

# Causal Inference: прозрение и практика

## Лекция 1. Основные понятия Causal Inference

Юрашку Иван Вячеславович

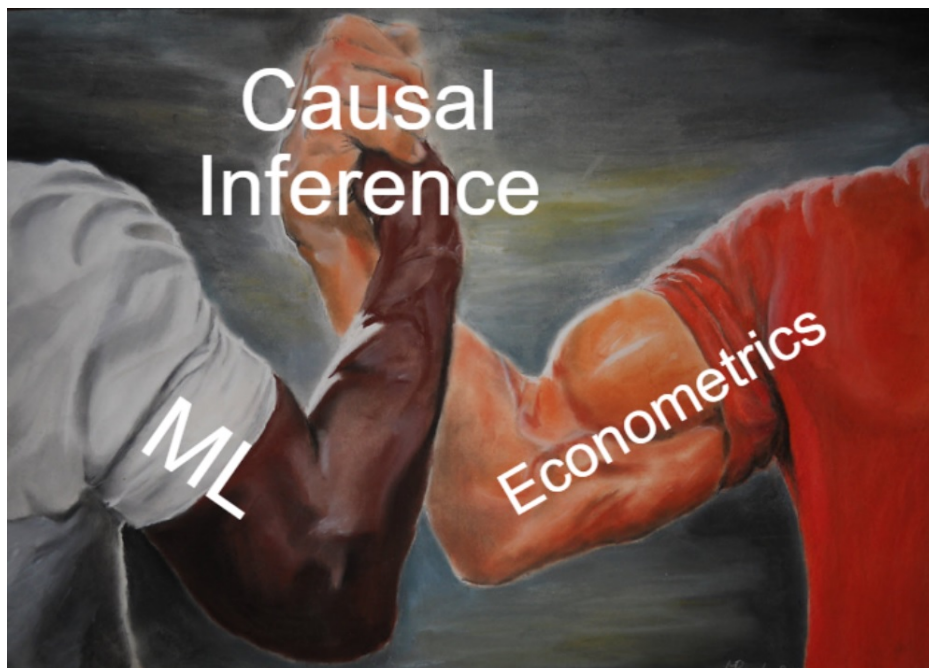
June 10, 2024

### Введение

В современном мире Data Science играет ключевую роль в анализе и использовании данных. Мы живем в эпоху, где информация - это золото, а умение извлекать из нее ценные знания - наша сила. Однако часто понятие Data Science ограничивается лишь алгоритмами машинного обучения или даже искусственным интеллектом, умаляя другие важные аспекты этой дисциплины.

Вот где начинается история о сближении двух мощных инструментов: эконометрики и Machine Learning. В разных эпохах они казались как бы двумя противоположными полярностями в анализе данных. Машинное обучение стремилось к высокой точности прогнозов, зачастую уступая интерпретируемости моделей. С другой стороны, эконометрика ставила акцент на интерпретируемость, понимание причинно-следственных связей, иногда уходя в тень из-за ограниченности моделей.

Однако со временем стало понятно, что для полного понимания данных нам нужно объединить эти подходы. И здесь на сцену выходит Causal Inference, или причинно-следственная связь. Этот инструмент помогает нам разгадывать причины за явлениями, объединяя преимущества как машинного обучения, так и эконометрики. Так, Judea Pearl в своей статье 2021 года подчеркивает важность CI как ключевого элемента для достижения баланса между эмпирическим и интерпретируемым.



Таким образом, Causal Inference — это область статистики и научных исследований, направленная на выявление и измерение причинно-следственных связей между переменными. Она помогает определить, какое воздействие оказывает изменение одной переменной на другую, отличая это воздействие от простых корреляций.

Погружение в мир причинно-следственной связи и машинного обучения не только расширит ваш кругозор, но и даст вам ключ к разгадке сложных и важных вопросов, стоящих перед современным обществом.

Допустим, вы владеете интернет-магазином и хотите понять, какие факторы влияют на продажи. С помощью методов причинно-следственной связи вы сможете определить, какие из ваших маркетинговых кампаний действительно приносят наибольший доход, и направить свои усилия и ресурсы в нужное русло.

Еще один пример - медицинская сфера. С помощью анализа причинно-следственных связей можно выявить, какие лечебные методы наиболее эффективны для конкретного заболевания, что позволит разрабатывать более точные и эффективные методики лечения.

Этот курс - не просто набор теории, он предлагает вам практические инструменты для анализа данных и принятия обоснованных решений. С его помощью вы сможете выйти за рамки обычных аналитических методов и раскрыть потенциал данных, лежащих у вас под рукой. Полученные знания не только помогут читателю в работе, но и дадут возможность вносить реальные изменения в мир вокруг нас.

