

CATE Overview

Meta Learners

인과추론의 근본적인 문제에 다가가기

CATE 식별

Meta Learners



2025 PseudoCon Coming Soon!

- 일시 : 25.05.17(토) 오후 1:00 ~ 5:30
- 장소 : 공덕 StartUp Hub 세미나홀
- 규모 : 200명
- 구성 : Talk / Booth / Ignite / Panel Talk

How to Contribute? Ignite Talk / Demo Booth

We're hiring!

KRAFTON

[< 채용 정보로 돌아가기](#)

신규

[Publishing Platform Div.] Data Analyst (5년 이상)

Seoul

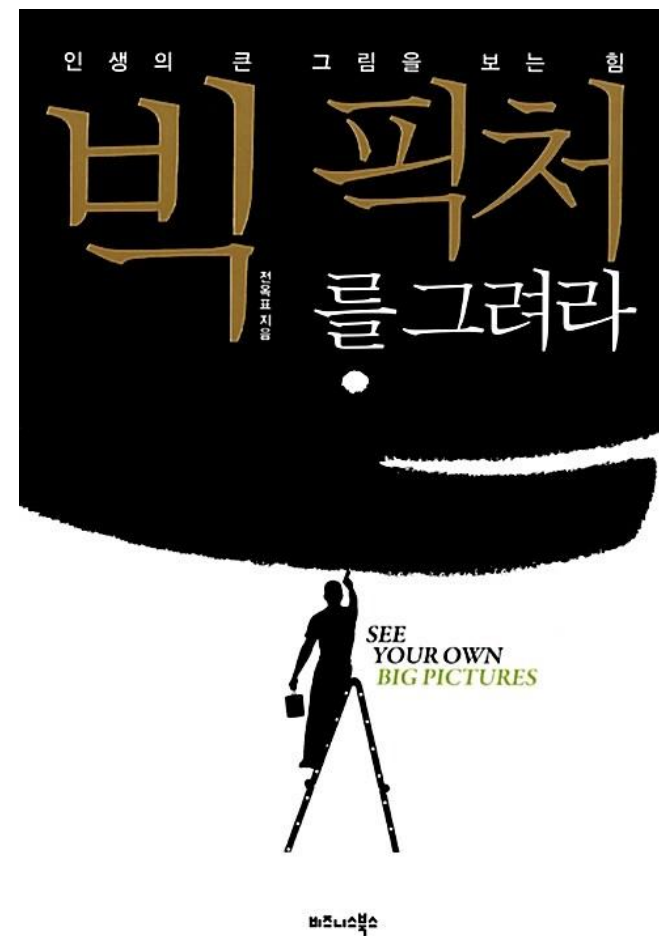
우리 팀과 함께할 미션을 소개합니다.

- 정의와 계산 기준이 일관된 **표준 KPI 프레임워크**를 설계하고 구현합니다.
- 분석 및 제품 팀과 협업하여, **비즈니스 로직을 탄탄한 데이터 아키텍처**로 전환합니다.
- 제품과 조직 전반에 걸쳐 **신뢰할 수 있고 확장 가능한 데이터 마트**를 구축하고 운영합니다.
- 표준화를 위한 **데이터 거버넌스 전략**을 수립 및 실행하고, 다양한 데이터 직군이 변화하는 환경에 적응할 수 있도록 가이드합니다.

이런 경험을 가진 분과 함께 성장하고 싶습니다! (필수요건)

- 데이터 분석 또는 데이터 엔지니어링 분야에서 **5년 이상의 실무 경험**
- 편의성과 확장성을 고려한 **데이터 모델링** 경험
- 가독성과 효율성을 고려한 **SQL 작성 능력**
- 기술·비기술적 이해관계자와 소통하며 **비즈니스 로직을 데이터 구조로 설계한 경험**
- Tableau, Power BI 등으로 **대시보드를 설계 및 개발한 경험**
- 컴퓨터공학, 데이터과학, 통계학, 수학 등 관련 전공의 학사 이상
- 해외 출장에 결격 사유가 없는 분

