ETH zürich

Interpretable Off-Policy Learning via Hyperbox Search

International Conference on Machine Learning (ICML) 2022

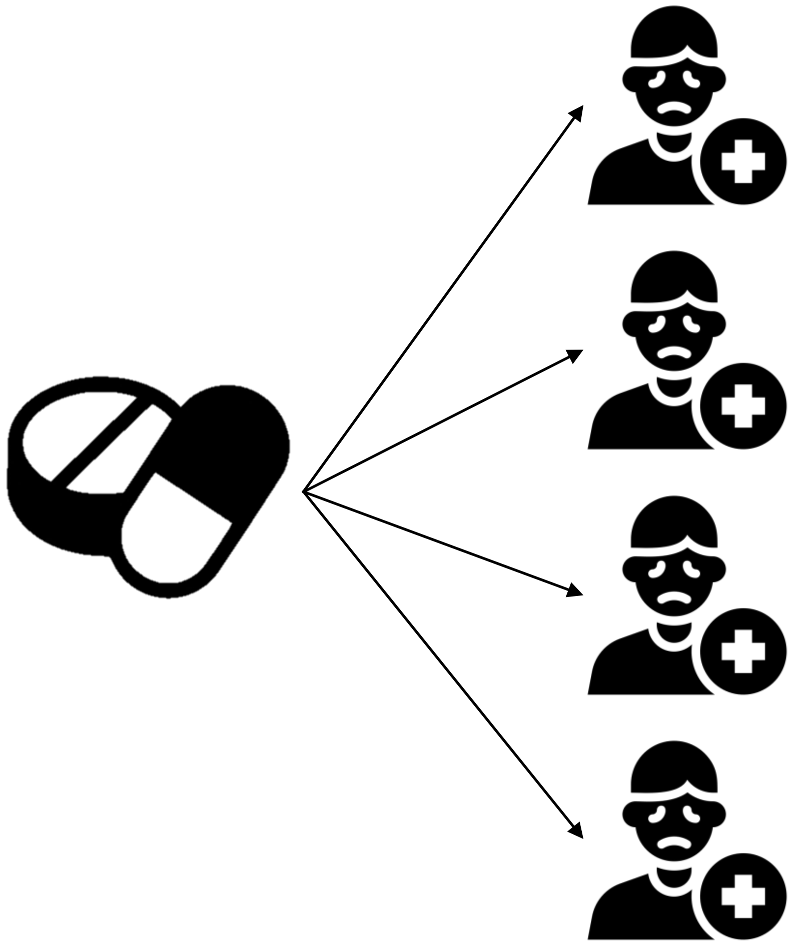
Baltimore, Maryland, USA

Daniel Tschernutter