## И все же Causal Inference - это не ML

Машинное обучение в настоящее время успешно решает задачи прогнозирования. Как подчеркивают Ajay Agrawal, Joshua Gans и Avi Goldfarb в книге "Prediction Machines":

"Новая волна искусственного интеллекта на самом деле приносит нам не интеллект, а важный компонент интеллекта - прогнозирование".

С машинным обучением мы можем совершать самые разнообразные и впечатляющие вещи. Главное требование заключается в том, чтобы сформулировать наши задачи как задачи прогнозирования. Хотите перевести текст с английского на португальский? Тогда создайте модель машинного обучения, которая предсказывает португальские предложения по английским. Хотите распознавать лица? Тогда разработайте модель машинного обучения, которая определяет наличие лица в определенной области изображения. Хотите создать автомобиль с автоматическим управлением? Тогда создайте модель машинного обучения, которая предсказывает направление поворота руля, а также давление на тормоза и акселератор при предоставлении изображений и данных с сенсоров, полученных из окружающей среды автомобиля.

Однако ML - не панацея. Он может производить чудеса в рамках строгих условий, но при этом может потерпеть крах, если данные немного отличаются от того, что модель привыкла видеть.

Машинное обучение известно своей неспособностью решать проблемы обратной причинности. Оно требует ответа на вопросы типа "а что, если", которые экономисты называют контрфактуальными. Как отмечается в "Prediction Machines", ML не справляется с такими задачами. Оно может предсказывать на основе данных, но не может оценить воздействие изменений. В качестве примера из книги "Prediction Machines":

"Во многих отраслях низкая цена ассоциируется с низкими продажами. Например, в гостиничной индустрии цены низки вне туристического сезона, а в период пикового спроса цены высоки и гостиницы полностью заполнены. Исходя из этих данных, наивное предположение может подсказать, что повышение цены приведет к увеличению числа проданных номеров".

По сути, ответ на вопросы о причинности является более сложной задачей, чем многие могут подумать. Это то, чему посвящен курс "Causal Inference: прозрение и практика". В нем мы исследуем, как использовать данные для изучения причинно-следственных связей и оценки воздействия вмешательств на результаты. Сперва определимся с тем, что конкретно мы хотим научиться делать.

## Постановка задачи и обозначения

Формализуем задачу следующим образом.

Пусть существует множество объектов, которые нас интересуют (обозначим его  $\mathcal{U}$  от слова universe). Изучаемым объектом может быть пациент, потенциальный клиент коммерческой компании, город — что угодно. Значение произвольного параметра  $X$  одного конкретного объекта  $i$  будем обозначать с добавлением верхнего индекса, а вектор значений, соответствующий этому параметру - будет употребляться без индекса:  $X^i$ ,  $\mathbf{X}$ .

Рассмотрим возможность воздействия на объект. В реальном мире это может включать в себя, например, лечение пациента, рекламную кампанию или введение административных ограничений, применяемых в определенных городах. Варианты воздействия практически неограничены. Мы будем представлять влияние в виде двоичного признака  $T_i$ , который может принимать значения 0 или 1, без учета интенсивности. Обозначим сравниваемые множества объектов как  $\mathcal{A}$  и  $\mathcal{B}$ .

В терминологии causal inference воздействие, которое исследуется, называется "treatment" (лечение), так как этот термин часто используется в медицинских испытаниях для оценки эффекта лечебного метода или медикамента на пациентов. Однако "treatment" может означать любое воздействие на часть исследуемой системы. В соответствии с нормами, обозначим этот параметр как treatment или  $T$  в формулах. Формально, воздействие - это разделение группы объектов на две части по бинарному признаку. На практике значимого воздействия может и не быть, и в таком случае мы имеем дело с фиктивным воздействием, что часто встречается.

Целевую переменную изучаемого объекта обозначим как "target" или  $Y$ . Обычно это вещественная величина, измеряемая в конкретный промежуток времени, часто в будущем. Например, при исследовании влияния сентябрьских SMS-оповещений на клиентов нас интересует, как это отразится на количестве их покупок в декабре. В этом случае количество покупок в декабре будет целевой переменной  $Y$ . Количество покупок в сентябре - это другая величина, называемая лаговым значением целевой переменной. Для определенности будем обозначать такие лаговые значения отдельным символом, например  $Y_{\text{lag } 3 \text{ months}}$ .

