회귀분석팀

6팀

조수미
김민지
손재민
박윤아
조웅빈

CONTENTS

1

다중공선성

다중공선성

다중공선성 Multicollinearity

설명변수 X_j 간에 서로 **선형적인 상관관계**가 존재

즉 설명변수가 서로 간의 선형결합으로 표현 가능

↓

설명변수 간 **독립적**이어야 한다는 가정을 위배

회귀의 기본 가정에 관한 내용은 2주차 클린업 내용 참고!

다중공선성

다중공선성 Multicollinearity

설명변수 X_j 간에 서로 **선형적인 상관관계**가 존재

즉 설명변수가 서로 간의 선형결합으로 표현 가능

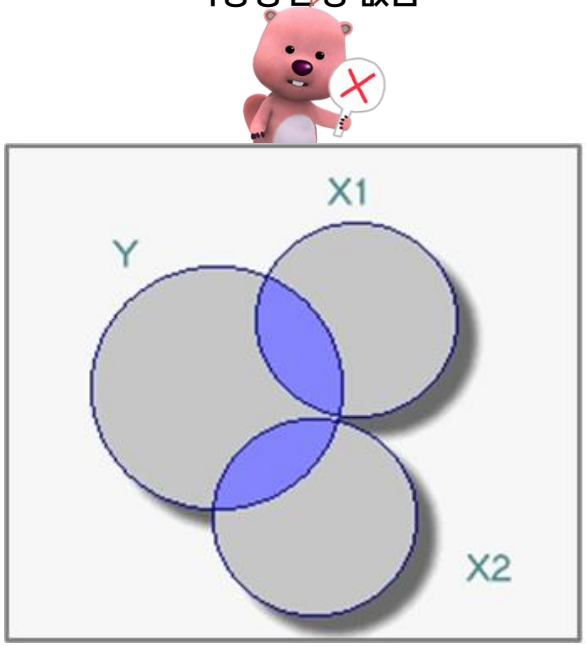
↓

설명변수 간 **독립적**이어야 한다는 가정을 위배

회귀의 기본 가정에 관한 내용은 2주차 클린업 내용 참고!

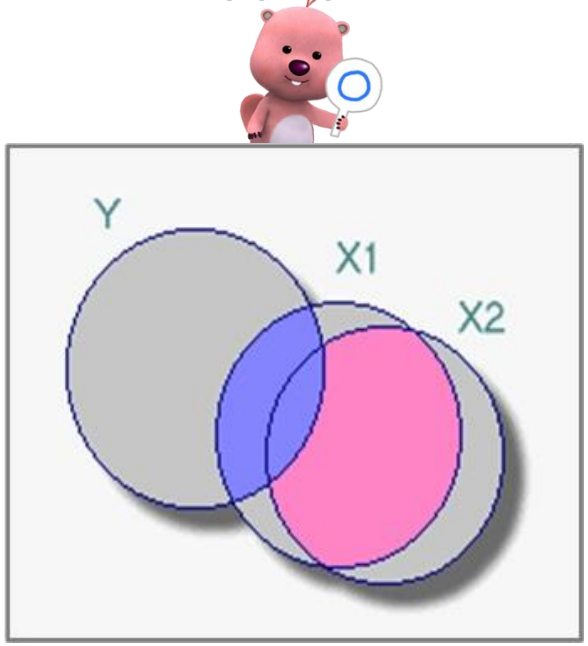
다중공선성

다중공선성 없음



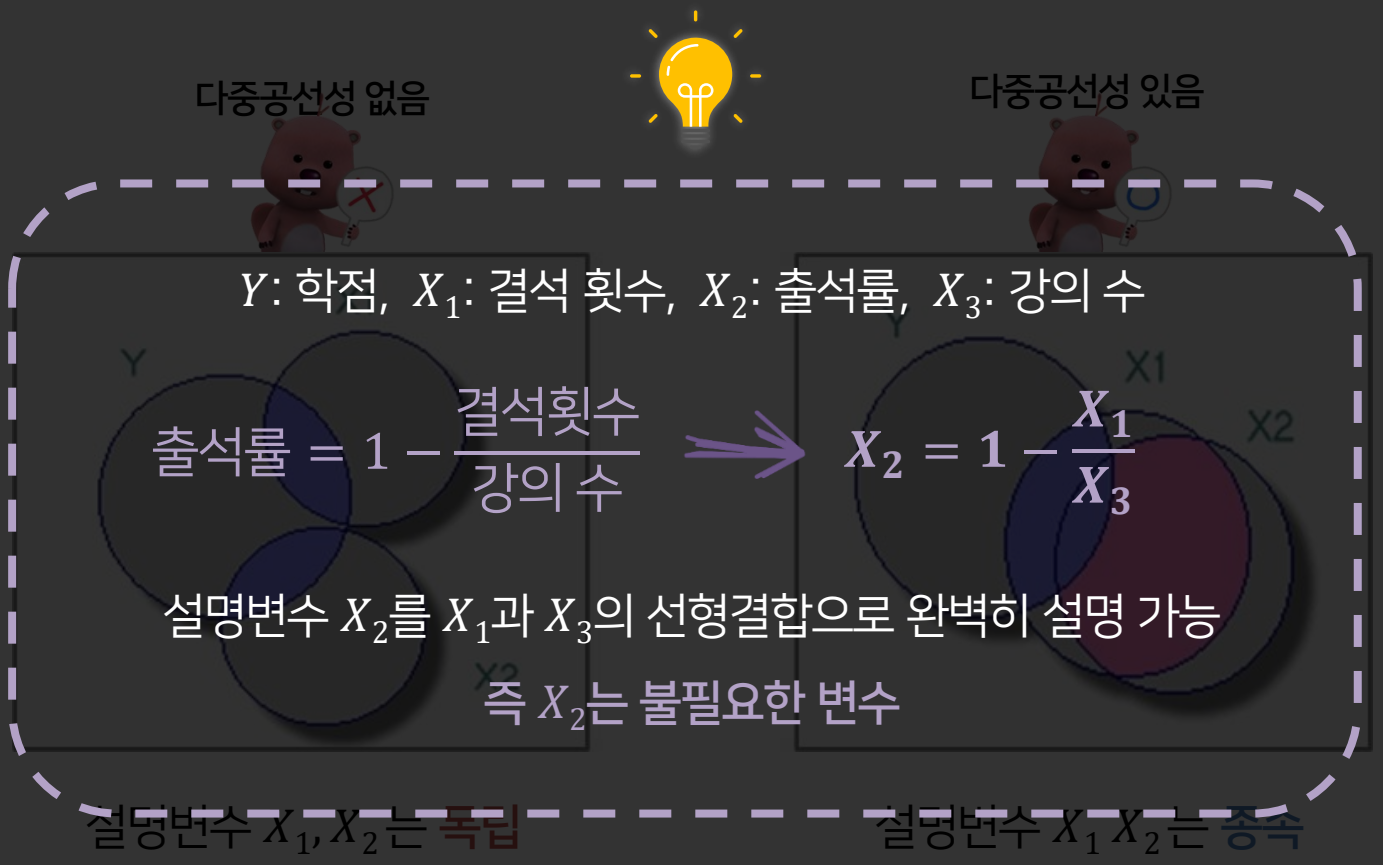
설명변수 X_1, X_2는 독립

다중공선성 있음



설명변수 X_1 X_2는 종속

다중공선성


다중공선성 없음
다중공선성 있음
Y: 학점, X_1: 결석 횟수, X_2: 출석률, X_3: 강의 수
출석률 $= 1 - \frac{\text{결석횟수}}{\text{강의 수}}$
$X_2 = 1 - \frac{X_1}{X_3}$
Y
X1
X2
설명변수 X_2를 X_1과 X_3의 선형결합으로 완벽히 설명 가능
즉 X_2는 불필요한 변수
설명변수 X_1, X_2는 독립
설명변수 X_1 X_2는 종속