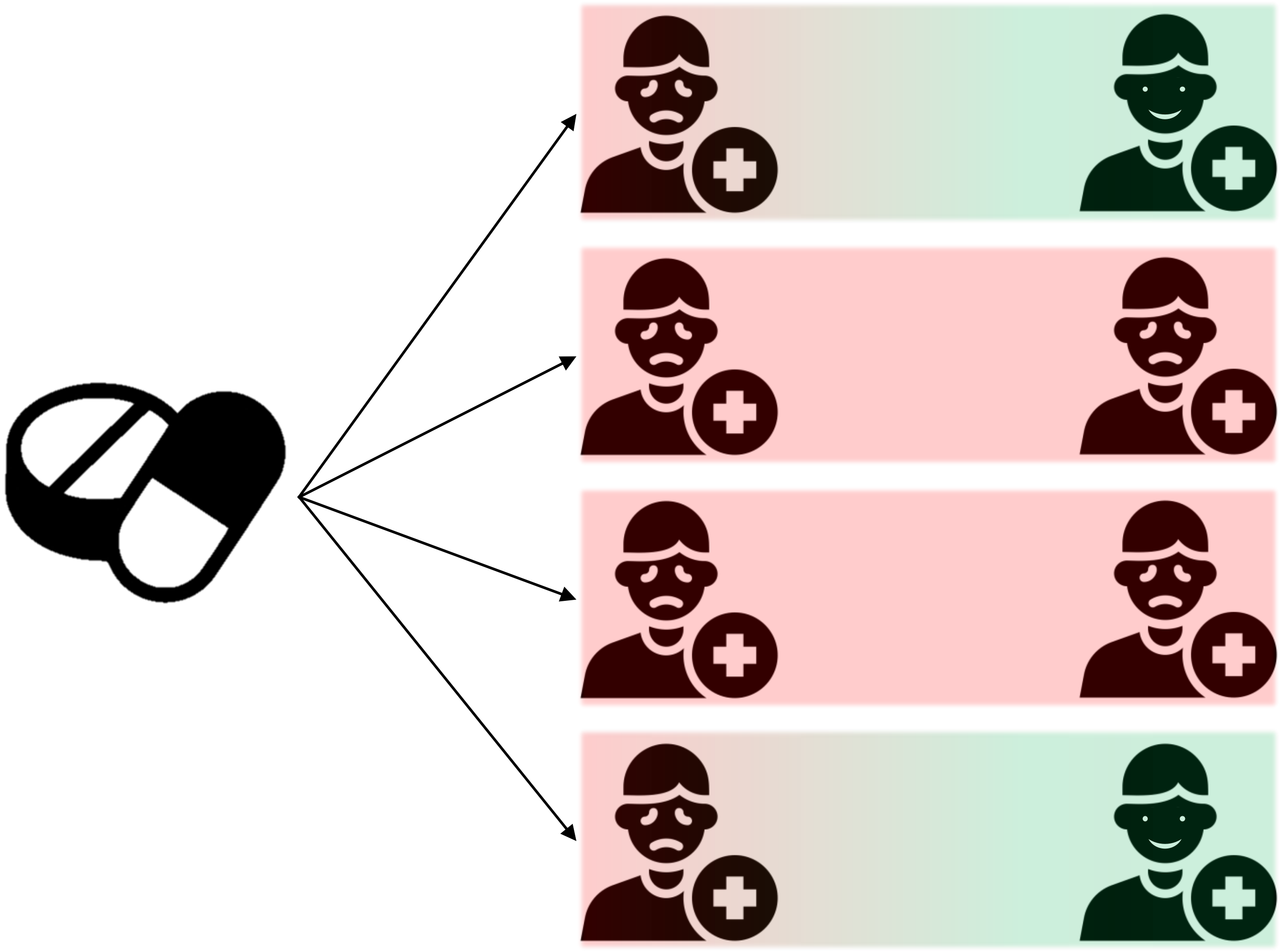
Tobias Hatt

Stefan Feuerriegel

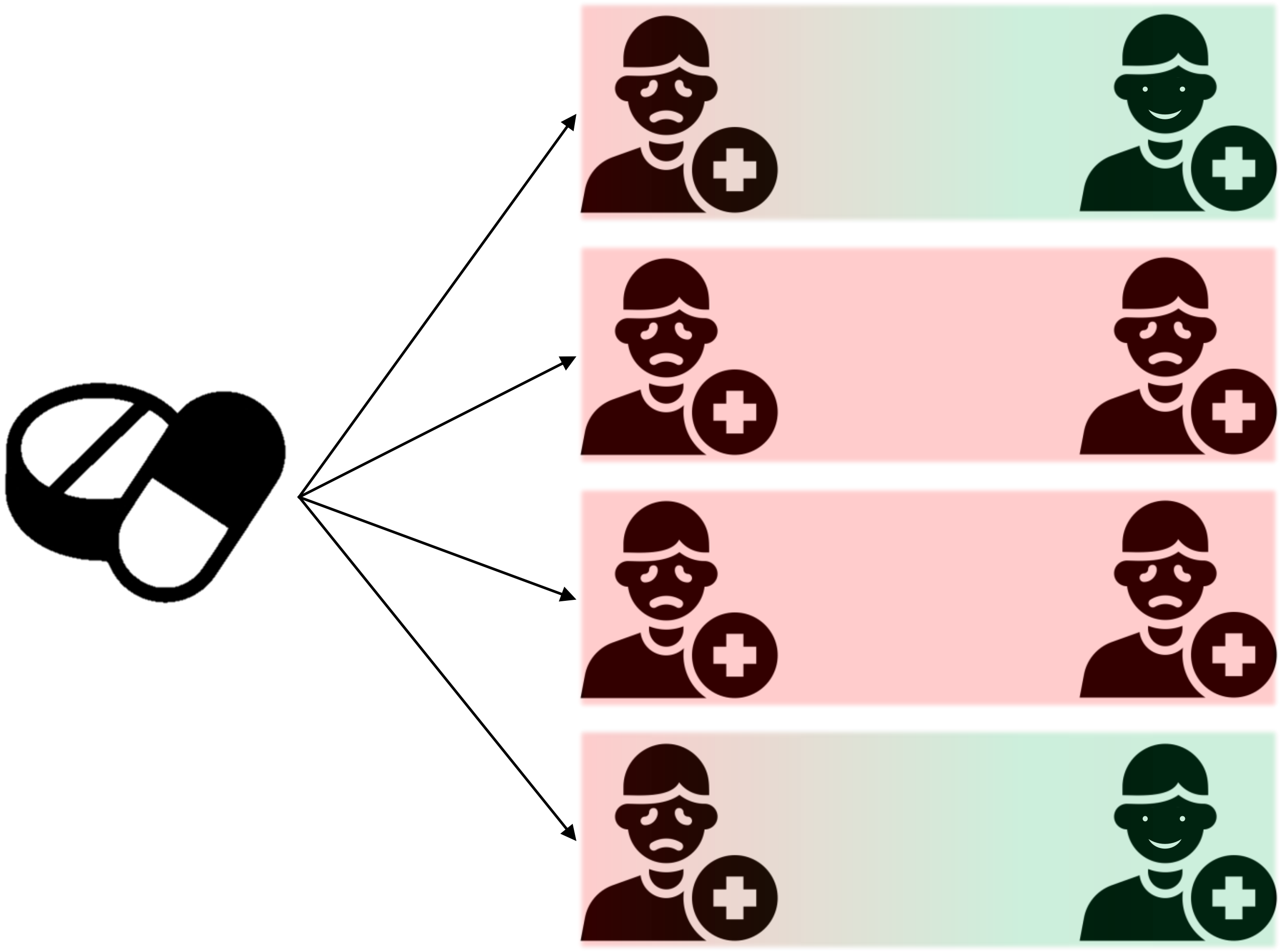
Why Off-Policy Learning?



Why Off-Policy Learning?

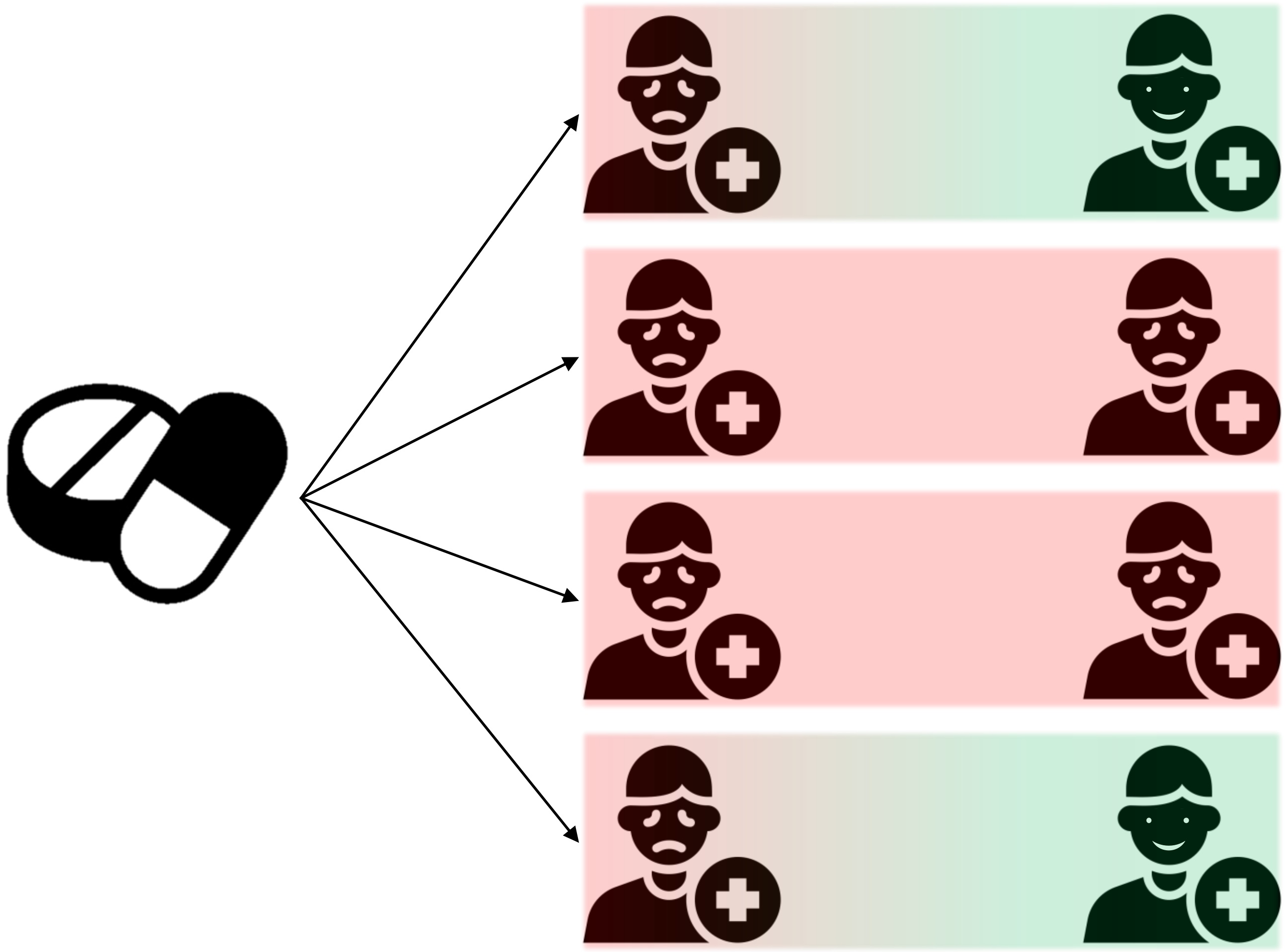


Why Off-Policy Learning?



