회귀분석팀

6팀

권남택
윤주희
진효주
한유진
황유나

INDEX

<변수선택의 장점>

- 다중공선성이 존재할 때

1) 높은 상관관계를 가지는 변수들 중 일부만을
선택하도록 해준다

2) 높은 상관관계를 가지는 변수들의 존재를
정당화 해줄 수 있다

- 다중공선성이 발견되지 않더라도, 변수선택법을 통해
최종모델에 대한 확신을 얻을 수 있다!

<Best Subset Selection의 한계>

<Best Subset Selection (All Possible Regression)>

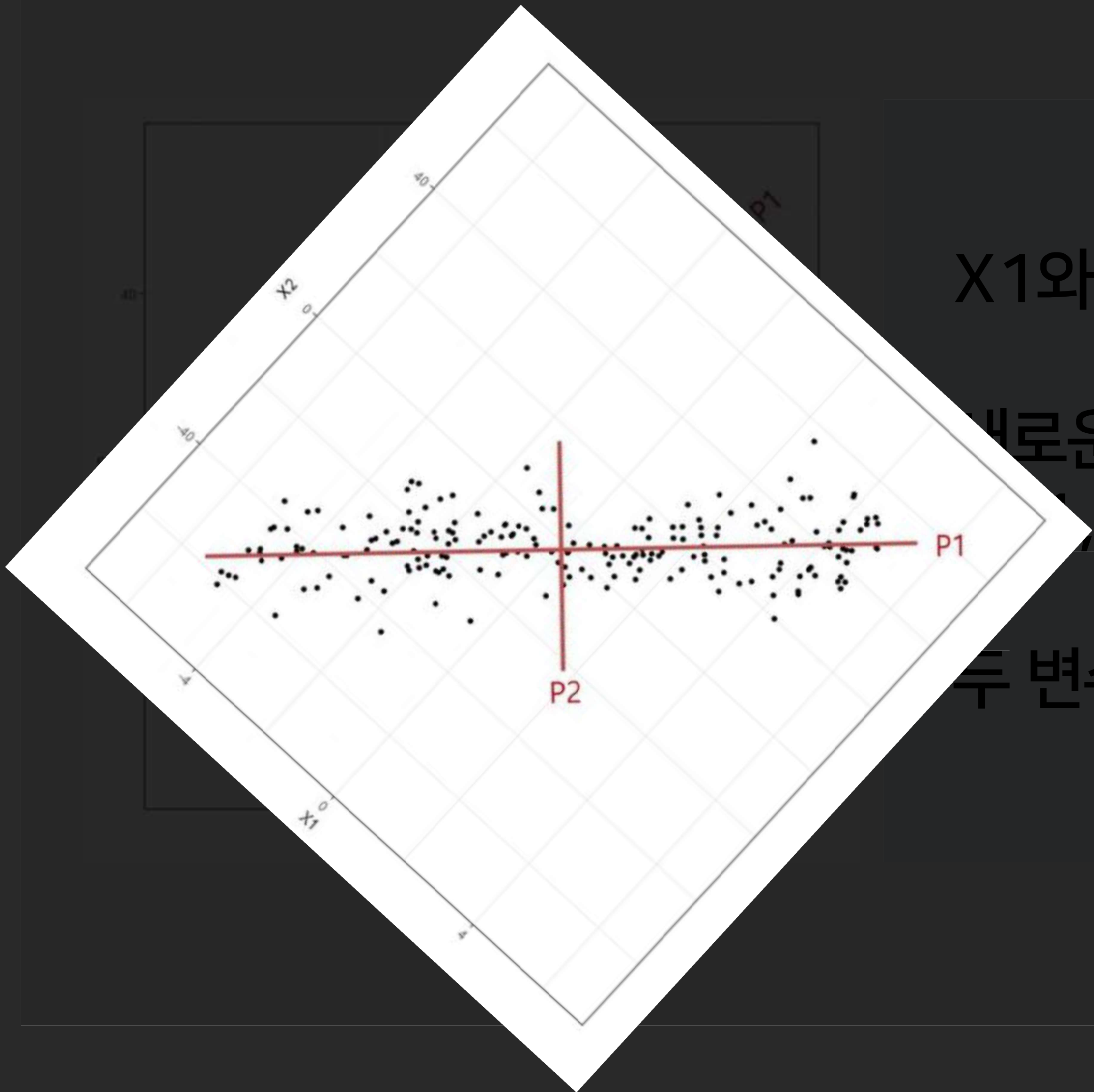
가능한 모든 조합의 모델을 고려하기 때문에…

- 가능한 모든 변수들의 조합을 고려 → 2^p 개의 모형
- Best Mo[illegible]

변수 개수 > 40인 경우 계산이 불가능

관측치가 많다면 계산 비용 증가!

