Part. 02

회귀분석

| 회귀계수 축소법 (Lasso, ElasticNet)

FASTCAMPUS
ONLINE
머신러닝과 데이터분석 A-Z|

강사. 이경택

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 RSS가 최소화되는 회귀계수를 추정

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

- 계수축소법에서는 위 식에 회귀계수를 축소하는 항을 추가

$$minimize\ SSE + f(\beta)$$

ㅣ회귀계수 축소법

Ridge 회귀

- Ridge 회귀에서는 $f(\beta)$에 회귀계수의 제곱의 합을 대입
- λ는 tuning parameter로 크면 클 수록 보다 많은 회귀계수를 0으로 수렴 (완전히 0이 되지는 않음)

회귀계수의 제곱의 합

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

$$\beta = (X'X + \lambda I)^{-1}X'y$$

다중공선성은 X들간의 강한 선형관계가 있을때 발생

$X'X$의 역행렬을 구할 수가 없음

Ridge는 $X'X$의 역행렬을 구할 수 있도록
강제로 작은 값을 diagonal term에 추가한 것!

I 회귀계수 축소법

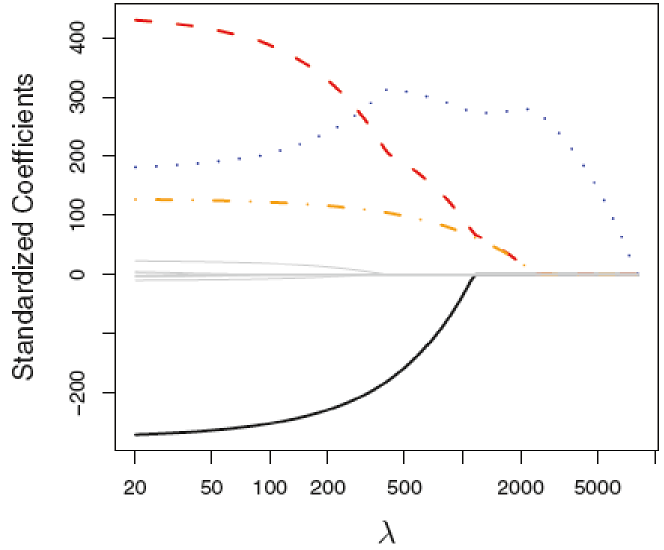
Lasso회귀

- Lasso 회귀에서는 $f(\beta)$에 회귀계수의 절대값의 합을 대입
- λ는 tuning parameter로 크면 클 수록 보다 많은 회귀계수를 0으로 수렴(완전히 0 이 됨)

회귀계수의 절대값의 합

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

- 오른쪽 그림은 λ가 커질수록 입력변수인 Income, LimitRating, Student가 0으로 수렴하는 것을 표현
- 적절한 λ의 값은 데이터마다 달라지며, 5000이넘는 값을 설정한 경우에 모든 입력 변수가 0이 됨
- Lasso의 β은 ridge와 다르게 한번에 구할 수 없음 β