- Aim is to **select treatments** that are effective for **individual patients**

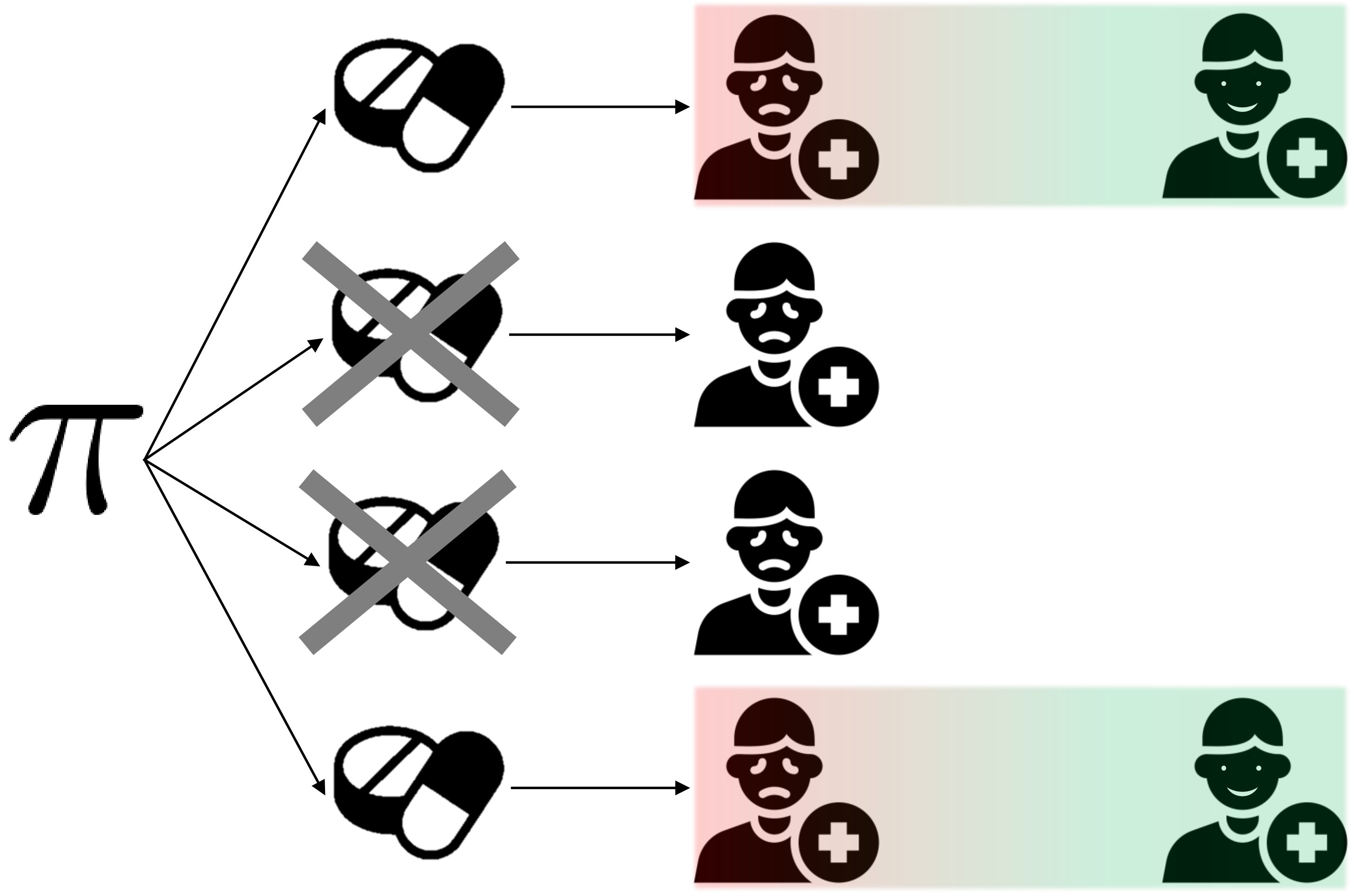
Why Off-Policy Learning?



- Aim is to **select treatments** that are effective for **individual patients**
- Personalized decision-making formalized via so called **policies**

$$\pi : \mathcal{X} \subseteq \mathbb{R}^d \rightarrow \{-1, 1\}$$

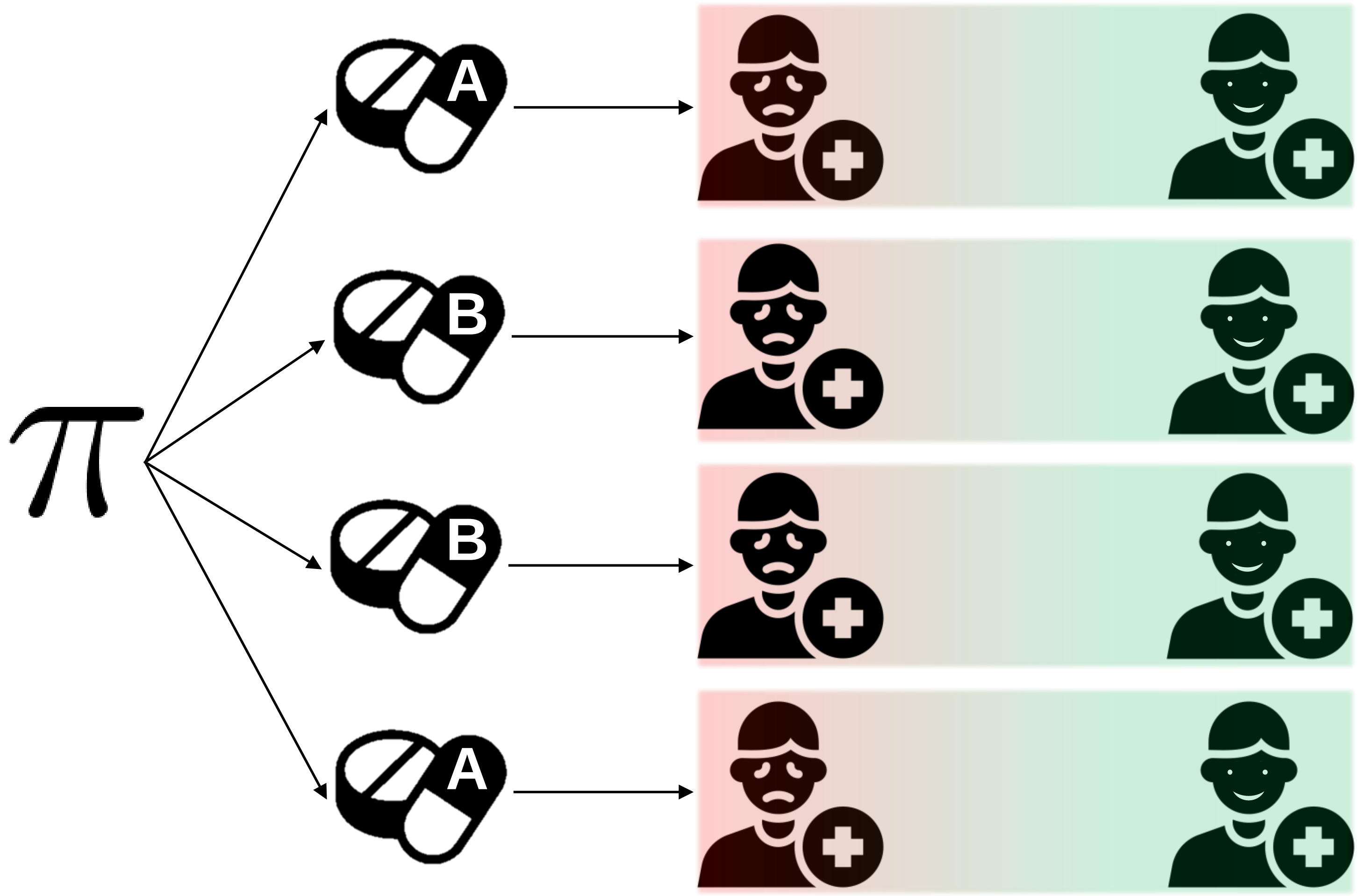
Why Off-Policy Learning?



- Aim is to **select treatments** that are effective for **individual patients**
- Personalized decision-making formalized via so called **policies**

$$\pi : \mathcal{X} \subseteq \mathbb{R}^d \rightarrow \{-1, 1\}$$

Why Off-Policy Learning?



