Модели оценки индивидуального эффекта лечения

Уткин Л.В.

Высшая школа технологий искусственного интеллекта

Задача оценки эффекта лечения

- Наблюдения в двух группах: контрольная (control) и после лечения (treatment)
- **Контрольная группа**: датасет $\mathcal{D}_0 = \{(\mathbf{x}_i^0, y_i^0, T_i = 0)\}_{i=1}^c$, $y_i \in \mathbb{R}$ - наблюдение (время до ремиссии).
- **Леченная группа**: датасет $\mathcal{D}_1 = \{(\mathbf{x}_j^1, y_j^1, T_j)\}_{j=1}^t$, $T \in \{1, ..., m\}$ - доза воздействия
- **Цель**: построить модель и обучить ее на $\mathcal{D}_0$ и $\mathcal{D}_1$, чтобы для нового пациента $\mathbf{x}$ оценить эффект лечения

Показатели эффекта лечения

- Средний эффект лечения (**ATE** - *Average Treatment Effect*):

$$ATE = \mathbb{E}_{P(X)}[Y_1 - Y_0]$$

- Индивидуальный эффект лечения (**CATE** - *Conditional Average Treatment Effect*):

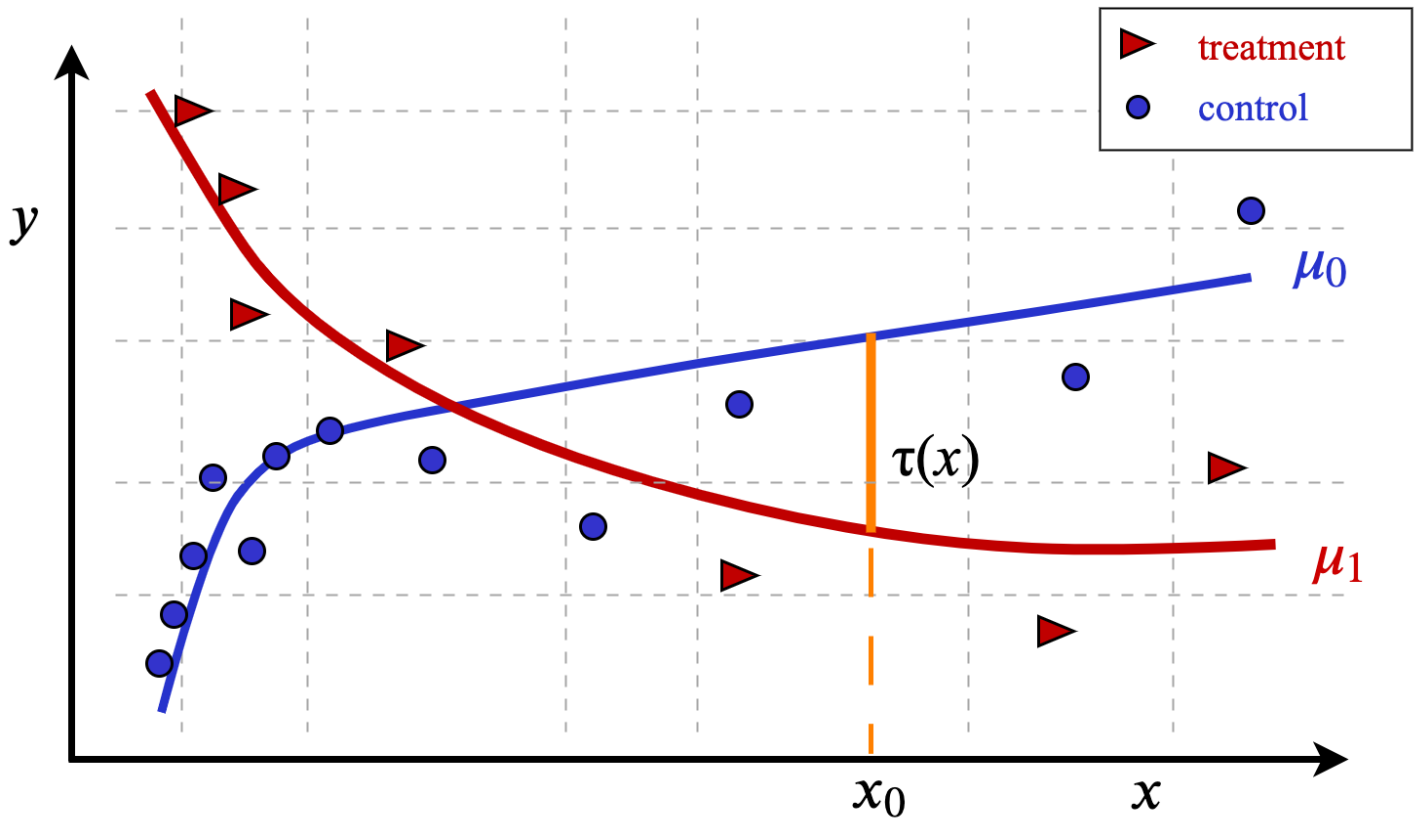
$$\tau(\mathbf{x}) = \mathbb{E}[Y_1 - Y_0 \mid \mathbf{X} = \mathbf{x}]$$

здесь мат.ож. берется по распределению потенциальных исходов внутри фиксированной подгруппы $\mathbf{X} = \mathbf{x}$

- Вероятность положительного лечения (**IPTB** - *Individual Probability of Treatment Benefit*):

$$\Psi(\mathbf{x}) = \Pr(Y_1 > Y_0 \mid \mathbf{X} = \mathbf{x})$$

Иллюстрация задачи оценки эффекта лечения



Проблема оценки эффекта лечения

- **Главная проблема**: мы никогда не видим индивидуальный эффект лечения:
  - Для одного человека мы знаем либо его результат с лечением, либо без него — но не оба одновременно
  - Это **фундаментальная** проблема причинно-следственного вывода (causal effect).
  - CATE - это разница $Y_1 - Y_0$ для конкретного человека с признаками $\mathbf{x}$, но Y_1 и Y_0 никогда не встречаются в данных вместе.

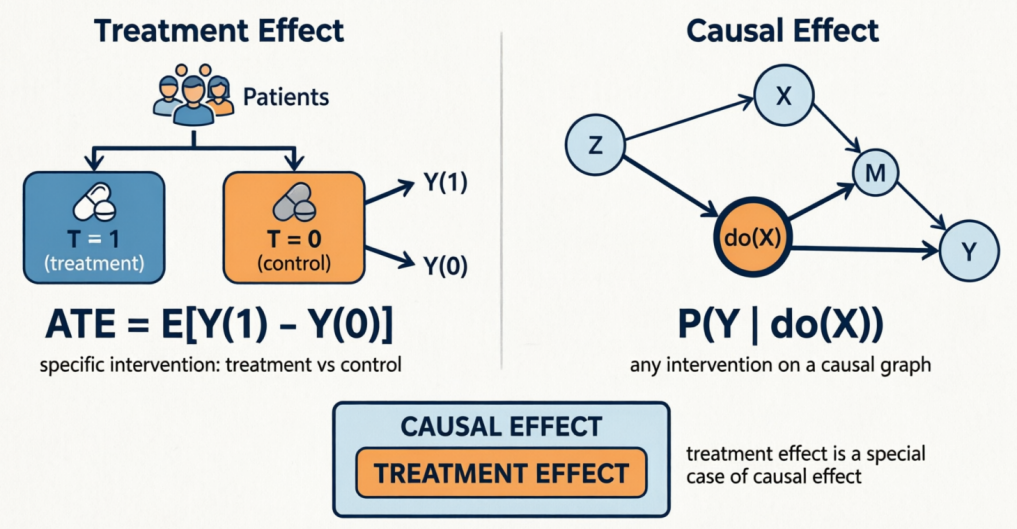
Фактические и контрфактические данные

- Пациент принял новое лекарство и через неделю выздоровел
  - **Фактические данные** *(наблюдаемый исход)*: он выздоровел после приема лекарства
  - **Контрфактические данные** *(ненаблюдаемый исход)*: это его состояние, если бы он не принимал лекарство
- **Варианты контрфакта**:
  - Он бы все равно выздоровел, например, без лекарства
    - **Вывод**: Лекарство не было причиной выздоровления
  - Он бы не выздоровел
    - **Вывод**: Лекарство, вероятно, было причиной выздоровления (истинный эффект).

Зачем знать фактические и контрфактические данные?

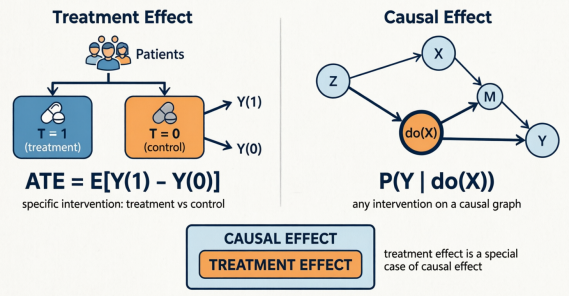
- Без контрфакта (counterfactual) нельзя отличить причину от совпадения, естественного хода событий или эффекта чего-то другого
- Факт показывает, что произошло
- Контрфакт показывает, благодаря чему это произошло
- Причинность - это всегда сравнение факта с тем, чего не было, но что могло бы быть

Treatment effect и Causal effect (небольшое отступление)



- Эффект лечения - это сравнение того, что случилось с группой, получившей лекарство, и с группой, которая его не получила (это приближение к контрфакту).
- В Causal Effect вместо пассивного наблюдения мы делаем интервенцию - оператор $do(X)$.
- $T \to X$, Y - исход, Z и M - посредники (mediators) или сопутствующие факторы

Что означает "делаем интервенцию do(X)"?

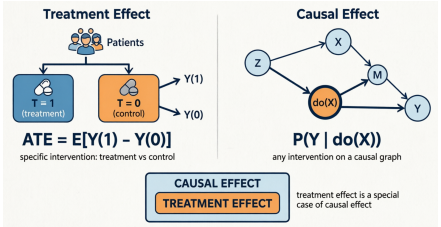


- Это означает: мы принудительно задаем значение X, **разрывая** все естественные связи, которые на него влияют.
- И смотрим на распределение исхода: $P(Y \mid do(X = x))$ - вероятность Y при условии, что мы насильно установили $X = x$.
- **Смысл**: Causal Effect - это влияние ЛЮБОГО вмешательства на ЛЮБУЮ переменную в любой сложной системе, а не только “таблетка vs плацебо”

Пример интервенции

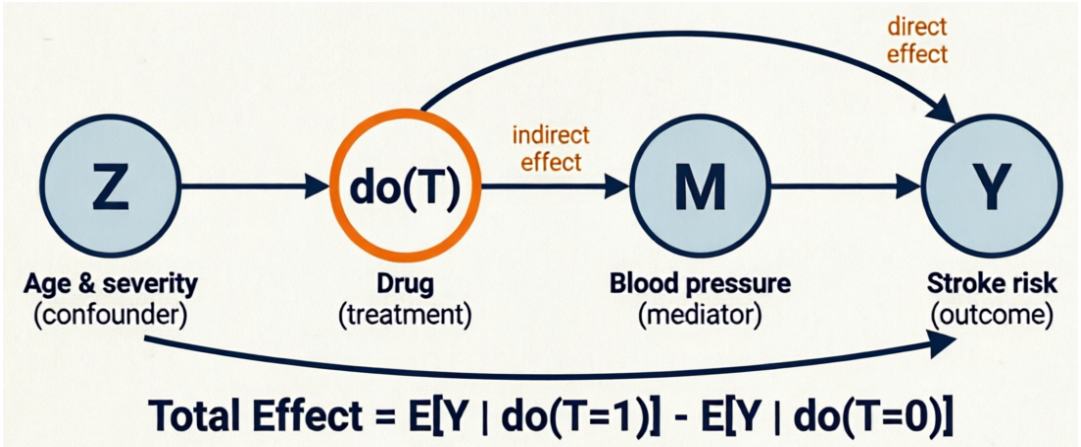
- P(шторм | барометр низкий) высокая: наблюдая низкий барометр, мы предсказываем шторм.
- P(шторм | do(барометр низкий)) = P(шторм): если принудительно перевести стрелку барометра, погода не изменится. Наблюдение использует **корреляцию** через общую причину (давление), а интервенция ее разрушает.
- M на графе — это «передатчик» влияния X на Y. Он показывает, как именно X вызывает Y
- Лекарство (X), артер-е давл-е (M), риск инсульта (Y): X снижает Y через снижение давления M

Что означает "разрезаем связи или граф"?