Objective

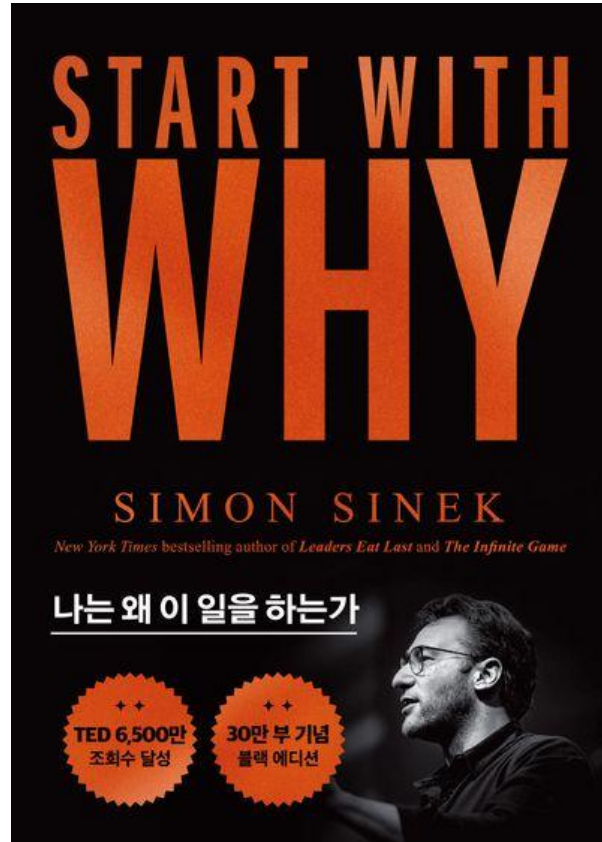


인과추론에 대한 **파편화된** 지식을 통합

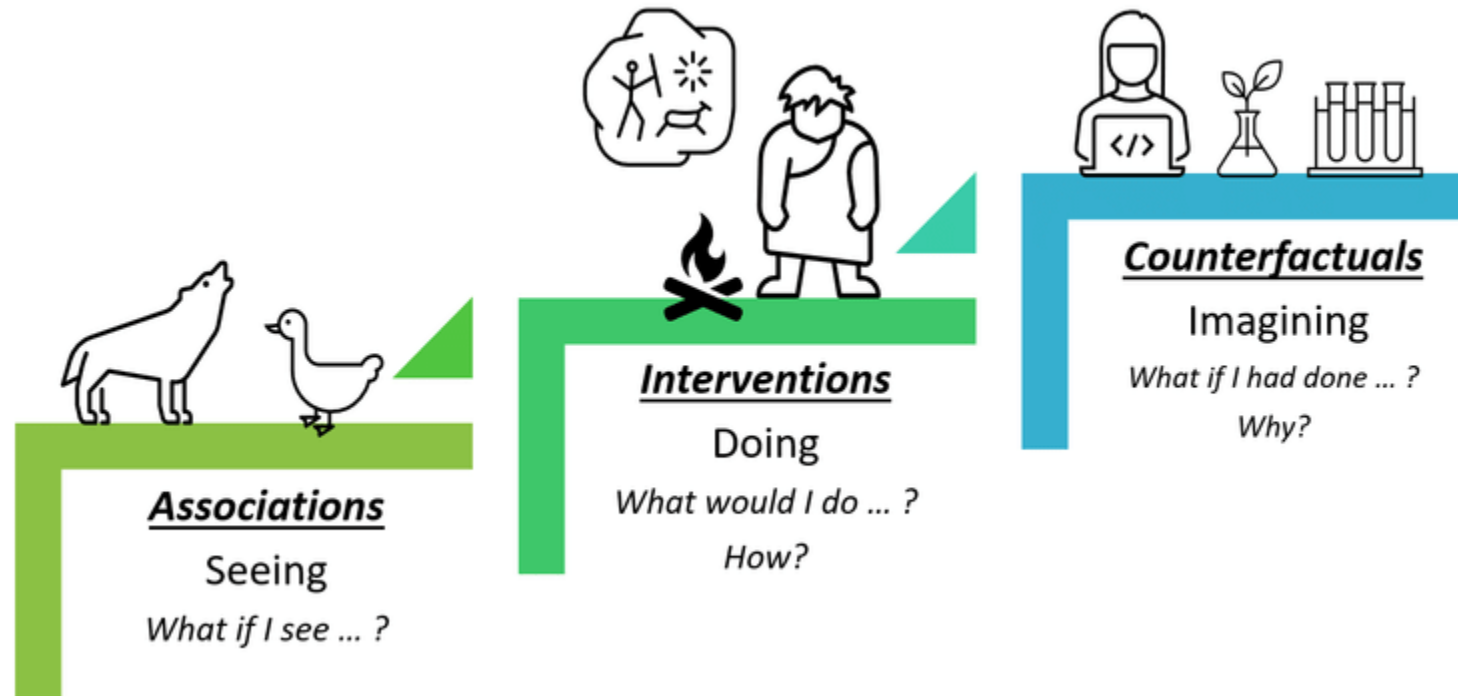
큰 그림에서 인과추론을 바라보기

이전 한 달 동안, 배운 내용을 떠올려보고

본인만의 인과추론 사고 흐름 & Framework 정리해보기



Ladder of Causal Inference



(1) 시간을 돌릴 수 없으므로, 동일 대상에 대해 서로 다른 처치를 동시에 관측 불가
(1.4.2절)

