

인과추론 스터디

8주차

김민수

pos02043@gmail.com

<https://github.com/minsoo1036>

디자인 vs. 모델 기반 식별(Identification)

디자인 기반 : T가 X에 따라 어떻게 배분되었는지를 모델링
(ex) 역확률 가중치(IPW), 프로슈-로-워 정리(회귀)

모델 기반 : 잠재적 결과 Y를 직접 모델링
(ex) 회귀 분석

지금까지 배운 인과추론 방법론은 총 세가지

1. 성향점수 매칭(PSM)
2. 역확률 가중치(IPW)
3. 회귀 분석 (OLS)

그렇다면 이중에 어떤것을 사용해야 할까?

이중 강건 추정 (Doubly Robust Estimator)

T~X 에 자신이 있다면 디자인 기반을,
Y~T+X 에 자신이 있다면 모델 기반을 사용하면 됨.

그러나 둘 중 뚜렷이 무엇이 더 좋은지 모르겠다면 ?
또는 두 방법론의 결과 보다 더욱 좋은 (표준오차가 작은) 추정치를 얻고 싶다면??

→ 두 방법을 섞어 보자!

IPW 추정량

선형회귀 추정량

$$\hat{\mu}_t^{DR}(\hat{m}, \hat{e}) = \frac{1}{N} \sum \frac{TY}{\hat{e}(X)} - \frac{1}{N} \sum \left[\frac{T - \hat{e}(X)}{\hat{e}(X)} \hat{m}(X) \right] = \frac{1}{N} \sum \hat{m}(X) + \frac{1}{N} \sum \left[\frac{T}{\hat{e}(x)} (Y - \hat{m}(X)) \right]$$

Y의 예전 T의 예전

$$ATE = \hat{\mu}_1^{DR}(\hat{m}, \hat{e}) - \hat{\mu}_0^{DR}(\hat{m}, \hat{e})$$

보정항

표준오차는 Bootstrap 통해서 추정!

처치모델이 쉬운 경우 ($T \sim X$)

$$P(T = 1 | X = x) = e(x) = \frac{1}{1 + e^{-(1+1.5x)}} \quad \text{<성향점수는 } X \text{에 대해 선형>}$$

BUT... 잠재적 결과는 x 에 대해 비선형 (상대적으로 모델링 어려움)

$$Y(1) = 1$$

$$Y(0) = 1 - \frac{1}{x^3}$$

이중강건추정은 정확도를 높일 2
번의 기회가 있음!

True ATE는 2

Regression ATE : 1.79

Propensity Score ATE : 2.00 [1.81, 2.23]

DR ATE : 2.00 [1.87, 2.15] ← 더 좁은 신뢰구간

결과 모델이 쉬운 경우 ($Y \sim T + X$)

$$P(T = 1 | X = x) = e(x) = \frac{1}{1 + e^{-2x + x^3}} \quad \text{<성향점수는 } X \text{에 대해 선형>}$$

BUT... 잠재적 결과는 x 에 대해 비선형 (상대적으로 모델링 어려움)

$$Y(1) = x$$

$$Y(0) = Y(1) + 1 \quad \text{\#실제 ATE는 -1}$$

True ATE는 -1

이중강건추정은 정확도를 높일 2
번의 기회가 있음!

Regression ATE : -1.00

Propensity Score ATE : -1.10 [-1.14, -1.07]

DR ATE : -1.00 [-1.04, -0.96] ← 더 좁은 신뢰구간

연속형 처치에서 성향점수 추정

$f(T|X)$ 즉, 조건부 확률 밀도를 추정해야함!

이를 위해서는 강한 조건이 필요함 !!

1. Mean의 함수에 대한 가정
2. Error의 Distribution에 대한 가정

→ 따라서 추천하지는 않음.
→ 머신러닝으로 시도해볼 수 는 있음.

NOTE

$$f(t_i) = \frac{\exp\left(-\frac{1}{2}\left(\frac{t_i - \mu_i}{\sigma}\right)^2\right)}{\sigma\sqrt{2} * \pi}$$

등 가정 필요 !!