- В у переменной X были естественные причины, например, Z (болезненность, рекомендации врача).
- $do(X)$ удаляет все входящие стрелки в X (перестаем учитывать, почему пациент принял лекарство).
- Фиксируем $X = 1$ (принял) или $X = 0$ (не принял) искусственно.
- Смотрим, как это меняет Y, уже **без влияния посторонних факторов**.
- Этим устраняем искажения (confounding), имитируя рандомизированный эксперимент.

Treatment effect и Causal effect



- Снижает ли препарат риск инсульта, и как именно?
- Граф отвечает на вопросы: работает ли препарат вообще (total effect), за счет какого механизма (непрямой эффект через M и прямой эффект без M).

Treatment effect и Uplift Modeling

Treatment Effect отвечает на вопрос: “На сколько именно изменится выручка/здоровье этого конкретного покупателя/пациента, если мы дадим ему скидку/лекарство?”
“Каким будет Y, если я применю лечение T к объекту с признаками X?”
Uplift Modeling:“Кому из 10 миллионов клиентов мы отправим промокод, чтобы максимизировать чистый доход маркетингового канала?”
“На сколько изменится Y исключительно из-за того, что я применил лечение?”

Следствия фундаментальной проблемы

- Лечение обычно получают не случайно (тяжелые пациенты лечатся чаще).
- Если сравнить средние у вылеченных и невылеченных, то оценивается не эффект лечения, а разница в состоянии до лечения:

$$\mathbb{E}[Y_1 \mid \mathbf{X} = \mathbf{x}] - \mathbb{E}[Y_0 \mid \mathbf{X} = \mathbf{x}] \neq \tau(\mathbf{x})$$

Проблема скрытых конфаундеров (Unobserved Confounder)

- **Конфаундер** (спутывающая переменная) - это фактор, который влияет одновременно на причину и на следствие, из-за чего возникает ложная связь между ними
- Скрытый конфаундер - это переменная U, которая одновременно влияет:
  1. На то, получит ли человек лечение T
  2. На исход болезни Y_1 or Y_0
- Переменной U нет в датасете, мы ее не знаем, не измеряли, модель ее не видит

Проблема скрытых конфаундеров (пример)

- Увеличивает ли прием витаминов (T) продуктивность сотрудников (Y)
- Группа А принимает витамины каждый день.
- Группа Б не принимает витамины.
- Собираем данные и видим: средняя продуктивность в группе А выше, чем в группе Б.
- **Вывод**: витамины повышают продуктивность!!!
- **Но** есть скрытый фактор U - уровень дисциплины
- Влияние U:
  - на лечение T: T растет - дисциплинированные не забывают пить витамины
  - на исход Y: Y растет - дисциплинированные не опаздывают, продуктивность растет и без витаминов
- В итоге модель увидит: Истинная причина эффекта - это скрытый фактор U

Условия для оценки CATE

- Чтобы оценить $\tau(\mathbf{x})$ нужно сравнить группы. Но как узнать, что различие между группами вызвана лечением, а не изначальными различиями людей?

***Условие Ignorability / Unconfoundedness**: в наблюдаемых данных лечение должно варьироваться случайным образом относительно ненаблюдаемых факторов:*

$$(Y_1, Y_0) \perp T \mid X$$

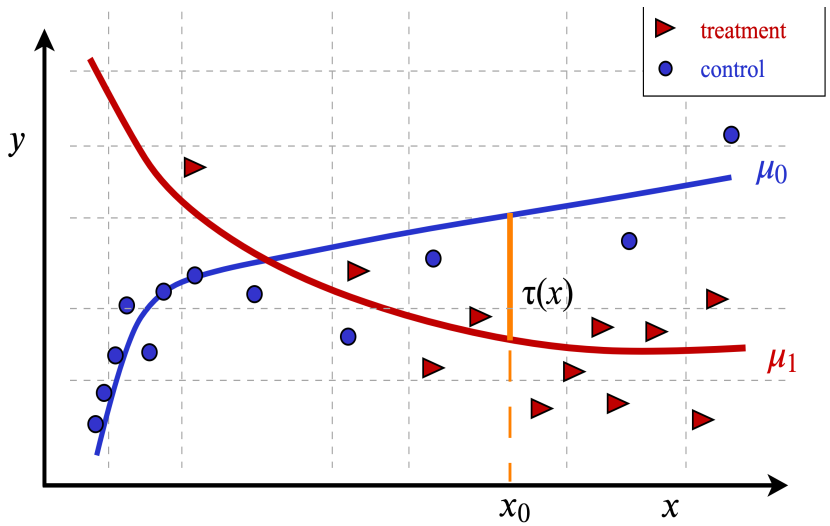
Y_1, Y_0 статистически не зависят от фактически назначенного лечения T при условии наблюдаемых призаков X

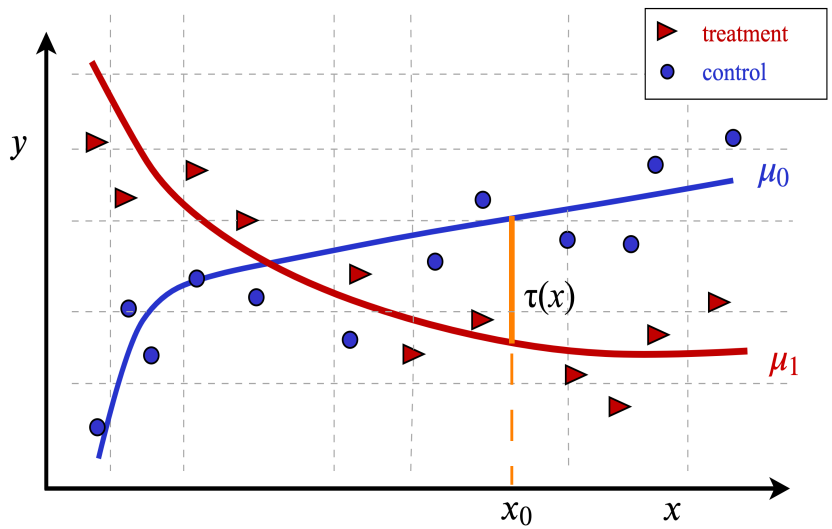
Условие Ignorability / Unconfoundedness другими словами

$$(Y_1, Y_0) \perp T \mid X$$

- Если зафиксировать признаки пациента X, то знание того, получил ли он лечение $T = 1$ или нет $T = 0$, не дает нам никакой доп. информации о том, какими были бы его потенциальные исходы Y_1 и Y_0
- Если условие нарушается (например, лечение назначают только тяжелым больным), то, сравнивая выживаемость, получим смещенную оценку

Условие Ignorability / Unconfoundedness





Propensity score

- Склонность (**propensity score**)
  $e(\mathbf{x}) = P(T = 1 \mid X = \mathbf{x})$
- **Условие перекрытия (Overlap)**: $\forall \mathbf{x},\ 0 < e(\mathbf{x}) < 1$.
- **Условная независимость** или **идентификация эффекта:** выполняется условие перекрытия и unconfoundedness, CATE выражается как:

$$\mathbb{E}[Y_1 - Y_0 \mid X = \mathbf{x}] = \mathbb{E}[Y \mid T = 1, X = \mathbf{x}] - \mathbb{E}[Y \mid T = 0, X = \mathbf{x}]$$

- Если лечение не варьируется (скажем, $T = 1$ для всех $\mathbf{x}$), то $\mathbb{E}[Y \mid T = 0, X = \mathbf{x}]$ не определено. Эффект математически не идентифицируем.

Что делает propensity score?

- **Балансировка**: используется $e(\mathbf{x}) = P(T = 1 \mid X = \mathbf{x})$
- Y - фактический наблюдаемый исход: $Y = T \cdot Y_1 + (1 - T) \cdot Y_0$
- Видим только один из двух исходов
- Вводим весовую переменную $w(T, \mathbf{x}) = \frac{T}{e(\mathbf{x})} + \frac{(1-T)}{1-e(\mathbf{x})}$

Утверждение: *Если* $e(\mathbf{x}) \in (0, 1)$ *известна, то для взвешенной по IPW выборке*

$$\mathbb{E}\left[\frac{T \cdot Y}{e(\mathbf{x})} \mid X = \mathbf{x}\right] = \mathbb{E}\left[Y_1 \mid X = \mathbf{x}\right]$$

$$\mathbb{E}\left[\frac{(1 - T) \cdot Y}{1 - e(\mathbf{x})} \mid X = \mathbf{x}\right] = \mathbb{E}\left[Y_0 \mid X = \mathbf{x}\right]$$

Еще про propensity score

- Оценка CATE (**Inverse Propensity Weighting** (IPW)):

$$\widehat{\tau}(\mathbf{x}) = \mathbb{E}\left[\frac{T \cdot Y}{e(\mathbf{x})} + \frac{(1-T)\cdot Y}{1-e(\mathbf{x})} \mid X = \mathbf{x}\right]$$

- Для каждого набора признаков X должны существовать оба варианта (и с лечением, и без)
- **Нюанс**: Оценка $\widehat{\tau}(\mathbf{x})$ корректна только при $n \to \infty$ и когда $e(\mathbf{x})$ известна точно

Используя $e(\mathbf{x})$ как вес, мы "обманываем"статистику, делая группы леченых и нелеченых максимально похожими, чтобы потом честно сравнить их исходы

Теорема Розенбаума-Рубина

- Если сравнить двух людей с одинаковым $e(\mathbf{x})$, то распределение всех их признаков X (возраст, пол, болезни) в среднем одинаково

Теорема Розенбаума-Рубина *(Rosenbaum & Rubin, 1983): Если выполнено условие игнорируемости (Unconfoundedness / Ignorability) и условие перекрытия (Overlap), то справедливо:* $Y_1, Y_0 \perp T \mid e(X)$

- $e(\mathbf{x})$ является "достаточной статистикой" для балансировки всех наблюдаемых признаков X
- Если уравняли группы леченых и нелеченых по $e(\mathbf{x})$, то это эквивалентно тому, что уравняли их по всем признакам X (возрасту, полу, болезням и т.д.) одновременно

О теореме Розенбаума-Рубина (1)

- Влияние препарата A (T) на давление (Y)
- В датасете 1000 пациентов с признаками X =(Возр., Вес, Уровень холестерина, Наслед-ть)
- **Контролировать** X - значит сравнивать пациентов строго одинаковых по ВСЕМ 4 признакам
- Двое мужчин: 45 лет, вес 80 кг, холестерин 6.0, наслед-ть есть. Одному дали A, второму - нет
- Сравниваем их давления и говорим: «A дает эффект для этой группы»
- Если признаков 4, это легко. Но если их 100, то найти двух человек с одинаковыми 100 признаками практически невозможно

О теореме Розенбаума-Рубина (2)

- Теперь можно не смотреть на признаки, а только на одно число - $e(\mathbf{x})$
- Допустим, обучили модель и выяснили, что для пациентов с набором признаков (45 лет, 80 кг, холестерин 6.0, наслед-ть) вероятность получить лечение составляет $e(\mathbf{x}) = 0.75$
- Теперь находим другого пациента, у которого совершенно другие признаки (60 лет, 100 кг, холестерин 4.0, без наследственности), НО ... у него тоже $e(\mathbf{x}) = 0.75$

Что об этом говорит теорема Розенбаума-Рубина

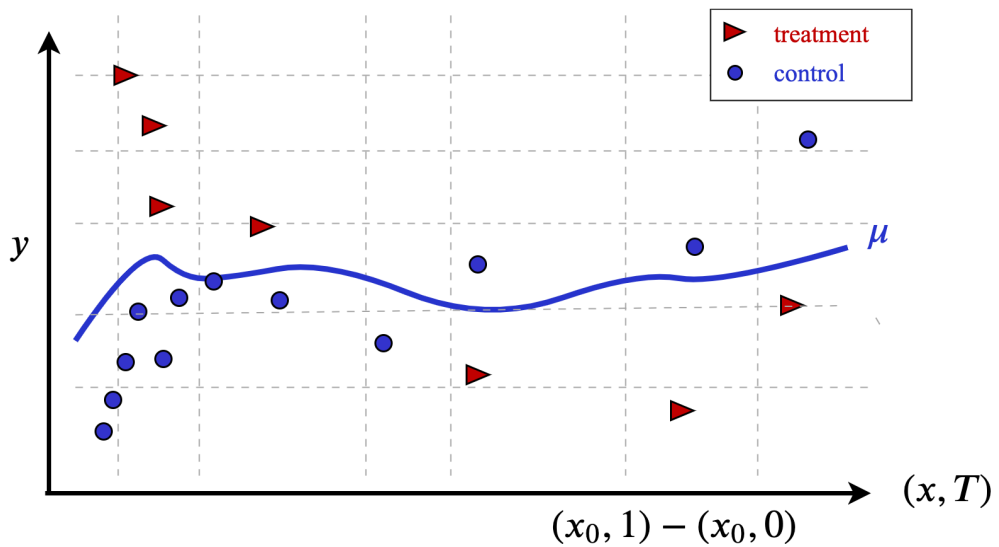
- Если набрать 1000 пациентов со склонностью 0.75, то среди них окажется примерно равное количество старых и молодых, курящих и некурящих. Усреднив их результаты, получим тот же ответ, как если бы сравнивали идеальных “близнецов”
- Если у двух людей одинаковый шанс получить лечение, значит логика назначения лечения для них была одинаковой. А раз логика была одинаковой, значит их совокупные характеристики (возраст, вес, болезни) в среднем уравновешены