다중공선성의 문제점

모델의 문제

모델의 검정 결과를 신뢰할 수 없게 됨

전체 회귀식(F-test)은 유의한데

개별 변수 중 유의한 것이 없는 말도 안되는 상황 발생 가능

회귀계수들의 분산이 커짐에 따라 t검정통계량이 작아지기 때문

다중공선성의 문제점

모델의 문제

모델의 검정 결과를 신뢰할 수 없게 됨

전체 회귀식(F-test)은 유의한데

개별 변수 중 유의한 것이 없는 말도 안되는 상황 발생 가능

회귀계수들의 분산이 커짐에 따라 t검정통계량이 작아지기 때문

다중공선성의 문제점

해석의 문제

개별 회귀계수 β에 관한 해석의 어려움 발생

다중공선성이 발생하면 x_j의 변화가 다른 설명변수를 변화시키므로
나머지 변수가 고정된 상황을 가정하기 힘들어짐

회귀계수 β_j는 설명변수 x_j를 제외한 변수들이 고정되어 있을 때
x_j가 종속변수 Y 에 미치는 영향으로 해석

다중공선성의 문제점

해석의 문제

개별 회귀계수 β에 관한 해석의 어려움 발생

다중공선성이 발생하면 x_j의 변화가 다른 설명변수를 변화시키므로
나머지 변수가 고정된 상황을 가정하기 힘들어짐

회귀계수 β_j는 설명변수 x_j를 제외한 변수들이 고정되어 있을 때
x_j가 종속변수 Y에 미치는 영향으로 해석

다중공선성 진단 | ① 직관적 판단

F-test는 유의했지만 개별 회귀계수들에 대한
t-test에서 귀무가설을 기각하지 못할 경우

상식적으로 유의한 회귀계수가 유의하지 않다고 나올 경우

추정된 회귀계수의 부호가 상식과 다를 경우

다중공선성

다중공선성 진단 | ① 직관적 판단

F-test는 유의했지만 개별 회귀계수들에 대한
t-test에서 귀무가설을 기각하지 못할 경우

다른 변수가 이미 해당 변수의 영향력을 설명하고 있기 때문에
상식적으로 유의한 회귀계수가 유의하지 않다고 나올 경우

다중공선성이 발생한다고 판단

추정된 회귀계수의 부호가 상식과 다를 경우

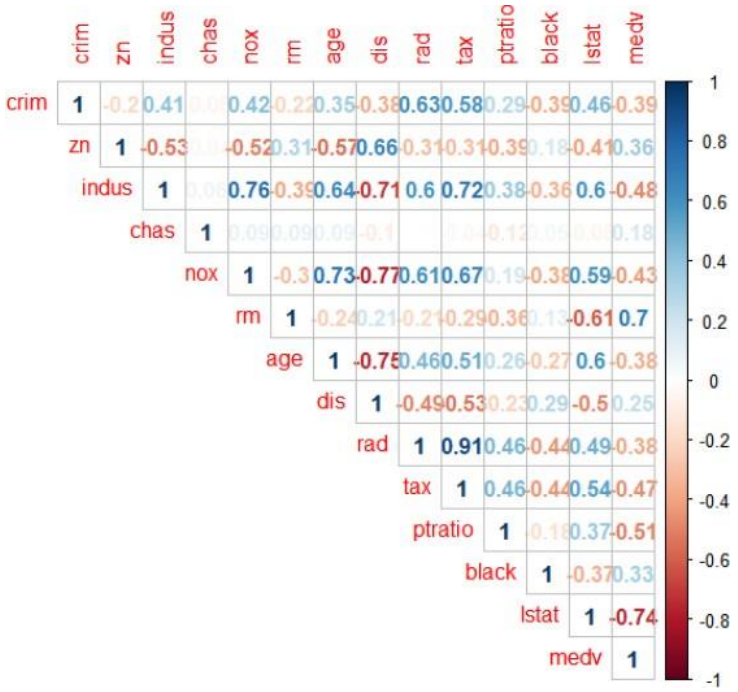
다중공선성

다중공선성 진단 | ② 상관계수 플랏

상관계수 플랏 Correlation Plot

상관계수 플랏을 통해 변수들 간 선형관계를 확인 가능

R의 corrplot 패키지 이용

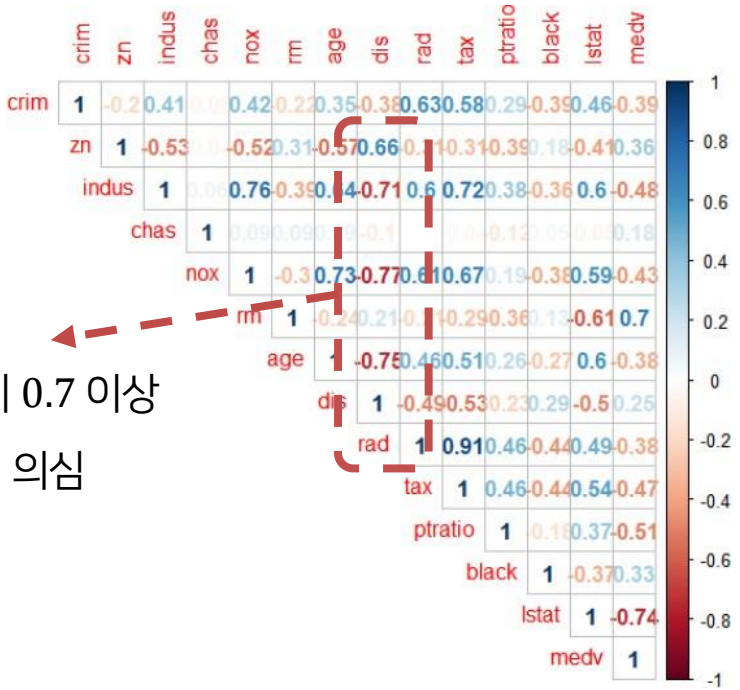


다중공선성 진단 | ② 상관계수 플랏

상관계수 플랏 Correlation Plot

상관계수 플랏을 통해 변수들 간 선형관계를 확인 가능

R의 corrplot 패키지 이용



계수의 절댓값이 0.7 이상

다중공선성 의심

다중공선성 진단 | ③ VIF

VIF 분산팽창인자

$$VIF_j = \frac{1}{1 - R_j^2}, j = 1, \dots, p$$

R_j^2 : $\mathbf{x_j = \gamma_1 x_1 + \cdots + \gamma_{j-1} x_{j-1} + \gamma_{j+1} x_{j+1} + \cdots + \gamma_p x_p}$

$\mathbf{x_j}$ 를 나머지 설명변수에 대해 적합했을 때의 결정계수

R_j^2가 크다면, 설명변수 $\mathbf{x_j}$가

나머지 변수들의 선형결합으로 충분히 표현됨을 의미

다중공선성 진단 | ③ VIF

VIF 분산팽창인자

$$VIF_j = \frac{1}{1 - R_j^2}, j = 1, \ldots, p$$

$$R_j^2 : \quad \mathbf{x_j} = \boldsymbol{\gamma_1 x_1} + \cdots + \boldsymbol{\gamma_{j-1} x_{j-1}} + \boldsymbol{\gamma_{j+1} x_{j+1}} + \cdots + \boldsymbol{\gamma_p x_p}$$

다중선형회귀모델을 적합했을 때의 결정계수

일반적으로 VIF_j 값이 10이상인 경우 $(R_j^2 \geq 0.9)$

심각한 다중공선성이 존재한다고 판단 가능

다중공선성 진단 | ③ VIF

VIF 분산팽창인자

$$VIF_j = \frac{1}{1 - R_j^2}, j = 1, \ldots, p$$

$$R_j^2 : \quad \mathbf{x_j} = \boldsymbol{\gamma_1 x_1} + \cdots + \boldsymbol{\gamma_{j-1} x_{j-1}} + \boldsymbol{\gamma_{j+1} x_{j+1}} + \cdots + \boldsymbol{\gamma_p x_p}$$

다중선형회귀모델을 적합했을 때의 결정계수

반대로, 다중공선성이 적을수록

VIF값은 1에 가까워짐 ($VIF \rightarrow 1 \; as \; R_j^2 \rightarrow 0$)

다중공선성 해결

Variable Selection

Normalization

Dimension Reduction

Filtering

다중공선성 해결

Variable Selection

Normalization

Dimension Reduction

Filtering

2

변수선택법

변수선택법

변수선택법 variable Selection

수 많은 변수들 중 **적절한 변수 조합**을 찾아내는 방법

서로 상관이 있는 독립변수들을 일부 제거하여 **다중공선성을 해결**

변수선택법을 통해 다중공선성을 완벽하게 제거할 수 없음

그러나 **최종 모델에 대한 확신**을 얻을 수 있음

변수선택법

변수선택법 variable Selection

수 많은 변수들 중 **적절한 변수 조합**을 찾아내는 방법

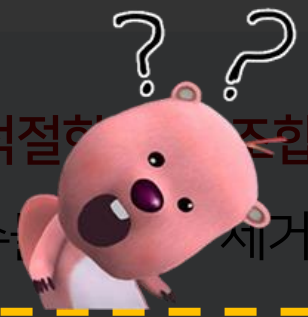
서로 상관이 있는 독립변수들을 일부 제거하여 **다중공선성을 해결**

변수선택법을 통해 다중공선성을 완벽하게 제거할 수 없음

그러나 **최종 모델에 대한 확신**을 얻을 수 있음

변수선택법

변수선택법 variable Selection



수 많은 변수들 중 **적절한** 조합을 찾아내는 방법

서로 상관이 있는 독립변수들 제거하여 **다중공선성을 해결**

어떤 지표를 기준으로 변수를 선택해야 할까?

변수선택법을 통해 다중공선성을 완벽하게 제거할 수 없음

그러나 **최종 모델에 대한 확신**을 얻을 수 있음

변수선택지표 | ① Partial F-test

Partial F-test

유의하지 않은 변수들을 없애는 방식으로 변수 선택

Full Model(FM)과 Reduced Model(RM)이
서로 내포 관계에 있어야 한다는 단점이 있음
즉 RM에 있는 모든 변수가 FM에 있어야 함

변수선택지표 | ① Partial F-test

Partial F-test

유의하지 않은 변수들을 없애는 방식으로 변수 선택

Full Model(FM)과 Reduced Model(RM)이

서로 내포 관계에 있어야 한다는 단점이 있음

즉 RM에 있는 모든 변수가 FM에 있어야 함

변수선택지표 | ① Partial F-test

Partial F-test

비교하는 두 모델이 Nested되어 있을 때만 사용 가능

유의하지 않은 변수들을 없애는 방식으로 변수 선택

	Model A	Model B
Nested O	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
Nested X	$y = \beta_0 + \beta_1 x_1 + \beta_4 x_4$	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

변수선택지표 | ① Partial F-test

Partial F-test

비교하는 두 모델이 Nested되어 있을 때만 사용 가능

유의하지 않은 변수들을 없애는 방식으로 변수 선택

	Model A	Model B
Nested O	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
Nested X	$y = \beta_0 + \beta_1 x_1 + \beta_4 x_4$	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

Model A ⊂ Model B 처럼 변수들 집합의 포함관계 성립

변수선택지표 | ① Partial F-test

Partial F-test

비교하는 두 모델이 Nested되어 있을 때만 사용 가능

유의하지 않은 변수들을 제거해는 방식으로 변수 선택

포함관계와 무관하게 일반적인 상황에서

모델 간의 비교를 가능하게 해주는 기준이 필요!

	Model A	Model B
Nested O	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
Nested X	$y = \beta_0 + \beta_1 x_1 + \beta_4 x_4$	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