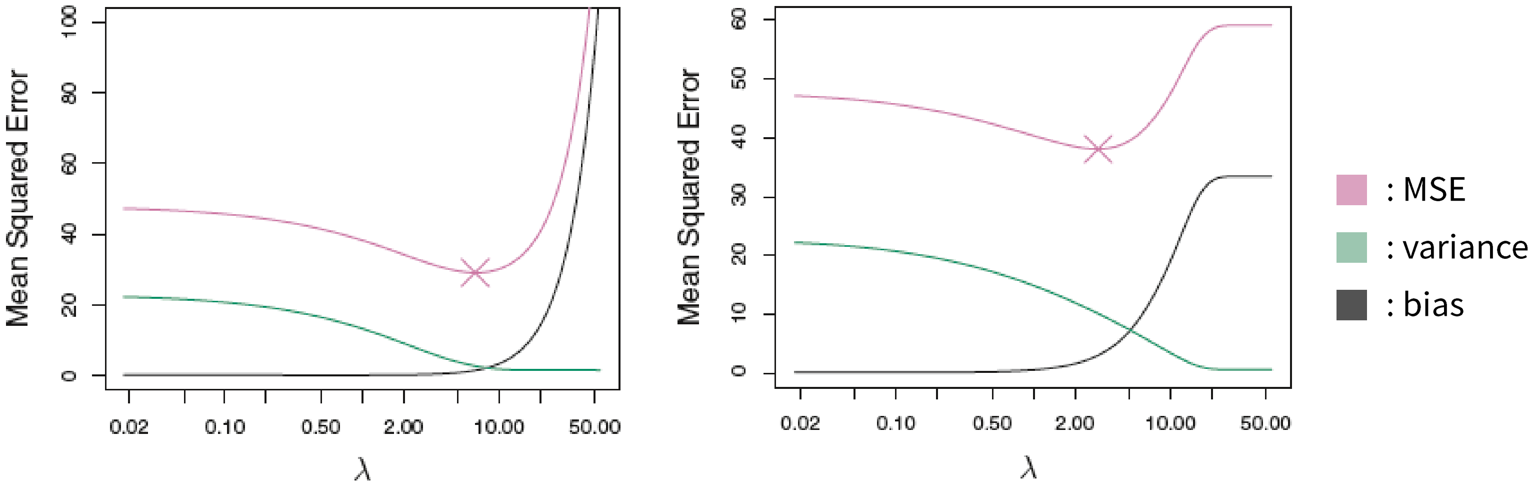
I 회귀계수 축소법

람다(lambda) 값의 설정

- 적절한 람다 값은 다음과 같은 방법으로 설정

람다 값을 변화시켜가며 MSE가 최소일 때의 람다를 탐색

- Ridge 회귀(좌측 그림)와 Lasso(우측 그림)의 람다에 따른 MSE의 변화는 아래 그림과 같음
- 한 모의 데이터에 적용하였으며 Ridge 회귀의 경우 8, Lasso의 경우 4 부근에서 MSE가 최소



James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: Springer.

I 회귀계수 축소법

계수축소법의 최적화 표현

- Ridge 회귀와 Lasso는 다음과 같은 최적화 방식으로 표현이 가능(상단: Ridge 회귀, 하단: Lasso)

$$minimize\left\{\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2\right\}\text{ subject to }\sum_{j=1}^{p}{\beta_j}^2\le s$$

$$minimize\left\{\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2\right\}\text{ subject to }\sum_{j=1}^{p}\left|\beta_j\right|\le s$$

- λ 대신 s를 사용하여 회귀계수의 크기를 제한
- 위 식은 라그랑지안(Lagrangian) 최적화 기법으로 최적 해(최적 회귀계수)를 구할 수 있음