Мета-модели оценки CATE

- **Мета-модели** (**meta-learners**) - семейство алгоритмов, которые решают задачу оценки CATE путем декомпозиции ее на несколько более простых подзадач, каждая из которых решается стандартными методами обучения с учителем.
- В качестве базовых моделей можно использовать любой алгоритм машинного обучения: от линейной регрессии до градиентного бустинга или нейросетей

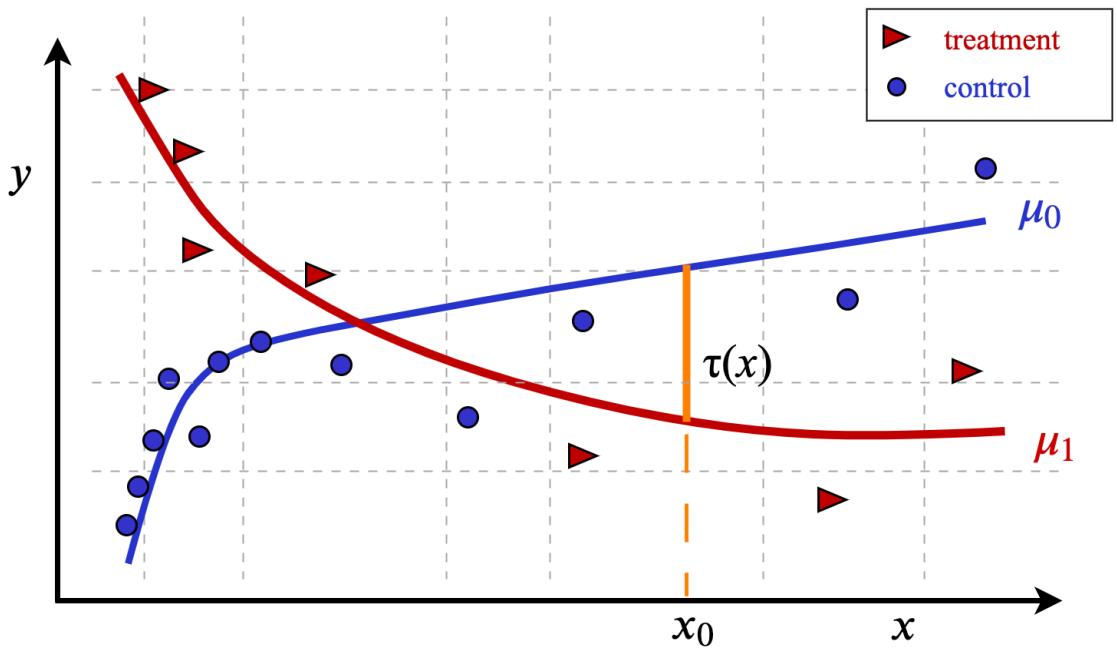
Мета-модели (S-learner)

- Kunzel, S.R. et al. Metalearners for estimating heterogeneous treatment effects using machine learning. 2019
- Обучается одна модель регрессии на всем датасете, лечение T включается как еще один признак наряду с признаками X: $\mu(X, T) = \mathbb{E}[Y \mid X, T]$
- Тогда $\tau(\mathbf{x}) = \mu(\mathbf{x}, 1) - \mu(\mathbf{x}, 0)$



Мета-модели (T-learner)

- Kunzel, S.R. et al. Metalearners for estimating heterogeneous treatment effects using machine learning. 2019
- Обучаются две отдельные модели: $\mu_0(\mathbf{x})$ на controls и $\mu_1(\mathbf{x})$ на treatments
- Тогда $\tau(\mathbf{x}) = \mu_1(\mathbf{x}) - \mu_0(\mathbf{x})$

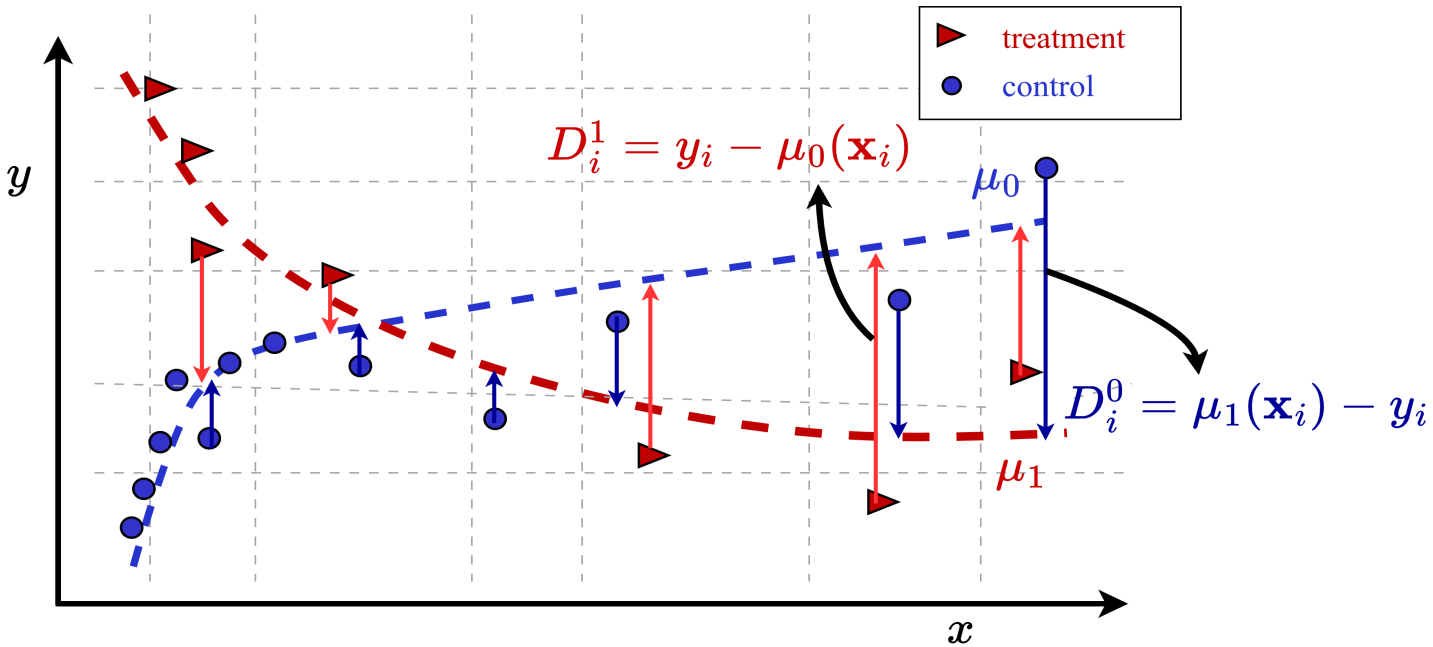


Мета-модели (X-learner)

- Kunzel, S.R. et al. Metalearners for estimating heterogeneous treatment effects using machine learning. 2019
- X-learner (Cross-learner): обучаются две модели: $\mu_0(\mathbf{x})$ и $\mu_1(\mathbf{x})$
- Строятся псевдо-эффекты:
  - $D_i^0 = \mu_1(\mathbf{x}_i) - y_i$ (что было бы, если бы контрольный получил лечение)
  - $D_i^1 = y_i - \mu_0(\mathbf{x}_i)$ (что было бы, если бы леченный не получил лечение)
- Обучаются модели $\tau_0(\mathbf{x})$ и $\tau_1(\mathbf{x})$ на $(\mathbf{x}_i, D_i^0)$ и $(\mathbf{x}_i, D_i^1)$
- Финальная оценка - взвешенная комбинация:

$$\tau(\mathbf{x}) = e(\mathbf{x}) \cdot \tau_0(\mathbf{x}) + (1 - e(\mathbf{x})) \cdot \tau_1(\mathbf{x})$$

Иллюстрация X-learner



Мета-модели (R-learner, Residual learner)

- A General Method for Estimating Heterogeneous Treatment Effects using Residuals (Nie & Wager, 2018)
- Данные $\mathcal{D}$ делятся на две части $\mathcal{D}_1$ (для обучения вспомогательных моделей) и $\mathcal{D}_2$ (для обучения финальной модели CATE)
- Датасет $\mathcal{D}_1$:
  - Обучается модель: $\mu(\mathbf{x}) = \mathbb{E}[Y \mid X = \mathbf{x}]$ для предсказания исхода по признакам
  - Обучается модель: $e(\mathbf{x}) = P(T = 1 \mid X = \mathbf{x})$ для предсказания склонности

R-learner, второй этап алгоритма

- Датасет $\mathcal{D}_2$:
  - Вычисляются остатки:
    - $R_Y(\mathbf{x}_i) = y_i - \mu(\mathbf{x}_i)$ - остаток исхода
    - $R_T(\mathbf{x}_i) = T_i - e(\mathbf{x}_i)$ - остаток лечения
- В остатке $R_Y(\mathbf{x})$ содержится эффект лечения независимо от лечения
- Остаток $R_T(\mathbf{x})$ говорит, насколько решение о лечении для данного человека было "необъяснимо" его признаками.
- Обучается модель, минимизирующая:

$$\tau(\mathbf{x}) = \arg\min_{\tau \in \mathcal{F}} \frac{1}{n_2} \sum_{i=1}^{n_2} \left(R_Y(\mathbf{x}_i) - \tau(\mathbf{x}_i) \cdot R_T(\mathbf{x}_i)\right)^2 + \Omega(\tau)$$

Мета-модели (DR-learner)

- Kennedy, E.H., 2023. Towards optimal doubly robust estimation of heterogeneous causal effects // Electronic Journal of Statistics
- Doubly Robust-learner: данные делятся на 2 части.
- На первой части обучаются

$$\mu_i(\mathbf{x}) = \mathbb{E}[Y \mid T = i, X = \mathbf{x}], \ i \in \{0, 1\}$$
$$e(\mathbf{x}) = P(T = 1 \mid X = \mathbf{x})$$

- Для каждого элемента $\mathbf{x}$ из второй части вычисляется псевдо-исход $\psi(\mathbf{x})$

$$\psi(\mathbf{x}) = \mu_1(\mathbf{x}) - \mu_0(\mathbf{x}) + \frac{T\,(y - \mu_1(\mathbf{x}))}{e(\mathbf{x})} - \frac{(1 - T)\,(y - \mu_0(\mathbf{x}))}{1 - e(\mathbf{x})}$$

- Первая часть $\mu_1(\mathbf{x}) - \mu_0(\mathbf{x})$ - наивная оценка CATE, вторая часть - поправка

Функция помех (nuisance function)

- Затем модель (бустинг, нейросеть и т.д.) обучается предсказывать

$$\tau(\mathbf{x}) = \mathbb{E}[\psi(\mathbf{x}) \mid X = \mathbf{x}]$$

- В DR-learner используются две функции помех:
  - Propensity Score: $e(\mathbf{x}) = P(T = 1 \mid X = \mathbf{x})$
  - Функции регрессии (Outcome Regression Model):

$$\mu_i(\mathbf{x}) = \mathbb{E}[Y \mid T, X = \mathbf{x}], i \in \{0, 1\}$$

- Неймановская ортогональность (Neyman Orthogonality)