- Aim is to **select treatments** that are effective for **individual patients**
- Personalized decision-making formalized via so called **policies**

$$\pi : \mathcal{X} \subseteq \mathbb{R}^d \to \{-1, 1\}$$

Off-Policy Learning

1 Formalize objective via policy value

$$V(\pi) = \mathbb{E}\left[Y(1)\ I(\pi(X) = 1) + Y(-1)\ I(\pi(X) = -1)\right]$$

Off-Policy Learning

1 Formalize objective via policy value

$$V(\pi) = \mathbb{E}\left[Y(1)\, I(\pi(X) = 1) + Y(-1)\, I(\pi(X) = -1)\right]$$

Potential outcomes

Policy value

Indicator function for treatment decisions

Off-Policy Learning

1 **Formalize objective via policy value**

2 **Estimate policy value from data**

$$V(\pi) = \mathbb{E}\left[Y(1)\ I(\pi(X) = 1) + Y(-1)\ I(\pi(X) = -1)\right]$$

$$\mathcal{J}(\pi) = \mathbb{E}\left[\psi I(T \neq \pi(X))\right]$$

$$\psi^{\mathrm{DM}} = T(\mu_{-1}(X) - \mu_1(X))$$

$$\psi^{\mathrm{IPS}} = \frac{-Y}{e_T(X)}$$

$$\psi^{\mathrm{DR}} = \psi^{\mathrm{DM}} + \psi^{\mathrm{IPS}} + \frac{\mu_T(X)}{e_T(X)}$$

$$\mu_t(x) = \mathbb{E}\left[Y(t) \mid X = x\right]$$

$$e_t(x) = \mathbb{P}\left(T = t \mid X = x\right)$$

Standard methods:

DM – direct method
IPS – inverse propensity score
DR – doubly robust method

Off-Policy Learning

1 Formalize objective via policy value

2 Estimate policy value from data

$$V(\pi) = \mathbb{E}\left[Y(1)\ I(\pi(X)=1) + Y(-1)\ I(\pi(X)=-1)\right]$$

$$\left.\begin{aligned}
\mathcal{J}(\pi) &= \mathbb{E}\left[\psi I(T \neq \pi(X))\right] \\
\psi^{\mathrm{DM}} &= T(\mu_{-1}(X) - \mu_1(X)) \\
\psi^{\mathrm{IPS}} &= \frac{-Y}{e_T(X)} \\
\psi^{\mathrm{DR}} &= \psi^{\mathrm{DM}} + \psi^{\mathrm{IPS}} + \frac{\mu_T(X)}{e_T(X)} \\
\mu_t(x) &= \mathbb{E}\left[Y(t) \mid X = x\right] \\
e_t(x) &= \mathbb{P}\left(T = t \mid X = x\right)
\end{aligned}\right\} \min_\pi V(\pi) \Leftrightarrow \min_\pi \mathcal{J}(\pi)$$

Off-Policy Learning

1 Formalize objective via policy value

$$V(\pi) = \mathbb{E}\left[Y(1)\ I(\pi(X)=1) + Y(-1)\ I(\pi(X)=-1)\right]$$

2 Estimate policy value from data

$$\mathcal{J}(\pi) = \mathbb{E}\left[\psi I(T \neq \pi(X))\right]$$

$$\psi^{\mathrm{DM}} = T(\mu_{-1}(X) - \mu_1(X))$$

$$\psi^{\mathrm{IPS}} = \frac{-Y}{e_T(X)}$$

$$\psi^{\mathrm{DR}} = \psi^{\mathrm{DM}} + \psi^{\mathrm{IPS}} + \frac{\mu_T(X)}{e_T(X)}$$

$$\mu_t(x) = \mathbb{E}\left[Y(t) \mid X = x\right]$$

$$e_t(x) = \mathbb{P}\left(T = t \mid X = x\right)$$

$$\min_\pi V(\pi) \Leftrightarrow \min_\pi \mathcal{J}(\pi)$$

$$\min_\pi \mathcal{J}_n(\pi)$$

Estimate from data

Off-Policy Learning

1 Formalize objective via policy value

$$V(\pi) = \mathbb{E}\left[Y(1)\ I(\pi(X)=1) + Y(-1)\ I(\pi(X)=-1)\right]$$

2 Estimate policy value from data

$$\mathcal{J}(\pi) = \mathbb{E}\left[\psi I(T \neq \pi(X))\right]$$
$$\psi^{\mathrm{DM}} = T(\mu_{-1}(X) - \mu_1(X))$$
$$\psi^{\mathrm{IPS}} = \frac{-Y}{e_T(X)}$$
$$\psi^{\mathrm{DR}} = \psi^{\mathrm{DM}} + \psi^{\mathrm{IPS}} + \frac{\mu_T(X)}{e_T(X)}$$
$$\mu_t(x) = \mathbb{E}\left[Y(t) \mid X = x\right]$$
$$e_t(x) = \mathbb{P}\left(T = t \mid X = x\right)$$

$$\min_\pi V(\pi) \Leftrightarrow \min_\pi \mathcal{J}(\pi)$$

$$\min_\pi \mathcal{J}_n(\pi)$$

3 Optimize over pre-specified policy class *Π*

$$\min_{\pi \in \Pi} \mathcal{J}_n(\pi)$$

Off-Policy Learning

1 Formalize objective via policy value

$$V(\pi) = \mathbb{E}\left[Y(1)\ I(\pi(X)=1) + Y(-1)\ I(\pi(X)=-1)\right]$$

2 Estimate policy value from data

$$\mathcal{J}(\pi) = \mathbb{E}\left[\psi I(T \neq \pi(X))\right]$$

$$\psi^{\text{DM}} = T(\mu_{-1}(X) - \mu_1(X))$$

$$\psi^{\text{IPS}} = \frac{-Y}{e_T(X)}$$

$$\psi^{\text{DR}} = \psi^{\text{DM}} + \psi^{\text{IPS}} + \frac{\mu_T(X)}{e_T(X)}$$

$$\mu_t(x) = \mathbb{E}\left[Y(t) \mid X = x\right]$$

$$e_t(x) = \mathbb{P}\left(T = t \mid X = x\right)$$

$$\min_\pi V(\pi) \Leftrightarrow \min_\pi \mathcal{J}(\pi)$$

$$\min_\pi \mathcal{J}_n(\pi)$$

3 Optimize over pre-specified policy class *Π*

$$\min_{\pi \in \Pi} \mathcal{J}_n(\pi)$$

Requirements from clinical practice

Requirements in Clinical Practice

$$\pi \in \Pi$$

Sufficiently Rich

- It should contain a **large class of policies**
- Mitigate the risk of model misspecification
- Directly related to policies that yield **low regrets**, i.e., the difference between clinical outcome across the population of the learned policy and the a priori best policy

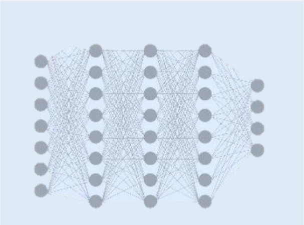
Interpretable

- Decisions can be explained in **understandable terms** to humans
- Clinical practitioners need to understand **which** treatment is chosen **when**
- Important for **debugging**, detecting **biases**, and for patients to gain **trust** in the algorithm

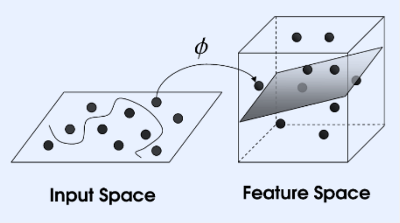
Requirements in Clinical Practice

$$\pi \in \Pi$$

Sufficiently Rich



Neural Networks



Kernel Methods

Interpretable

if f_1 **then**
  output b_1
else if f_2 **then**
  output b_2
...
else if f_r **then**
  output b_r

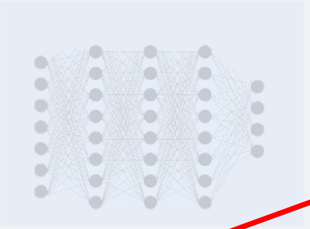
Decision Lists

Linear Eligibility Scores

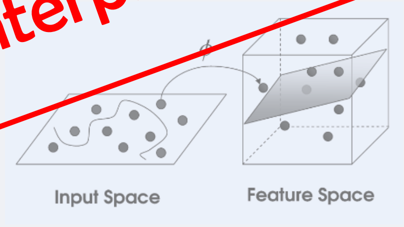
Requirements in Clinical Practice

$$\pi \in \Pi$$

Sufficiently Rich



Neural Networks



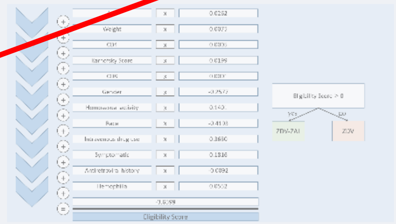
Kernel Methods

Not Interpretable

Interpretable

if f_1 then
    output b_1
else if f_2 then
    output b_2
...
else if f_r then
    output b_r

