Semi-supervised D-Learning for Optimal Individual Treatment Regimes

Xintong Li

East China Normal University

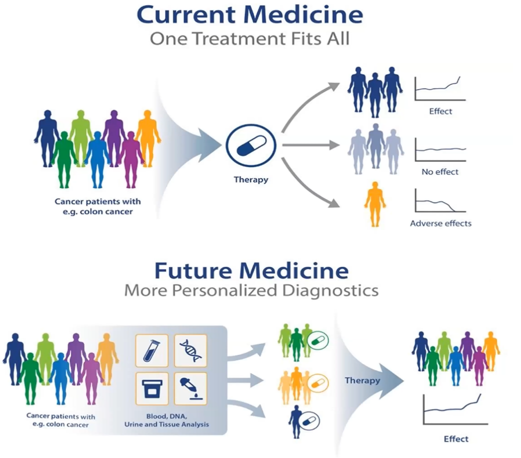
Joint work with Shuyi Zhang and Yong Zhou

July 13, JCSDS 2025

Outline

Precision Medicine

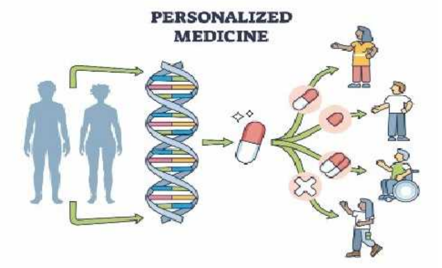
- **Heterogeneity**: different patients respond differently to the same treatment.
  - Positive treatment effects;
  - Side effects.
- **One-size-fits-all** $\rightarrow$ **Precision Medicine**
- **Advantages**:
  - Improve patient adherence;
  - Reduce unnecessary treatments and side effects;
  - Promote recovery;
  - Enhance quality of care and quality of life;
  - Optimize allocation of medical resources;
  - Lower overall healthcare costs;
  - . . . . . .



Precision Medicine - Personalized Decision-Making

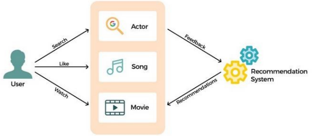
- **Goal**: Find the optimal mapping from individual characteristics $\mathbf{X} \in \mathcal{X}$ to treatments $A \in \mathcal{A}$, i.e. $d^{opt}(\mathbf{X})$, to maximize the expected clinical outcome $E[Y^*(d(\mathbf{X}))]$.
  - $\mathbf{X}$: demographics, clinical features, genetic information, environmental factors, etc.;
  - A: drug choice, dosage, surgery, specific dietary or exercise recommendations, etc.;
  - Y: biomarker levels, survival time, disease progression or remission status, quality of life scores, etc.

- **Applications**:
  - **Disease management**: Recommend the optimal drug dosage based on patient characteristics to optimize treatment efficacy;
  - **Smart health monitoring**: Use wearable devices and biosensors for personalized health management;
  - **Personalized medical intervention**: Combine multimodal data to predict disease risk and enable early intervention.



Personalized Decision-Making - Beyond Precision Medicine

- **Computer Science**: context-aware recommender systems that improve accuracy by incorporating time, location, and social context.



- **Finance**: provide personalized investment advice and wealth management plans based on consumption habits and risk preferences.

- **Public Management**: improve the overall effectiveness of policies through personalized interventions targeting individuals with high social connectivity.

Traditional Methods

- **Q-learning** (Qian and Murphy, 2011; Watkins, 1989; Watkins and Dayan, 1992)
  Define the Q-function: $Q(\mathbf{x}, a) := E[Y|\mathbf{X} = \mathbf{x}, A = a]$, and specify a model $Q(\mathbf{X}, A; \boldsymbol{\beta})$.

$$\widehat{d}^{opt}(\mathbf{X}) = \arg\max_{a \in \mathcal{A}} Q(\mathbf{X}, a; \widehat{\boldsymbol{\beta}}),$$

  where $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} (Y_i - Q(\mathbf{X}_i, A_i; \boldsymbol{\beta}))^2$.

- **A-learning** (Blatt et al., 2004; Murphy, 2003; Robins, 2004)
  Define the contrast function: $C(\mathbf{X}) = Q(\mathbf{X}, 1) - Q(\mathbf{X}, 0)$, then $d^{opt}(\mathbf{X}) = I(C(\mathbf{X}) \geqslant 0)$.