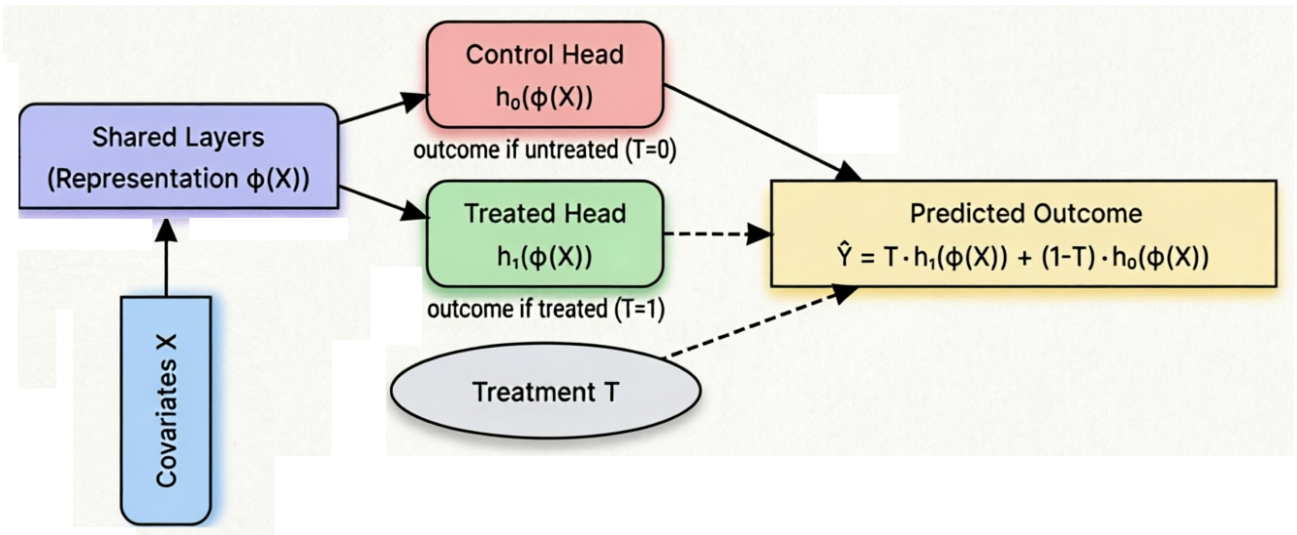
Как обучить propensity score

- Это задача бинарной классификации, где предсказывается вероятность получения лечения $T = 1$ по признакам $\mathbf{x}$
- Отличие от обычной классификации: не интересует точность предсказания сама по себе, т.е. нужна не модель, которая лучше всех угадает, кто получил лечение, а модель, которая правильно выдаст вероятности (калибровка)
- Часто используют логистическую регрессию с регуляризацией как базовую модель

TARNet

- Treatment-Agnostic Representation Network (TARNet)

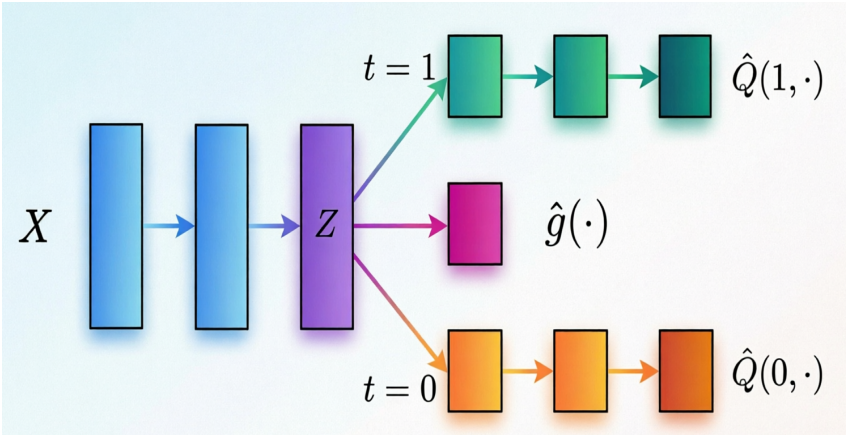
U. Shalit, F.D. Johansson, D. Sontag, Estimating individual treatment effect: generalization bounds and algorithms // Proc. of ICML, 2017



DragonNet

- Нейросеть с тремя “головами”: для прогноза исхода без лечения (h0), для прогноза исхода с лечением (h1), для прогноза склонности (g).

C. Shi, D. Blei, V. Veitch. Adapting neural networks for the estimation of treatment effects // Advances in NIPS, 2019.



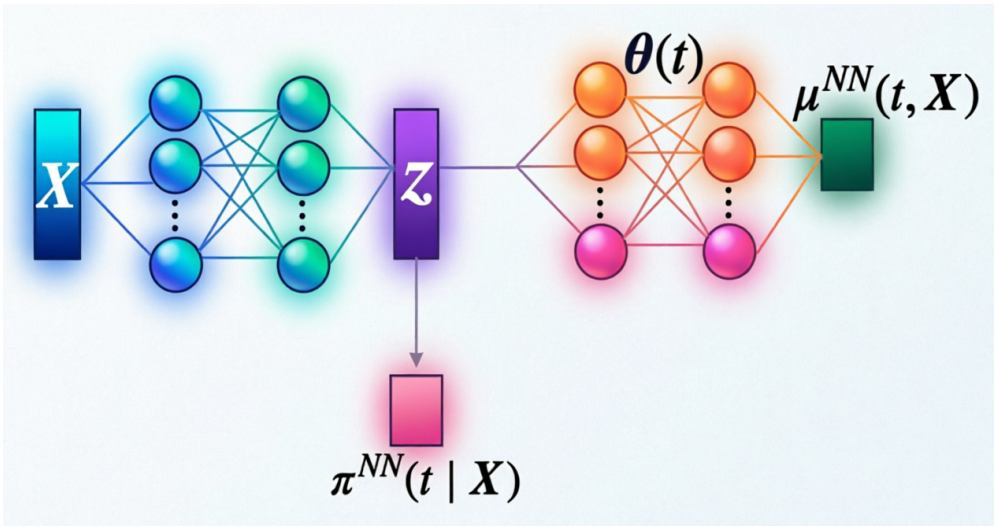
- Model parameters θ are trained with loss:

$$L = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\widehat{Q}(t_i, \mathbf{x}_i, \theta) - y_i\right)^2 + \alpha \mathrm{CE}\left(\widehat{g}(\mathbf{x}_i, \theta), t_i\right)\right]$$

VCNet (1)

- **VCNet (Varying Coefficient Neural Network)** решает задачу оценку эффекта, когда лечение является непрерывным (например, разная доза)
- VCNet обучает единую нейросеть, где веса сети являются функцией от дозы лечения.

L. Nie et al. A VCNet and Functional Targeted Regularization For Learning Causal Effects of Continuous Treatments // arXiv:2103.07861, 2021.



VCNet (2)

- Модель обучается с функцией потерь:

$$L = \frac{1}{n}\sum_{i=1}^{n}\left[(\mu(t_i, \mathbf{x}_i) - y_i)^2 + \frac{\alpha}{n}\log\left(\pi(t_i \mid \mathbf{x}_i)\right)\right]$$

- Для разных доз t используется одна и та же нейросеть, но с разными специально вычисляемыми весами $\theta(t)$.
- Чтобы веса менялись плавно, в VCNet они задаются с помощью математических сплайнов
- $\pi(t \mid \mathbf{x})$ - это обобщенная склонность (generalized propensity score): для непрерывного лечения это условная плотность вероятности получить конкретную дозу t при данных признаках $\mathbf{x}$
- Как реализовать $\pi(t \mid \mathbf{x})$?

VCNet (модели для обобщенной склонности)

- Параметрическая модель (нормальное распределение): $T \mid X \sim \mathcal{N}(\mu_\pi(\mathbf{x}), \sigma^2_\pi(\mathbf{x}))$
  - Голова предсказывает 2 параметра $\mu_\pi(\mathbf{x}), \sigma^2_\pi(\mathbf{x})$, а π - плотность с этими параметрами
- Смесь гауссиан (Gaussian Mixture)
- Непараметрический (оценка плотности ядром)
- Обучение - Maximum Likelihood (для каждого объекта известна его доза t_i)

$$L_\pi = \sum_{i=1}^{n} \log\left(\pi(t_i \mid \mathbf{x}_i)\right)$$

Модель TNW-CATE

- **TNW-CATE**(**T**rainable **N**adaraya-**W**atson regression for **CATE**)
- Контрольные пациенты $\mathcal{C} = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_c, y_c)\}$
- Для каждого i из $\{1, ..., c\}$ определим N подмн-в $\mathcal{C}_{i,r}$, $r = 1, ..., N$, из n случ-х примеров из $\mathcal{C} \backslash (\mathbf{x}_i, y_i)$:

$$\mathcal{C}_{i,r} = \{(\mathbf{x}_1^{(r)}, y_1^{(r)}), ..., (\mathbf{x}_n^{(r)}, y_n^{(r)})\}, \ r = 1, ..., N$$

- Каждое подмн-во $\mathcal{C}_{i,r}$ с $(\mathbf{x}_i, y_i)$ образует обучающий пример для контрольной сети:

$$\mathbf{a}_i^{(r)} = \left(\mathbf{x}_1^{(r)}, ..., \mathbf{x}_n^{(r)}, \mathbf{x}_i, y_1^{(r)}, ..., y_n^{(r)}, y_i\right)$$

- То же самое можно сделать для **леч. пациентов**

Konstantinov A.V., Kirpichenko S.R., Utkin L.V. Heterogeneous treatment effect with trained kernels of the Nadaraya–Watson regression // Algorithms. 2023

Регрессия Надарая-Уотсона

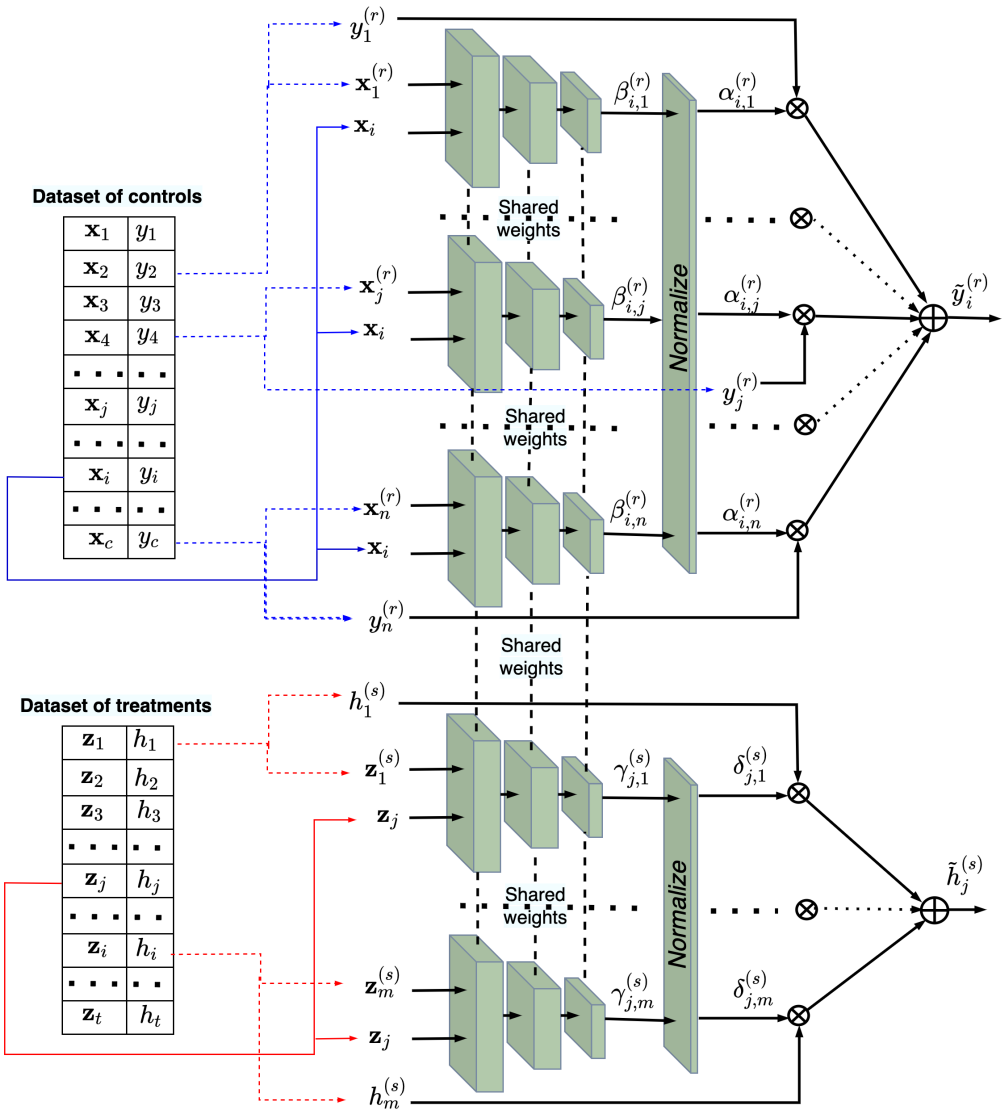
$$\tilde{y}_i^{(r)} = g_0(\mathbf{x}_i) = \sum_{j=1}^{n} \alpha(\mathbf{x}_i, \mathbf{x}_j^{(r)}) y_j^{(k)},$$

$$\tilde{h}_j^{(s)} = g_1(\mathbf{z}_j) = \sum_{j=1}^{n} \delta(\mathbf{z}_j, \mathbf{z}_i^{(s)}) h_j^{(s)}$$

Веса внимания:

$$\alpha(\mathbf{x}_i, \mathbf{x}_j^{(r)}) = \frac{K\left(\mathbf{x}_i, \mathbf{x}_j^{(r)}\right)}{\sum_{k=1}^{n} K\left(\mathbf{x}_i, \mathbf{x}_k^{(r)}\right)}, \; \delta(\mathbf{z}_j, \mathbf{z}_i^{(s)}) = \frac{K\left(\mathbf{z}_j, \mathbf{z}_i^{(s)}\right)}{\sum_{k=1}^{m} K\left(\mathbf{z}_j, \mathbf{z}_k^{(s)}\right)}$$

Модель TNW-CATE



Особенности TNW-CATE

- Использует регрессию Надарая-Уотсона, где обучаемые веса внимания реализуются в виде нейронных сетей
- Предполагается, что области векторов признаков из групп контрольных и леченных пациентов аналогичны, Это делает модель аналогичной модели transfer learning
- В тестовом режиме для нового $\mathbf{x}$ сеть не обучается. Множества c пар $(\mathbf{x}, \mathbf{x}_1), ..., (\mathbf{x}, \mathbf{x}_c)$ и t pairs $(\mathbf{x}, \mathbf{z}_1), ..., (\mathbf{x}, \mathbf{z}_t)$ подаются на вход сети и вычисляются веса внимания в регрессии Надарая-Уотсона.