Decision Lists



Linear Eligibility Scores

Not Sufficiently Rich

Idea: Interpretable Policy Class via Hyperboxes

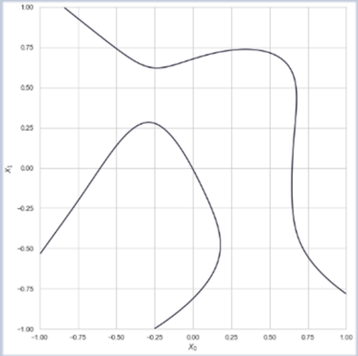
$$\min_{\pi} V(\pi) = \min_{\pi} \mathbb{E}\left[Y(1)\ I(\pi(X) = 1) + Y(-1)\ I(\pi(X) = -1)\right]$$

ETH zürich

Idea: Interpretable Policy Class via Hyperboxes

$$\min_{\pi} V(\pi) = \min_{\pi} \mathbb{E}\left[Y(1)\ I(\pi(X) = 1) + Y(-1)\ I(\pi(X) = -1)\right]$$

$$\pi : \mathcal{X} \subseteq \mathbb{R}^d \to \{-1, 1\}$$

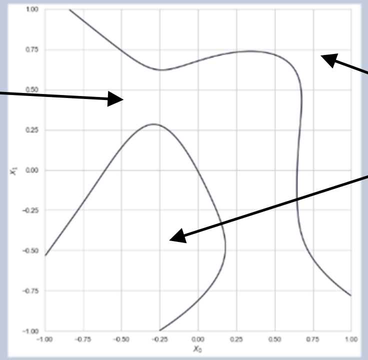


Idea: Interpretable Policy Class via Hyperboxes

$$\min_{\pi} V(\pi) = \min_{\pi} \mathbb{E}\left[Y(1)\ I(\pi(X)=1) + Y(-1)\ I(\pi(X)=-1)\right]$$

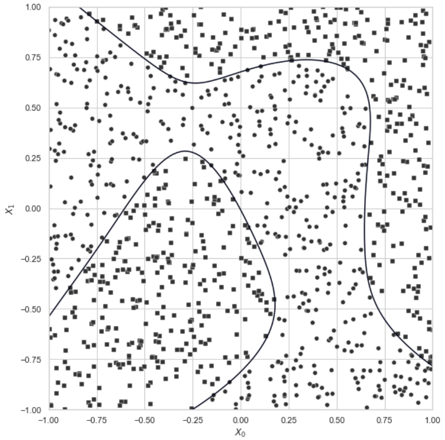
$$\pi : \mathcal{X} \subseteq \mathbb{R}^d \rightarrow \{-1, 1\}$$

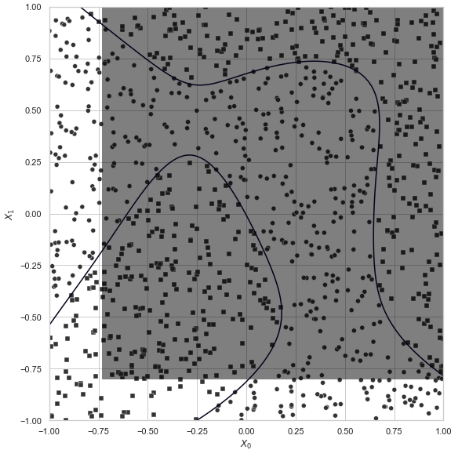
Do not treat patient



Treat patient

Policy Class Π_H^M: Why Hyperboxes?





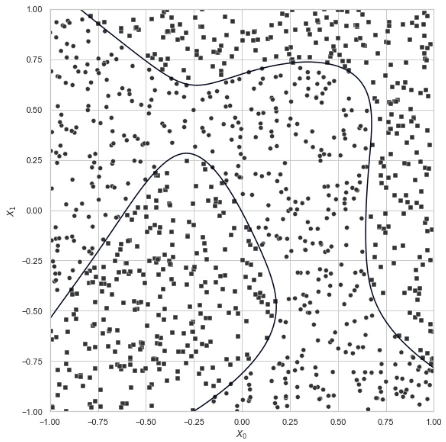
$M = 1$

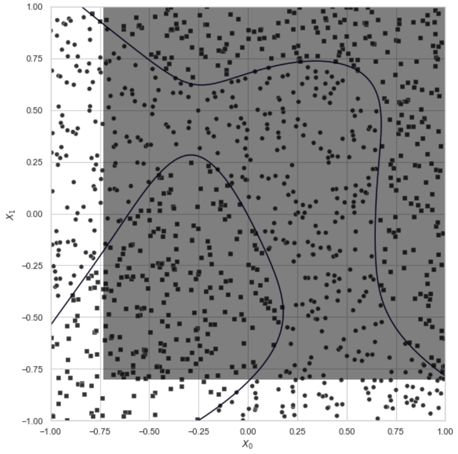
$M = 3$

$M = 5$

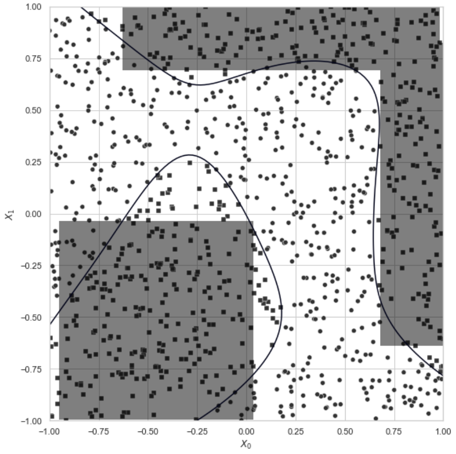
$M = 10$

Policy Class Π_H^M: Why Hyperboxes?





$M = 1$

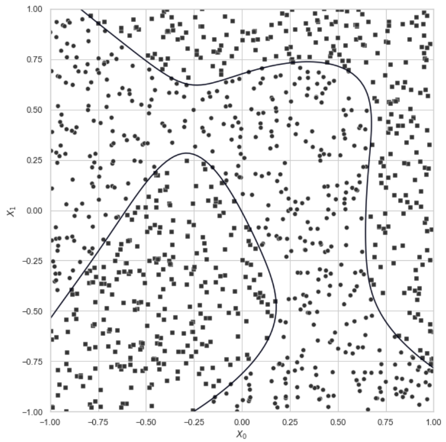


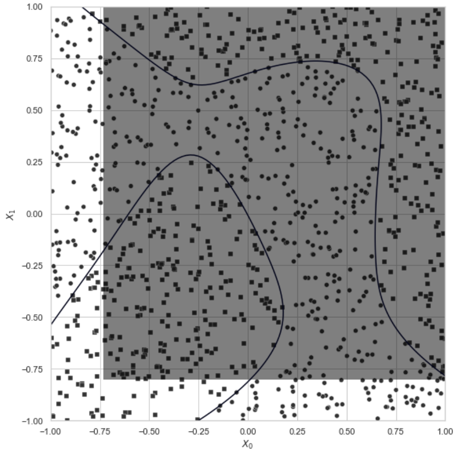
$M = 3$

$M = 5$

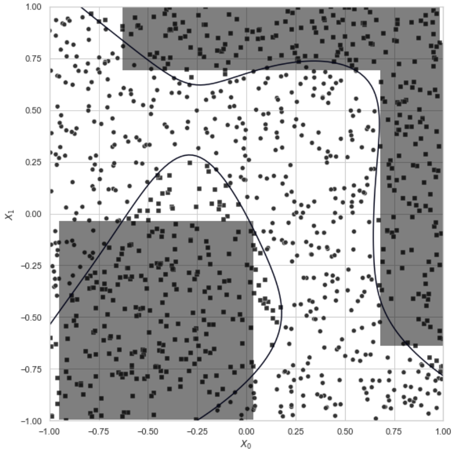
$M = 10$

Policy Class Π_H^M: Why Hyperboxes?