  Doubly robust A-learning: Let $\nu(\mathbf{X}) = E[Y|\mathbf{X}]$ and $\pi(\mathbf{X}) = E[A|\mathbf{X}]$, with corresponding estimators $\widehat{\nu}(\mathbf{X})$ and $\widehat{\pi}(\mathbf{X})$. Specify a model for the contrast function $C(\mathbf{X}; \boldsymbol{\theta})$, then

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{n} \sum_{i=1}^{n} \{Y_i - \widehat{\nu}(\mathbf{X}_i) - [A_i - \widehat{\pi}(\mathbf{X}_i)] C(\mathbf{X}_i; \boldsymbol{\theta})\}^2.$$

Traditional Methods

- **Direct search methods** (Chu et al., 2023; Zhang et al., 2012)
  Denote value function $V(d(\mathbf{X})) := E[Y(d(\mathbf{X}))]$, then $d^{opt}(\mathbf{X}) = \arg\max_{d(\mathbf{X})\in\mathcal{D}} V(d(\mathbf{X}))$.
  - IPW-based estimator: $\widehat{V}_{IPW}(d(\mathbf{X})) = P_n\left[\frac{I(A=d(\mathbf{X}))}{\widehat{\pi}(\mathbf{X},A)}Y\right]$.
  - AIPW-based estimator: $\widehat{V}_{AIPW}(d(\mathbf{X})) = P_n\left[\frac{I(A=d(\mathbf{X}))}{\widehat{\pi}(\mathbf{X},A)}Y - \frac{I(A=d(\mathbf{X}))-\widehat{\pi}(\mathbf{X},A)}{\widehat{\pi}(\mathbf{X},A)}\widehat{Q}(\mathbf{X},d(\mathbf{X}))\right]$, where $Q(\mathbf{X},d(\mathbf{X})) = Q(\mathbf{X},1)I(d(\mathbf{X})=1) + Q(\mathbf{X},0)I(d(\mathbf{X})=0)$.
- **D-learning** (Qi et al., 2020; Qi and Liu, 2018; Shah et al., 2023)

$$d^{\text{opt}}(\mathbf{X}) = \text{sign}\{E[Y|\mathbf{X},A=1] - E[Y|\mathbf{X},A=-1]\} = \text{sign}\left\{E\left[\frac{AY}{\pi(A,\mathbf{X})}\middle|\mathbf{X}\right]\right\} := \text{sign}\{f^{\text{opt}}(\mathbf{X})\}.$$

$$f^{\text{opt}}(\mathbf{X}) \in \arg\min_{f(\mathbf{X})\in\mathcal{F}} E\left[\frac{1}{\pi(\mathbf{X},A)}(2AY - f(\mathbf{X}))^2\right].$$

Chanlleges

In the literature, **linear decision classes** are particularly favored by researchers due to their **simple structure** and **good interpretability** (Chu et al., 2023; Fan et al., 2017; Li et al., 2025).

- **Misspecification** $\rightarrow$ Suboptimal decisions (Maronge et al., 2023; Zhao et al., 2012).
  - Song et al. (2017) proposed a novel method to estimate optimal ITR under a semiparametric additive single-index model and the link function was estimated by B-spline. But it suffers from model mis-specification.
  - Qi and Liu (2018) handled the nonlinear ITR by kenel-based and machine learning methods.
- **Scarcity of labeled data** $\rightarrow$ Underutilization of large amounts of unlabeled data.

Chanlleges

In the literature, **linear decision classes** are particularly favored by researchers due to their **simple structure** and **good interpretability** (Chu et al., 2023; Fan et al., 2017; Li et al., 2025).

- **Misspecification** $\rightarrow$ Suboptimal decisions (Maronge et al., 2023; Zhao et al., 2012).
  - Song et al. (2017) proposed a novel method to estimate optimal ITR under a semiparametric additive single-index model and the link function was estimated by B-spline. But it suffers from model mis-specification.
  - Qi and Liu (2018) handled the nonlinear ITR by kenel-based and machine learning methods.
- **Scarcity of labeled data** $\rightarrow$ Underutilization of large amounts of unlabeled data.

