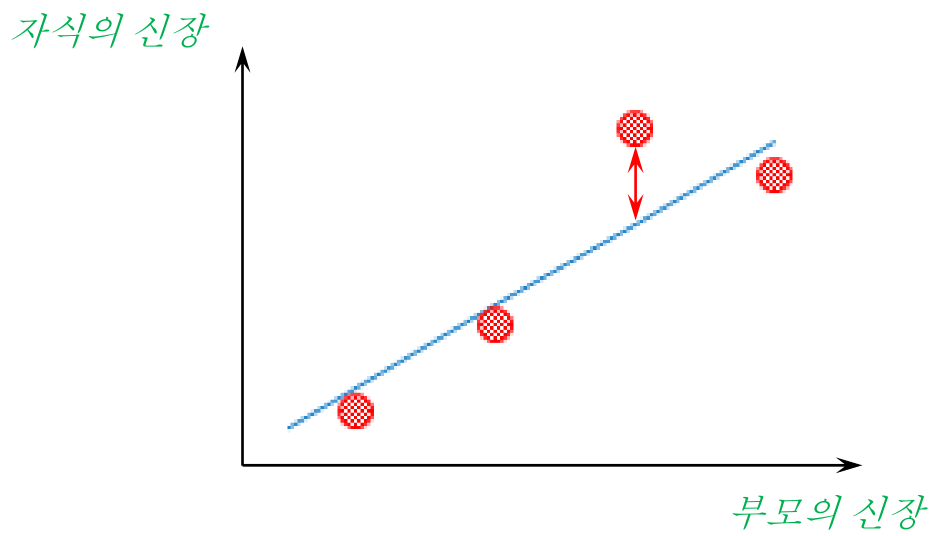
2. 독립변수 값의 변동에 따른 종속변수의 변동을 예측함.

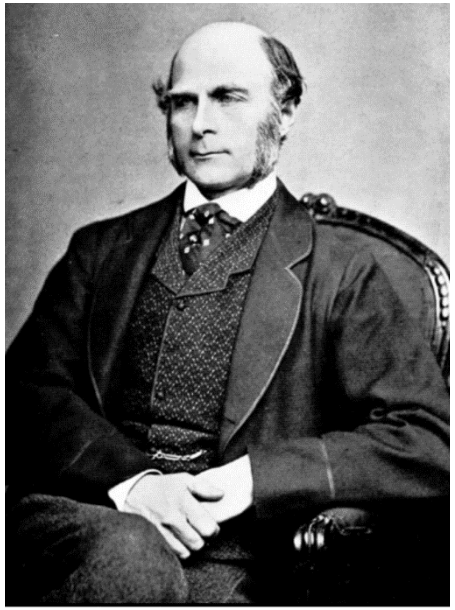
예). 학습된 모형으로 아파트의 적정 가격을 알아 맞춘다.

I 역사적 배경

- 19세기 영국의 우생학자인 Francis Galton이 평균으로 돌아간다라는 의미의 "회귀"라는 용어를 처음 사용함.

⇒ 신장에 있어서 부모와 자식 사이의 유전적 관계를 연구함.





I 선형회귀 원리

- 회귀모형:

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

I 선형회귀 원리

- 회귀모형:

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

종속변수

(데이터로 값이 주어짐/예측의 대상)

I 선형회귀 원리

- 회귀모형:

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

독립변수

(데이터로 값이 주어짐)

FAST CAMPUS
ONLINE
장순용 강사.

l 선형회귀 원리

- 회귀모형:

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

계수

(학습을 통해서 밝혀냄)

FAST CAMPUS
ONLINE
장순용 강사.

I 선형회귀 원리

- 회귀모형:

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

계수가 데이터 속의 패턴을 담음.

I 선형회귀 원리

- 회귀모형의 예 #1:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

MPG

I 선형회귀 원리

- 회귀모형의 예 #1:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

MPG

N# of Cylinders

I 선형회귀 원리

- 회귀모형의 예 #1:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

MPG N# of Cylinders HP

I 선형회귀 원리

- 회귀모형의 예 #1:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

MPG   N# of Cylinders   HP   Weight

I 선형회귀 원리

- 회귀모형의 예 #1:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

Y	X_1	X_2	X_3	X_4
MPG	N# of Cylinders	HP	Weight	Auto or Manual

원리에 대해서 자세히 알아 본다.

I 선형회귀 원리

- 오차변수 (white noise):

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

평균[ε] = 0

표준편차[ε] = σ_ε

I 선형회귀 원리

- 공선성을 피해야 함:

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

$$\mathrm{Cor}(X_i, X_j) \approx 0$$

$$\text{for } i \neq j$$

I 선형회귀 원리

- 공선성을 피해야 함:

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

공선성은 계수의 "분산 인플레"

문제를 일으킬 수 있다.

I 선형회귀 원리

- 회귀계수:

$$\Delta Y = \beta_1 \Delta X_1 + \cdots + \beta_i \Delta X_i + \cdots + \beta_K \Delta X_K$$

X변수가 ΔX만큼 변동한다면

$\Rightarrow$ Y변수는 ΔY만큼 반응한다

I 선형회귀 원리

- 회귀계수:

$$\Delta Y = \beta_1 \Delta X_1 + \cdots + \beta_i \Delta X_i + \cdots + \beta_K \Delta X_K$$

1

β_i 는 다른 X 변수는 그대로 있으면서 X_i만 1 증가할 때의 ΔY.

I 선형회귀 원리

- 회귀계수:

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

절편
(Intercept)

l 선형회귀 원리

- 회귀계수:

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

절편
(Intercept)

If $X_i = 0$

I 선형회귀 원리

- 회귀계수:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + \varepsilon = 0$$

절편
(Intercept)

I 선형회귀 원리

- 회귀계수:

$$Y = \beta_0$$

"베이스"의 역할.

절편
(Intercept)

I 선형회귀 원리

- 회귀모형의 예 #2:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

급여

I 선형회귀 원리

- 회귀모형의 예 #2:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

급여

$X_1 =$ 교육

$X_2 =$ 경력

I 선형회귀 원리

- 회귀모형의 예 #2:

$\beta_1 > 0$

$\beta_2 > 0$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

급여

$X_1 =$ 교육

$X_2 =$ 경력

I 선형회귀 원리

- 회귀모형의 예 #2:

$\beta_1 > 0$

$\beta_2 > 0$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

급여

$X_1 =$ 교육

$X_2 =$ 경력

$\beta_0 =$ 베이스 급여

Life Changing Education

I 끝.

감사합니다.

FAST CAMPUS
ONLINE
장순용 강사.