

Causal Inference: прозрение и практика

Лекция 1. Основные понятия Causal Inference

Юрашку Иван Вячеславович

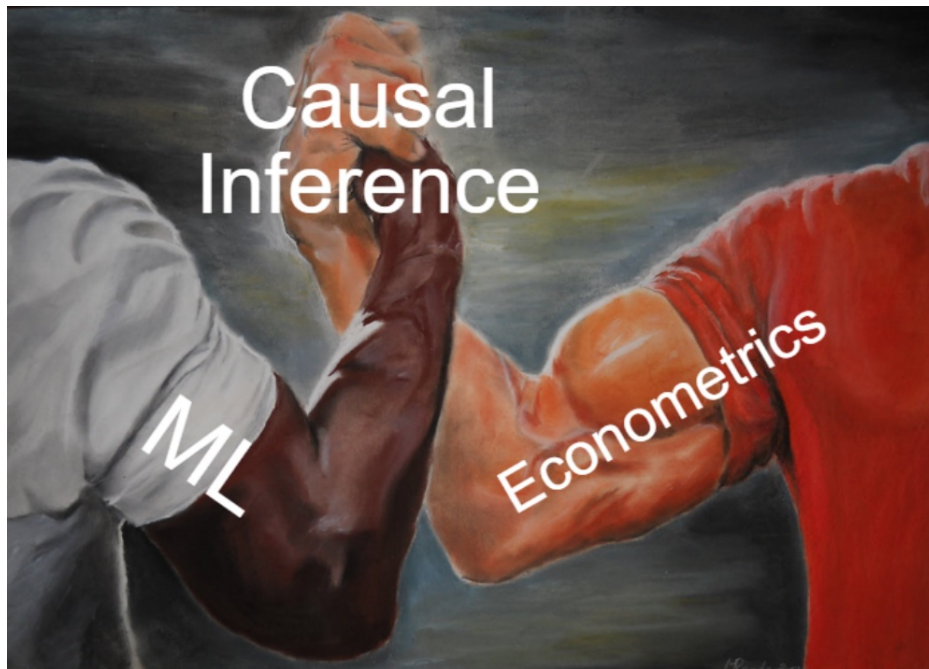
1 апреля 2024 г.

Что такое Causal Inference.

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.



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 Аяу 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. Поехали!

"Correlation is not causation"