Semi-supervised learning: Leverage information from both labeled and unlabeled data to enhance estimation efficiency and robustness.

Outline

Methodology

Notations $(\mathbf{X}, A, Y)$

$\mathbf{X} = (1, X_1, \ldots, X_p) \in \mathcal{X} \subseteq \mathbb{R}^{p+1}$: p-dimensional covariates $\mathbf{X}^- = (X_1, \ldots, X_p)$ including the interception term, with bounded support $\mathcal{X}$, and positive definite variance Var$(\mathbf{X}^-)$;

$A \in \mathcal{A} = \{-1, 1\}$: the binary treatment indicator;

$Y \in \mathcal{Y} \subseteq \mathbb{R}$: the outcome variable, larger values are better.

Observations $\mathcal{L} \cup \mathcal{U}$

$\mathcal{L} = \{(\mathbf{X}_i, A_i, Y_i) : i = 1, 2, \ldots, n\}$: n iid labeled observations;

$\mathcal{U} = \{\mathbf{X}_i : i = n+1, n+2, \ldots, n+N\}$: N iid unlabeled observations.

Methodology

Semi-supervised assumptions

a. $\mathcal{L} \perp \mathcal{U}$;

b. Observations in $\mathcal{L}$ and $\mathcal{U}$ potentially follow the same distribution;

c. $\frac{n}{N} \to 0$ as $n, N \to \infty$.

Identifiability Assumptions

Let $Y^*(a)$ be the potential outcome.

a. SUTVA: $Y = \sum_{a \in \mathcal{A}} Y^*(a) I(A = a)$;

b. Ignoreability: $A \perp \{Y^*(-1), Y^*(1)\} \mid \mathbf{X}$;

c. Positivity: $0 < P(A = a | \mathbf{X} = \mathbf{x}) < 1$.

Methodology

Optimal ITR: $d^{\text{opt}}(\mathbf{X}) = \arg\max_d V(d) = \arg\max_d E\left[\frac{YI(A=d(\mathbf{X}))}{\pi(A,\mathbf{X})}\right]$.

D-learning - Equivalent representation of optimal ITR

$$d^{\text{opt}}(\mathbf{X}) = \text{sign}\{E[Y|\mathbf{X},A=1] - E[Y|\mathbf{X},A=-1]\} = \text{sign}\left\{E\left[\frac{AY}{\pi(A,\mathbf{X})}\middle|\mathbf{X}\right]\right\} := \text{sign}\{f^{\text{opt}}(\mathbf{X})\}.$$

Structural Nested Mean Model (SNMM)

$$Y = \mu_0(\mathbf{X}) + A\delta(\mathbf{X}) + e, E[e] = 0 \Longrightarrow f^{\text{opt}}(\mathbf{X}) = 2\delta(\mathbf{X}).$$

Linear decision function class: $f \in \mathcal{F} = \{f(\mathbf{X}) = \mathbf{X}^T\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^{p+1}\}$.

Supervised estimation: $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} P_n\left\{\frac{AY}{\pi(A,\mathbf{X})} - \mathbf{X}^T\boldsymbol{\beta}\right\}^2$.

Assume $\pi(1,\mathbf{X}) = 0.5$ in a RCT setting. Define $\boldsymbol{\beta}_0$ as the solution to the equation $E[\mathbf{X}(2AY - \mathbf{X}^T\boldsymbol{\beta})] = 0$. The supervised estimator is obtained by solving

$$P_n\mathbf{X}(2AY - \mathbf{X}^T\boldsymbol{\beta}) = 0.$$

Misspecification of linear decision classes $\Longrightarrow$ Semi-supervised D-learning (SSDL)

SSDL estimator based on fully nonparametric imputation: $\widehat{\beta}_{np}$

Let $m(\mathbf{X}) = E[2AY|\mathbf{X}] = E[2AY|\mathbf{X}^-]$, and its corresponding kernel estimator is

$$\widehat{m}(\mathbf{X}_j) = \frac{(nh^p)^{-1} \sum_{i=1}^n H_h(\mathbf{X}_i^-, \mathbf{X}_j^-) \times 2A_i Y_i}{(nh^p)^{-1} \sum_{i=1}^n H_h(\mathbf{X}_i^-, \mathbf{X}_j^-)},$$

where $H_h(u, v) = H(\frac{u-v}{h})$ with kernel function $H : \mathbb{R}^p \to \mathbb{R}$ and bandwidth h. Then $\widehat{\beta}_{np}$ be obtained by the solution of

$$\frac{1}{N} \sum_{j=n+1}^{n+N} \mathbf{X}_j (\widehat{m}(\mathbf{X}_j) - \mathbf{X}_j^T \boldsymbol{\beta}) = 0.$$