주성분분석(PCA) 주성분회귀(PCR)



이렇게 고래프를 돌려보면

주성분 분석을 통해
상관계수가 0이 되는 것을
한 눈에 알 수 있어요!

두 변수의 상관계수는 0 이 된다.

고차원의 데이터를
상관관계가 없는
저차원의 공간으로 축소!

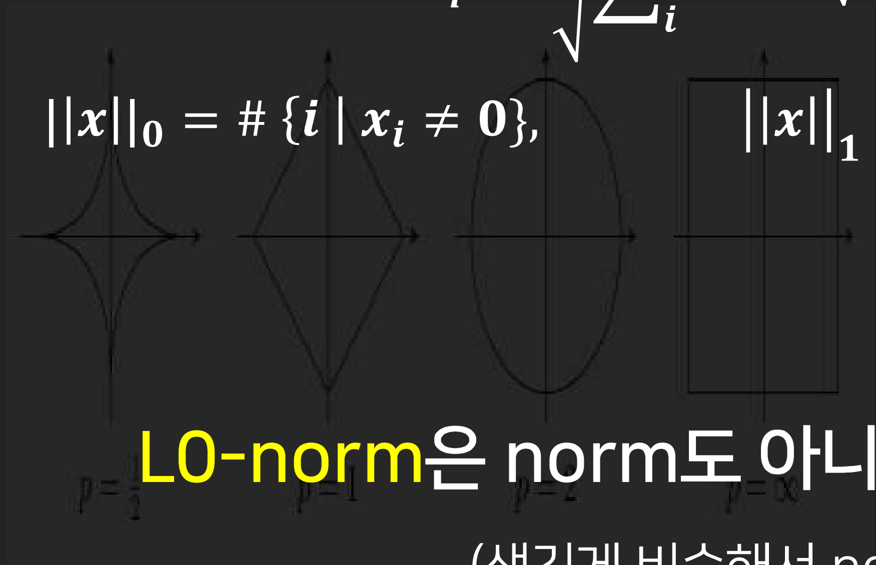
어떤 함수가 컨벡스 함수인가?

Lp-norm이란?

3. 모든 R^n 상의 norm

$$||x||_p = \sqrt[p]{\sum_i x_i{}^p} = \sqrt[p]{x_1{}^p + x_2{}^p + \cdots + x_n{}^p}$$

$$||x||_0 = \#\{i \mid x_i \neq 0\}, \qquad ||x||_1 = \sum_i |x_i|, \qquad ||x||_2^2 = \sum_i x_i{}^2$$



← Lp norm에서 p값에따른 시각화

L0-norm은 norm도 아니고 컨벡스 함수도 아니다!

(생긴게 비슷해서 norm처럼 취급함!)

하지만 nonconvex한 함수라면?

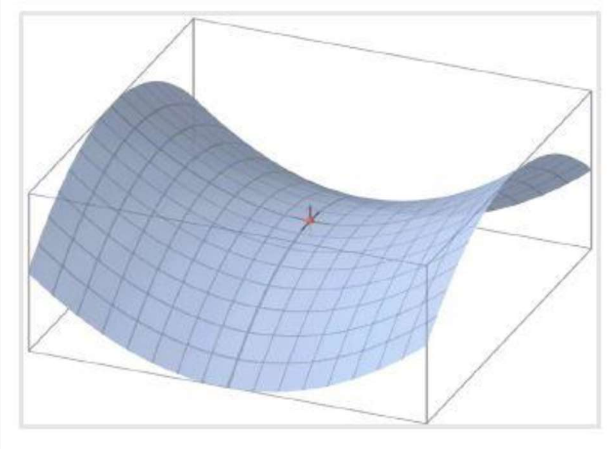
:컨벡스한 형태로 완화 (Convex Relaxation)