Model A $\subset$ Model B 처럼 변수들 집합의 포함관계 성립

변수선택지표 | ② 수정결정계수

수정결정계수 adjusted R-squared

설명력을 담당하는 **결정계수**와 **변수 개수 패널티**가

수정결정계수를 구하는 식에 들어감

$$R^2_{adj} = 1 - \frac{SSE/(n-p-1)}{SST/(n-1)}$$

변수선택지표 | ② 수정결정계수

수정결정계수 adjusted R-squared

설명력을 담당하는 **결정계수**와 **변수 개수 패널티**가
수정결정계수를 구하는 식에 들어감

$$R^2_{adj} = 1 - \frac{SSE/(n-p-1)}{SST/(n-1)}$$

변수선택지표 | ② 수정결정계수

수정결정계수 adjusted R-squared

설명력을 담당하는 **결정계수**와 **변수 개수 패널티**가

수정결정계수를 구하는 식에 들어감

변수선택법의 핵심은

"적은 변수 안에서 데이터를 제일 잘 설명해야 한다"는 것

변수가 너무 적으면 간결하더라도 예측력은 떨어짐

변수가 너무 많으면 과적합될 가능성이 높음

이러한 trade-off를 고려한

$$R^2_{adj} = 1 - \frac{SSE/(n-p-1)}{SST/(n-1)}$$

AIC & BIC

변수선택지표

AIC Akaike Information Criterion

$$AIC = -2\log(Likelihood) + 2k$$

변수선택지표

AIC Akaike Information Criterion

$$AIC = -2\log(Likelihood) + 2k$$

k는 모델의 모수 개수로, **변수 개수에 따른 패널티**를 부과한 것

변수선택지표

AIC Akaike Information Criterion

$$AIC = -2\log(Likelihood) + 2k$$

$Likelihood$는 값이 커질수록 모델이 데이터를 잘 설명함

$likelihood$가 커지면 AIC는 작아지는데

AIC가 낮을수록 더 좋은 모형

변수선택지표

BIC Bayesian Information Criterion

$$BIC = -2\log(Likelihood) + klog(n)$$

변수선택지표

BIC Bayesian Information Criterion

$$BIC = -2\log(Likelihood) + klog(n)$$

AIC와 다르게 **데이터의 개수를 모수의 개수에 곱함**으로써

AIC보다 더 큰 패널티를 부과

BIC 역시 낮을수록 더 좋은 모형

변수선택지표

BIC Bayesian Information Criterion

$$BIC = -2\log(Likelihood) + klog(n)$$

BIC는 **AIC**보다 변수 증가에 더 민감하므로

변수의 개수가 작은 것이 우선순위라면

AIC보다 BIC를 참고하는 게 좋음

AIC와 다르게 데이터의 개수를 모수의 개수에 곱함으로써

AIC보다 더 큰 패널티를 부과

BIC 역시 낮을수록 더 좋은 모형

변수선택지표

BIC Bayesian Information Criterion

$$BIC = -2\log(Likelihood) + k\log(n)$$

고차원 데이터에서는 정확성이 떨어질 수 있고,

AIC와 BIC 모두 각각 문제 발생

따라서 둘을 종합적으로 고려해서 모형을 선택해야 함

AIC보다 더 큰 패널티를 부과

BIC 역시 낮을수록 더 좋은 모형

변수선택법

Best
Subset Selection

Forward
Selection

Backward
Selection

Stepwise
Selection

변수선택법 | ① Best Subset Selection

Best Subset Selection

가능한 모든 변수들의 조합을 다 고려하는 방법

All Possible Regression

↓

변수의 개수가 p개라면, 2^p개의 모형을 모두 적합하고 비교

모든 조합을 다 고려해서 결과를 내기 때문에

Best Model에 대한 더 신뢰할 수 있는 결과를 산출

변수선택법 | ① Best Subset Selection

Best Subset Selection

가능한 모든 변수들의 조합을 다 고려하는 방법

All Possible Regression

변수의 개수가 p개라면, **2^p개의 모형을 모두 적합하고 비교**

모든 조합을 다 고려해서 결과를 내기 때문에

Best Model에 대한 더 **신뢰할 수 있는 결과를 산출**

변수선택법 | ① Best Subset Selection

Best Subset Selection

가능한 모든 변수들의 조합을 다 고려하는 방법

All Possible Regression

Best Subset Selection Algorithm

1) $M_1, \cdots, M_p$개의 모형 적합

2) $(M_1 \sim M_p)$ p개의 모형 중 AIC 또는 BIC가 가장 작은 모형 선택

3) 만약 AIC/BIC가 가장 작은 모형이 서로 다를 경우 다른 근거에 의해 선택

$M_k (k = 1,2, \ldots, p)$란 변수의 개수를 k개로 적합했을 때 적합한 회귀식 중 MSE가 제일 작은 식

변수선택법 | ① Best Subset Selection

Best Subset Selection

가능한 모든 변수들의 조합을 다 고려하는 방법

All Possible Regression

Best Subset Selection Algorithm

1) $M_1, \cdots, M_p$개의 모형 적합

2) $(M_1 \sim M_p)$ p개의 모형 중 AIC 또는 BIC가 가장 작은 모형 선택

3) 만약 AIC/BIC가 가장 작은 모형이 서로 다를 경우 다른 근거에 의해 선택

$M_k(k = 1,2, \ldots, p)$란 변수의 개수를 k개로 적합했을 때 적합한 회귀식 중 MSE가 제일 작은 식

변수선택법 | ① Best Subset Selection

Best Subset Selection

가능한 모든 변수들의 조합을 다 고려하는 방법

All Possible Regression

Best Subset Selection Algorithm

1) $M_1, \cdots, M_p$개의 모형 적합

2) $(M_1 \sim M_p)$ p개의 모형 중 AIC 또는 BIC가 가장 작은 모형 선택

3) 만약 AIC/BIC가 가장 작은 모형이 서로 다를 경우 다른 근거에 의해 선택

$M_k (k = 1,2, \ldots, p)$란 변수의 개수를 k개로 적합했을 때 적합한 회귀식 중 MSE가 제일 작은 식

변수선택법 | ① Best Subset Selection

장점

모든 조합을 다 고려해서
결과를 내기 때문에
Best Model에 대한
더 신뢰할 수 있는 결과를 산출

단점

1) $p > 40$ 인 경우에는
사용할 수 없음
2) 모든 가능성을
직접 찾기 때문에 계산 비용이 큼

변수선택법 | ② 전진선택법

전진선택법 Forward Selection

Null Model($y = \beta_0$)에서 시작해 변수를 하나씩 추가하는 방법

Forward Selection Algorithm

1) Null Model에서 시작해 X_1부터 X_p까지의 변수들 중
AIC와 BIC를 낮추는 변수를 선택해 추가

2) 만약 X_1이 선택되면 $y = \beta_0 + \beta_1 x_1$의 식에서
X_2부터 X_p까지의 변수들 중 AIC와 BIC를 낮추는 변수 추가

3) 위 과정을 반복하며 AIC와 BIC가 더 이상 낮아지지 않으면 중단

변수선택법 | ② 전진선택법

전진선택법 Forward Selection

Null Model($y = \beta_0$)에서 시작해 변수를 하나씩 추가하는 방법

Forward Selection Algorithm

1) Null Model에서 시작해 X_1부터 X_p까지의 변수들 중 AIC와 BIC를 낮추는 변수를 선택해 추가

2) 만약 X_1이 선택되면 $y = \beta_0 + \beta_1 x_1$의 식에서 X_2부터 X_p까지의 변수들 중 AIC와 BIC를 낮추는 변수 추가

3) 위 과정을 반복하며 AIC와 BIC가 더 이상 낮아지지 않으면 중단

변수선택법 | ② 전진선택법

전진선택법 Forward Selection

Null Model($y = \beta_0$)에서 시작해 변수를 하나씩 추가하는 방법

Forward Selection Algorithm

1) Null Model에서 시작해 X_1부터 X_p까지의 변수들 중
AIC와 BIC를 낮추는 변수를 선택해 추가

2) 만약 X_1이 선택되면 $y = \beta_0 + \beta_1 x_1$의 식에서
X_2부터 X_p까지의 변수들 중 AIC와 BIC를 낮추는 변수 추가

3) 위 과정을 반복하며 AIC와 BIC가 더 이상 낮아지지 않으면 중단

변수선택법 | ② 전진선택법

전진선택법 Forward Selection

Null Model($y = \beta_0$)에서 시작해 변수를 하나씩 추가하는 방법

Forward Selection Algorithm

1) Null Model에서 시작해 X_1부터 X_p까지의 변수들 중
AIC와 BIC를 낮추는 변수를 선택해 추가

2) 만약 X_1이 선택되면 $y = \beta_0 + \beta_1 x_1$의 식에서
X_2부터 X_p까지의 변수들 중 AIC와 BIC를 낮추는 변수 추가

3) 위 과정을 반복하며 AIC와 BIC가 더 이상 낮아지지 않으면 중단

변수선택법 | ② 전진선택법

장점

1) Best Subset Selection에 비해 계산이 매우 빠름

2) 변수의 개수가 관측치보다 많은 경우에도 사용 가능

단점

선택된 모형이 최적의 모형이라고 할 수 없음 (가능한 모든 변수 조합을 고려X)

변수선택법 | ③ 후진제거법

후진제거법 Backward Elimination

Forward selection의 반대

Full Model $(y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$ 에서 시작해
변수를 하나씩 제거하는 방법

Backward Selection Algorithm

1) Full Model에서 시작해 X_1부터 X_p까지의 변수 중
가장 AIC와 BIC를 낮추는 변수를 선택해 제거

2) 위 과정을 반복하며 AIC와 BIC가 더 이상 낮아지지 않으면 중단

변수선택법 | ③ 후진제거법

후진제거법 Backward Elimination

Forward selection의 반대

Full Model $(y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$ 에서 시작해

변수를 하나씩 제거하는 방법

Backward Selection Algorithm

1) Full Model에서 시작해 X_1부터 X_p까지의 변수 중

가장 AIC와 BIC를 낮추는 변수를 선택해 제거

2) 위 과정을 반복하며 AIC와 BIC가 더 이상 낮아지지 않으면 중단

변수선택법 | ③ 후진제거법

후진제거법 Backward Elimination

Forward selection의 반대

Full Model $(y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$ 에서 시작해

변수를 하나씩 제거하는 방법

Backward Selection Algorithm

1) Full Model에서 시작해 X_1부터 X_p까지의 변수 중

가장 AIC와 BIC를 낮추는 변수를 선택해 제거

2) 위 과정을 반복하며 AIC와 BIC가 더 이상 낮아지지 않으면 중단

변수선택법 | ③ 후진제거법

장점

Best Subset Selection에
비해 계산이 매우 빠름

단점

1) $p > 40$ 인 경우에는
사용할 수 없음
2) 선택된 모형이 최적의
모형이라고 할 수 없음

변수선택법 | ④ 단계적 선택법

단계적 선택법 Stepwise Selection

Forward Selection과 Backward Elimination 과정을 섞은 방법

Stepwise Selection Algorithm

1) Forward selection 과정을 통해 가장 유의한 변수들을 모델에 추가

2) 나머지 변수들에 대해 Backward Elimination을 적용

3) 제거된 변수는 다시 모형에 포함되지 않으며,
모형에 유의하지 않은 설명변수가 존재하지 않을 때까지 1)번과 2)번 과정을 반복

변수선택법 | ④ 단계적 선택법

단계적 선택법 Stepwise Selection

Forward Selection과 Backward Elimination 과정을 섞은 방법

Stepwise Selection Algorithm

1) Forward selection 과정을 통해 가장 유의한 변수들을 모델에 추가

2) 나머지 변수들에 대해 Backward Elimination을 적용

3) 제거된 변수는 다시 모형에 포함되지 않으며,
모형에 유의하지 않은 설명변수가 존재하지 않을 때까지 1)번과 2)번 과정을 반복

변수선택법 | ④ 단계적 선택법

단계적 선택법 Stepwise Selection

Forward Selection과 Backward Elimination 과정을 섞은 방법

Stepwise Selection Algorithm

1) Forward selection 과정을 통해 가장 유의한 변수들을 모델에 추가

2) 나머지 변수들에 대해 Backward Elimination을 적용

3) 제거된 변수는 다시 모형에 포함되지 않으며,
모형에 유의하지 않은 설명변수가 존재하지 않을 때까지 1)번과 2)번 과정을 반복

변수선택법 | ④ 단계적 선택법

단계적 선택법 Stepwise Selection

Forward Selection과 Backward Elimination 과정을 섞은 방법

Stepwise Selection Algorithm

1) Forward selection 과정을 통해 가장 유의한 변수들을 모델에 추가

2) 나머지 변수들에 대해 Backward Elimination을 적용

3) 제거된 변수는 다시 모형에 포함되지 않으며,
모형에 유의하지 않은 설명변수가 존재하지 않을 때까지 1)번과 2)번 과정을 반복

변수선택법 | ④ 단계적 선택법

장점

Best Subset Selection에

비해 계산이 매우 빠름

단점

선택된 모형이

최적의 모형이라고 할 수 없음

변수선택법 | ④ 단계적 선택법

Best Subset Selection을 제외한

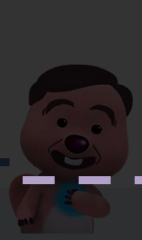
장점

Best Subset Selection에
비해 계산이 매우 빠름

단점

나머지 방법들이 계산이 빠르다고 했지만 이는 상대적인 것
선택된 모형이
위 방법 모두 계산비용이 굉장히 많이 소모됨
최적의 모형이라고 할 수 없음

변수선택법 | ④ 단계적 선택법



Forward Selection, Backward Elmination, Stepwise Selection은

장점 모든 경우의 수를 고려하지 않기 때문에 **단점**

기계적으로 변수를 제거하는 것은 매우 위험

Best Subset Selection에 비해 계산이 매우 빠름

선택된 모형이 최적의 모형이라고 할 수 없음

"정규화"