(2) 즉, 반사실적 결과(Counterfactual Outcome)가 존재한다

→ 반사실적 결과 : 실험군이 처치 받지 않았을 때의 결과

→ 사실적 결과 : 실험군이 처치 받았을 때의 결과

- 우리가 할 수 있는 최선은? 대조군을 실험군의 반사실과 최대한 비슷하게 모델링
- 인과추론에 대한 근본적인 문제가 있으니, 가정이 늘 수반됨

- 우리가 할 수 있는 최선은? **대조군을 실험군의 반사실과 최대한 비슷하게 모델링**

(1) 비슷하게 모델링 = 편향을 제거하기 (1.5절)

(2) 편향을 제거하는 이유? 처치 (Treatment)에 따른 인과효과를 식별

(3) 식별을 하기 위해서는 데이터 생성 과정을 이해하고 있어야 함

$$E[Y \mid T = 1] - E[Y \mid T = 0] = E[Y_1 - Y_0 \mid T = 1] + \left\{ E[Y_0 \mid T = 1] - E[Y_0 \mid T = 0] \right\}$$

관측된 처치 효과 인과 효과 편향

- 편향을 바라보는 시각
 - (1) 잠재적 결과 : 앞서 봤던 편향 (Missing Value)
 - (2) 구조적 인과 모형 : Backdoor Path
 - (3) 회귀분석 : 내생성 (Endogeneity)
- 이 책은 Potential Outcome Framework 중심으로 쓰여진 책
→ (4장) Potential Outcome Modeling → Meta

- 우리가 할 수 있는 최선은? **대조군을 실험군의 반사실과 최대한 비슷하게 모델링**

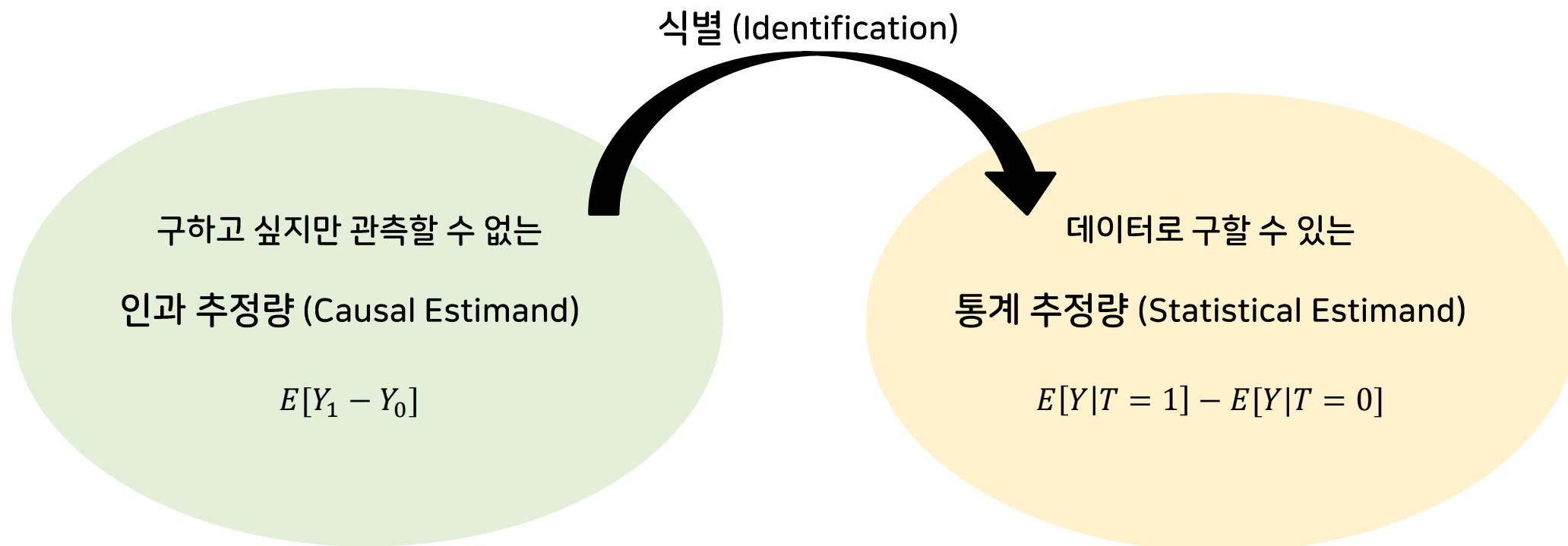
편향 제거 전

$$\underbrace{E[Y \mid T=1] - E[Y \mid T=0]}_{\text{관측된 처치 효과}} = \underbrace{E[Y_1 - Y_0 \mid T=1]}_{\text{인과 효과}} + \underbrace{\left\{ E[Y_0 \mid T=1] - E[Y_0 \mid T=0] \right\}}_{\text{편향}}$$