Best Subset Selection

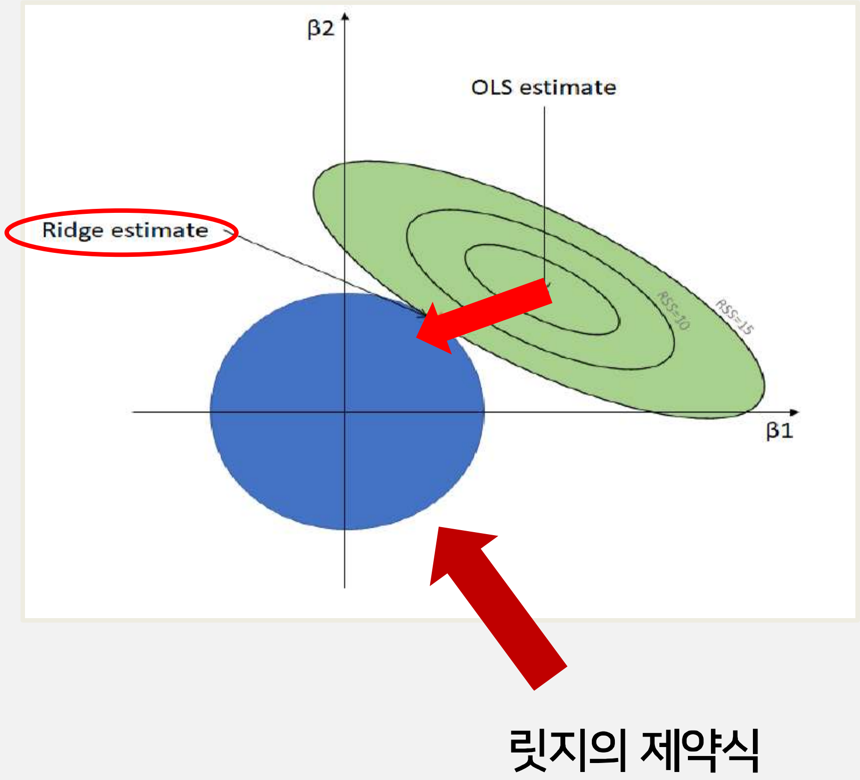
: $min_{\beta} \sum_i (y_i - {x_i}^t \beta)^2$,

$s.t. \quad ||\beta||_0 \leq t \quad (\#\{i \mid x_i \neq 0\} \leq t)$

0이 아닌 베타의 개수



- Primal Problem



- 기존의 LSE보다 RSS는 커짐
- 최소제곱 추정량보다 목적함수의 결과값은 커질 수 있지만, 베타 값에 대한 제약 때문에 베타 값이 커지는 것을 막음
- 기존 LSE의 beta1, beta2보다 작다!

- Dual Problem

$$\hat{\beta}^{ridge} = min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} {\beta_j}^2, \qquad \lambda \geq 0$$

- 현재 목적함수와 제약식이 모두 Convex하기 때문에

  Primal과 Dual의 해는 완전히 같다.

- Dual Problem – Properties of Ridge

$$\hat{\beta}^{ridge} = (X^tX + \lambda I)^{-1}X^ty$$

- 미분이 가능하다
- X^tX가 full rank가 아니어도 unique beta solution이 존재한다

But..

- 릿지 추정량은 개별 베타값을 0에 가깝게 만들지만,

  정확히 0으로 만들지는 않는다.

컨벡스하지 않은 L0-norm을 L1-norm 으로 변형

$$\|\beta\|_0 = \#\{i|\beta_i \neq 0\} \quad \Rightarrow \quad \|\beta\|_1 = \sum_i |\beta_i|$$

L0-norm	L1-norm
0이 아닌 베타의 개수	β_i의 절댓값의 합

컨벡스 완화 원문제 **쌍대문제** 특성

- Dual Problem

$$\hat{\beta}^{Lasso} = min_{\beta} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 \ + \lambda \sum_{j=1}^{p} |\beta_j| , \qquad \lambda \geq 0$$

- 기존의 최소제곱 목적함수도 최소화하면서,
  개별 베타의 절대값의 합도 동시에 작게 만듦

- 목적함수와 제약식이 모두 컨벡스하기 때문에
  원문제와 쌍대문제의 해는 완전히 같다

- **Lasso의 특징**

③ Lasso는 Ridge와 마찬가지로 관측치보다 변수개수가 많은 경우에도 유일한 해를 갖는다.

④ Lasso는 정확히 0이 되는 β가 있어서 변수선택의 효과가 있다.

⑤ Lasso는 비교적 합리적으로 변수선택법을 대체할 수 있는 방법으로 많이 쓰이며 최적화 방법에 의해 유일해에 접근하므로 훨씬 빠르다.