3

정규화

정규화

정규화 Regularization

회귀계수가 가질 수 있는 값에 **제약 조건**을 부여함으로써

계수들을 작게 만들거나 0으로 만드는 방법

다중공선성은 OLS 추정량의 분산을 크게 증가시킴

정규화는 OLS 추정량의 **불편성 포기**

But **분산을 줄이는 효과**가 있음

정규화

정규화 Regularization

회귀계수가 가질 수 있는 값에 **제약 조건**을 부여함으로써

계수들을 작게 만들거나 0으로 만드는 방법

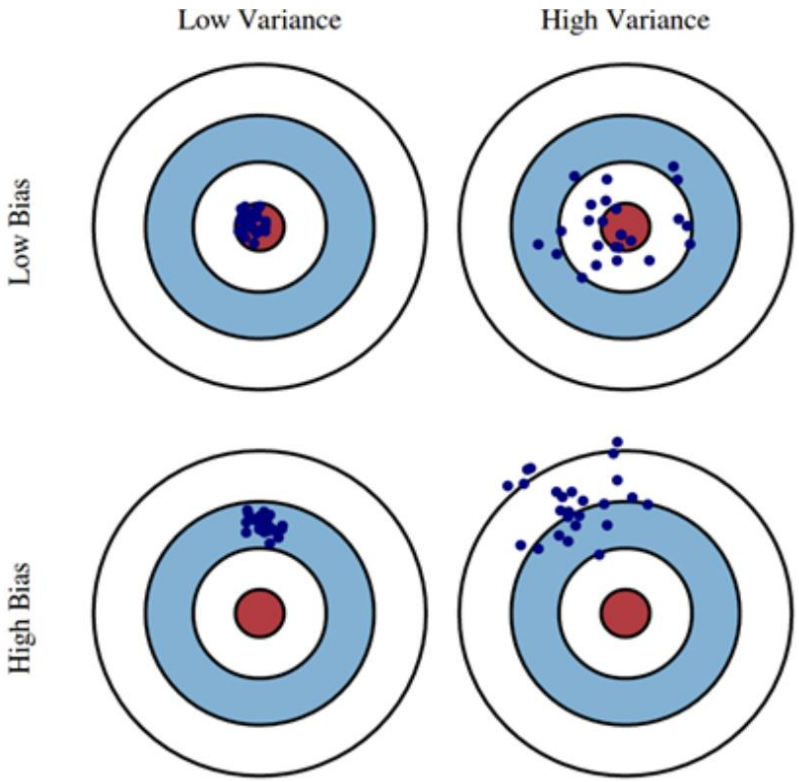
다중공선성은 OLS 추정량의 분산을 크게 증가시킴

정규화는 OLS 추정량의 **불편성 포기**

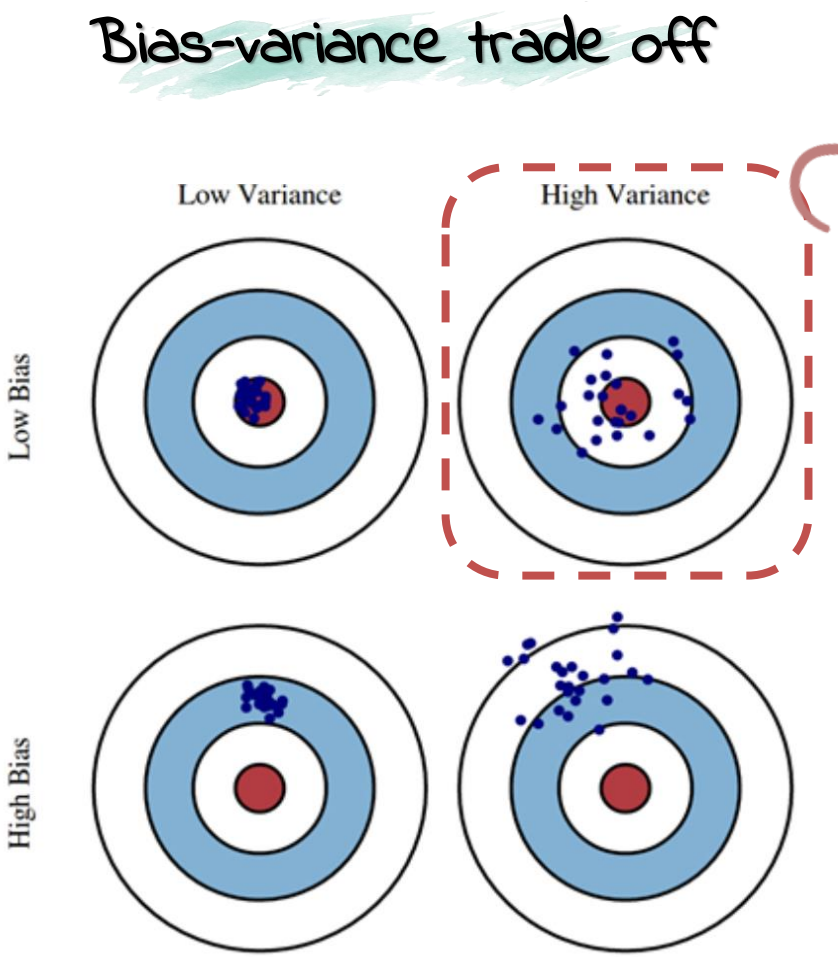
But **분산을 줄이는 효과**가 있음

정규화

Bias-variance trade off



정규화



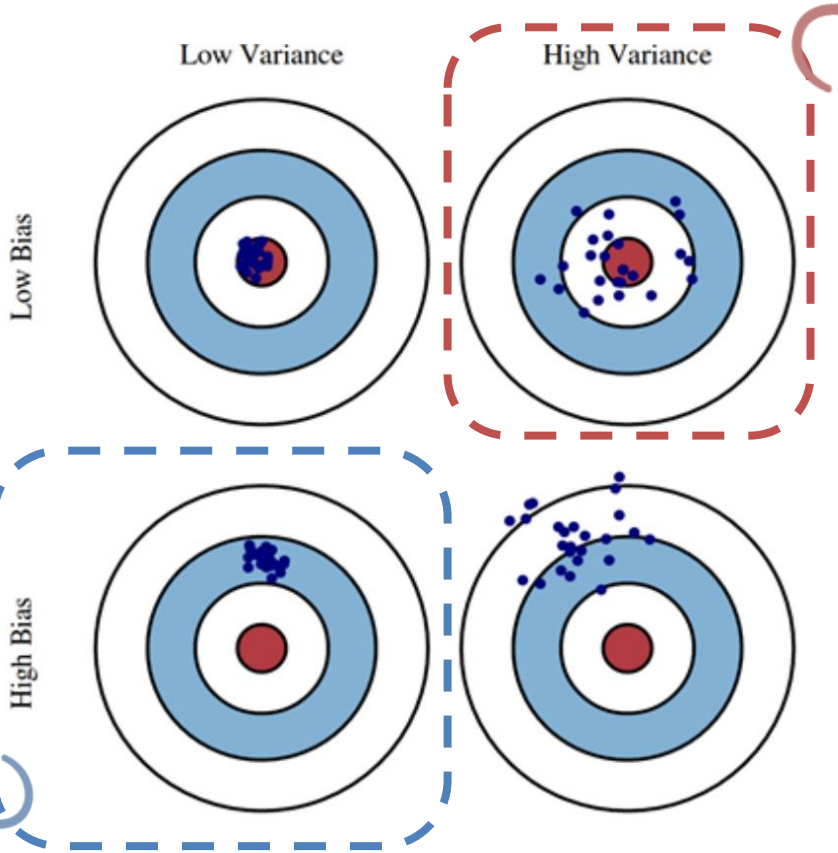
정규화 이전

다중공선성이 존재하는 경우

회귀계수들의 OLS 추정량

정규화

Bias-variance trade off



정규화 이전

다중공선성이 존재하는 경우

회귀계수들의 OLS 추정량

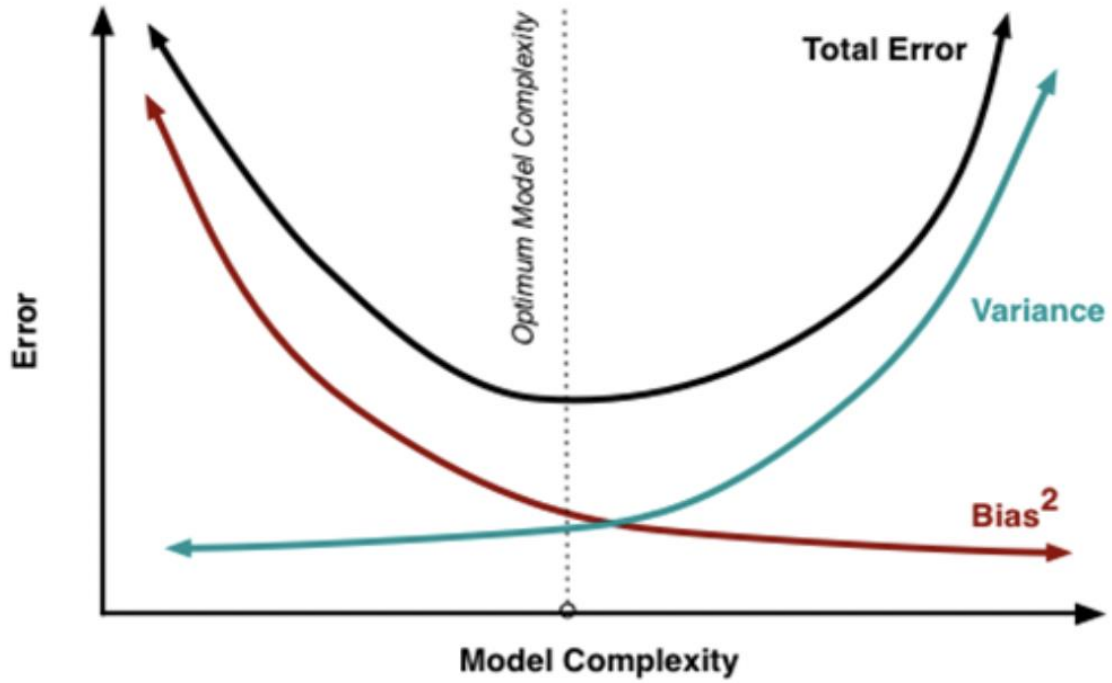
정규화 이후

정규화로 불편성을 포기하고

분산을 줄이는 효과

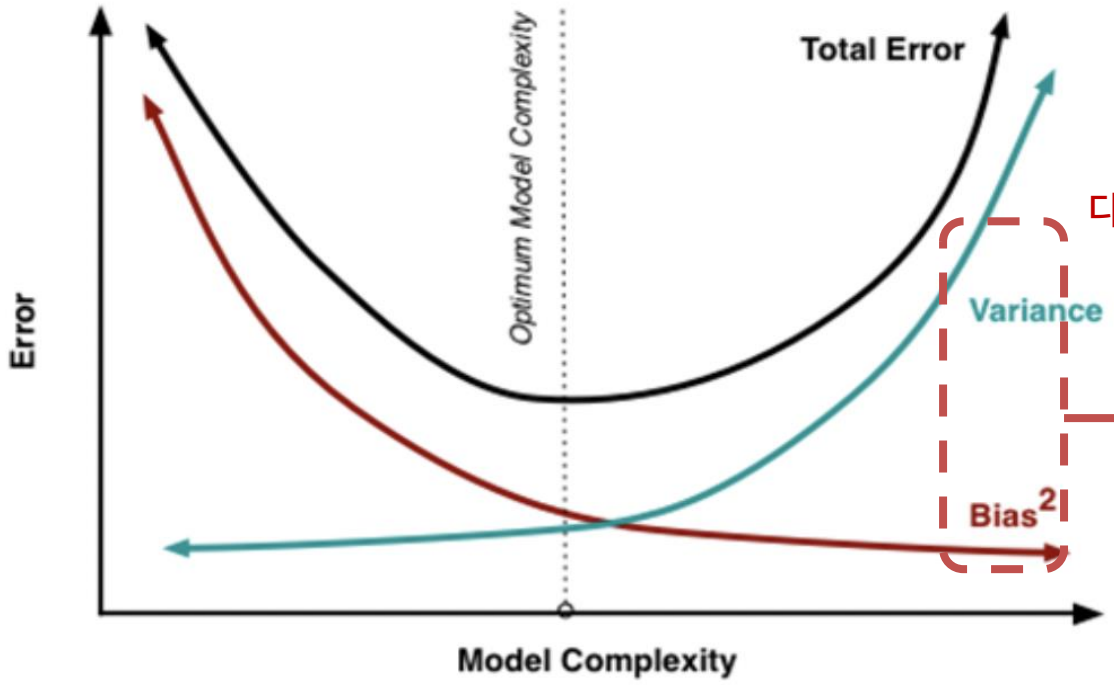
정규화

Bias-variance trade off



정규화

Bias-variance trade off



정규화 이전

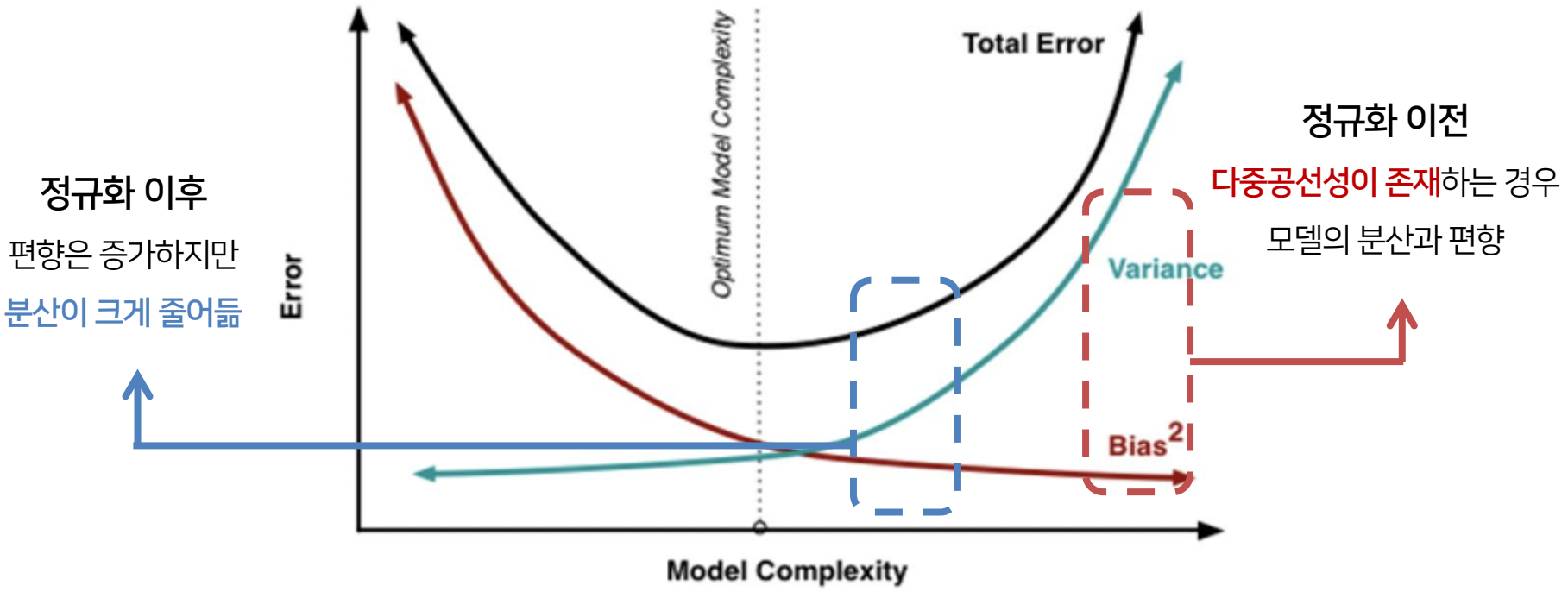
다중공선성이 존재하는 경우

모델의 분산과 편향

정규화

정규화

Bias-variance trade off



정규화 이후

편향은 증가하지만

분산이 크게 줄어듦

정규화 이전

다중공선성이 존재하는 경우

모델의 분산과 편향

Ridge

Ridge L2 Regularization

SSE를 최소화하면서 회귀계수 β에 제약 조건을 거는 방법

제약 조건식 L2-norm 형태 → L2 Regularization

L2-norm은 선형대수학 2주차 클린업 참고!

목적함수

$$\hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right)^2 \text{ subject to } \sum_{j=1}^{p}\beta_j^2 \le s$$

$$\Leftrightarrow \hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

Ridge

Ridge L2 Regularization

SSE를 최소화하면서 회귀계수 β에 제약 조건을 거는 방법

제약 조건식 L2-norm 형태 → L2 Regularization

L2-norm은 선형대수학 2주차 클린업 참고!

목적함수

$$\hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right)^2 \text{ subject to } \sum_{j=1}^{p}\beta_j^2 \le s$$