편향 제거 후 → 제거함으로써, 원하는 인과추정량을 식별하고 이를 추정할 수 있음

$$E[Y \mid T=1] - E[Y \mid T=0] = ATT = ATE = E[Y_1 - Y_0]$$

- 인과추론에서 식별 = 실험의 **내적 타당성 (Internal Validity)**를 확보하기 위한 과정



인과추론의 근본적인 문제에 다가가기

- 인과 추정량 이해하기 (1.4.8절)
- 개별 처치효과 (ITE, Individual Treatment Effect)
- 조건부 평균 처치효과 (CATE, Conditional Treatment Effect)

$$CATE = E[Y_{1i} - Y_{0i} \mid X = x]$$
- 평균 처치효과 (ATE, Average Treatment Effect)

$$ATE = E[Y_{1i} - Y_{0i}]$$

개인화



추정 비용/어려움



- 우리가 할 수 있는 최선은? 대조군을 실험군의 반사실과 최대한 비슷하게 모델링
- 인과추론에 대한 근본적인 문제가 있으니, **가정이 늘 수반됨 (1.6절)**

- 인과추론에 대한 근본적인 문제가 있으니, **가정이 늘 수반됨** (1.6절)
- **CATE**를 추정하는 경우, 다음과 같은 가정이 기본적으로 필요 (3.4절)
 - (1) Conditional Independence (=Unconfoundedness)
 - (2) Positivity – 통제하는 변수에 대한 비교 그룹의 실험 대상이 반드시 존재
 - (3) SUTVA + Consistency

- 조건부 독립성이 만족하는 경우, 떠올려야 할 내용은? **Adjustment Formula**
- 교란변수를 통제하면, ATE는 비교 그룹 간 그룹 내 차이의 **가중 평균**으로 **식별** 가능

$$ATE = E_X[E[Y \mid T = 1] - E[Y \mid T = 0]]$$

$$\begin{aligned} ATE &= \sum_x \{(E[Y \mid T = 1, X = x] - E[Y \mid T = 0, X = x])P(X = x)\} \\ &= \sum_x \{E[Y \mid T = 1, X = x]P(X = x) - E[Y \mid T = 0, X = x]P(X = x)\} \end{aligned}$$

- Why? 모델이 필요한 이유 (4.1절)
- 공변량 내에서 여러 그룹 내에서 ATE를 추정하고 각 결과의 평균을 구한다면 발생하는 문제 → **Sparsity (차원의 저주)**
- How? 잠재적 결과 ($E[Y|X, T]$)를 모델링

$$ATE = E_X[E[Y \mid T = 1] - E[Y \mid T = 0]]$$

$$\begin{aligned} ATE &= \sum_x \{(E[Y \mid T = 1, X = x] + E[Y \mid T = 0, X = x])P(X = x)\} \\ &= \sum_x \{E[Y \mid T = 1, X = x]P(X = x) - E[Y \mid T = 0, X = x]P(X = x)\} \end{aligned}$$

- 잠재적 결과 모델링 시, 가장 기본이 되는 모델 = Regression (회귀분석)
- 인과추론에서 회귀분석을 바라보는 관점 = Bias Adjustment
→ 이 책에서는 공변량 X 가 주어진 상황 (조건부)으로 회귀분석을 설명
- 회귀분석이 매력적인 이유 (4.4절)
 - Design-based Identification = Orthogonalization (=FWL)
 - Model-based Identification = Potential Outcome Modeling

- 디자인 기반 식별 - **처치 배정 매커니즘** 활용 (5.3.4절)
- 앞서 배운 2가지 방법 - IPW & Regression
 - **가중치 부여 방식**이 다르나, 공변량 주어졌을 때, CATE를 식별하는 방법
- IPW : Propensity Score 활용 (SCM 관점에 조금 더 가깝지 않나..?)
- Regression : 조건부 효과 추정 (Potential Outcome)

- Design-based vs Model-based Identification(5.4절)

→ 디자인 기반 식별 : $E[T|X]$ 모델링

→ 모델 기반 식별 : $E[Y|X]$ 모델링

$$\mu_0(x) = E[Y \mid T = 0, X]$$

$$\mu_1(x) = E[Y \mid T = 1, X]$$

- Mixture of both Identification = Doubly Robust Method (5.5절)

→ Debiased/Orthogonal Machine Learning (= R-learner)