Misspecification of linear decision classes $\Longrightarrow$ Semi-supervised D-learning (SSDL)

SSDL estimator based on fully nonparametric imputation: $\widehat{\beta}_{np}$

Let $m(\mathbf{X}) = E[2AY|\mathbf{X}] = E[2AY|\mathbf{X}^-]$, and its corresponding kernel estimator is

$$\widehat{m}(\mathbf{X}_j) = \frac{(nh^p)^{-1}\sum_{i=1}^n H_h(\mathbf{X}_i^-, \mathbf{X}_j^-) \times 2A_iY_i}{(nh^p)^{-1}\sum_{i=1}^n H_h(\mathbf{X}_i^-, \mathbf{X}_j^-)},$$

where $H_h(u, v) = H(\frac{u-v}{h})$ with kernel function $H : \mathbb{R}^p \to \mathbb{R}$ and bandwidth h. Then $\widehat{\beta}_{np}$ be obtained by the solution of

$$\frac{1}{N}\sum_{j=n+1}^{n+N} \mathbf{X}_j(\widehat{m}(\mathbf{X}_j) - \mathbf{X}_j^T\beta) = 0.$$

Curse of dimensionality $\Longrightarrow$ Projection (dimensionality reduction) + Refitting (debias)

SSDL estimator based on semiparametric imputation: $\widehat{\beta}_{sp}$

Projection: Define $m(\mathbf{X}^T\boldsymbol{\beta}) = E[2AY|\mathbf{X}^T\boldsymbol{\beta}]$, and its corresponding kernel estimator is

$$\widehat{m}(\mathbf{X}_j^T\boldsymbol{\beta}) = \frac{(nh)^{-1}\sum_{i=1}^{n} K_h(\mathbf{X}_i^T\boldsymbol{\beta}, \mathbf{X}_j^T\boldsymbol{\beta}) * 2A_iY_i}{(nh)^{-1}\sum_{i=1}^{n} K_h(\mathbf{X}_i^T\boldsymbol{\beta}, \mathbf{X}_j^T\boldsymbol{\beta})},$$

where $K_h(u,v) = K(\frac{u-v}{h})$ with kernel function $K : \mathbb{R} \to \mathbb{R}$ and bandwidth h.

Refitting: Define $\boldsymbol{\theta}_0$ as the solution of $E[\mathbf{X}(2AY - m(\mathbf{X}^T\boldsymbol{\beta}_0) - \mathbf{X}^T\boldsymbol{\theta})] = 0$, and $\widehat{\boldsymbol{\theta}}$ is estimated by the estimating equation that

$$P_n\mathbf{X}(2AY - \widehat{m}(\mathbf{X}^T\widehat{\boldsymbol{\beta}}) - \mathbf{X}^T\boldsymbol{\theta}) = 0.$$

Denote the imputation function after refitting as $\nu(\mathbf{X};\boldsymbol{\beta},\boldsymbol{\theta}) = m(\mathbf{X}^T\boldsymbol{\beta}) + \mathbf{X}^T\boldsymbol{\theta}$, and its semiparametric (SP) estimator is

$$\widehat{\nu}(\mathbf{X};\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}}) = \widehat{m}(\mathbf{X}^T\widehat{\boldsymbol{\beta}}) + \mathbf{X}^T\widehat{\boldsymbol{\theta}}.$$

Then $\widehat{\beta}_{sp}$ can be obtained by solving

$$\frac{1}{N}\sum_{j=n+1}^{n+N} \mathbf{X}_j(\widehat{\nu}(\mathbf{X}_j;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}}) - \mathbf{X}_j^T\boldsymbol{\beta}) = 0.$$

Overfitting? $\Longrightarrow$ $\mathbb{K}$-fold cross-validation (CV)

SSDL estimator based on semiparametric imputation with $\mathbb{K}$-fold CV: $\widehat{\boldsymbol{\beta}}_{sp,\mathbb{K}}$

Let $\mathcal{L}_k$ be the k-th random disjoint partition of $\mathcal{L}$ with sample size $n_{\mathbb{K}} = \frac{n}{\mathbb{K}}$ and index set $\mathcal{I}_k$ for $k \in \{1, \ldots, \mathbb{K}\}$. Let the set excluding the k-th partition be $\mathcal{L}_k^- = \mathcal{L} - \mathcal{L}_k$ with sample size $n_{\mathbb{K}}^- = n - n_{\mathbb{K}}$ and index set $\mathcal{I}_k^-$.