$$\Leftrightarrow \hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

정규화

Ridge

Ridge L2 Regularization

SSE를 최소화하면서 회귀계수 β에 제약 조건을 거는 방법

제약 조건식 L2-norm 형태 → L2 Regularization

L2-norm은 선형대수학 2주차 클린업 참고!

목적함수를 최소화함으로써 Ridge Estimator 추정 가능

이차식 형태이므로 미분을 통해 추정량 계산

설명변수들은 표준화된 상태여야 함!

목적함수

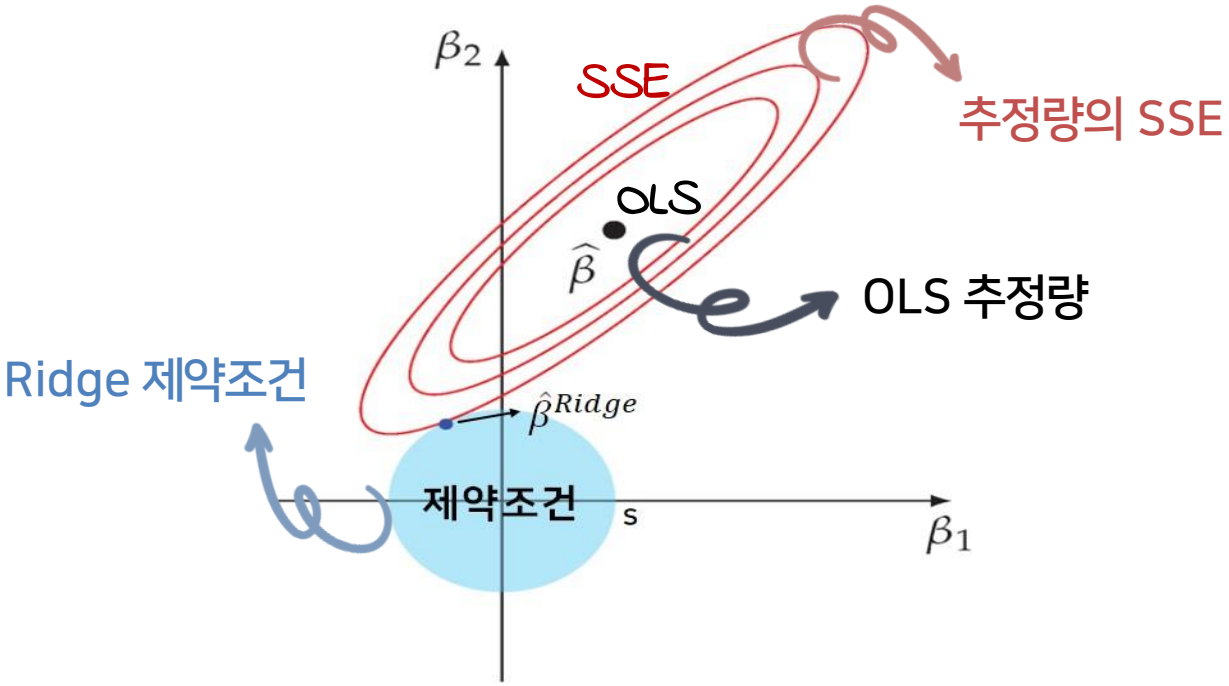
$$\hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right) \text{ subject to } \sum_{j=1}^{p}\beta_j^2 \le s$$

$$\Leftrightarrow \hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

정규화

목적함수

$$\hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right)^2 \text{ subject to } \sum_{j=1}^{p}\beta_j^2 \le s$$



정규화

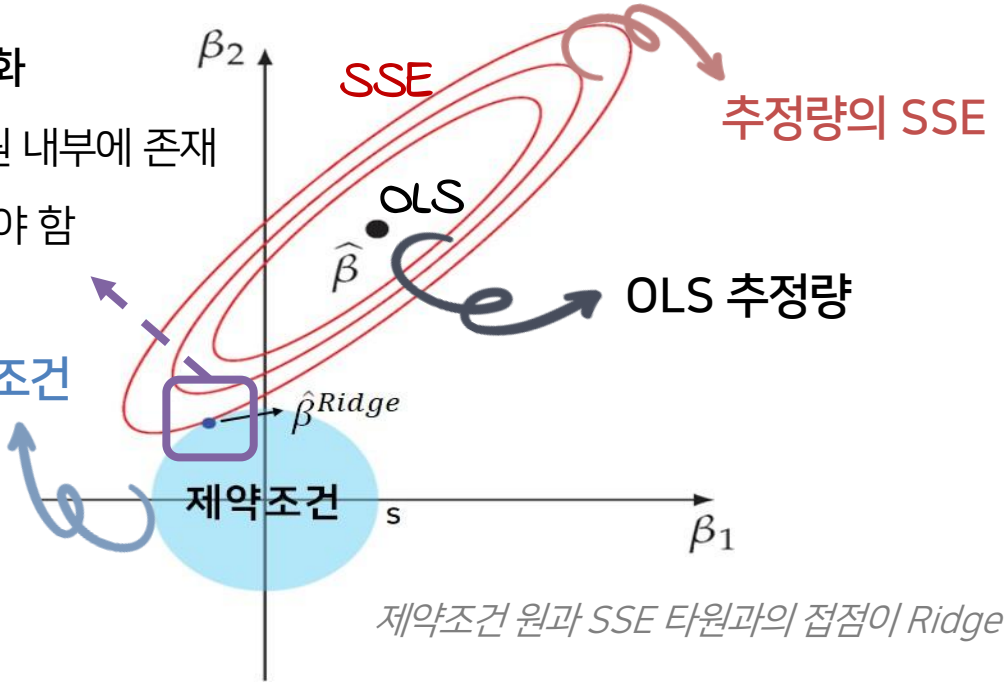
목적함수

$$\hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right)^2 \text{ subject to } \sum_{j=1}^{p}\beta_j^2 \le s$$

!! **회귀계수의 최소화**

① 회귀계수 $\hat{\beta}$는 제약조건 원 내부에 존재

② SSE를 최소화해야 함



제약조건 원과 SSE 타원과의 접점이 Ridge Estimatior

정규화

목적함수

$$\hat{\beta}^{ridge} = \text{argmin} \sum_{} \left(\quad \right)^2 \text{subject to} \sum_{j=1}^{p} \beta_j^2 \le s$$

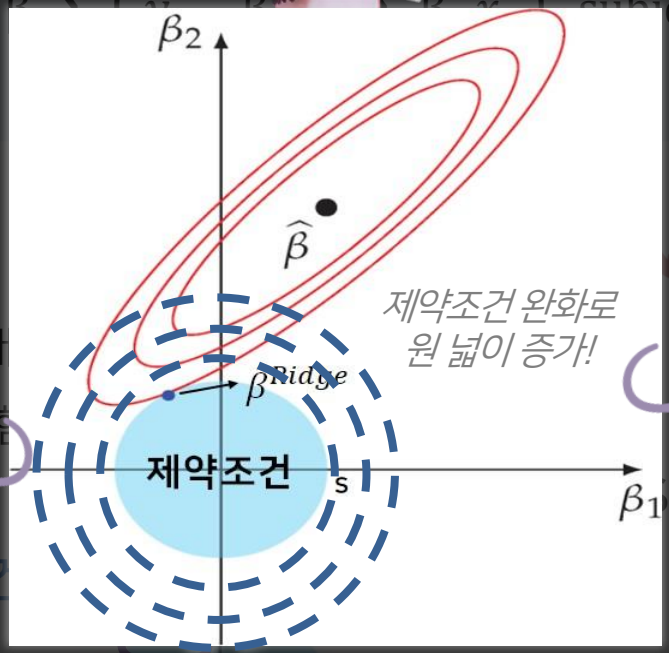
!!

s가 증가하면서
회귀계수의 최소화
원의 넓이 증가

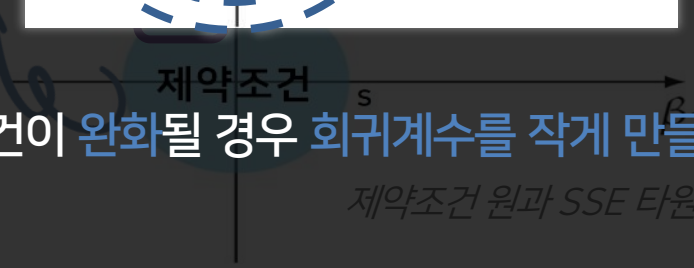
① 회귀계수 β는 제약조건 원 내

② SSE를 최소화해야 함

원이 타원을 밀어내며
추정량이 0에서 멀어짐
추정량의 SSE



Ridge 제약조건



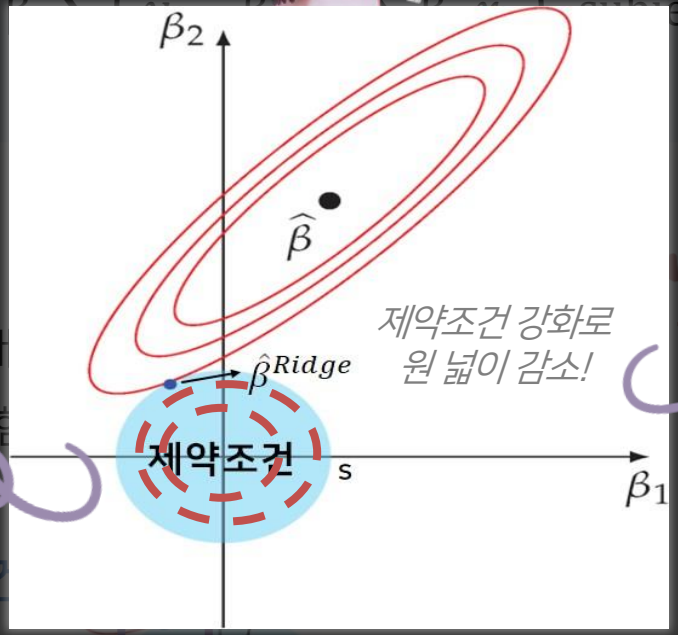
제약 조건이 완화될 경우 회귀계수를 작게 만들 수 없음

제약조건 원과 SSE 타원과의 접점이 Ridge Estimatior

3 정규화

목적함수

$$\hat{\beta}^{ridge} = \text{argmin} \sum_{} \left(\right)^2 \text{subject to} \sum_{j=1}^{p} \beta_j^2 \leq s$$



!! s가 감소하면서 회귀계수의 최소화

원의 넓이 감소

① 회귀계수 β는 제약조건 원 내

② SSE를 최소화하야 함

제약조건 강화로 원 넓이 감소!