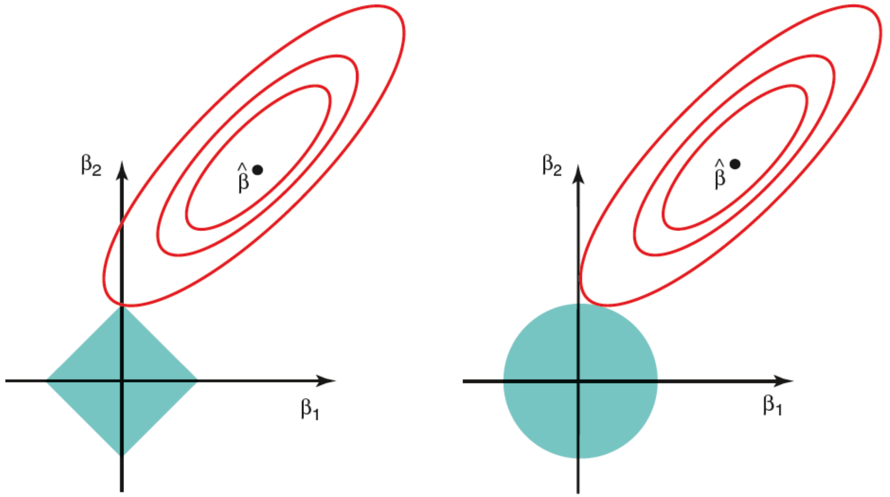
I 회귀계수 축소법

Ridge 회귀와 Lasso의 차이점

- Ridge 회귀와 Lasso의 가장 큰 차이점은 Ridge는 계수를 축소하되 0에 가까운 수로 축소하는 반면, Lasso는 계수를 완전히 0으로 축소함
- Ridge 회귀: 입력 변수들이 전반적으로 비슷한 수준으로 출력 변수에 영향을 미치는 경우에 사용
- Lasso 회귀: 출력 변수에 미치는 입력 변수의 영향력 편차가 큰 경우에 사용

- 초록색 그림: 회귀계수가 가질 수 있는 영역(feasible region)
- 빨간색 원: SSE가 같은 지점을 연결한 그림
- Ridge 회귀와 Lasso 회귀 모두 SSE를 희생하여 계수를 축소하는 방법
- Lasso 회귀의 경우 회귀계수가 0이 될 수 있지만, Ridge 회귀는 불가능

〈Lasso 회귀〉 〈Ridge 회귀〉



James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: Springer.

I 회귀계수 축소법

Elastic-Net 회귀

- ElasticNet 회귀는 Lasso 와 Ridge 회귀의 하이브리드(정규화) 회귀 모델
- Lasso에 적용된 회귀계수의 절대값의 합과 Ridge에 적용된 회귀계수의 제곱의 합을 모두 $f(\beta)$ 에 대입

$$minimize \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda_1\sum_{j=1}^{p}\beta_j^2 + \lambda_2\sum_{j=1}^{p}|\beta_j|$$

- λ_1, λ_2는 Ridge와 Lasso 속성에 대한 강도를 조절
- Lasso의 변수 축소 효과로 sparse model 생성
- Ridge의 정규화 속성으로 변수의 grouping effect 유도 및 Lasso 의 sparsity를 안정화

- Grouping effect : Lasso는 상관관계가 있는 다수의 변수들 중 하나를 무작위로 선택하여 계수를 축소하는 반면, elastic-net은 상관성이 높은 다수의 변수들을 모두 선택(또는 제거)

- 따라서, 다수의 변수 간에 상관관계가 존재할 때 효과적

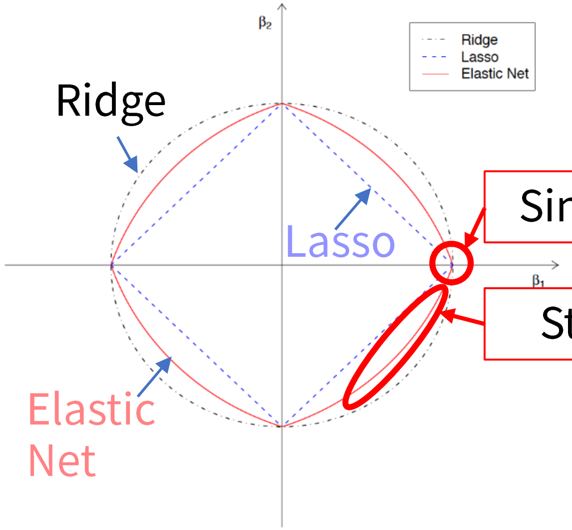
I 회귀계수 축소법

Elastic-Net 회귀 geometry illustration

$$minimize \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda_1 \sum_{j=1}^{p}\beta_j^2 + \lambda_2 \sum_{j=1}^{p}|\beta_j|$$

회귀계수의 절대값의 합 (Lasso)

회귀계수 제곱의 합 (Ridge)



Singularities at the vertexes (necessary for sparsity)

Strict convex edges (necessary for grouping)

Part. 02

회귀분석

Feature selection

FASTCAMPUS
ONLINE
머신러닝과 데이터분석 A-Z

강사. 이경택