

Part.02  
회귀분석

# | 회귀계수 축소법

FASTCAMPUS  
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택

# I 회귀계수 축소법

## ■ 분석용 데이터의 이상적 조건

- 독립변수 X 사이에 상관성이 작아야 이상적임
- 반면에 독립변수 X와 종속변수 Y의 상관성은 커야 함
- 위 두 성질을 만족하는 소수의 독립변수 집합
- 많은 양질의 데이터(결측치와 노이즈가 없는 깨끗한 데이터)

	독립변수	종속변수
	X	Y
데이터		

Long and Thin

## ■ 변수선택(Variable Selection):

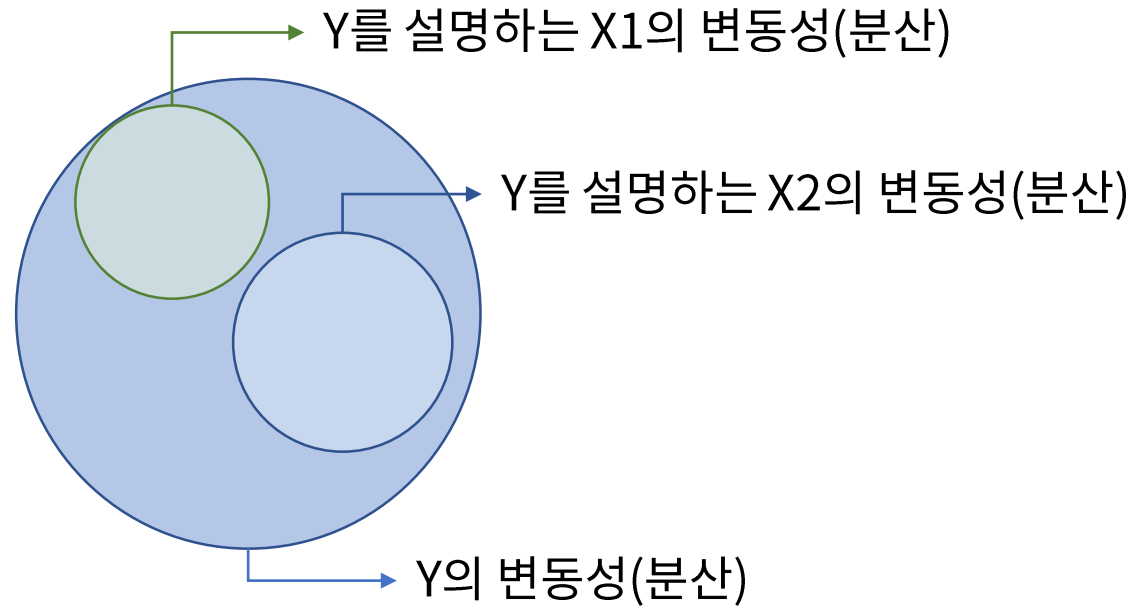
독립변수 X간에는 상관성이 적고(Minimal Redundancy), X와 종속변수 Y간에는 상관성이 큰 (Maximal Relevance) 독립변수만을 추출(e.g. 사람의 키를 예측하는데 있어서 다리길이와 팔길이는 중복성이 존재)

	독립변수	종속변수
	X	Y
데이터		

Short and Fat

# I 회귀계수 축소법

좋은 변수는?



- Y의 변동성을 잘 설명하면서 X들끼리는 상관관계가 없는 변수들이 좋은 변수이다.

# I 회귀계수 축소법

## 회귀계수를 축소하는 이유

- 영향력이 없는 입력 변수의 계수를 0에 가깝게 가져간다면, 모형에 포함되는 입력 변수의 수를 줄일 수 있음
- 입력 변수의 수를 줄이면 크게 세 가지 장점이 있음

- ① 잡음(noise)을 제거해 모형의 정확도를 개선
- ② 모형의 연산 속도가 빨라짐
- ③ 다중공선성의 문제를 완화시켜 모형의 해석 능력을 향상

많은 모형에서 입력 변수들끼리 독립임을 가정하지만, 입력 변수들끼리 상관관계를 가지는 경우가 대부분

입력 변수가 나이, 잔고액, 생년인 경우 나이와 생년은 같은 의미를 갖기 때문에 둘 중 하나를 제거

# I 회귀계수 축소법

## 계수축소법의 종류

- 계수축소법은 기본적으로 다중선형회귀와 유사
- 다중선형회귀에서 잔차를 최소화했다면, 계수축소법에서는 잔차와 회귀계수를 최소화
- 계수축소법에는 크게 3 가지의 방법이 있음: Ridge 회귀, Lasso 회귀, Elastic-Net 회귀
- 아래 식은 다중선형회귀의 SSE이며, 다중선형회귀에서는 SSE가 최소화되는 회귀계수를 추정

$$\text{minimize } SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2$$

- 계수축소법에서는 위 식에 회귀계수를 축소하는 항을 추가

$$\text{minimize } SSE + f(\beta)$$

Part.02  
회귀분석

# | 회귀계수 축소법 (Ridge)

FASTCAMPUS  
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택