Вероятность положительного лечения (IPTB)

- **Individual Probability of Treatment Benefit (IPTB):**

$$\Psi(\mathbf{x}) = \Pr\left(Y_1 > Y_0 \mid \mathbf{X} = \mathbf{x}\right)$$

- При условии exchangeability и positivity, данные имеют условные маргинальные распределения
  $F_1(t \mid \mathbf{x}) = \Pr\{Y_1 \leq t \mid X = \mathbf{x}\}$ и
  $F_0(t \mid \mathbf{x}) = \Pr\{Y_0 \leq t \mid X = \mathbf{x}\}$
- Но совместное распределение не известно. Его можно представить, используя копулу:
  $F_{0,1}(y_0, y_1 \mid \mathbf{x}) = C_{\theta,\mathbf{x}}\left(F_0(t \mid \mathbf{x}), F_1(t \mid \mathbf{x})\right)$
- Используем $C(u, v) = u \cdot v$

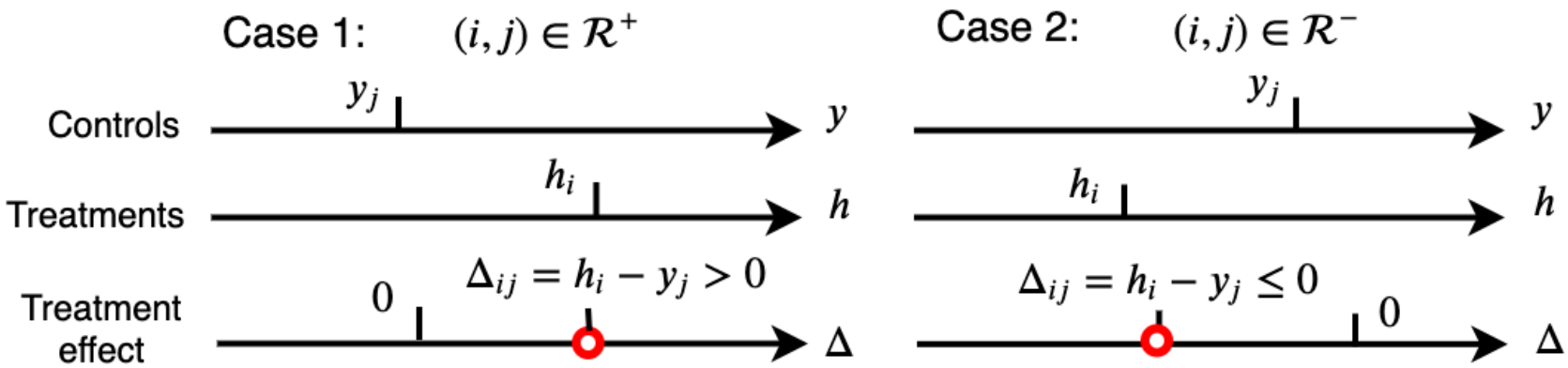
Что предлагается

- Предлагаем трюк: берем $(\mathbf{z}_i, h_i)$ и $(\mathbf{x}_j, y_j)$ из двух различных групп
- для каждой пары определять эффект лечения (псевдо-метки) $\Delta_{ij} = h_i - y_j$
- Тогда задача определить вероятность

$$p^+(\mathbf{z}, \mathbf{x}) = \Pr\{\Delta > 0 \mid \mathbf{Z} = \mathbf{z}, \mathbf{X} = \mathbf{x}\}$$

используя все пары $(\mathbf{z}_i, h_i)$ и $(\mathbf{x}_j, y_j)$.

Что предлагается



Решение - механизм внимания

Рассматривается задача бинарной классификации с вероятностями

$$p^+(\mathbf{z}, \mathbf{x}) = \sum_{(l,k) \in \mathcal{R}^+ \cup \mathcal{R}^-} a((\mathbf{z}, \mathbf{x}), (\mathbf{z}_l, \mathbf{x}_k)) \cdot \mathbf{1}[\Delta_{lk} > 0]$$

$$= \sum_{(i,j) \in \mathcal{R}^+} a((\mathbf{z}, \mathbf{x}), (\mathbf{z}_i, \mathbf{x}_j)) \cdot 1$$

$$+ \sum_{(r,s) \in \mathcal{R}^-} a((\mathbf{z}, \mathbf{x}), (\mathbf{z}_r, \mathbf{x}_s)) \cdot 0$$

и

$$\sum_{(r,s) \in \mathcal{R}^+ \cup \mathcal{R}^-} a((\mathbf{z}, \mathbf{x}), (\mathbf{z}_r, \mathbf{x}_s)) = 1$$

Веса внимания

- Веса внимания:

$$a((\mathbf{z},\mathbf{x}),(\mathbf{z}_i,\mathbf{x}_j)) = \frac{K((\mathbf{z},\mathbf{x}),(\mathbf{z}_i,\mathbf{x}_j))}{\sum_{(r,s)\in\mathcal{R}^+\cup\mathcal{R}^-} K((\mathbf{z},\mathbf{x}),(\mathbf{z}_r,\mathbf{x}_s))}$$

- Для гауссова ядра:

$$a((\mathbf{z},\mathbf{x}),(\mathbf{z}_i,\mathbf{x}_j)) = \operatorname{softmax}\left(-\frac{\|(\mathbf{z},\mathbf{x})-(\mathbf{z}_i,\mathbf{x}_j)\|^2}{\tau}\right)$$

Обучение модели

- Функция потерь

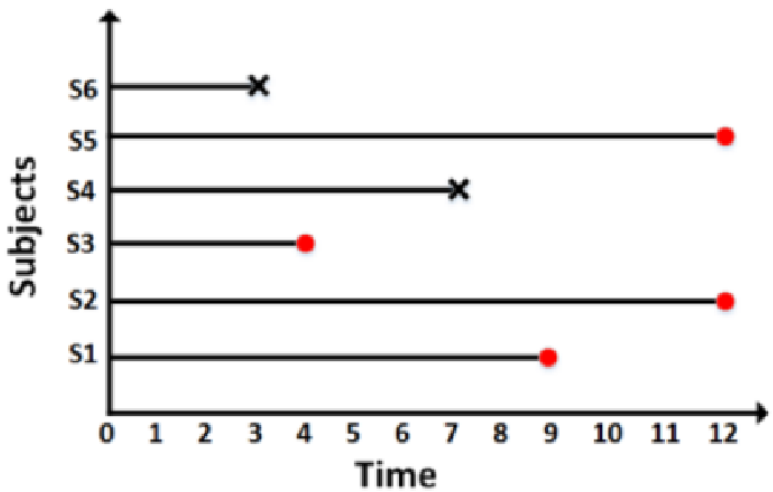
$$\mathcal{L}(\mathbf{p} \mid \mathbf{W}_Q, \mathbf{W}_K) = -\sum_{(i,j)\in\mathcal{R}^+} \log\left(p^+(\mathbf{z}_i, \mathbf{x}_j)\right)$$
$$-\sum_{(i,j)\in\mathcal{R}^-} \log\left(1 - p^+(\mathbf{z}_i, \mathbf{x}_j)\right) + \eta\left(\|\mathbf{W}_Q\|^2 + \|\mathbf{W}_K\|^2\right)$$

Особенности модели

- Модель обходит проблему отсутствия контрфактических данных для одного пациента за счет формирования пар наблюдений и оценки разности их исходов Δ_{ij}.
- Подход эффективен в условиях малого числа примеров в леченной группе и задействует все возможные пары объектов для обучения.
- Веса внимания позволяют модели адаптивно оценивать вклад различных пар объектов в зависимости от их схожести.

Анализ выживаемости (Survival Analysis)

- Анализ выживаемости является инструментом для модел-я данных типа “время до наступления события”.
- Цензурированные наблюдения — это случаи, в которых известно, что интересующее нас событие не произошло в течение определенного периода времени, однако точное время его наступления остается неизвестным.

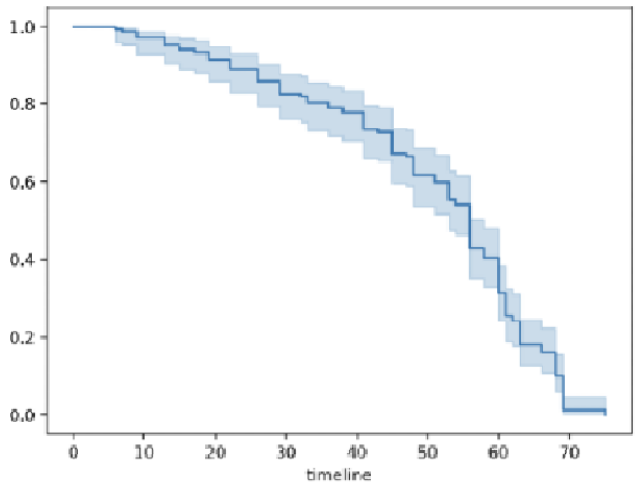


Элементы анализа выживаемости (1)

- **Датасет**: $\mathcal{A} = \{(\mathbf{x}_1, \delta_1, T_1), ..., (\mathbf{x}_N, \delta_N, T_N)\}$. $\mathbf{x}_i \in \mathbb{R}^d$;
- Время T_i соответствует одному из двух типов наблюдений за событием (нецензурированному и цензурированному);
- Индикар цензурирования $\delta = 1$ для нецензурированных событий; 0 - для цензурированных.
- **Цель**: Модель выживаемости обучается на $\mathcal{A}$, чтобы предсказывать вероятностные хар-ки для нового вектора $\mathbf{x}$.

