Purity's Test

Phuc Nguyen

Questions about Causal Inference

Data Scientist

December 22, 2024

Outline

Question 1

How to estimate optimal treatment assignment policies, as well as tools to evaluate the benefits of such policies?

Policies

A policy is a rule or method by which we decide who, on the basis of their characteristics, will be targeted to receive the intervention: $\pi : X \rightarrow W$

Heterogeneous treatment effects

Heterogeneous treatment effects (HTE) refer to variations in the effects of an intervention, treatment, or policy between different subgroups of a population or individuals within the population. These differences arise because the impact of a treatment may not be uniform; instead, it can depend on characteristics such as age, sex, socioeconomic status, baseline risk, or other observed and unobserved factors.

4 steps using machine learning for heterogeneous treatment effects

- Estimate conditional average treatment effects (CATE): estimate conditional average treatment effects using machine learning algorithms
- Create subgroups: create subgroups based on the predicted treatment effect strength
- Compare group characteristics: observe how subgroups vary across a range of characteristics
- Choosing an optimal policy: Is personalization worthwhile?

Tools to evaluate the benefits of such policies

Evaluating the benefits of an estimated policy in causal inference involves comparing the outcomes under the policy to a baseline or alternative policy. The evaluation ensures the policy improves desired outcomes while minimizing costs or adverse effects.

Tools to evaluate the benefits of such policies

Statistical and Machine Learning Libraries Here are some libraries but not all in Python

- EconML
- CausalML
- scikit-learn

Tools to evaluate the benefits of such policies

Specialized Causal Inference Platforms

- CausalNex
- Zepheira's Causal AI Framework
- Tetrad

Tools to evaluate the benefits of such policies

Visualization and Reporting Tools

- Tableau and Power BI
- Matplotlib/Seaborn/Plotly
- ggplot2 (R)

Tools to evaluate the benefits of such policies

Simulation and Synthetic Data Tools

- SimPy
- AnyLogic
- Synthetic Data Vault (SDV)

Tools to evaluate the benefits of such policies

Experimental and A/B Testing Tools

- Optimizely
- Google Optimize
- Microsoft Experimentation Platform (ExP)

Tools to evaluate the benefits of such policies

Cloud Platforms

- AWS Sagemaker Clarify
- Google Cloud AI
- Azure Machine Learning

Question 2

How to build a recommendation system aim to efficiently recommend policies for the personalized targeting needs?

We can use experiments to estimate optimal policies. The experiment is implemented in the following steps:

- Firstly, collect customer data such as weight, height, age, sex, shopping behavior, shopping budget, etc.
- Base on these customer's characteristics, we apply some machine learning algorithms to divide them into different subgroups
- For each subgroup, apply some policies
- Evaluate the benefit of each policy in every subgroup to find the optimal policies
- Collect more data and evaluate outcome effects and improve the system

Referrences

During doing this test, I have referred to these articles, YouTube courses and ChatGPT

- Miguel Hernán, Jamie Robins (April 26, 2024). *Causal Inference: What If*.
- https://www.youtube.com/watch?v=carmc4abRzA
- https://www.youtube.com/watch?v=eUGwd_pmLaA