추정량이 0으로 수렴함

(0은 될 수 없음)

추정량의 SSE

Ridge 제약조건

제약조건

s

β_1 추정량

제약 조건이 강화될 경우 회귀계수를 작게 만들 수 있음

제약조건 원과 SSE 타원과의 접점이 Ridge Estimation

정규화

목적함수 | 라그랑지안 승수법

Lagrangian

$$\hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j\, x_i\right)^2 + \lambda \sum_{j=1}^{p}\beta_j^2$$

정규화

목적함수 | 라그랑지안 승수법

Lagrangian

$$\hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j\, x_i\right)^2 + \lambda \sum_{j=1}^{p}\beta_j^2$$

오차제곱합(SSE) 최소화

정규화

목적함수 | 라그랑지안 승수법

regularization term을 통해
회귀계수 크기 조정

Lagrangian

$$\hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

오차제곱합(SSE) 최소화

정규화

목적함수 | 라그랑지안 승수법

regularization term을 통해
회귀계수 크기 조정

Lagrangian

$$\hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right)^2 + \lambda \sum_{j=1}^{p}\beta_j^2$$

오차제곱합(SSE) 최소화

음수가 아닌 튜닝 파라미터

최적의 모델을 찾는 과정에서 직접 CV를 통해 조정해주는 모수

제약조건의 크기를 결정 (s와는 반대 관계)

목적함수 | 라그랑지안 승수법

regularization term을 통해
회귀계수 크기 조정

Lagrangian

$$\hat{\beta}^{ridge} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right)^2 + \lambda \sum_{j=1}^{p}\beta_j^2$$

[Λ가 커지는 경우]

λ 의 영향력이 증가하므로
$\sum_{j=1}^{p}\beta_j^2$ 은 작아져야 함

오차제곱합(SSE) 최소화
∴ 개별 회귀계수 작아짐
음수가 아닌 튜닝 파라미터
$\Lambda \to \infty$ 이면, 개별 회귀계수 ≈ 0
최적의 모델을 찾는 과정에서 직접 CV를 통해 조정해주는 모수

제약조건의 크기를 결정 (s와는 반대 관계)

목적함수 | 라그랑지안 승수법

regularization term을 통해
회귀계수 크기 조정

Lagrangian

$$\hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_i \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

[λ가 작아지는 경우]

λ 의 영향력이 감소하므로
상대적으로 $\sum_{j=1}^{p} \beta_j^2$ 의 영향력 커짐

오차제곱합(SSE) 최소화
∴ 개별 회귀계수 커짐

음수가 아닌 튜닝 파라미터

$\lambda = 0$ 이면, OLS 추정량과 동일

최적의 모델을 찾는 과정에서 직접 CV를 통해 조정해주는 모수

제약조건의 크기를 결정 (s와는 반대 관계)

정규화

특징

Scaling

회귀계수는 변수 단위에 큰 영향을 받음

단위의 영향을 제거, 순수 영향력만 사용

주로 standard scaling 사용

Cost

regularization term 덕분에

미분 가능

λ를 바꾸며 미분과 함께 행렬 연산

Prediction

상관관계가 높은 변수들이

모델에 존재할 경우

높은 예측 성능 보임

다중공선성이 존재할 경우

Variable Selection

영향력을 줄일 뿐 변수 잔존

다중공선성을 일으키는 변수 제거 불가

Ridge를 통한 해석력 증가는 어려움

정규화

특징

Scaling

회귀계수는 변수 단위에 큰 영향을 받음

단위의 영향을 제거, 순수 영향력만 사용

주로 standard scaling 사용

Cost

regularization term 덕분에

미분 가능

λ를 바꾸며 미분과 함께 행렬 연산

Prediction

상관관계가 높은 변수들이

모델에 존재할 경우

높은 예측 성능 보임

다중공선성이 존재할 경우

Variable Selection

영향력을 줄일 뿐 변수 잔존

다중공선성을 일으키는 변수 제거 불가

Ridge를 통한 해석력 증가는 어려움

정규화

특징

Scaling

회귀계수는 변수 단위에 큰 영향을 받음

단위의 영향을 제거, 순수 영향력만 사용

주로 standard scaling 사용

Cost

regularization term 덕분에

미분 가능

λ를 바꾸며 미분과 함께 행렬 연산

Prediction

상관관계가 높은 변수들이

모델에 존재할 경우

높은 예측 성능 보임

다중공선성이 존재할 경우

Variable Selection

영향력을 줄일 뿐 변수 잔존

다중공선성을 일으키는 변수 제거 불가

Ridge를 통한 해석력 증가는 어려움

3 정규화

특징

Scaling

회귀계수는 변수 단위에 큰 영향을 받음

단위의 영향을 제거, 순수 영향력만 사용

주로 standard scaling 사용

Cost

regularization term 덕분에

미분 가능

λ를 바꾸며 미분과 함께 행렬 연산

Prediction

상관관계가 높은 변수들이

모델에 존재할 경우

높은 예측 성능 보임

다중공선성이 존재할 경우

Variable Selection

영향력을 줄일 뿐 변수 잔존

다중공선성을 일으키는 변수 제거 불가

Ridge를 통한 해석력 증가는 어려움

Lasso

Lasso L1 Regularization

SSE를 최소화하면서 회귀계수 β에 제약 조건을 거는 방법

제약 조건식 L1-norm 형태 → L1 Regularization

L1-norm은 선형대수학 1주차 클린업 참고!

목적함수

$$\hat{\beta}^{Lasso} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j\, x_i\right)^2 \text{ subject to } \sum_{j=1}^{p}|\beta_j| \le s$$

$$\Leftrightarrow \hat{\beta}^{Lasso} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j\, x_i\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

Lasso

Lasso L1 Regularization

SSE를 최소화하면서 회귀계수 β에 제약 조건을 거는 방법

제약 조건식 L1-norm 형태 → L1 Regularization

L1-norm은 선형대수학 1주차 클린업 참고!

목적함수

$$\hat{\beta}^{Lasso} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j\, x_i\right)^2 \text{ subject to } \sum_{j=1}^{p}|\beta_j| \le s$$

$$\Leftrightarrow \hat{\beta}^{Lasso} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j\, x_i\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

Lasso

Lasso L1 Regularization

SSE를 최소화하면서 회귀계수 β에 제약 조건을 거는 방법

제약 조건식 L1-norm 형태 → L1 Regularization

L1-norm은 선형대수학 1주차 클린업 참고!

목적함수를 최소화함으로써 Lasso Estimator 추정 가능

Lasso는 미분 불가능 → 수치적인 방법

설명변수들은 표준화된 상태여야 함!

목적함수

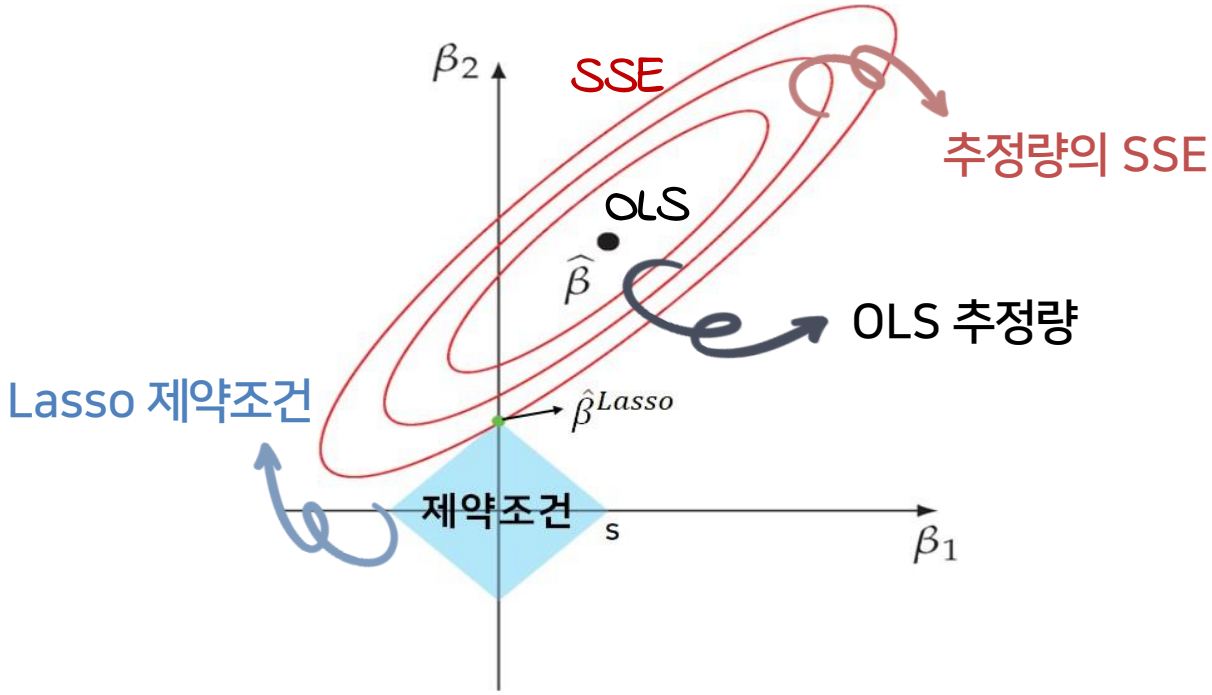
$$\hat{\beta}^{Lasso} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right) \text{ subject to } \sum_{j=1}^{p}|\beta_j| \le s$$

$$\Leftrightarrow \hat{\beta}^{Lasso} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

정규화

목적함수

$$\hat{\beta}^{Lasso} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j\, x_i\right)^2 \text{ subject to } \sum_{j=1}^{p}|\beta_j| \le s$$



정규화

목적함수

$$\hat{\beta}^{Lasso} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_i\right)^2 \text{ subject to } \sum_{j=1}^{p}|\beta_j| \le s$$