Denote the OLS and nonparametric imputation estimator under dimension reduction based on $\mathcal{L}_k^-$ as $\widehat{\boldsymbol{\beta}}_k$ and $\widehat{m}_k(\mathbf{X}^T\boldsymbol{\beta})$ respectively. Then $\widehat{\boldsymbol{\theta}}_{\mathbb{K}}$ is obtained by solving

$$\frac{1}{n}\sum_{k=1}^{\mathbb{K}}\sum_{i\in\mathcal{I}_k}\mathbf{X}_i(2A_iY_i - \widehat{m}_k(\mathbf{X}_i^T\widehat{\boldsymbol{\beta}}_k) - \mathbf{X}_i^T\boldsymbol{\theta}) = 0,$$

and the semiparametric imputation function estimation is $\widehat{\nu}(\mathbf{X}; \widehat{\boldsymbol{\beta}}_k, \widehat{\boldsymbol{\theta}}_{\mathbb{K}}) = \frac{1}{\mathbb{K}}\sum_{k=1}^{\mathbb{K}} \widehat{m}_k(\mathbf{X}^T\widehat{\boldsymbol{\beta}}_k) + \mathbf{X}^T\widehat{\boldsymbol{\theta}}_{\mathbb{K}}$.

Then $\widehat{\boldsymbol{\beta}}_{sp,\mathbb{K}}$ can be obtained by solving

$$\frac{1}{N}\sum_{j=n+1}^{n+N}\mathbf{X}_j(\widehat{\nu}(\mathbf{X}_j; \widehat{\boldsymbol{\beta}}_k, \widehat{\boldsymbol{\theta}}_{\mathbb{K}}) - \mathbf{X}_j^T\boldsymbol{\beta}) = 0.$$

Outline

Asymptotic properties

Theorem 1

Under certain regularity conditions, we have

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{sp} - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(\mathbf{Z}_i) + O_p(r_{n,N}),$$

where the influence function $\varphi(\mathbf{Z}_i) = E[\mathbf{X}\mathbf{X}^T]^{-1}\{\mathbf{X}_i[2A_iY_i - \nu(\mathbf{X}_i; \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)]\}$ and $r_{n,N} = O_p\left(\frac{n}{N}\right)^{\frac{1}{2}} + O_p(b_n)$ with $b_n = n^{-\frac{2q-3}{2(2q+1)}}$.

Thus $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{sp} - \boldsymbol{\beta}_0) \xrightarrow{d} N_{p+1}(\mathbf{0}, \Sigma)$ with positive definite $(p+1) \times (p+1)$ matrix $\Sigma = E[\mathbf{X}\mathbf{X}^T]^{-1} E\{\mathbf{X}\mathbf{X}^T[2AY - \nu(\mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\theta}_0)]^2\} E[\mathbf{X}\mathbf{X}^T]^{-1}$.

Theorem 2

When the fold of CV is fixed and satisfies $\mathbb{K} \geqslant 2$, under the conditions same as Theorem 1, we have

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{sp,\mathbb{K}} - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(\mathbf{Z}_i) + O_p(\gamma_{n,N}),$$

where $\gamma_{n,N} = O_p\left(\frac{n}{N}\right)^{\frac{1}{2}} + O_p(a_{n_{\mathbb{K}-}})$ with $a_n = n^{-\frac{q}{2q+1}}\sqrt{\log n}$. Thus $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{sp,\mathbb{K}} - \boldsymbol{\beta}_0) \xrightarrow{d} N_{p+1}(\mathbf{0}, \Sigma)$.