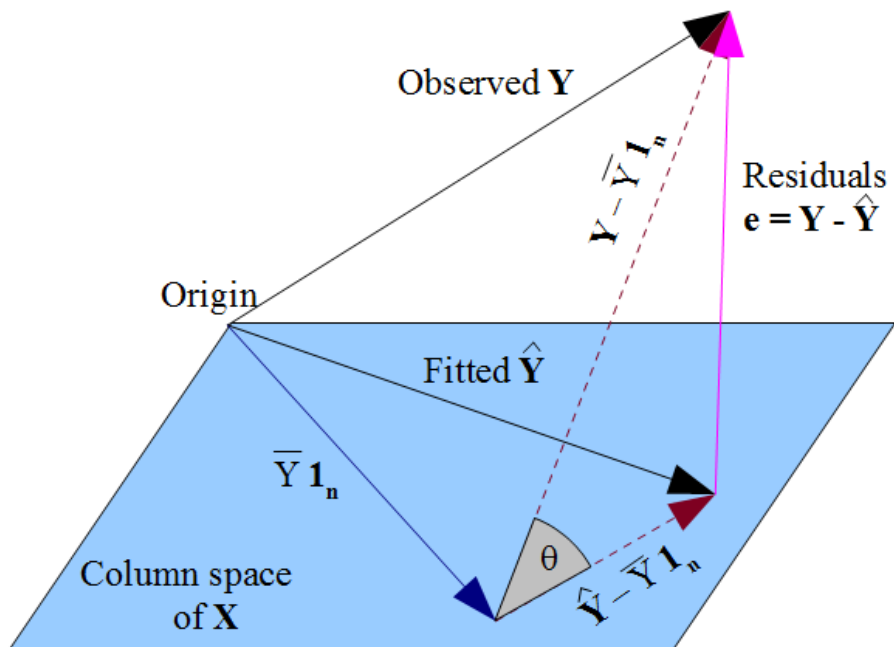
Doubly Robust Estimator (5.5절)

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

Double Machine Learning

$$Y_i - \hat{\mu}_y(X_i) = \tau \cdot (T_i - \hat{\mu}_t(X_i)) + \epsilon_i$$

- Design-based



- Model-based
(= Missing value Imputation)

$$ATE = \frac{1}{N} \sum (\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i))$$

S-Learner

$$\hat{\tau}(x)_i = M_s(X_i, T=1) - M_s(X_i, T=0)$$

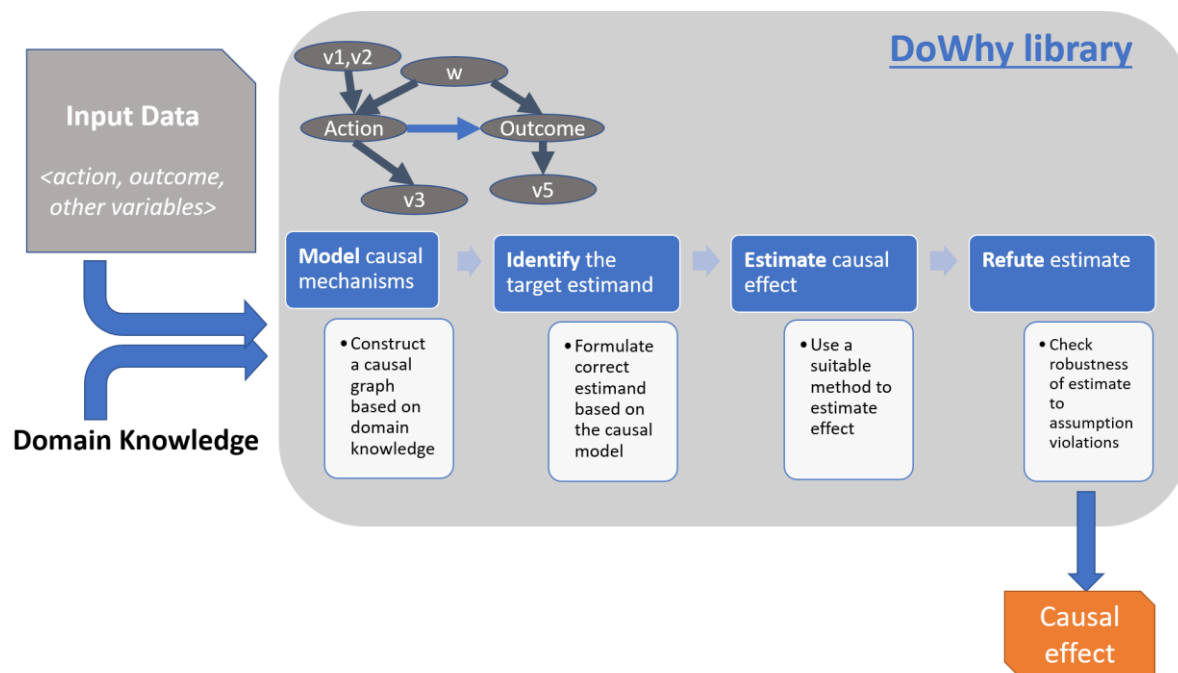
- 굳이 회귀분석으로만 잠재적 결과를 모델링해야 하나요? NO
- Adjustment Formula를 보면, 모델에 대한 제약을 하고 있지 않음!
- 우회적 (Regularization / Dimensionality Reduction 등)으로
 - (1) 보정 공식의 차원문제를 효과적으로 다룰 수 있는
 - (2) 유연한 모델을 활용할 수 있는 Machine Learning 활용이 가능!
- 메타몽처럼, Machine Learning을 활용할 수 있는 Meta-learners

- CATE 식별 Flow

(당연히 앞 단계의 문제 정의와 가설 설정은 필수적)

1. **데이터 생성 과정** - 공변량을 조건부로 두면, 편향을 제거할 수 있는가?
2. **식별** - 교란 요인을 통제하고, 인과효과를 식별을 위해서는 어떠한 가정이 필요한가?
3. **디자인** - 인과 추정량을 식별하려면, 어떠한 방법을 활용해야 적절한가?
4. **추정** - 데이터에 맞는 식별 방법을 선택했을 때, 통계적으로 추정할 수 있는가?