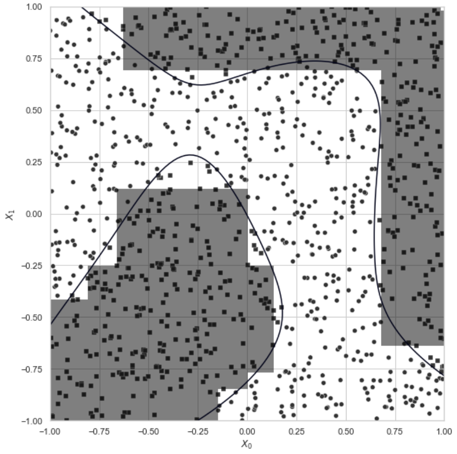




$M = 1$



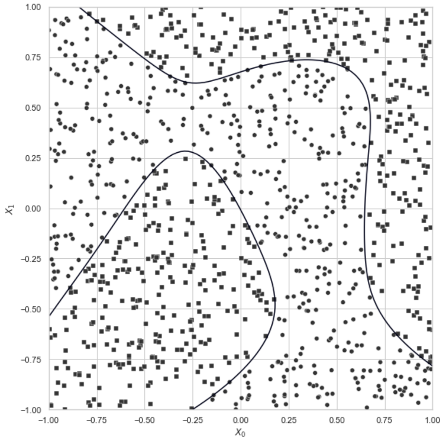
$M = 3$

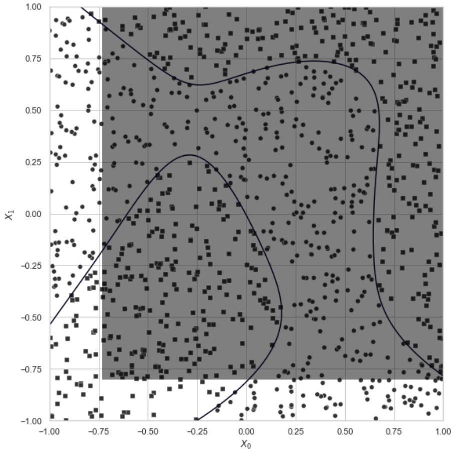


$M = 5$

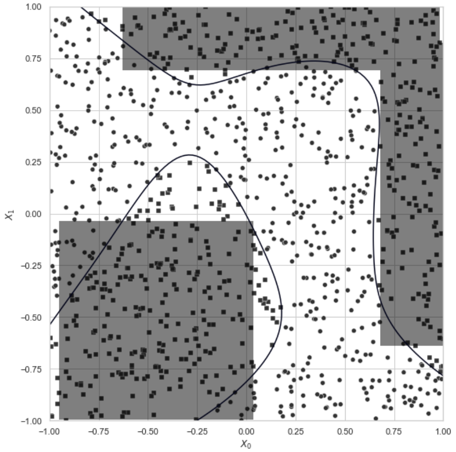
$M = 10$

Policy Class Π_H^M: Why Hyperboxes?

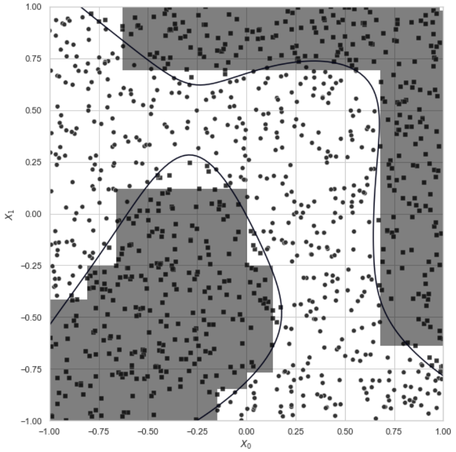




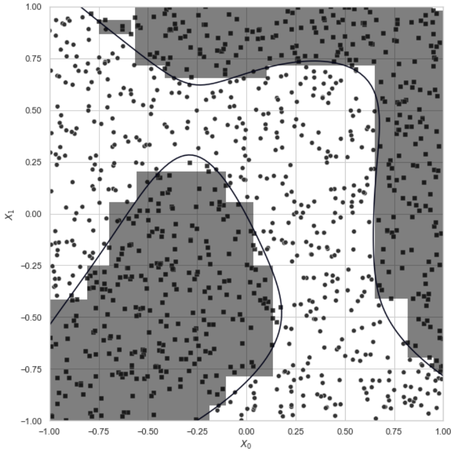
$M = 1$



$M = 3$

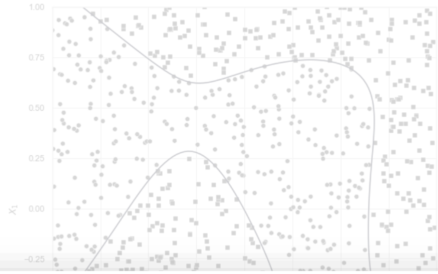


$M = 5$



$M = 10$

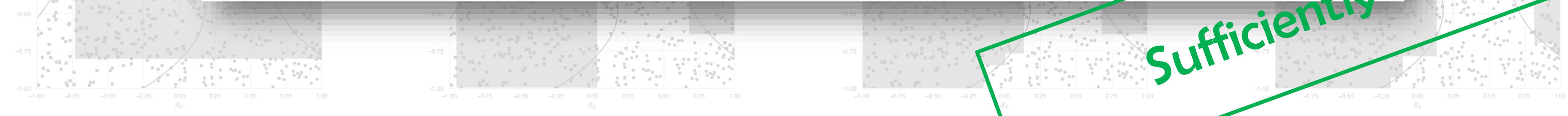
Policy Class Π_H^M: Why Hyperboxes?



Theorem 1. *Let $1 \leq p < \infty$ and $\pi^* : \mathcal{X} \to \{-1, 1\}$ be any Lebesgue measurable function*[4]*. Then, for every $\delta \in (0, 1)$ and $\epsilon > 0$, there exists a sample size $n_{\delta,\epsilon} \in \mathbb{N}$ and $M \in \mathbb{N}$ sufficiently large, as well as, a policy $\pi_{\mathcal{D}^*} \in \Pi_H^M$ as defined in (2.3), such that*

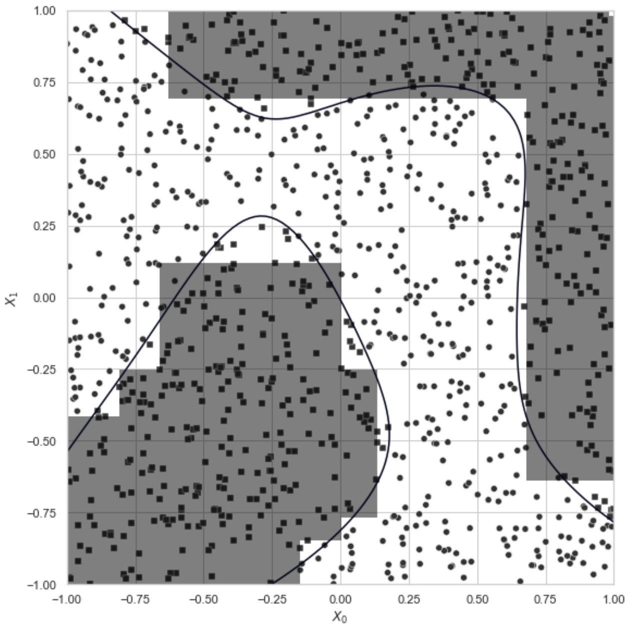
$$\|\pi^* - \pi_{\mathcal{D}^*}\|_p < \epsilon \tag{2.15}$$

holds with probability at least $1 - \delta$.



