

Shrinkage Methods & Methods Using Derived Input Directions

-The Elements of Static Learning Data Mining, Inference, and Prediction-

장예훈

18.05.28.Mon

Shrinkage Methods

- Shrinkage Methods
 - 모델을 모든 변수로 fitting을 하되, 계수들을 0 으로 constrain 혹은 regularize
 - 추정된 계수들의 Variance를 대폭 줄여줌!!
- Ridge
- LASSO

Ridge Regression

- RSS: OLS(Ordinary Least Square: $y = ax + b$)가 최소화할 식

$$RSS = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2$$

($\beta_0 = \text{bias}$, $\beta_j = \text{weight vector}$)

- LSE(Least Square Error)방식을 통해 추정 → 일반적인 OLS 도출
- Ridge는 RSS를 최소화 + 계수에 **Penalty 부과**

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2$$

Ridge Regression

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2$$

- λ 가 0이면 일반적인 OLS와 동일 \rightarrow unbiased but high variance
- λ 가 무한대로 수렴하면 계수에 Penalty 부과 \rightarrow biased but low variance
- λ 를 tuning parameter로 설정 후, 적절한 λ 를 찾기 위한 Cross Validation을 구하기도 함
- λ 를 크게 둔 Ridge를 통해 생성되는 계수는 0에 근사 \rightarrow 0은 X

LASSO

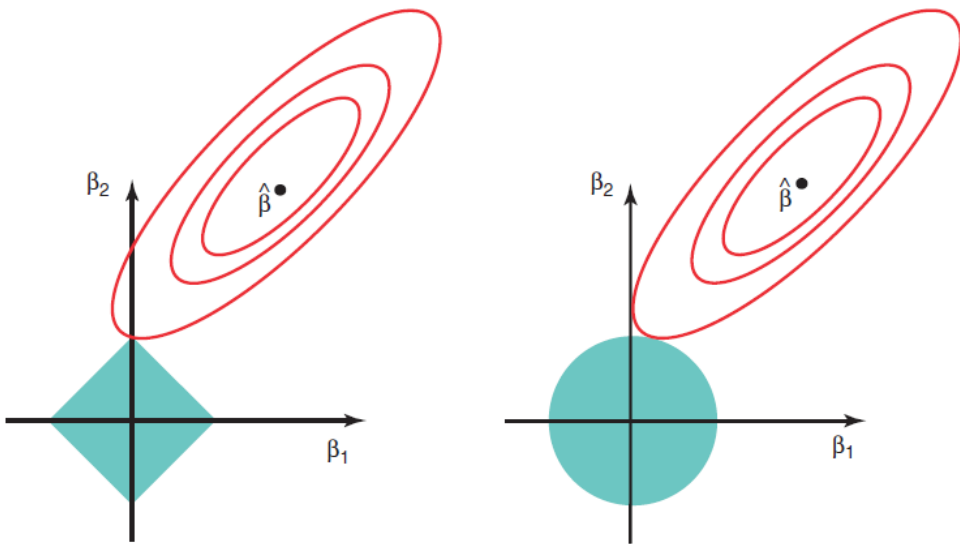
- LASSO (Least Absolute Shrinkage and Selection Operator)

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j|$$

- λ 가 클수록 계수가 0 \rightarrow Ridge와 다르게 식 추정과 더불어 변수선택 가능

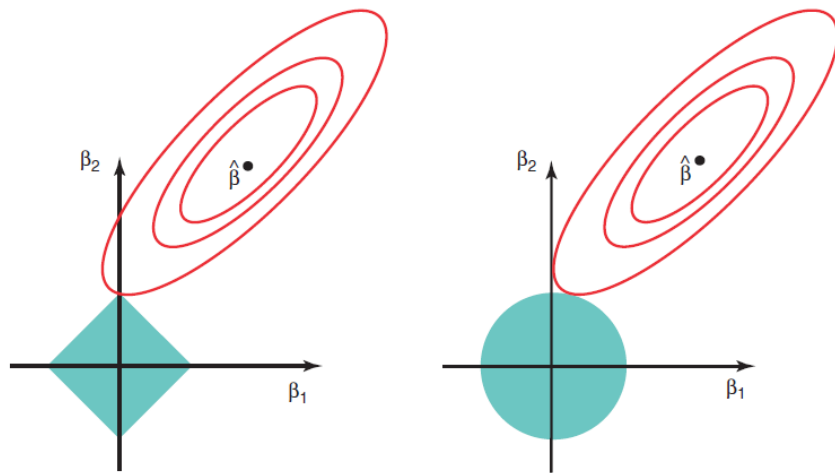
Ridge Regression VS LASSO

- 두 방식의 가장 큰 차이점: 보정 RSS의 최소화를 어떤 방식으로 접근하는 지



왼쪽 : LASSO, 오른쪽: Ridge

Ridge Regression VS LASSO



- 추정식(등고선)과 초록색 도형($\text{norm} \leq s$)이 직선상에서 교점 여부
- 변수선택과 norm의 반비례 관계

LASSO-변수선택을 하지만 norm은 적음 / Ridge-원의 도형, 많은 변수를 설명하려고 노력

Performance

- 독립변수와 종속변수 간의 독립성 유지
→ LASSO > Ridge
- 독립변수와 종속변수 간의 독립성 X
→ LASSO < Ridge

Dimension Reduction Methods

- Dimension Reduction(차원 축소): 변수 자체를 변환하여 모델에 적합한 방법
- 차원 축소
 - 기존의 p 개의 변수 X_p 가 아닌, 기존의 변수들의 linear combinatio으로 만들어진 M 개의 새로운 변수 Z_m 을 생성 ($M < p$)

$$Z_m = \sum_{j=1}^p \phi_{jm} X_j$$

- M 개의 변수로 기존과 동일하게 model fitting

Principal Components Regression

- Z를 이용한 Least Square Error 방정식을 정립할 수 있음
 - 상대적으로 많은 변수를 가진 모델의 variance를 낮출 수 있음
 - 하지만 variance와 상충관계인 bias는 높아짐
 - **주성분 회귀분석의 특징**
- 주성분 회귀분석: Supervised + Unsupervised 혼재
 - PCA(Supervised) + 차원축소(Unsupervised)