DoWhy Library Framework



- 데이터 타입에 따라 활용 전략이 달라질 수 있음 (7.1 / 7.2절)
 - (1) **Discrete** : S (Single Model), T (Two Model), X (Cross)
 - (2) **Continuous** : S, Double (=Orthogonal) ML
(QUIZ) 왜 Orthogonal ML 일까요?
- 기본적인 Framework
 - (1) 연구자가 선택한 Learner를 통해, 실험군 / 대조군 조건부 기댓값을 계산
 - (2) 각 그룹에서 구한 결과를 바탕으로, 조건부 기댓값의 차이 (CATE) 계산

X (Cross) Learner

1. 1st Step : T 러너와 동일하며, 반응 함수를 추정

$$\mu_0(x) = \mathbb{E}[Y(0)|X = x], \text{ and}$$

$$\mu_1(x) = \mathbb{E}[Y(1)|X = x],$$

2. 위 적합된 모형을 사용하여, 각 그룹에 대한 잠재적 결과 Imputation

$$\tilde{D}_i^1 := Y_i^1 - \hat{\mu}_0(X_i^1), \text{ and}$$

$$\tilde{D}_i^0 := \hat{\mu}_1(X_i^0) - Y_i^0,$$

3. 두 모델 중, CATE를 잘 근사한 모델에게 가중치 (성향점수) 부여하여 결합

$$\hat{\tau}(x) = g(x)\hat{\tau}_0(x) + (1 - g(x))\hat{\tau}_1(x),$$

Uber Use Case

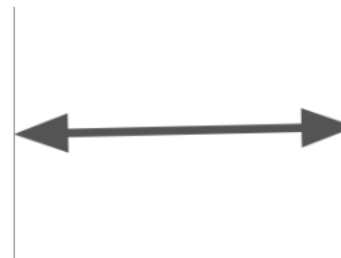
Business Problem

How do we connect business goals to marketplace decisions?

Uber Announces Results for First Quarter 2024

*Trips grew 21% year-over-year; MAPCs and monthly trips per MAPC grew 15% and 6% year-over-year, respectively
Gross Bookings grew 20% year-over-year and 21% year-over-year on a constant currency basis
Income from operations of \$172 million; Adjusted EBITDA of \$1.4 billion, up 82% year-over-year
Operating cash flow of \$1.4 billion; Free cash flow of \$1.4 billion*

SAN FRANCISCO – May 8, 2024 – Uber Technologies, Inc. (NYSE: UBER) today announced financial results for the quarter ended March 31, 2024.



Two-sided marketplace

Strategy

1. Demand Side – Dynamic Pricing / Demand Balance (Forecasting) / CX
2. Supply Side – Driver Retention (Availability Forecasting) / Rejection Behaviour
3. Operation Side – Matching / Incentive Design / Fraud
4. Governance (Policy) Side – Multi-Homing Behavior / Regional Laws

Background

How to allocate limited budget?

1. Marketplace Lever (to affect behaviour of drivers & riders)
2. Region (Promotion)
3. Previous Operation
 - (1) Manually Operated
 - (2) Unclear Objective
 - (3) Quantitatively immeasurable
 - (4) Lack of scalability

Business goal with constrained factors

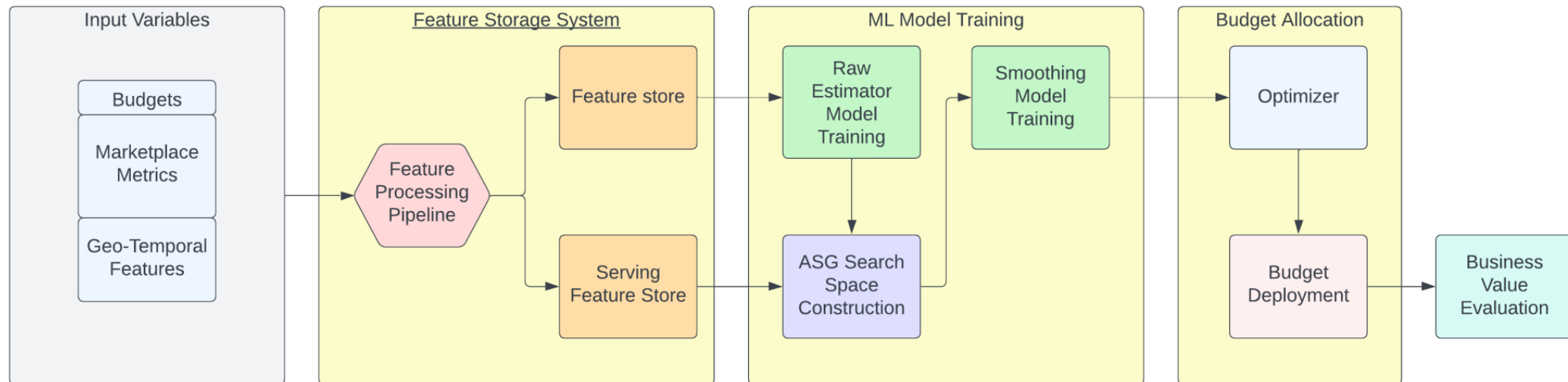
“제한된 예산 안에서 지역/시간/인센티브 유형별로 자원을 배분하여,
전체적으로 가장 큰 causal uplift를 얻는 최적의 조합을 찾는다”