Sufficiently Rich

Policy Class Π_H^M: Why are Hyperboxes Interpretable?



If Age in [13 , 62] **AND** weight in [47.8 , 87.09] **AND** CD4 in [122 , 411] **AND** Karnofsky score in [90 , 100] **AND** CD8 in [116 , 2130] → **ZDV-ZAL**

OR if Age in [25 , 54] **AND** weight in [49.44 , 111.13] **AND** CD4 in [437 , 771] **AND** Karnofsky score in [80 , 100] **AND** CD8 in [1059 , 2739] **AND** no drug abuse **AND** no hemophilia → **ZDV-ZAL**

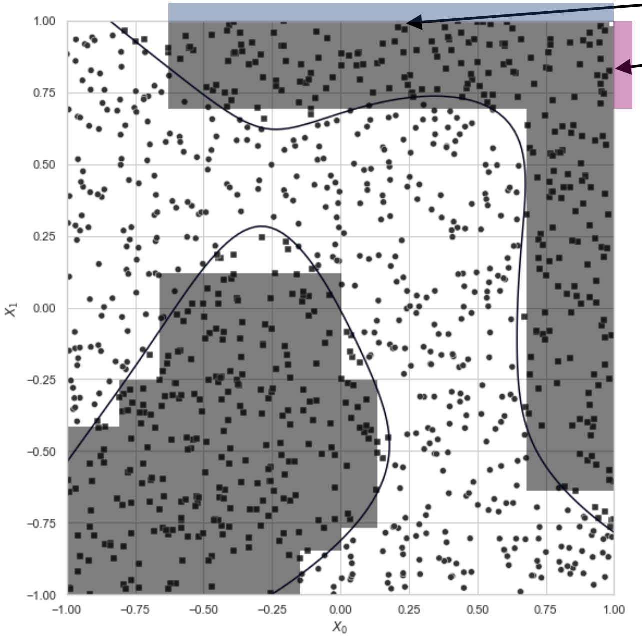
OR if Age in [23 , 40] **AND** weight in [52.39 , 83.0] **AND** CD4 in [450 , 630] **AND** Karnofsky score in [90 , 100] **AND** CD8 in [370 , 978] → **ZDV-ZAL**

OR if Age in [17 , 43] **AND** weight in [92.7 , 149.0] **AND** CD4 in [143 , 770] **AND** Karnofsky score in [80 , 100] **AND** CD8 in [418 , 3190] → **ZDV-ZAL**

OR if Age in [31 , 66] **AND** weight in [81.8 , 89.81] **AND** CD4 in [415 , 663] **AND** Karnofsky score in [90 , 100] **AND** CD8 in [752 , 2248] **AND** no drug abuse → **ZDV-ZAL**

ELSE → **ZDV**

Policy Class Π_H^M: Why are Hyperboxes Interpretable?



If Age in [13 , 62] **AND** weight in [47.8 , 87.09] **AND** CD4 in [122 , 411] **AND** Karnofsky score in [90 , 100] **AND** CD8 in [116 , 2130] → **ZDV-ZAL**

OR if Age in [25 , 54] **AND** weight in [49.44 , 111.13] **AND** CD4 in [437 , 771] **AND** Karnofsky score in [80 , 100] **AND** CD8 in [1059 , 2739] **AND** no drug abuse **AND** no hemophilia → **ZDV-ZAL**

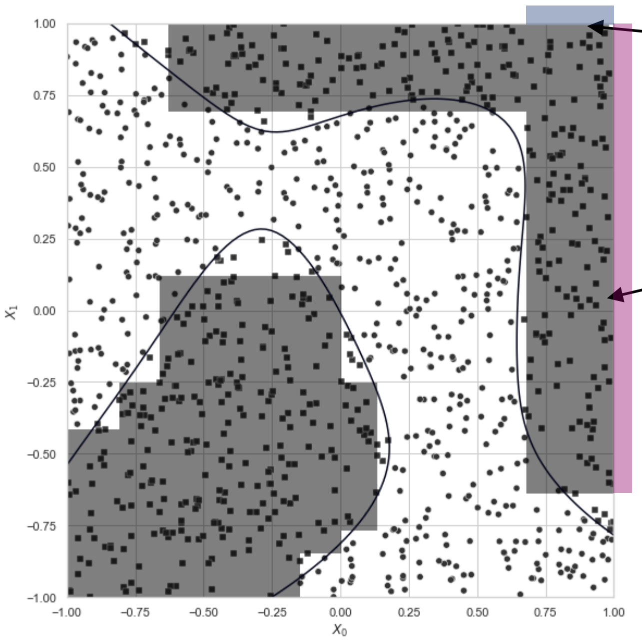
OR if Age in [23 , 40] **AND** weight in [52.39 , 83.0] **AND** CD4 in [450 , 630] **AND** Karnofsky score in [90 , 100] **AND** CD8 in [370 , 978] → **ZDV-ZAL**

OR if Age in [17 , 43] **AND** weight in [92.7 , 149.0] **AND** CD4 in [143 , 770] **AND** Karnofsky score in [80 , 100] **AND** CD8 in [418 , 3190] → **ZDV-ZAL**

OR if Age in [31 , 66] **AND** weight in [81.8 , 89.81] **AND** CD4 in [415 , 663] **AND** Karnofsky score in [90 , 100] **AND** CD8 in [752 , 2248] **AND** no drug abuse → **ZDV-ZAL**

ELSE → **ZDV**

Policy Class Π_H^M: Why are Hyperboxes Interpretable?



If Age in [13 , 62] **AND** weight in [47.8 , 87.09] **AND** CD4 in [122 , 411] **AND** Karnofsky score in [90 , 100] **AND** CD8 in [116 , 2130] → **ZDV-ZAL**

OR if Age in [25 , 54] **AND** weight in [49.44 , 111.13] **AND** CD4 in [437 , 771] **AND** Karnofsky score in [80 , 100] **AND** CD8 in [1059 , 2739] **AND** no drug abuse **AND** no hemophilia → **ZDV-ZAL**

OR if Age in [23 , 40] **AND** weight in [52.39 , 83.0] **AND** CD4 in [450 , 630] **AND** Karnofsky score in [90 , 100] **AND** CD8 in [370 , 978] → **ZDV-ZAL**

OR if Age in [17 , 43] **AND** weight in [92.7 , 149.0] **AND** CD4 in [143 , 770] **AND** Karnofsky score in [80 , 100] **AND** CD8 in [418 , 3190] → **ZDV-ZAL**

OR if Age in [31 , 66] **AND** weight in [81.8 , 89.81] **AND** CD4 in [415 , 663] **AND** Karnofsky score in [90 , 100] **AND** CD8 in [752 , 2248] **AND** no drug abuse → **ZDV-ZAL**

ELSE → **ZDV**

Policy Class Π_H^M: Why are Hyperboxes Interpretable?

