

Part.02
회귀분석

| 변수선택법

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택

I 변수선택법

- 모델 선택(변수선택)

- 변수가 여러 개 일 때 최적의 변수 조합을 찾아내는 기법
- 변수의 수가 p 개일 때 변수의 총 조합은 2^p 으로 변수 수가 증가함에 따라 변수 조합의 수는 기하급수적으로 증가
- 총 변수들의 조합 중 최적의 조합을 찾기 위한 차선의 방법

(optimal은 아님, optimal한 조합을 찾는 방법은 모든 경우의 수 조합을 다 해보는 것)

1) Feedforward Selection 방법

2) Backward Elimination 방법

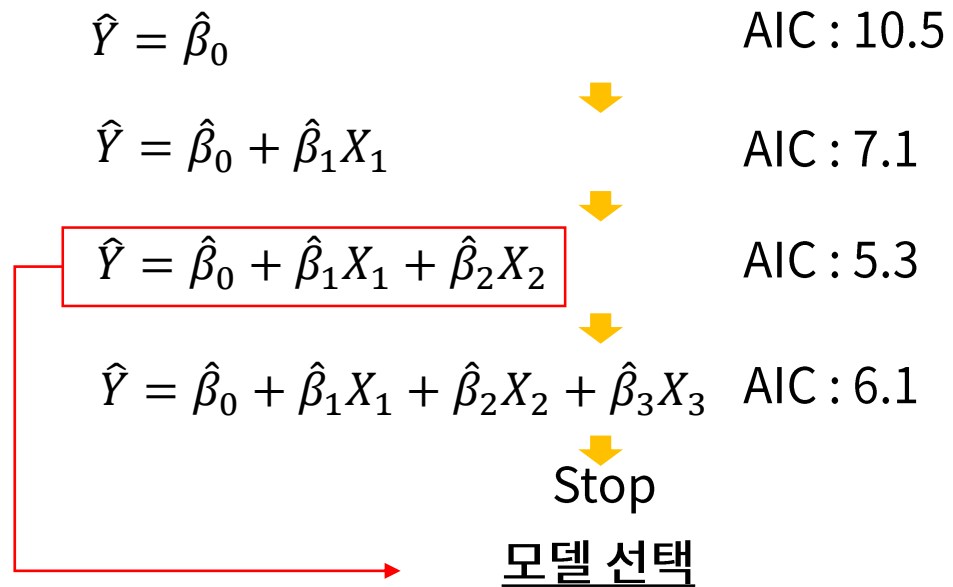
3) Stepwise 방법

I 변수선택법

■ 모델 선택(변수선택)

1) Feedforward Selection 방법 : 변수를 추가해가며 성능지표를 비교해가는 방법

ex) 총 변수 4개



I 변수선택법

■ 모델 선택(변수선택)

2) Backward Elimination 방법 : 변수를 제거해가며 성능지표를 비교해가는 방법

ex) 총 변수 4개

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 \quad \text{AIC : 5.5}$$



$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 \quad \text{AIC : 5.1}$$



$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 \quad \text{AIC : 5.3}$$

Stop

모델 선택

I 변수선택법

- 모델 선택(변수선택)

3) Stepwise 방법 :

- 가장 유의한 변수를 추가하거나 유의하지 않는 변수를 제거해나가는 방법
- 전진선택법을 사용할 때 한 변수가 선택되면 이미 선택된 변수 중 중요하지 않은 변수가 있을 수 있음
- 전진선택법의 각 단계에서 이미 선택된 변수들의 중요도를 다시 검사하여 중요하지 않은 변수를 제거하는 방법
- 일반적으로 가장 널리 쓰이는 방법
 1. 변수 입력/제거를 위한 p-value 임계치 설정
 2. Forward selection을 통한 변수 선정
 3. 선택된 변수 중 유의미한 변수를 남기고 제거, 2-3반복
 4. 변수가 추가되거나 제거할 케이스가 없는 경우 종료

Part.02
회귀분석

| 교호작용

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택