Outline

Numerical Simulations

Data generation: $Y = \mu_0(\mathbf{X}) + A\delta(\mathbf{X}) + e$; $P(A = 1) = P(A = -1) = 0.5$, $e \sim N\left(0, 0.5^2\right)$,

$$\mathbf{X} = (1, X_1, X_2, \ldots, X_p)': X_i \ (i = 1, \ldots, p) \sim U(-5, 5).$$

Model settting:

- Linear (Lin): $\delta(\mathbf{X}) = 20\mathbf{X}^T\alpha$,
- Nonlinear 1 (NL1): $\delta(\mathbf{X}) = 0.2(\mathbf{X}^T\alpha)^3$,
- Nonlinear 2 (NL2): $\delta(\mathbf{X}) = \mathbf{X}^T\alpha + 0.2(\mathbf{X}^T\alpha)^3 + \sin(\mathbf{X}^T\alpha)$.

- Quadratic: $\mu_0^Q(\mathbf{X}) = \mathbf{X}^T\omega_1 + (\mathbf{X}^T\omega_2)^2$,
- Cubic: $\mu_0^C(\mathbf{X}) = 0.1(\mathbf{X}^T\omega_2)^3$.

Parameters setting:

- $\alpha = \alpha^{(a)} = (0, -\mathbf{1}_{p/2}^T, \mathbf{1}_{p/2}^T)^T$,
- $\alpha = \alpha^{(b)} = (0, \mathbf{1}_p^T)^T$

- $\omega_1 = (0, \mathbf{1}_{p/2}^T, -\mathbf{1}_{p/2}^T)^T$,
- $\omega_2 = (0, \mathbf{1}_p^T)^T$.

Table: The average of RE and PCD for $p = 10$

	$\alpha = \alpha^{(a)}$				$\alpha = \alpha^{(b)}$			
	Quadratic $\mu_0^Q(\mathbf{X})$		Cubic $\mu_0^C(\mathbf{X})$		Quadratic $\mu_0^Q(\mathbf{X})$		Cubic $\mu_0^C(\mathbf{X})$	
Lin	RE	PCD	RE	PCD	RE	PCD	RE	PCD
SUP	1	96.20%	1	92.69%	1	96.85%	1	94.28%
NP	0.19	92.64%	0.38	87.75%	0.19	92.99%	0.38	88.45%
SP	0.99	96.16%	1.00	92.64%	1.06	96.94%	1.10	94.55%
SP.CV	0.98	96.14%	0.98	92.64%	0.94	96.81%	0.92	94.18%
KRLS	1.01	96.20%	1.07	92.79%	1.02	96.80%	1.06	94.30%
KRLS.CV	0.94	96.10%	0.95	92.54%	0.94	96.71%	0.95	94.14%
NL1	RE	PCD	RE	PCD	RE	PCD	RE	PCD
SUP	1	96.60%	1	95.60%	1	96.84%	1	96.24%
NP	0.20	92.51%	0.24	91.24%	0.19	91.87%	0.21	89.81%
SP	3.05	97.97%	1.76	96.40%	3.12	98.46%	2.10	97.63%
SP.CV	3.24	97.99%	1.89	96.57%	2.94	98.23%	1.62	97.27%
KRLS	0.91	97.28%	0.98	96.06%	0.86	97.56%	0.94	96.62%
KRLS.CV	1.54	97.08%	1.18	95.76%	1.52	97.34%	1.16	96.40%

Table: The average of SE and CP of proposed estimators with $p = 10$

SE(CP %)	Quadratic $\mu_0^Q(\mathbf{X})$			Cubic $\mu_0^C(\mathbf{X})$		
	SP	SP.CV	SP.DCV	SP	SP.CV	SP.DCV
Lin $\alpha^{(a)}$	5.300 (93.5)	5.539 (94.6)	5.728 (95.2)	10.48 (93.8)	10.90 (94.6)	11.28 (95.6)
Lin $\alpha^{(b)}$	4.931 (91.8)	5.525 (93.5)	5.711 (94.1)	9.297 (89.5)	10.69 (92.2)	11.06 (93.0)
NL1 $\alpha^{(a)}$	5.590 (81.1)	7.214 (92.5)	7.473 (93.5)	10.75 (89.1)	12.10 (94.3)	12.53 (95.1)
NL1 $\alpha^{(b)}$	4.827 (73.2)	7.415 (90.4)	7.682 (91.4)	8.494 (81.6)	11.98 (90.4)	12.41 (91.2)
NL2 $\alpha^{(a)}$	5.579 (81.1)	7.204 (92.5)	7.462 (93.5)	10.74 (89.1)	12.09 (94.4)	12.52 (95.1)
NL2 $\alpha^{(b)}$	4.820 (73.2)	7.406 (90.4)	7.671 (91.4)	8.490 (81.6)	11.98 (90.4)	12.40 (91.2)

Outline

Real Data Analysis: MIMIC-IV

Subjects: 9,052 adult patients with sepsis who were first admitted to the ICU in the MIMIC-IV data.