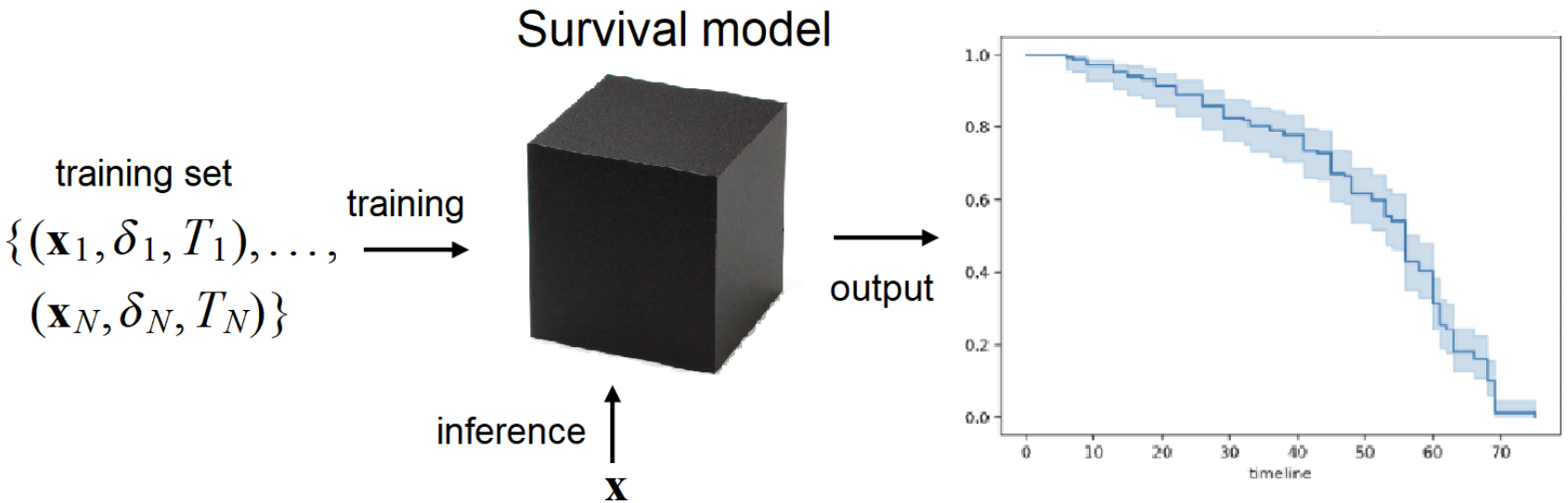
Элементы анализа выживаемости (2)

- Функция выживаемости (SF) $S(t \mid \mathbf{x}) = \Pr\{T > t \mid \mathbf{x}\}$ - вероятность дожития до t



- Функция риска $\lambda(t \mid \mathbf{x}) = -\frac{\mathrm{d}}{\mathrm{d}t} \ln S(t \mid \mathbf{x})$ интенсивность наступления события в момент t при условии, что до t событие не произошло
- Кумулятивная функция риска $H(t \mid \mathbf{x}) = \int_0^t \lambda(r \mid \mathbf{x}) \mathrm{d}r$, $H(t \mid \mathbf{x}) = -\ln\left(S(t \mid \mathbf{x})\right)$

Модели выживаемости



Как сравнивать модели выживаемости

- **C-index** (Harrell et al. 1982) оценивает вероятность, что предсказанные ожидаемые времена событий пары примеров правильно ранжированы

$$C = \frac{\sum_{(i,j)\in\mathcal{J}} \mathbf{1}[\widehat{T}_i < \widehat{T}_j]}{\sum_{(i,j)\in\mathcal{J}} 1}$$

- **Integrated Brier Score** (IBS) определяется как:

$$IBS = \frac{1}{t_{\max}} \int_0^{t_{\max}} \mathbb{E}\left[\left(\mathbf{1}\left(T_{new} > t\right) - \hat{S}\left(t \mid \mathbf{x}_{new}\right)\right)^2\right] \mathrm{d}t$$

Типы моделей выживаемости

1. Параметрические: известны законы распределений
2. Непараметрические: модель Каплана-Мейера (безусловная), оценка Берана (условная)
3. Полупараметрические: модель Кокса

$$H(t|\mathbf{x},\mathbf{b}) = H_0(t)\exp\left(\mathbf{b}^{\mathrm{T}}\mathbf{x}\right)$$

Существующие модели

- Случайный лес выживаемости RSF (Ishwaran, Kogalur 2007); GBM Cox (Ridgeway 1999); GBM AFT (Barnwal et al. 2022)
- Survival neural networks (Chen et al., 2024)
- Survival transformers (Arroyo et al. 2024, Hu et al. 2021, Li et al. 2022, Zhang et al. 2025)

Оценка Берана (Beran estimator)

Пусть $t_1 < t_2 < \ldots < t_n$ - упорядоченная послед-ть времен,

$$S(t \mid \mathbf{x}) = \prod_{t_i \le t} \left\{ 1 - \frac{\alpha(\mathbf{x}, \mathbf{x}_i)}{1 - \sum_{j=1}^{i-1} \alpha(\mathbf{x}, \mathbf{x}_j)} \right\}^{\delta_i}$$

$$\alpha(\mathbf{x}, \mathbf{x}_i) = \frac{K(\mathbf{x}, \mathbf{x}_i)}{\sum_{j=1}^{n} K(\mathbf{x}, \mathbf{x}_j)}$$

Если гауссово ядро, то

$$\alpha(\mathbf{x}, \mathbf{x}_i) = \operatorname{softmax}\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{\tau} \right)$$

CATE и анализ выживаемости

- Проблема оценки CATE: целевая переменная - время до события
- Определения CATE в анализе выживаемости:
  - Разность **ожидаемых времен до событий**:

$$\tau(\mathbf{x}) = \mathbb{E}\{T_1 - T_0 \mid X = \mathbf{x}\}$$

  - Разность функций выживаемости:

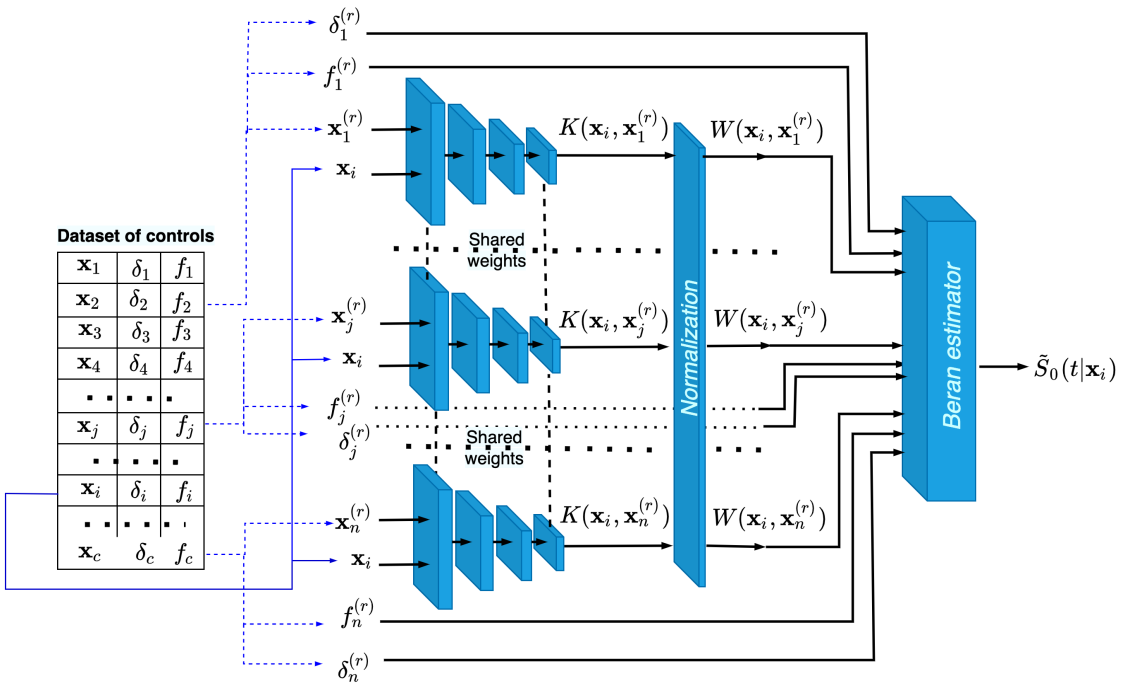
$$\tau(t, \mathbf{x}) = S_1(t \mid \mathbf{x}) - S_0(t \mid \mathbf{x})$$

  - Отношение рисков:

$$\tau(t, \mathbf{x}) = h_1(t \mid \mathbf{x}) / h_0(t \mid \mathbf{x})$$

BENK: A new CATE model (training)

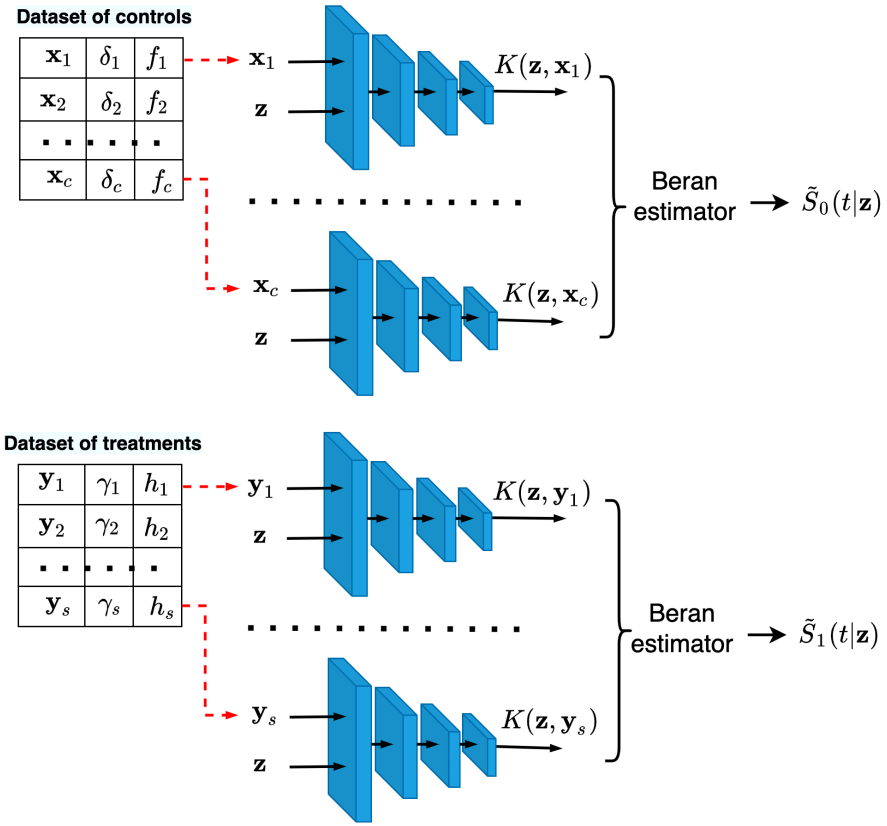
Idea: network representation of kernels in the Beran model



Kirpichenko S.R., Utkin L.V., Konstantinov A.V., Muliukha V.A. BENK: The Beran Estimator with Neural Kernels for Estimating the Heterogeneous Treatment Effect // **Algorithms**. 2024

BENK: A new CATE model (inference)

Inference for vector $\mathbf{z}$:



Surv-IPTB

$$\mathcal{D}_0 = \{(\mathbf{x}_i, y_i, \xi_i)\}_{i=1}^{c}, \ \mathcal{D}_1 = \{(\mathbf{z}_j, h_j, \delta_j)\}_{j=1}^{t}$$

$$\Psi(\mathbf{x}) = \Pr\left(H^* > Y^* \mid \mathbf{X} = \mathbf{x}\right)$$

При условии exchangeability и positivity, данные определяют маргинальные расп-я

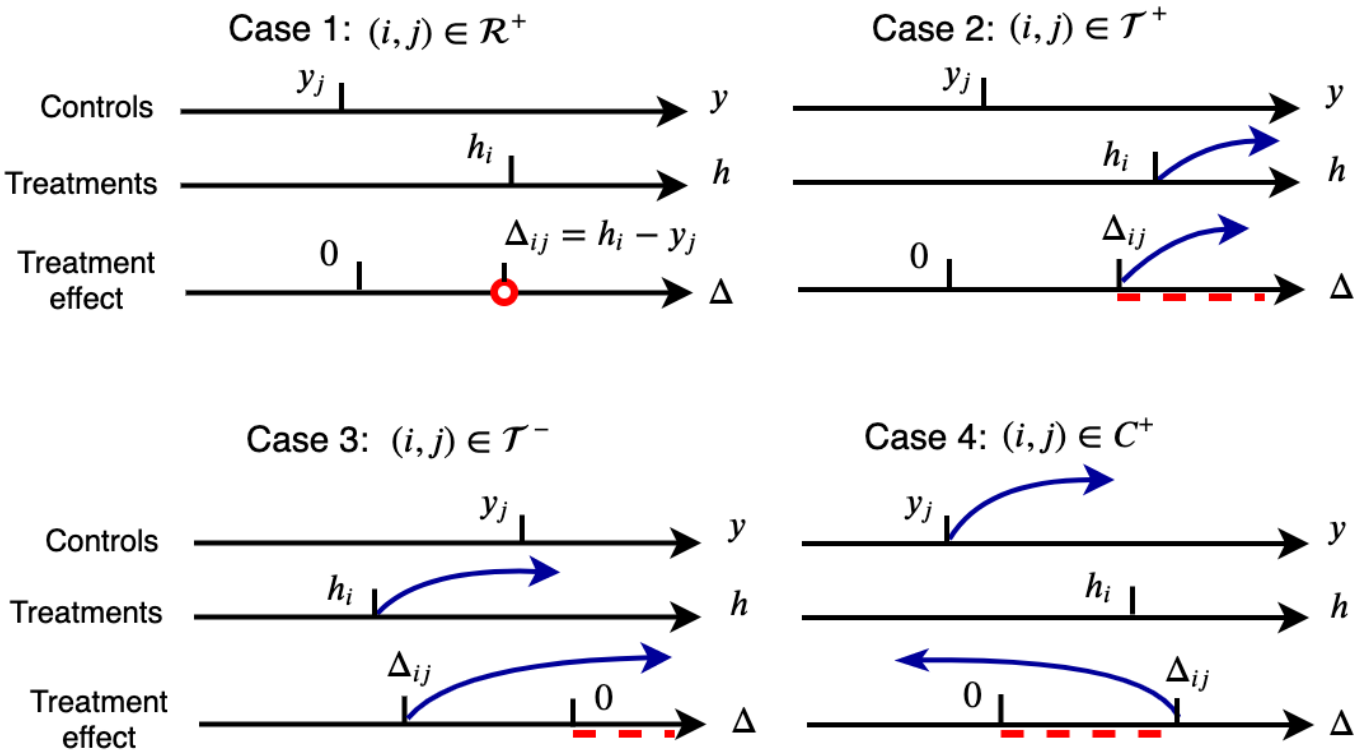
$$F_{H^*}(t \mid \mathbf{x}) = \Pr\{H^* \leq t \mid X = \mathbf{x}\}$$
$$F_{Y^*}(t \mid \mathbf{x}) = \Pr\{Y^* \leq t \mid X = \mathbf{x}\}$$