$$\begin{aligned} &\text{maximize} && \sum_{w \in \text{weeks}, c \in \text{cities}} \text{obj}_{w,c}(\mathbf{b}_{w,c}) \\ &\text{subject to} && \sum_{w \in \text{weeks}, c \in \text{cities}, l \in \text{levers}} b_{w,c,l} = B \\ & && b_{w,c,l} \geq h_{w,c,l} \\ & && b_{w,c,l} \leq g_{w,c,l} \end{aligned}$$

- **Treatment** : Levers (Continuous)
- **Covariate** : Spatial / Temporal / Context Features
- **Outcome** : Business KPI (ex. ROI / Supply Surge Rate)
- **Granularity** : City / Week

Method

Present end-to-end causal machine learning to enhance budget allocation

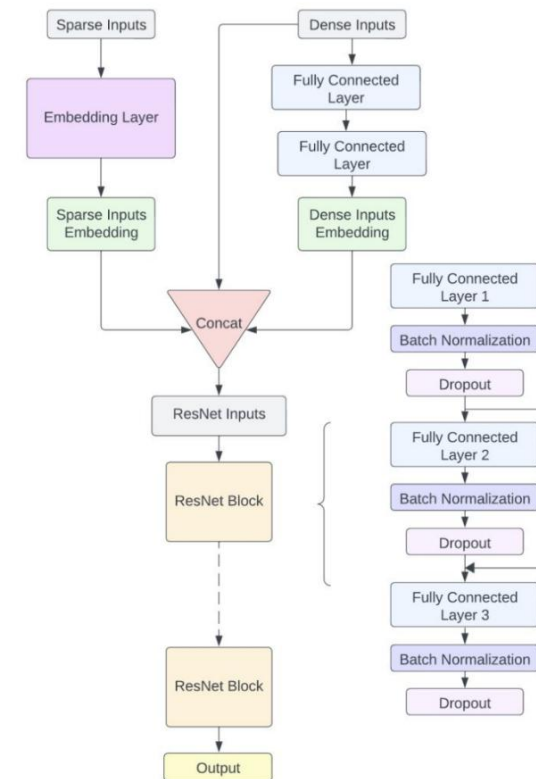


Approach

Two-stage Modeling Approach

Stage 1: User Causal Effect Estimator

- S-learner that models treated-unit effects
- Leverage A/B data which contains the information of user behaviors between treatment and control groups
- Model Structure:
 - Sparse and dense embeddings, concatenate them into trainable dense vectors as model input
 - ResNet blocks, better handle vanishing gradients and reduce the risk of overfitting

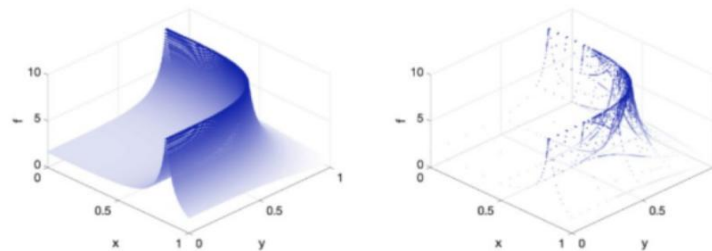


Approach

Two-stage Modeling Approach

Stage 2: Smoothing Regression Model

- Surrogate model learned from adaptive search grid (ASG) in control variable space
 - ASG allows us to learn multi-dimensional model with sublinear dependence on control dimensions instead of naive exponential
 - Marginalizing to control variables speeds up inference during optimization
- B-spline-based model allows us to easily impose shape control using both heuristics and experimental data
- Also allows us to control extrapolation behavior



2025 가짜연구소 인과추론팀 Conference

 **Meetup** 
for **Product & Data**

25.03.29(토) 오후 1시 30분
한빛미디어 리더스홀

Product / Data / Decision Making에
관심이 있는 실무자라면 참여 가능!