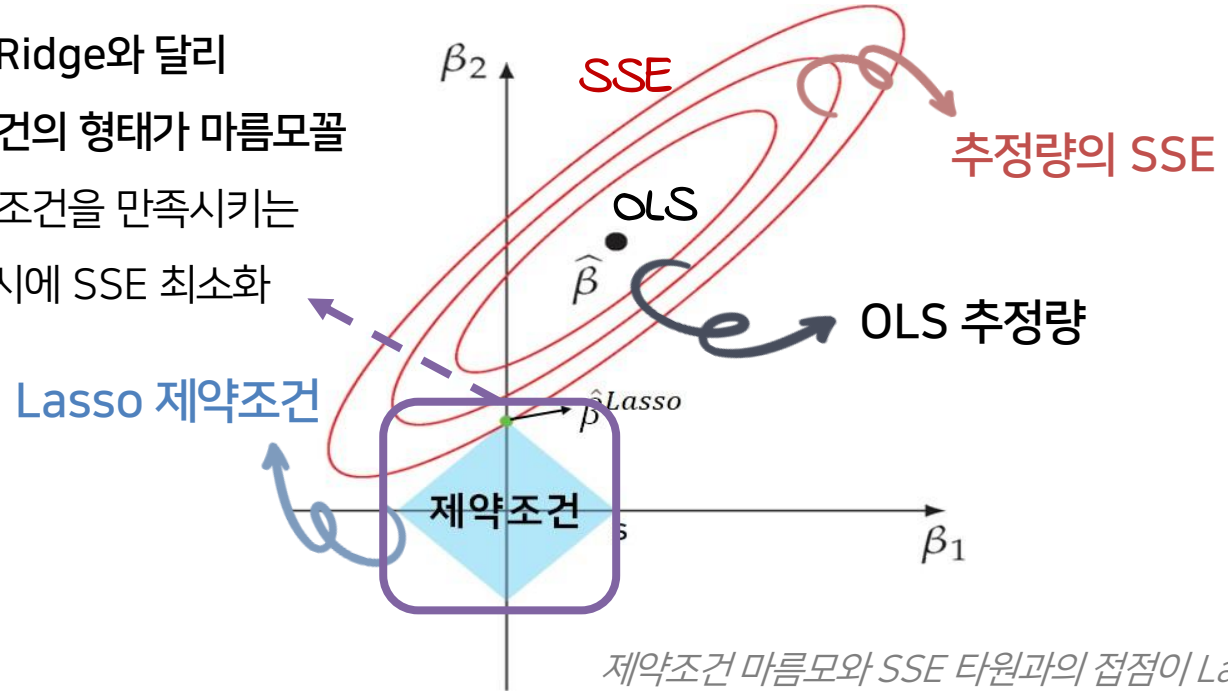
!!

Ridge와 달리

제약 조건의 형태가 마름모꼴

제약조건을 만족시키는

동시에 SSE 최소화



제약조건 마름모와 SSE 타원과의 접점이 Lasso Estimatior

정규화

목적함수

$$\hat{\beta}^{Lasso} = \text{argmin}_{\beta} \sum_{} \left(\right)^2 \text{subject to } \sum_{j=1}^{p} |\beta_j| \le s$$

!!

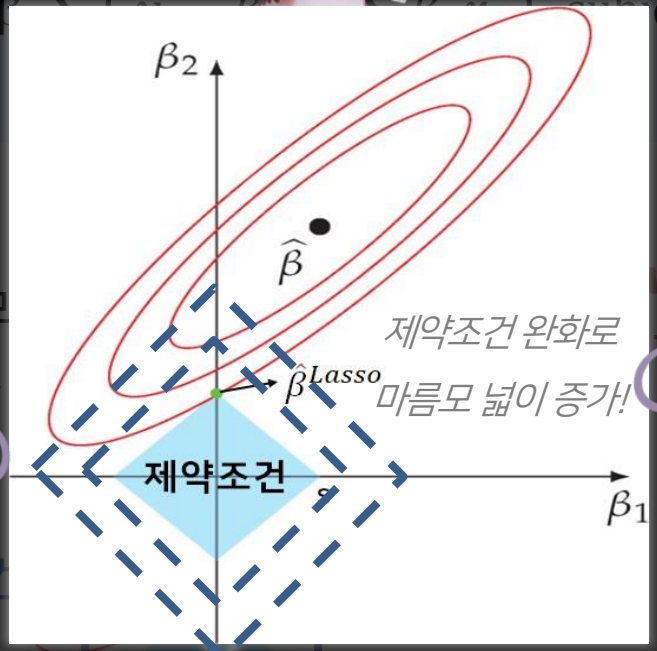
s가 증가하면서

제약조건의 확대가 마름모

마름모의 넓이 증가

제약조건을 만족시키는

동시에 SSE 최소화



마름모가 타원을 밀어내며

추정량이 0에서 멀어짐

추정량의 SSE

제약조건 완화로

마름모 넓이 증가!

추정량

Lasso 제약조건

제약 조건이 완화될 경우 회귀계수를 작게 만들 수 없음

제약조건 마름모와 SSE 타원과의 접점이 Lasso Estimation

정규화

목적함수

$$\hat{\beta}^{Lasso} = \text{argmin}_{\beta} \sum_{i=1}^{n} \left(\quad \right)^2 \text{subject to } \sum_{j=1}^{p} |\beta_j| \le s$$

!!

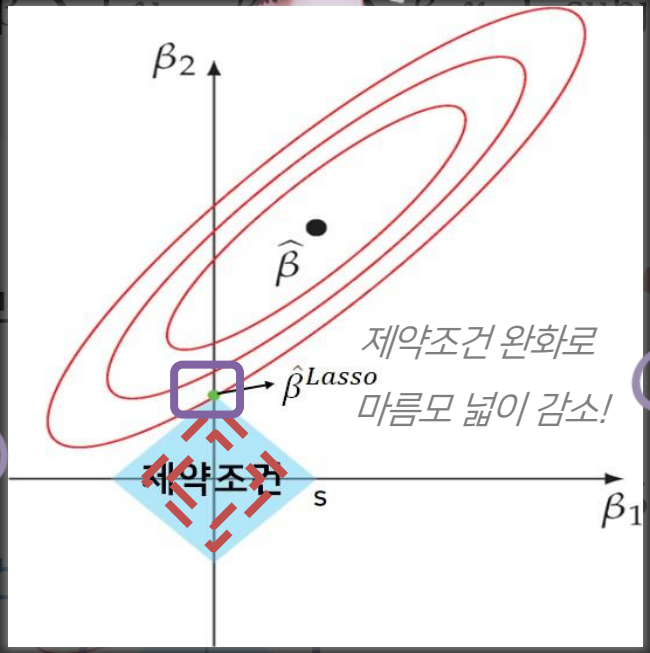
s가 감소하면서

마름모의 넓이 감소

제약조건을 만족시키는

동시에 SSE 최소화

Lasso 제약조건



추정량이 0으로 수렴함

(0이 될 수 있음)

추정량의 SSE

추정량

제약조건

β_1

제약 조건이 강화될 경우 회귀계수를 작게 만들 수 있음

제약조건 마름모와 SSE 타원과의 접점이 Lasso Estimation

정규화

목적함수 | 라그랑지안 승수법

regularization term을 통해
회귀계수 크기 조정

Lagrangian

$$\hat{\beta}^{Lasso} = \text{argmin}\,\beta \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j\, x_i\right)^2 + \lambda \sum_{j=1}^{p}|\beta_j|$$

오차제곱합(SSE) 최소화

튜닝 파라미터

최적의 모델을 찾는 과정에서 직접 CV를 통해 조정해주는 모수

제약조건의 크기를 결정 (s와는 반대 관계)

정규화

목적함수 | 라그랑지안 승수법

regularization term을 통해
회귀계수 크기 조정

Lagrangian

$$\hat{\beta}^{Lasso} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_i \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

[λ가 커지는 경우]

λ 의 영향력이 증가하므로

$\sum_{j=1}^{p} |\beta_j|$은 작아져야 함

오차제곱합(SSE) 최소화

∴ 개별 회귀계수 작아짐

최적의 모델을 찾는 과정에서 직접 CV를 통해 조정해주는 모수

제약조건의 크기를 결정 (s와는 반대 관계)

정규화

목적함수 | 라그랑지안 승수법

regularization term을 통해
회귀계수 크기 조정

Lagrangian

[λ가 작아지는 경우]

$$\hat{\beta}^{Lasso} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_i \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

λ 의 영향력이 감소하므로

상대적으로 $\sum_{j=1}^{p} |\beta_j|$의 영향력 커짐

오차제곱합(SSE) 최소화

∴ 개별 회귀계수 커짐

최적의 모델을 찾는 과정에서 직접 CV를 통해 조정해주는 모수

제약조건의 크기를 결정 (s와는 반대 관계)

정규화

목적함수 | 라그랑지안 승수법

regularization term을 통해
회귀계수 크기 조정

Lagrangian

$$\hat{\beta}^{Lasso} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_i \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

$[\lambda = 0]$

regularization term이 사라지므로

기존 OLS 추정량을 도출

오차제곱합(SSE) 최소화

튜닝 파라미터

최적의 모델을 찾는 과정에서 직접 CV를 통해 조정해주는 모수

제약조건의 크기를 결정 (s와는 반대 관계)

정규화

목적함수

큰 λ 값	작은 λ 값
적은 변수	많은 변수
간단한 모델	복잡한 모델
해석 쉬움	해석 어려움
높은 학습 오차	낮은 학습 오차

underfitting 위험 증가

overfitting 위험 증가

3 정규화

특징

Scaling

개별 변수에 대한 scaling

변수 단위에 의한 영향력 제거

주로 standard scaling 사용

하지만 변수 간 상관관계가 높으면 변수 선택 성능이 떨어짐

0이 되는 계수의 존재로 인해 sparsity(희박성)을 지님

Variable Selection

Ridge와 달리 λ값이 0이 되는

회귀계수가 존재하므로 변수 선택 가능

변수 해석 가능성 증가

Prediction

상관관계가 높은 변수들이 모델에 존재할 경우

예측에 유의미한 변수들을 0으로 만들 수 있음

∴ Ridge에 비해 상대적으로 **예측 성능 떨어짐**

다중공선성이 존재할 경우

3 정규화

특징

하지만 변수 간 상관관계가 높으면 변수 선택 성능이 떨어짐

0이 되는 계수의 존재로 인해 sparsity(희박성)을 지님

Scaling

개별 변수에 대한 scaling

변수 단위에 의한 영향력 제거

주로 standard scaling 사용

Variable Selection

Ridge와 달리 λ값이 0이 되는

회귀계수가 존재하므로 변수 선택 가능

변수 해석 가능성 증가

Prediction

상관관계가 높은 변수들이 모델에 존재할 경우

예측에 유의미한 변수들을 0으로 만들 수 있음

∴ Ridge에 비해 상대적으로 **예측 성능 떨어짐**

다중공선성이 존재할 경우

3 정규화

특징

하지만 변수 간 상관관계가 높으면 변수 선택 성능이 떨어짐

0이 되는 계수의 존재로 인해 sparsity(희박성)을 지님

Scaling

개별 변수에 대한 scaling

변수 단위에 의한 영향력 제거

주로 standard scaling 사용

Variable Selection

Ridge와 달리 λ값이 0이 되는

회귀계수가 존재하므로 변수 선택 가능

변수 해석 가능성 증가

Prediction

상관관계가 높은 변수들이 모델에 존재할 경우

예측에 유의미한 변수들을 0으로 만들 수 있음

∴ Ridge에 비해 상대적으로 **예측 성능 떨어짐**

다중공선성이 존재할 경우

Elastic-Net

Elastic-Net

상관성이 있는 변수를 모두 제거하거나 선택하여 성능 보완(Grouping Effect)

변수 간 상관관계가 존재할 때

LASSO의 성능이 떨어지는 한계를 보완하기 위한 방법

목적함수

$$\hat{\beta}^{elastic} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_i \right)^2 \; s.t. \; \underbrace{t_1 \sum_{j=1}^{p} |\beta_j|}_{\text{LASSO}} + \underbrace{t_1 \sum_{j=1}^{p} \beta_j^2}_{\text{RIDGE}} \leq s$$

!!