Совместное распределение не определено, но может быть представлена через копулу (copula) $C_{\theta,\mathbf{x}}$:

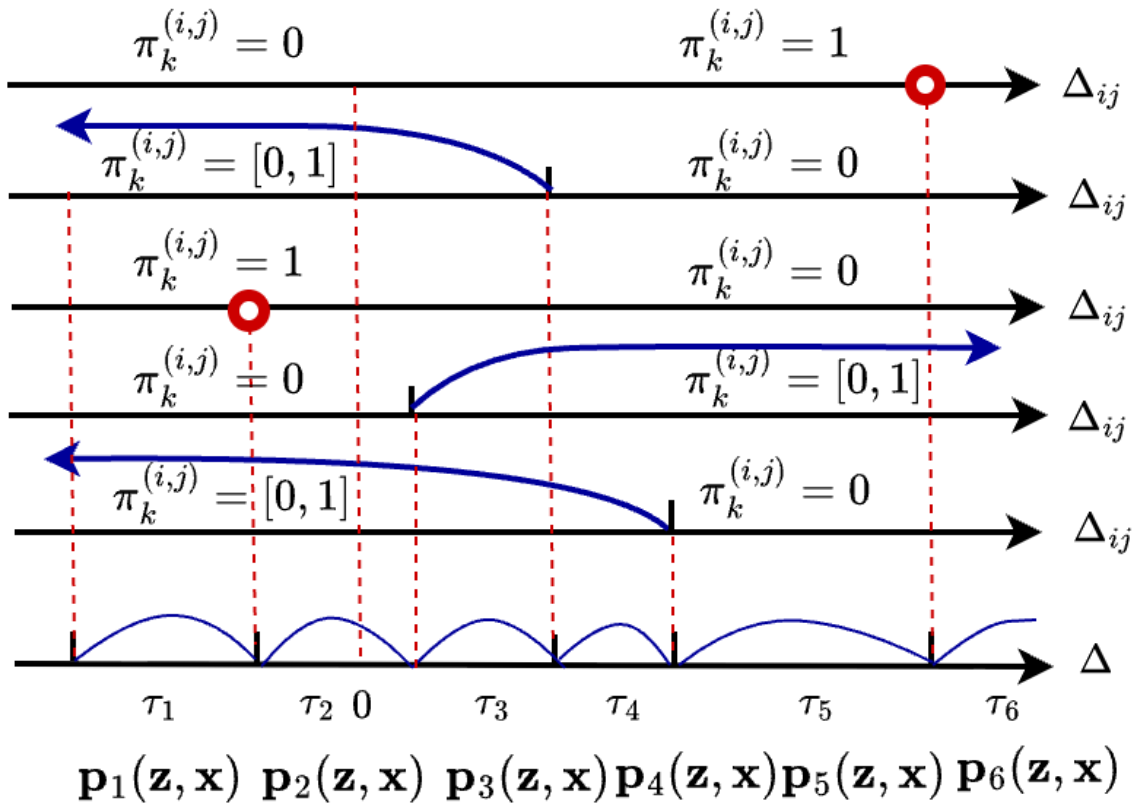
$$F_{H^*,Y^*}(h, y \mid \mathbf{x}) = C_{\theta,\mathbf{x}}\left(F_{H^*}(h \mid \mathbf{x}), F_{Y^*}(y \mid \mathbf{x})\right)$$

Для пар используем независимую копулу $C(u, v) = u \cdot v$.

Случаи положительного эффекта



Распределение вероятностей интервалов



Искомая вероятность

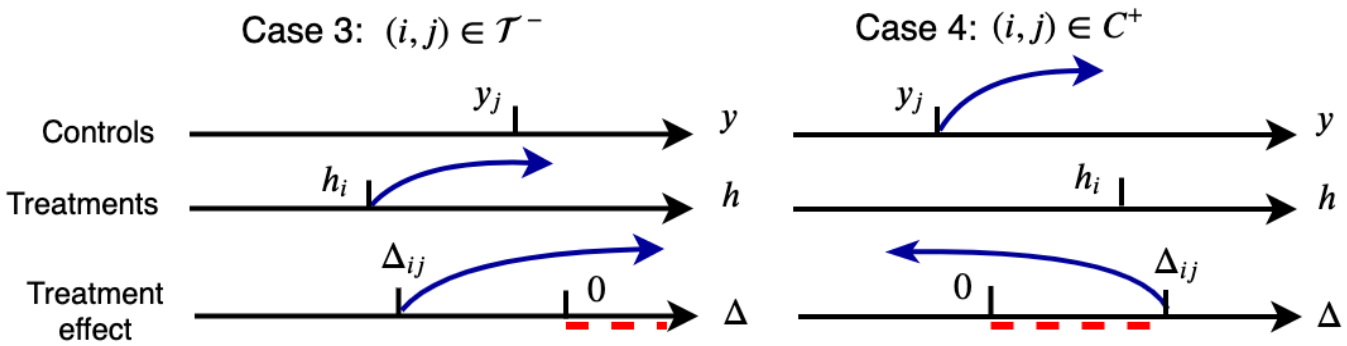
$$p^+(\mathbf{z},\mathbf{x}) = \sum_{(i,j)\in\mathcal{R}^+} a(\mathbf{z},\mathbf{x},\mathbf{z}_i,\mathbf{x}_j)\cdot 1 + \sum_{(l,k)\in\mathcal{T}^+} a(\mathbf{z},\mathbf{x},\mathbf{z}_l,\mathbf{x}_k)\cdot 1$$
$$+ \sum_{(v,w)\in\mathcal{T}^-} a(\mathbf{z},\mathbf{x},\mathbf{z}_v,\mathbf{x}_w)\cdot [0,1]$$
$$+ \sum_{(r,s)\in\mathcal{C}^+} a(\mathbf{z},\mathbf{x},\mathbf{z}_r,\mathbf{x}_s)\cdot [0,1]$$

Веса внимания:

$$a(\mathbf{z},\mathbf{x},\mathbf{z}_i,\mathbf{x}_j) = \frac{K(\mathbf{z},\mathbf{z}_i)\cdot K(\mathbf{x},\mathbf{x}_j)}{\sum_{(r,s)\in\mathcal{G}} K(\mathbf{z},\mathbf{z}_r)\cdot K(\mathbf{x},\mathbf{x}_s)}$$

Soft метки

$$p^+(\mathbf{z},\mathbf{x}) = \sum_{(i,j)\in\mathcal{R}^+\cup\mathcal{T}^+} a(\mathbf{z},\mathbf{x},\mathbf{z}_i,\mathbf{x}_j) + \sum_{(r,s)\in\mathcal{T}^-\cup\mathcal{C}^+} a(\mathbf{z},\mathbf{x},\mathbf{z}_r,\mathbf{x}_s)\cdot\pi^{(r,s)}$$



$$\phi_{i,j} = \Pr\{H^* \geq y_j \mid H^* > h_i\} = \frac{S_1(y_j)}{S_1(h_i)}$$

$$\rho_{i,j} = \Pr\{y_j \leq Y^* \leq h_i \mid Y^* \geq y_j\} = \frac{S_0(y_j) - S_0(h_i)}{S_0(y_j)}$$

Функции потерь

$$L(\mathcal{R}^+ \cup \mathcal{T}^+) = - \sum_{(i,j) \in \mathcal{R}^+ \cup \mathcal{T}^+} \log \left(p^+(\mathbf{z}_i, \mathbf{x}_j) \right),$$

$$L(\mathcal{R}^- \cup \mathcal{S}^-) = - \sum_{(i,j) \in \mathcal{R}^- \cup \mathcal{S}^-} \log \left(p^-(\mathbf{z}_i, \mathbf{x}_j) \right),$$

$$L(\mathcal{T}^-) = - \sum_{(r,s) \in \mathcal{T}^-} \phi_{rs} \log \left(p^+(\mathbf{z}_r, \mathbf{x}_s) \right),$$

$$L(\mathcal{C}^+) = - \sum_{(r,s) \in \mathcal{C}^+} \rho_{rs} \log \left(p^+(\mathbf{z}_r, \mathbf{x}_s) \right),$$

Функции потерь

$$L(\mathcal{Q}^-) = -\sum_{(r,s)\in\mathcal{Q}^-} (1-\phi_{rs}) \log\left(p^-(\mathbf{z}_r, \mathbf{x}_s)\right),$$

$$L(\mathcal{C}^-) = -\sum_{(r,s)\in\mathcal{C}^-} (1-\rho_{rs}) \log\left(p^-(\mathbf{z}_r, \mathbf{x}_s)\right),$$

Регуляризвции

$$L(\mathbf{W}) = \|\mathbf{W}_Q\|^2 + \|\mathbf{W}_K\|^2,$$

$$L(\pi) = \sum_{(r,s)\in\mathcal{T}^-\cup\mathcal{C}^+} \pi^{(r,s)} \log\left(\pi^{(r,s)}\right),$$

От статического лечения к динамическим стратегиям

- CATE отвечает на вопрос: “Что лучше: препарат А или Б?”
- Однако в реальности врачи принимают решения последовательно, адаптируя терапию под изменяющееся состояние пациента
- **Формальная постановка**:
  - время: $t = 0, ..., T$
  - состояние: $L(t)$ - вектор показателей пациента (анализы, симптомы) на момент времени t
  - действие: $A(t)$ - назначенное лечение в момент t.
  - исход: Y - финальный результат (например, выживаемость)

Цель

- **Цель**: Оценить эффект не одного статичного действия, а всей последовательности решений (динамического режима лечения) или оценить эффект всей истории лечения $\widetilde{A}(T) = (A(0), ..., A(T))$ на Y
- Например, мы хотим сравнить две стратегии:
  - Начинать терапию, если $L(t) < 200$ (Стратегия d_1)
  - Начинать терапию, если $L(t) < 350$ (Стратегия d_2)
- **Сложность**: решение о лечении $A(t)$ зависит от предыдущего состояния $L(t)$, которое само зависит от прошлого лечения. Это создает временное смешение (time-varying confounding)

Метод: динамическая регрессия

- **Динамические регрес. модели** оценивают эффект, “замораживая” состояние пациента. Ответ на вопрос: “Если у двух пациентов одинаковый уровень сахара ($L(t)$), как повлияет назначение инсулина на исход?”
- Модель для исхода Y на основе истории лечения и текущего состояния:

$$\mathbb{E}[Y \mid \mathbf{x}, A(t), L(t)] = \beta_0 + \beta_1 \cdot \mathbf{x} + \beta_2 \cdot A(t) + \gamma \cdot A(t) \cdot L(t)$$

  - $\beta_1 = (\beta_{1,1}, ..., \beta_{1,d})$, $\mathbf{x} = (x_1, ..., x_d)^{\mathrm{T}}$ - признаки
  - β_2 - эффект лечения при $L(t) = 0$: как изменится исход Y при лечении $A(t) = 1$ по сравнению с $A(t) = 0$ при условии $L(t) = 0$ (базовый эффект)
  - γ - дин. эффект: как изменяется эф-ть лечения от тяжести состояния пациента

Веса пациентов (1)

- Для корректной оценки веса пациентов пересчитываются с помощью обратной вер-ти назначения лечения (IPW):
  - Вес$(t) = 1/P(A(t)\,|$история_ пациента$)$
- История_ пациента $= \{V, L(0), A(0), L(1), ..., L(t)\}$
- Строим модель (лог.регр.), которая выдает число в [0,1]
- Врач решает, назначать ли инсулин $(A(t) = 1)$ или нет $(A(t) = 0)$. По истории пациента:
  - Пац.1: сахар 15 ммоль/л и он диабетик 10 лет, вер-ть назначения инсулина 0.95
  - Пац.2: сахар 5.5 ммоль/л и он пришел с легким преддиабетом, вер-ть назначения инсулина 0.05

Веса пациентов (2)

- Если модель говорит, что назначение инсулина Пац.1 было предсказуемо (0.95), то его вес будет $1/0.95 \approx 1.05$. Пац. 1 получает небольшой вес, т.к. он “типичный” для своей группы
- Если модель говорит, что назначение инсулина Пац. 2 было маловероятно (0.05), то его вес будет $1/0.05 = 20$. Пац. 2 получает большой вес, т.к.он “редкий случай” (получил инсулин, хотя по всем признакам не должен был его получить)
- Таким образом: Вес(t) показывает, насколько “удивительно”для статистики было решение врача

А где время?