Outcome: lactate clearance within 48 hours of ICU admission.

Tratments: $A = -1$: intravenous fluid resuscitation; $A = 1$: vasopressor therapy.

Covariates: age (years), admission weight (kilograms), blood urea nitrogen (BUN) amount (mg/dL), creatinine (mg/dL), white blood cell count (WBC) count (K/μL), and heart rate (HR) (bpm)

Sample size: labeled dataset: n=184 samples; unlabeled dataset: N=7,623 samples.

Table: Treatment recommendation

Treatment	Methods					
	SUP	NP	SP	SP.CV	KRLS	KRLS.CV
A=-1: IV Fluid Resuscitation	5113	5113	5112	5132	5150	5142
A=1: Vasopressors	2694	2694	2695	2675	2657	2665

Outline

Conclusions

- We propose a novel **semi-supervised D-learning framework** for estimating optimal individualized treatment regime (ITR).
- Theoretical results:
  - Estimators converge asymptotically to normality at $\sqrt{n}$ rate, depending on labeled sample size.
  - Under model misspecification, semi-supervised estimator achieves lower asymptotic variance than supervised estimator.
- Numerical studies:
  - When linear decision model is correct, unlabeled data does not improve efficiency.
  - When misspecified, semi-supervised estimator significantly improves performance.
  - Remains robust with increasing covariate dimension p; outperforms fully nonparametric methods and KRLS-based estimators.
- Our method mitigates the curse of dimensionality, and maintains robustness and efficiency in multi-dimensions.

Reference I

Blatt, D., Murphy, S. A., and Zhu, J. (2004). A-learning for approximate planning. *Ann Arbor*, 1001:48109–2122.

Chu, J., Lu, W., and Yang, S. (2023). Targeted optimal treatment regime learning using summary statistics. *Biometrika*, 110(4):913–931.

Fan, C., Lu, W., Song, R., and Zhou, Y. (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(5):1565–1582.

Li, C., Zeng, D., and Zhu, W. (2025). A robust covariate-balancing method for learning optimal individualized treatment regimes. *Biometrika*, 112(1):asae036.

Maronge, J. M., D HULING, J., and Chen, G. (2023). A reluctant additive model framework for interpretable nonlinear individualized treatment rules. *The annals of applied statistics*, 17(4):3384.

Reference II

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(2):331–355.

Qi, Z., Liu, D., Fu, H., and Liu, Y. (2020). Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various outcomes. *Journal of the American Statistical Association*, 115(530):678–691.

Qi, Z. and Liu, Y. (2018). D-learning to estimate optimal individual treatment rules. *Electronic Journal of Statistics*, 12(2):3601–3638.

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180–1210.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In Lin, D. Y. and Heagerty, P. J., editors, *Proceedings of the Second Seattle Symposium in Biostatistics*, volume 179 of *Lecture Notes in Statistics*, pages 189–326, New York. Springer.

Reference III

Shah, K. S., Fu, H., and Kosorok, M. R. (2023). Stabilized direct learning for efficient estimation of individualized treatment rules. *Biometrics*, 79(4):2843–2856.

Song, R., Luo, S., Zeng, D., Zhang, H. H., Lu, W., and Li, Z. (2017). Semiparametric single-index model for estimating optimal individualized treatment strategy. *Electronic journal of statistics*, 11(1):364.

Watkins, C. (1989). *Learning from delayed rewards*. PhD thesis, King's College.

Watkins, C. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3):279–292.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118.

Thank you for listening.

More details can be referred to “Li, X., Zhang, S., and Zhou, Y. (2025). Semisupervised D-Learning for Optimal Individualized Treatment Regimes. Stat, 14(2), e70063."