If Age in [13 , 62] **AND** weight in [47.8 , 87.09] **AND** CD4 in [122 , 411] **AND** Karnofsky score in [90 , 100] **AND** CD8 in [116 , 2130] → ZDV-ZAL

OR if Age in [25 , 54] **AND** weight in [49.44 , 111.13] **AND** CD4 in [437 , 771] **AND** Karnofsky score in [80 , 100] **AND** CD8 in [1059 , 2739] **AND** no drug abuse **AND** no hemophilia → **ZDV-ZAL**

OR if Age in [23 , 40] **AND** weight in [52.39 , 83.0] **AND** CD4 in [450 , 630] **AND** Karnofsky score in [90 , 100] **AND** CD8 in [370 , 978] → **ZDV-ZAL**

OR if Age in [17 , 43] **AND** weight in [92.7 , 149.0] **AND** CD4 in [143 , 770] **AND** Karnofsky score in [80 , 100] **AND** CD8 in [418 , 3190] → **ZDV-ZAL**

OR if Age in [31 , 66] **AND** weight in [81.8 , 89.81] **AND** CD4 in [415 , 663] **AND** Karnofsky score in [90 , 100] **AND** CD8 in [752 , 2248] **AND** no drug abuse → **ZDV-ZAL**

ELSE → **ZDV**

Interpretable

IOPL Algorithm

MILP Formulation of Off-Policy Learning

$$
\begin{aligned}
&\min \frac{1}{n}\sum_{i=1}^{n}\psi_i\xi_i \\
&\text{s.t. } \xi_i + \sum\nolimits_{j\in\mathcal{K}_i} s_j \geq 1 && \text{for } i \in I_1 \cap \mathcal{P}, \\
&\quad \xi_i \geq s_j && \text{for } i \in I_{-1} \cap \mathcal{P} \text{ and } j \in \mathcal{K}_i, \\
&\quad \xi_i \leq 1 - s_j && \text{for } i \in I_1 \cap \mathcal{N} \text{ and } j \in \mathcal{K}_i, \\
&\quad \xi_i \leq \sum\nolimits_{j\in\mathcal{K}_i} s_j && \text{for } i \in I_{-1} \cap \mathcal{N}, \\
&\quad \sum\nolimits_{j=1}^{N} s_j \leq M, \\
&\quad s_j \in \{0,1\}, \xi_i \in [0,1].
\end{aligned}
$$

IOPL Algorithm

MILP Formulation of Off-Policy Learning

$$
\begin{aligned}
&\min \frac{1}{n}\sum_{i=1}^{n}\psi_i\xi_i && \\
&\text{s.t. } \xi_i + \sum\nolimits_{j\in\mathcal{K}_i} s_j \geq 1 && \text{for } i \in I_1 \cap \mathcal{P}, \\
&\quad \xi_i \geq s_j && \text{for } i \in I_{-1} \cap \mathcal{P} \text{ and } j \in \mathcal{K}_i, \\
&\quad \xi_i \leq 1 - s_j && \text{for } i \in I_1 \cap \mathcal{N} \text{ and } j \in \mathcal{K}_i, \\
&\quad \xi_i \leq \sum\nolimits_{j\in\mathcal{K}_i} s_j && \text{for } i \in I_{-1} \cap \mathcal{N}, \\
&\quad \sum\nolimits_{j=1}^{N} s_j \leq M, && \\
&\quad s_j \in \{0,1\}, \xi_i \in [0,1]. &&
\end{aligned}
$$

Interpretable **O**ff-**P**olicy **L**earning

Algorithm 1: IOPL

Input: Initial working set $\mathcal{W}_0$
Output: Optimal subset $\mathcal{K}^* \subseteq \mathcal{K}$, optimal solution s^*
Initialize the list of active subproblems $\mathcal{L} \leftarrow \{\text{RMILP}(\mathcal{W}_0, \emptyset)\}$
Initialize iteration counter $l \leftarrow 0$
while $\mathcal{L}$ not empty **do**
  Select and remove first subproblem $\text{RMILP}(\mathcal{W}, \mathcal{C})$ from $\mathcal{L}$
  Perform column generation to get the current relaxed solution and the new working set $s', v', \mathcal{W}' \leftarrow \text{ColumnGeneration}(\text{RMILP}(\mathcal{W}, \mathcal{C}))$
  if $l = 0$ **then**
    Solve restricted integer problem to get current optimal integer solution and objective value $s^*, v^* \leftarrow \text{Solve}(\text{MILP}(\mathcal{W}'))$
    Update optimal subset $\mathcal{W}^* \leftarrow \mathcal{W}'$
  end
  if $v' \leq v^*$ **then**
    if s' integral **then**
      Update new optimal integer solution $(\mathcal{W}^*, s^*) \leftarrow (\mathcal{W}', s')$
    else
      Set j' according to branching rule
      Branch by updating $\mathcal{L} \leftarrow \mathcal{L} \cup \{\text{RMILP}(\mathcal{W}', \mathcal{C} \cup \{(j', 1)\}), \text{RMILP}(\mathcal{W}', \mathcal{C} \cup \{(j', 0)\})\}$
      Solve restricted problem $(s', v') \leftarrow \text{Solve}(\text{MILP}(\mathcal{W}'))$
      if $v' \leq v^*$ **then**
        Update new optimal integer solution $(\mathcal{W}^*, s^*) \leftarrow (\mathcal{W}', s')$
      end
    end
  end
  Increase counter $l \leftarrow l + 1$
end

IOPL is a highly efficient **branch-and-price algorithm**, i.e., a column generation procedure within a branch-and-bound framework

IOPL Algorithm

MILP Formulation of Off-Policy Learning

$$
\begin{aligned}
\min\ & \frac{1}{n}\sum_{i=1}^{n} \psi_i \xi_i \\
\text{s.t. } & \xi_i + \sum\nolimits_{j\in\mathcal{K}_i} s_j \geq 1 && \text{for } i \in I_1 \cap \mathcal{P}, \\
& \xi_i \geq s_j && \text{for } i \in I_{-1} \cap \mathcal{P} \text{ and } j \in \mathcal{K}_i, \\
& \xi_i \leq 1 - s_j && \text{for } i \in I_1 \cap \mathcal{N} \text{ and } j \in \mathcal{K}_i, \\
& \xi_i \leq \sum\nolimits_{j\in\mathcal{K}_i} s_j && \text{for } i \in I_{-1} \cap \mathcal{N}, \\
& \sum\nolimits_{j=1}^{N} s_j \leq M, \\
& s_j \in \{0,1\}, \xi_i \in [0,1].
\end{aligned}
$$

IOPL is a highly efficient **branch-and-price algorithm**, i.e., a column generation procedure within a branch-and-bound framework

Interpretable **O**ff-**P**olicy **L**earning

Algorithm 1: IOPL

Input: Initial working set $\mathcal{W}_0$
Output: Optimal subset $\mathcal{K}^* \subseteq \mathcal{K}$, optimal solution s^*
Initialize the list of active subproblems $\mathcal{L} \leftarrow \{\text{RMILP}(\mathcal{W}_0, \emptyset)\}$
Initialize iteration counter $l \leftarrow 0$
while $\mathcal{L}$ not empty **do**
  Select and remove first subproblem $\text{RMILP}(\mathcal{W}, \mathcal{C})$ from $\mathcal{L}$
  Perform column generation to get the current relaxed solution and the new working set $s', v', \mathcal{W}' \leftarrow \text{ColumnGeneration}(\text{RMILP}(\mathcal{W}, \mathcal{C}))$
  if $l = 0$ **then**
    Solve restricted integer problem to get current optimal integer solution and objective value $s^*, v^* \leftarrow \text{Solve}(\text{MILP}(\mathcal{W}'))$
    Update optimal subset $\mathcal{W}^* \leftarrow \mathcal{W}'$
  end
  if $v' \leq v^*$ **then**
    if s' integral **then**
      Update new optimal integer solution $(\mathcal{W}^*, s^*) \leftarrow (\mathcal{W}', s')$
    else
      Set j' according to branching rule
      Branch by updating $\mathcal{L} \leftarrow \mathcal{L} \cup \{\text{RMILP}(\mathcal{W}', \mathcal{C} \cup \{(j', 1)\}), \text{RMILP}(\mathcal{W}', \mathcal{C} \cup \{(j', 0)\})\}$
      Solve restricted problem $(s', v') \leftarrow \text{Solve}(\text{MILP}(\mathcal{W}'))$
      if $v' \leq v^*$ **then**
        Update new optimal integer solution $(\mathcal{W}^*, s^*) \leftarrow (\mathcal{W}', s')$
      end
    end
  end
  Increase counter $l \leftarrow l + 1$
end

For more details on our algorithm, experiments with baselines, more theoretical results, and proofs, see the paper

Thank you!

Interpretable Off-Policy Learning via Hyperbox Search

Daniel Tschernutter[1] Tobias Hatt[1] Stefan Feuerriegel[1,2]

1 ETH zürich

2 LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN