회귀분석팀

6팀

심은주
진수정
문병철
이수정
임주은

INDEX

1. 다중공선성

2. 변수선택법

3. 축소 추정

- **다중공선성이란?**

예측변수 X들 간의 선형관계가 존재하는 경우

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon_i$$

나머지 X 변수들이 고정되었을 때,
x_1이 1단위 증가하면 y는 평균적으로 β_1만큼 증가함을 의미

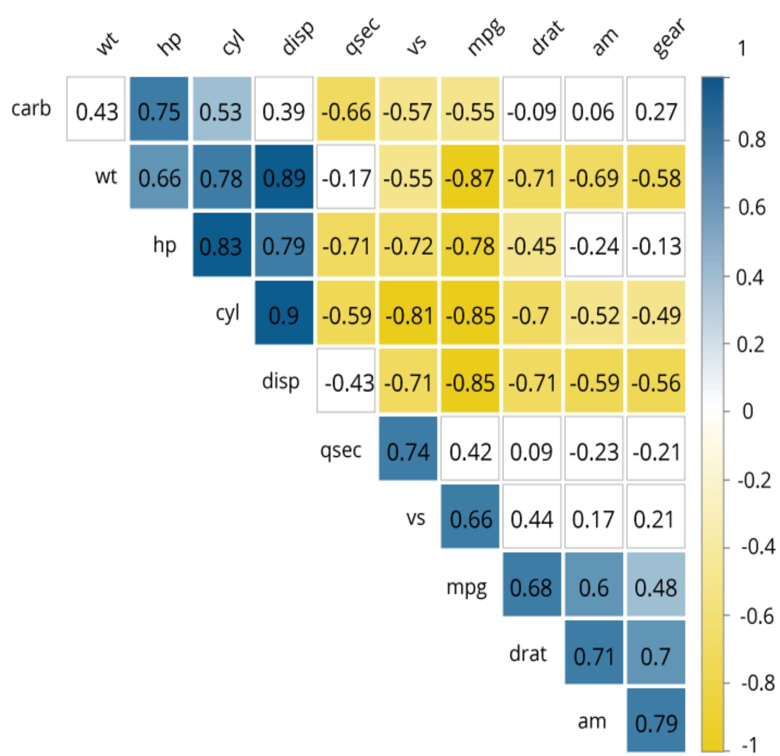
개별 변수 해석 시, '다른 변수를 **고정**한 상태에서 해당 X의 증분'

Uncorrelated한 경우만 가능

정확한 회귀분석을 위해선 다중공선성이 크면 안된다!

- **다중공선성의 판별법**

상관계수 플랏



- 절댓값 기준
  상관계수가 0.7이상일 경우
  다중공선성 의심

- R의 ‘Corrplot’ 패키지 이용

- **다중공선성의 판별법**

VIF(Variance Inflation Factor, 분산팽창인자)

$$VIF = \frac{1}{1-{R^2}_j}, \quad j = 1, \dots, p$$

${R^2}_j$: x_j를 $x_1 + \cdots + x_p$ 으로 회귀식을 적합했을 때, 도출되는 R^2 값

- VIF가 1이면, 다중공선성 없음 (${R^2}_j$=0이므로)

- VIF가 10이상이면, 심각한 다중공선성 존재

- **변수선택법이란?**

후보 변수들 중에서 불필요한 변수들을 제거하여 적절한 변수들의 집합을 찾는 방법

다중공선성이 존재할 때 많이 사용!

그래서 뭐가 좋은데..?

- 높은 상관관계를 가지는 변수들 중 일부만을 선택 가능
- 높은 상관관계를 가지는 변수들의 존재를 정당화

변수선택법을 통해 최종 모델에 대한 확신을 얻을 수 있음!

- **변수선택의 척도**

수정결정계수(R_a^2)

벌써 까먹은 건 아니쥐?

$$R_a^2 = \frac{SSR/p}{SST/(n-1)} = 1 - \frac{SSE/(n-p-1)}{SST/(n-1)}$$

회귀식의 설명력을 의미

수정결정계수가 더 큰 모델을 사용하자!

- **변수선택의 척도**

Mallows(C_p)

$$C_p = \frac{SSE_p}{\hat{\sigma}^2} + (2p - n)$$

p : RM에서 사용한 독립변수 개수

$\hat{\sigma}^2$: FM의 오차항의 분산의 추정값

n : 관측치의 개수

- 변수를 추가하면 SSE는 무조건 작아지기 때문에, 2p를 패널티로 넣음

- $C_p \approx p$ 일수록 bias가 작은, 좋은 모델

- C_p값을 이용하여 모델을 비교할 때에는
  동일한 독립변수의 전체 집합을 가진 모델일 경우에만 사용 가능

- **변수선택의 종류**

변수선택법은 경험적인 방법

명확한 답이 존재하는 것이 아니라 ,

직접 알고리즘에 따라 해당하는 모든 경우를 계산해서 가장 좋은 회귀식을 찾는다!