일시 2025.03.29 (토) 오후 1시 30분 ~ 5시 30분	프로그램 (1) 워크샵 - 알고싶다! 인과추론팀 & 후원사 (2) 세미나 - 알고싶다! 데이터의 모든 것 (마케팅 분석 / 데이터 엔지니어링 / 인과추론과 머신러닝 / 임상통계) (3) 네트워킹 : MBTI가 I인 사람도 할 수 있다!
장소 서울 서대문구 연희로2길 62 한빛미디어 사옥	
참가비 참가비가 없는 무료 행사!	
신청방법 (1) 인과추론팀 구성원 / 24년 행사 참여자라면? Eventus 바로 등록! (2) 인과추론팀이 처음이라면? 가짜연구소 Discord #인과추론과-실험 채널에 이름과 소속이 담긴 인사말 / 포스트 작성 후, Eventus 등록	

주관 : 가짜연구소 인과추론팀
후원 : 한빛미디어

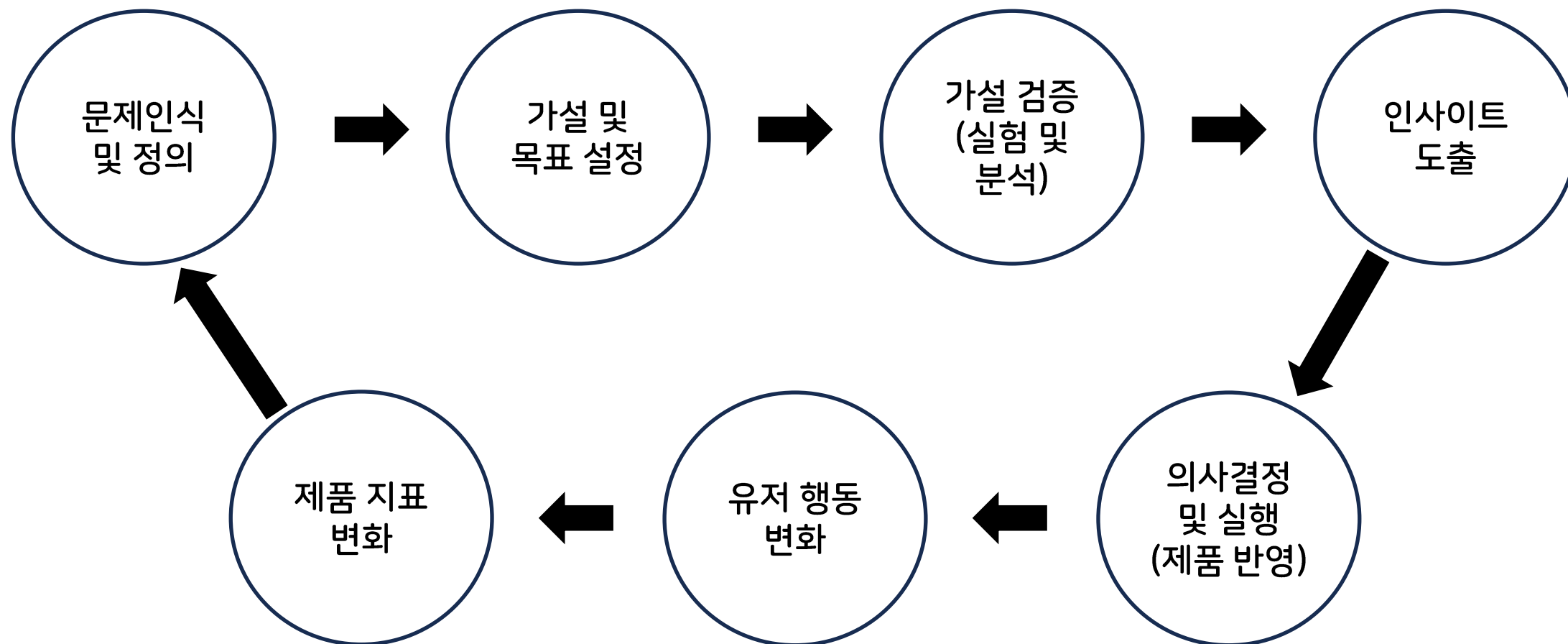
행사 FAQ

- 인과추론은 실무에서 **적용**이 어렵던데, 어떻게 활용하시나요?
- 배우는 단계인데요, 어디에 **적용**해야 할지 모르겠습니다.
- 인과추론으로 취업할 수 있나요?

인과추론은 문제 해결을 위해 올바른 질문을 할 수 있도록 안내

- 조직에서 문제를 해결하는 방법은 다양
- 가장 좋은 방법은 어느 누구의 리소스도 사용하지 않고 바로 해결할 수 있는 방법
- 만약 그렇지 않다면, 인과추론은 문제 해결을 위한 좋은 접근 방법 중 하나
- 인과추론은 단순 방법론 관점이 아닌, 문제 해결의 관점에서 활용되어야 함
- 인생은 늘 선택의 연속. 올바른 의사결정을 위해 인과적 사고는 도움되지 않을까요?

실무자 관점에서 문제 접근을 위한 Framework



Many Thanks!