- Модель смотрит на всю цепочку действий:

$$\mathbb{E}[Y \mid \mathbf{x}, \overline{A}(T)] = \beta_0 + \beta_1 \cdot \mathbf{x} + \sum_{i=1}^{T} \beta_2(t) \cdot A(t) + \sum_{i=1}^{T} \gamma(t) \cdot A(t) \cdot L(t)$$

- Модель суммирует эффект каждого приема лекарства в каждый месяц t.
- Коэффициенты зависят от времени, например:
  - $\beta_2(t = 1) = -2$ (в 1-й месяц лечение снижает риск)
  - $\beta_2(t = 6) = 0.5$ (на 6-й месяц лечение уже начинает вредить из-за побоч. эффектов)

Интерпретация коэффициентов

- $\beta_0 + \beta_1 \cdot \mathbf{x}$ - базовый эффект
- $\sum_{i=1}^{T} \beta_2(t) \cdot A(t)$ - эффект самого лечения
- $\sum_{i=1}^{T} \gamma(t) \cdot A(t) \cdot L(t)$ - динамич. эффект (лечение$\times$состояние)
- $\gamma(t)$ показывают, как состояние пац-та в момент лечения модифицирует эффект этого лечения
- $\gamma(t) < 0$, то лечение эфф-нее для пац-тов с высоким $L(t)$ (например, высокий сахар - лечение снижает его сильнее)
- $\gamma(t) > 0$, то лечение вредит пациентам с высоким $L(t)$

Q-Learning

- Поиск оптимальной стратегии через “обратную индукцию”
- Если динамическая регрессия оценивает эффект правила, то Q-Learning ищет оптимальное правило, максимизирующее долгосрочный выигрыш (исход). Он работает с конца, используя “псевдо-исходы”

Q-Learning - алгоритм

1. **Шаг 2 (последний)**: Строим регрессию, предсказывающую исход Y на основе истории H_2 и лечения A_2:

$$Q_2(H_2, A_2) = \beta_2 H_{2,0} + \psi_2 \cdot A_2$$

ψ_2 - эффект лечения на последнем шаге
$H_{2,0}$ - это вектор признаков, полученных из истории пациента H_2, но без учета переменной лечения, т.е.

$$\beta_2 H_{2,0} = \beta_{20} + \beta_{21}\mathbf{x}_1 + ... + \beta_{2k}\mathbf{x}_k,$$

2. **Расчет псевдо-исхода**: Для каждого пациента вычисляем, какой результат он мог бы иметь, если бы на втором шаге выбрал лучшее лечение:

$$\widetilde{Y}_1 = \max(Q_2(H_2, A_2 = 0),\ Q_2(H_2, A_2 = 1))$$

Q-Learning - алгоритм и итог

3. **Шаг 1**: Строим модель для Шага 1, где предсказываем уже не Y, а вычисленный $\widetilde{Y}_1$:

$$Q_1(H_1, A_1) = \beta_1 H_{1,0} + \psi_1 \cdot A_1$$

Результат: На основе полученных коэффициентов можем сказать:
"Назначай лекарство сейчас, если $Q_1(A_1 = 1) > Q_1(A_1 = 0)$".
В Q-learning $H(t)$ вся информация, доступная до назначения лечения в момент времени t:

$$H(t) = \{V, L(0), A(0), L(1), A(1), ..., L(t)\}$$

здесь есть $L(t)$ (текущее состояние, которое врач видит прямо сейчас), но нет $A(t)$

Q-Learning - итог алгоритма

Правило: На основе полученных коэффициентов можем сказать:
“Назначай лекарство сейчас, если
$Q_1(A_1 = 1) > Q_1(A_1 = 0)$”.
В Q-learning $H(t)$ - вся информация, доступная до назначения лечения в момент времени t:

$$H(t) = \{V, L(0), A(0), L(1), A(1), ..., L(t)\}$$

здесь есть $L(t)$ (текущее состояние, которое врач видит прямо сейчас), но нет $A(t)$

Q-Learning, когда много этапов

Для любого шага $t = 0, 1, ..., T$ Q-функция определяется как:

$$Q_t(H(t), A(t)) = \mathbb{E}\left[Y_{t+1} + \max_a Q_{t+1}(H(t+1), a) \mid H(t), A(t)\right]$$

$H(t)$ - история до назначения лечения на шаге t
$A(t)$ - лечение, назначенное на шаге t
Y_{t+1} - награда (или изменение состояния) между шагами t и $t+1$ (обычно это улучшение состояния или отсрочка плохого события)
$\max_a$ - максимум по всем возможным вариантам лечения на следующем шаге

Q-Learning в общем виде

- На последнем шаге T рекурсия останавливается, и Q-функция определяется как простое ожидание финального исхода:

$$Q_T(H(T), A(T)) = \mathbb{E}[Y \mid H(T), A(T)]$$

Y - итоговый результат (например, выживаемость).

- Q-функция на шаге t:

$$Q_t(H(t), A(t)) = \mathbb{E}\left[R_t + \max_{a_{t+1}} \max Q_{t+1}(H(t+1), a_{t+1}) \mid H(t), A(t)\right.$$

R_t - непосредственная награда за действие $A(t)$

Q-Learning, линейная аппроксимация

- Для каждого шага t аппроксимируем Q-функцию линейной регрессией:

$$Q_t(H(t), A(t)) = \beta_t H_{t,0} + (\psi_t \cdot H_{t,1}) \cdot A(t)$$

$H_{t,0}$ - базовые признаки (состояния) на шаге t
$H_{t,1}$ - признаки, с которыми лечение взаимодействует (обычно это же состояние)
β_t и ψ_t - коэффициенты, которые оцениваем отдельно для каждого шага

Q-Learning, алгоритм для T+1 шагов

- Для $t = T, T-1, ..., 0$:
  1. Если $t = T$, целевая переменная - Y (реальный исход).
  2. Если $t < T$, целевая переменная -

$$\widetilde{Y}_t = \max_a Q_{t+1}(H(t+1), a)$$

  3. Строим регрессию:

$$target_t = \beta_t H_{t,0} + (\psi_t \cdot H_{t,1}) \cdot A(t) + \varepsilon_t$$

  4. Находим β_t и ψ_t методом наименьших квадратов
  5. Переходим к следующему t (в обратном порядке)

Q-Learning, пример (1)

- **Болезнь**: Рак молочной железы
- **Исход**: Размер опухоли через 9 месяцев. Меньше - лучше
- **Три этапа лечения** (каждые 3 месяца):
  - t=0 (0 мес): Диагностика. Измеряем размер опухоли $L(0)$. Решение: назначать ли химию $A(0) \in \{0, 1\}$
  - t=1 (3 мес): Измеряем $L(1)$. Решение: $A(1)$ (продолжать химию или нет)
  - t=2 (6 мес): Измеряем $L(2)$. Решение: $A(2)$ (химию или нет)
  - t=3 (9 мес): Финальный размер опухоли Y.

Q-Learning, пример (2)

Данные пациентки M

t=0	L(0) (размер опухоли)	45 мм
t=0	A(0) (химиотерапия?)	1 (да)
t=1	L(1)	35 мм
t=1	A(1)	1 (да)
t=2	L(2)	25 мм
t=2	A(2)	0 (нет, перешли на гормоны)
t=3	Y(финальный размер)	28 мм

Q-Learning, пример (3)

Шаг 1. Оценка Q-функции для $t = 2$ (финал)

- Обучаем модель на пациентах, достигших состояние 2:

$$Y = \beta_{20} + \beta_{21} \cdot L(2) + \psi_2 \cdot A(2) + \varepsilon_2$$

- После обучения: $\beta_{20} = 5$, $\beta_{21} = 0.8$, $\psi_2 = -6$
- Для пациента M с $L(2) = 25$:

$$Q_2(H_2, A_2 = 0) = 5 + 0.8 \cdot 25 + (-6) \cdot 0 = 25$$
$$Q_2(H_2, A_2 = 1) = 5 + 0.8 \cdot 25 + (-6) \cdot 1 = 19$$

$$target_1 = \min(Q_2(\cdot, 0), Q_2(\cdot, 1)) = \min(25, 19) = 19$$

- Интерпретация: Если бы на этапе 2 продолжили химию, то к 9-му месяцу размер опухоли мог бы стать 19 мм

Q-Learning, пример (4)

Step 2. Оценка модели для $t = 1$

- Используем $target_1$ как выход:

$$target_1 = \beta_{10} + \beta_{11} \cdot L(1) + \psi_1 \cdot A(1) + \varepsilon_1$$

- Для пациента M с $L(1) = 35, A(1) = 1$:

$$19 = \beta_{10} + \beta_{11} \cdot 35 + \psi_1 \cdot 1 + \varepsilon_1$$

- После обучения: $\beta_{10} = 2, \beta_{11} = 0.5, \psi_1 = -4$
- Предсказание модели для пациентки M:

$$\widetilde{Y} = 2 + 0.5 \cdot 35 + (-4) \cdot 1 = 15.5$$

- Ошибка: $\varepsilon_1 = target_1 - \widetilde{Y} = 19 - 15.5 = 3.5$

Q-Learning, пример (5)

Шаг 3. Вычисление $target_0$ для $t = 0$

- Подставляем оба варианта лечения на этапе 1 для пациентки M:

$$Q_1(H_1, A_1 = 0) = 2 + 0.5 \cdot 35 + (-4) \cdot 0 = 19.5$$
$$Q_1(H_1, A_1 = 1) = 2 + 0.5 \cdot 35 + (-4) \cdot 1 = 15.5$$

- Берем минимум (т.к. меньший размер опухоли лучше):

$$target_0 = \min(19.5, 15.5) = 15.5$$

- Интерпретация: Если бы мы начали химиотерапию с самого начала и продолжили ее на этапе 1, то к 9 месяцам размер опухоли мог бы стать 15.5 мм

Q-Learning, пример (6)

Шаг 4. Оценка модели для $t = 0$:

$$target_0 = \beta_{00} + \beta_{01} \cdot L(0) + \psi_0 \cdot A(0) + \varepsilon_0$$

- Для пациентки M с $L(0) = 45, A(0) = 1$:

$$15.5 = \beta_{00} + \beta_{01} \cdot 45 + \psi_0 \cdot 1 + \varepsilon_0$$

- После обучения: $\beta_{00} = 1, \beta_{01} = 0.3, \psi_0 = -2$
- Предсказание:

$$\widetilde{Y} = 1 + 0.3 \cdot 45 + (-2) \cdot 1 = 12.5$$

- Ошибка:

$$\varepsilon_0 = target_0 - \widetilde{Y} = 15.5 - 12.5 = 3.0$$

Q-Learning, пример (7)

Результаты для пациентки M

t	$target_t$	Как получен	Предсказание	Ошибка ε_t
$t = 2$	$Y = 28$	Actual tumor size	-	-
$t = 1$	19	$\min(Q_2(0), Q_2(1))$	15.5	3.5
$t = 0$	15.5	$\min(Q_1(0), Q_1(1))$	12.5	3.0

Q-Learning, пример (8)

Применение к новой пациентке

- Пусть новая пациентка имеет $L(0) = 60$ мм. Подставляем в модель для $t = 0$:

$$Q_0(A_0 = 0) = 1 + 0.3 \cdot 60 + (-2) \cdot 0 = 19$$
$$Q_0(A_0 = 1) = 1 + 0.3 \cdot 60 + (-2) \cdot 1 = 17$$

- Так как $17 < 19$, **рекомендуем химиотерапию на старте**

Weighted Q-Learning

Применяется, чтобы сделать алгоритм более устойчивым к искажениям в наблюдаемых данных

- **Двойная устойчивость (Double-Robustness)**: взвесить каждое наблюдение с помощью обратной вероятности назначения лечения (Inverse Probability of Treatment Weighting, IPTW)
- Пациенту присваивается вес 1/P(A(t) | История) так, что “редкие” случаи (пациенты, которые получили нестандартное для себя лечение) имеют больший вклад в обучение
- **Doubly-robust**: (1) модель найдет правильную оптимальную стратегию лечения, даже если ошиблись в построении Q-функции; (2) ошибки в модели весов будут компенсированы, если модель исхода верна

Спасибо за внимание

?