Best Subset Selection

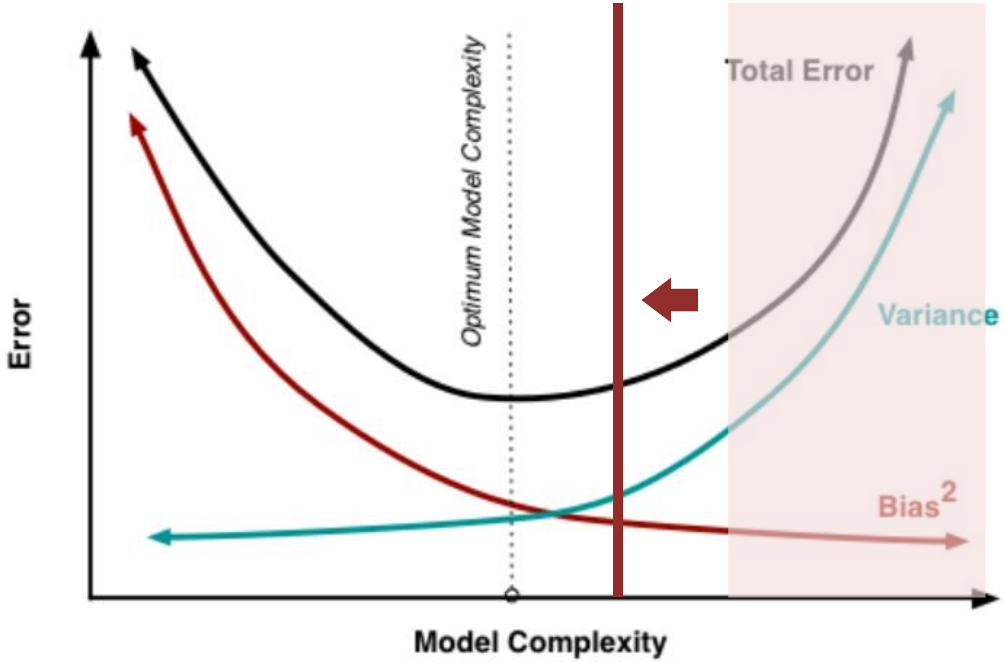
전진선택법

후진제거법

단계적 선택법

- **축소 추정이란?**

각각 개별 베타 추정량을 **0으로 수축**시키는 방법



불편성을 포기하되,

전체 MSE($Bias^2 + Variance$)를 더 작게 하는

추정량 얻을 수 있다!

- Ridge Regression (L2 Regularization)

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

- 위의 식을 최소화하는 β값을 가짐
- 앞 부분: SSE, 회귀식이 데이터에 잘 적합하여 SSE를 작게 만드는 계수 추정치를 찾음
- 뒤 부분: shrinkage penalty로 β값들이 0에 가까울 수록 작아짐

  → **이 항이 계수 추정치들을 0으로 축소하는 영향을 가짐**

- Ridge Regression (L2 Regularization)

$$\hat{\beta}^{ridge} = \underset{\beta}{argmin} \sum_i (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

조절 모수 λ: SSE항과 패널티항을 조절하는 역할

좋은 λ 값 선택 중요

주로 CV를 통해 튜닝

- λ=0 이면, 패널티 효과가 없어 최소제곱추정치 생성
- $\lambda \to \infty$ 이면, 패널티 효과가 커져 계수 추정치가 0에 가까워짐

∴ Ridge regression은 각각의 λ 값에 따라서 다른 추정치 집합들을 만듦

- Lasso Regression (L1 Regularization)

$$\hat{\beta}^{lasso} = \underset{\beta}{argmin} \sum_{i} (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Ridge와 비슷하게 패널티를 주어 계수 추정치들을 0으로 축소하는 방법

- Penalty term이 절댓값으로 들어감
- Ridge와 달리 변수 선택의 효과가 있음

변수 잘 가고~~

- **Elastic Net**

Lasso와 Ridge를 절충하여 각각의 단점을 보완한 정규화 방법

Ridge part Lasso part

$$L_{enet}(\hat{\beta}) = \frac{\sum_i (y_i - x_i^T \beta)^2}{2n} + \lambda\left(\left(\frac{1-\alpha}{2}\right)\sum_j \beta_j^2 + \alpha \sum_j |\beta_j|\right)$$

α 는 Ridge와 Lasso의 가중치로,

$\alpha = 0$이면 식은 Ridge가 되고, $\alpha = 1$이면 Lasso가 됨

- Adaptive Lasso Regression

$$\widehat{\beta^*}^{(n)} = \underset{\beta}{argmin} \left\| y - \sum_j x_j \beta_j \right\|^2 + \lambda_n \sum_j \widehat{w}_j |\beta_j|$$

$$where \;\; \widehat{w}_j = \frac{1}{|\hat{\beta}|^{\gamma}} \qquad \hat{\beta}\text{: initial estimate (ex. OLS)}$$

- 계수 추정량이 클수록 작은 가중치를 줌으로써, Sparsity는 유지하되 Bias를 줄일 수 있음

- λ를 적절히 선택한다면, Oracle Properties를 만족하게 됨