Представим, что для каждого изучаемого объекта существуют две параллельные вселенные, различающиеся только наличием воздействия на этот объект. Пусть мы можем узнать значения целевой переменной как при  $T = 1$ , так и при  $T = 0$ . Обозначим эти величины как  $Y_{i0} = Y_i|T=0$  и  $Y_{i1} = Y_i|T=1$ . Их разность называется "treatment effect" ( $TE_i = Y_{i1} - Y_{i0}$ ), которая представляет собой реальное отражение эффекта воздействия на объект  $i$ .

Кроме того, одна из этих вымышленных вселенных совпадает с реальной. Реальные значения называются factual, а параллельные им – counterfactual. Например, если на объект  $i_1$  в реальности воздействовали, то он принадлежит множеству  $\mathcal{A}$ , и его значения  $Y_{i_1 1}$  и  $Y_{i_1 0}$  будут  $Y_{i_1}^{\text{factual}}$  и  $Y_{i_1}^{\text{counterfactual}}$  соответственно.

Для каждого объекта существует свой  $TE_i$ . В рамках решаемой задачи наш истинный интерес заключается в том, чтобы оценить так называемую величину **среднего эффекта от воздействия** на множество  $\mathcal{M}$ , или average treatment effect ( $ATE_{\mathcal{M}}$ ).

$$ATE_{\mathcal{M}} = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} TE_i = \mathbb{E} TE = \mathbb{E} (Y_1 - Y_0).$$

Поскольку множество  $\mathcal{M}$  (matter objects), по которому мы усредняем, не всегда совпадает с множеством  $\mathcal{U}$  всех изучаемых объектов, будем уточнять множество усреднения с помощью нижнего индекса, когда это необходимо. Кроме того, существуют прижившиеся понятия и обозначения:

- $\mathcal{U}$  — множество всех объектов нашего исследования.
- $T$  — бинарная переменная — индикатор принадлежности объекта к тестовой группе.
- $Y$  — целевая метрика
- $\mathcal{A}$  — контрольная группа,  $T = 0$ .

- $\mathcal{A}$  — целевая группа,  $T = 1$ .
- $Y_i^{(0)}$  — отклик объекта  $i$ , если бы он находился в тестовой группе.
- $Y_i^{(1)}$  — отклик объекта  $i$ , если бы он находился в целевой группе.
- $ATE = ATE_{\mathcal{U}}$
- $ATC = ATE_{\mathcal{A}}$  - эффект на контрольную группу
- $ATT = ATE_{\mathcal{B}}$  - эффект на целевую группу

## Если мы знаем *contrfactual*

Рассмотрим синтетический пример.

\*\*\* пример в табличном виде с примером импорта \*\*\*

Как мы можем здесь увидеть,

$$ATE_{\mathcal{A}} = -50$$

$$ATE_{\mathcal{B}} = -50$$

$$ATE_{\mathcal{A} \cup \mathcal{B}} = -50$$

Нам повезло: имея доступ к *contrfactual* значениям, мы легко и точно определили  $ATE$ . Однако в реальности мы не можем измерить величины из параллельных вселенных. Поэтому были разработаны методы аппроксимации  $ATE$  на основе доступных данных. Рассмотрим самый наивный из этих методов.

Нам повезло. Имея в наличии *contrfactual* значения мы разобрались, что к чему. Легко и идеально точно решили задачу поиска  $ATE$ . К сожалению, в действительности, нам недоступен замер величин из параллельных вселенных. Поэтому были изобретены методы аппроксимации  $ATE$  по имеющимся данным. Рассмотрим самый наивный из них.

## Simple mean difference method

Давайте рассмотрим классический пример, иллюстрирующий, что иногда очевидные выводы оказываются ошибочными. Возьмём два госпиталя: один действовал уже много лет, когда был построен второй. Новый госпиталь был оснащён передовыми технологиями и привлёк лучших специалистов. Однако в процессе времени выяснилось, что средний уровень смертности во втором госпитале значительно превысил показатели первого.

\*\* пример данных и вычислений  $ATE$  наивным методом \*\*

Причиной этого стало то, что новый медицинский центр привлекал преимущественно пациентов с более тяжёлыми формами заболевания.

Это статистическое смещение (*bias*) в распределении пациентов искажало общую картину. Из-за чего нельзя было делать выводы на основании прямого сравнения смертностей в медицинских центрах.

Давайте попробуем устранить эти различия в распределениях, сделав данные однородными по тяжести заболеваний, и повторим наш эксперимент.

\*\*\* повторяем эксперимент \*\*\*

Видим, что значение  $ATE$  стало более правдоподобным. В следующей статье мы более детально рассмотрим класс подобных методов.