제약식에 RIDGE의 L2 term과 LASSO의 L1 term이 모두 반영된 모형

정규화

Elastic-Net

Elastic-Net

상관성이 있는 변수를 모두 제거하거나 선택하여 성능 보완(Grouping Effect)

변수 간 상관관계가 존재할 때

LASSO의 성능이 떨어지는 한계를 보완하기 위한 방법

목적함수

$$\hat{\beta}^{elastic} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_i \right)^2 \; s.t. \; \underbrace{t_1 \sum_{j=1}^{p} |\beta_j|}_{\text{LASSO}} + \underbrace{t_1 \sum_{j=1}^{p} \beta_j^2}_{\text{RIDGE}} \leq s$$

제약식에 RIDGE의 L2 term과 LASSO의 L1 term이 모두 반영된 모형

Elastic-Net

Elastic-Net

상관성이 있는 변수를 모두 제거하거나 선택하여 성능 보완(Grouping Effect)

변수 간 상관관계가 존재할 때

LASSO의 성능이 떨어지는 **한계를 보완**하기 위한 방법

목적함수

$$\hat{\beta}^{elastic} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_i \right)^2 \; s.t. \; t_1 \sum_{j=1}^{p} |\beta_j| + t_1 \sum_{j=1}^{p} \beta_j^2 \leq s$$

LASSO RIDGE

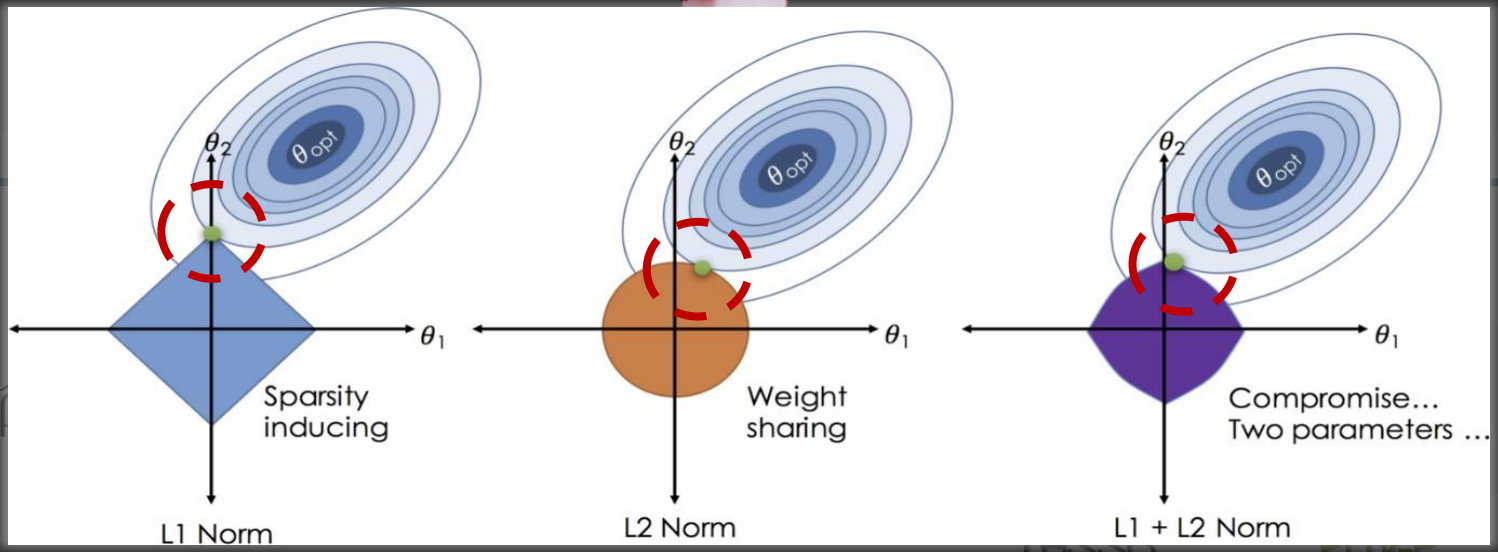
!! 제약식에 RIDGE의 L2 term과 LASSO의 L1 term이 모두 반영된 모형

Elastic-Net

Elastic-Net

상관성이 있는 변수 제거하거나 선택하여 성능 보완(Grouping Effect)

변수 간 상관 존재할 때



제약조건이 변함에 따라 추정량이 만들어지는 공간이 변화

제약식에 RIDGE의 L2 term과 LASSO의 L1 term이 모두 반영된 모형

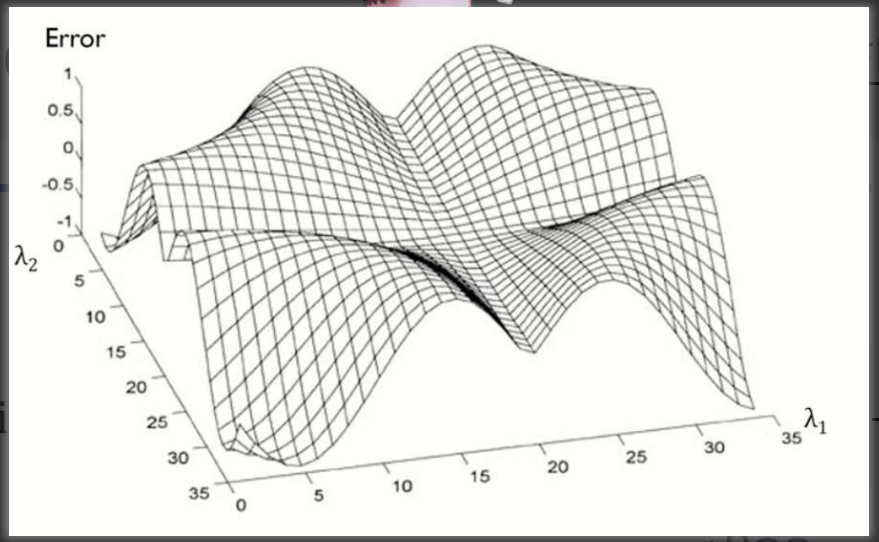
Elastic-Net

Elastic-Net

상관성이 있는 변수 제거하거나 선택하여 성능 보완(Grouping Effect)

변수 간 상관 존재할 때

LASSO 방법



$$\hat{\beta}^{elastic} = \underset{\beta}{\operatorname{argmi}} \quad + t_1 \sum_{j=1}^{p} \beta_j^2 \leq s$$

LASSO RIDGE

그리드 서치 벙법를 통해 error가 최소화되는 λ_1, λ_2조합을 선정

제약식에 RIDGE의 L2 term과 LASSO의 L1 term이 모두 반영된 모형

Fused Lasso

Fused Lasso

변수들 사이의 물리적 거리가 존재한다는

사전지식을 활용한 모델

목적함수

$$\hat{\beta}^{FL} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_i \right)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} |\beta_j - \beta_{j-1}|$$

LASSO　　New term

!!

인접한 변수들의 회귀 계수를 비슷하게 추정하도록 만드는 역할

Fused Lasso

Fused Lasso

변수들 사이의 물리적 거리가 존재한다는

사전지식을 활용한 모델

목적함수

$$\hat{\beta}^{FL} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_i \right)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} |\beta_j - \beta_{j-1}|$$

LASSO　　New term

!!

인접한 변수들의 회귀 계수를 비슷하게 추정하도록 만드는 역할

Fused Lasso

Fused Lasso

변수들 사이의 물리적 거리가 존재한다는

사전지식을 활용한 모델

목적함수

$$\hat{\beta}^{FL} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_i \right)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} |\beta_j - \beta_{j-1}|$$

LASSO

New term

!!

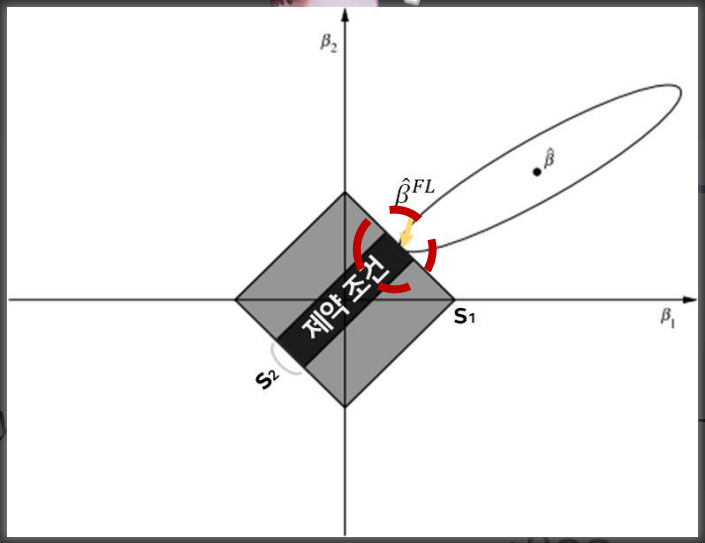
인접한 변수들의 회귀 계수를 **비슷하게 추정**하도록 만드는 역할

정규화

Fused Lasso

Fused Lasso

변수들 사이의 [illegible]가 존재한다는



$$\hat{\beta}^{FL} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \Bigg(y \qquad \lambda_2 \sum_{j=1}^{p} |\beta_j - \beta_{j-1}|$$

LASSO

New term

변수들의 차이에 관한 제약으로 인접한 변수의 값을 비슷하게 추정

!!

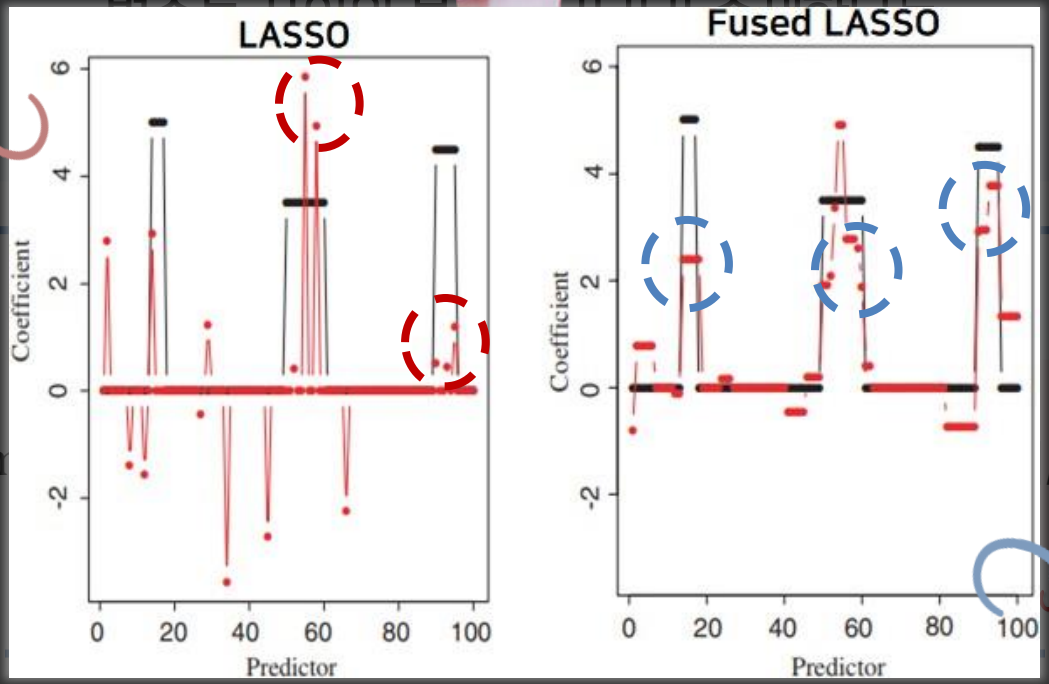
인접한 변수들의 회귀 계수를 비슷하게 추정하도록 만드는 역할

정규화

Fused Lasso

Lasso의 경우

인접한 변수들의 값에 차이가 발생



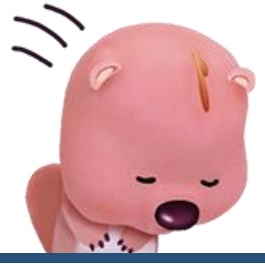
$$\hat{\beta}^{FL} = \underset{\beta}{\operatorname{argm}} \qquad \beta_j - \beta_{j-1}|$$

!!

인접한 변수들의 회귀 계수를 **비슷하게 추정**하도록 만드는 영향

Fused Lasso의 경우

인접한 변수들의 값을 비슷하게 추정



감사합니다