Precision Medicine: Lecture 05
Outcome Weighted Learning

Michael R. Kosorok,
Nikki L. B. Freeman and Owen E. Leete

Department of Biostatistics
Gillings School of Global Public Health
University of North Carolina at Chapel Hill

Fall, 2021

Outline

Outcome Weighted Learning

Efficient Augmentation and Relaxation Learning (EARL)

Indirect vs. Direct ITR estimation

- Regression and policy methods are generally considered "indirect" methods because there is an intermediate step before estimating the optimal policy
  - Regression methods build models for each treatment
  - Policy methods use estimated utility to find a treatment policy with better estimated utility
- In several situations it is difficult to ensure that model assumptions are met for regression based methods
- By inverting the model to find the optimal treatment rule, regression based methods emphasize prediction accuracy of the clinical response model instead of decision rule optimality

Outcome Weighted Learning

- Outcome weighted learning (OWL) directly estimates the decision rule that maximizes clinical response by formulating ITR estimation as a weighted classification problem
  - This allows us to use powerful machine learning methods for classification
  - The directness of the approach allows for better ITR estimation while making full use of available information

Outcome Weighted Learning

- The data consist of the triple (X, A, R)
  - Prognostic variables, $X = (X_1, \ldots, X_d)^T \in \mathcal{X}$
  - Treatment assignments, $A \in \mathcal{A} = \{-1, 1\}$
  - Clinical outcome (reward), R is bounded
- We assume that $R > 0$ and that larger values are better
- We assume the data are collected from a two-arm randomized trial
  - The treatment probabilities, $\pi(A|X) = P(A = a|X = x)$, are known
  - Positivity and other causal assumptions can be guaranteed

Value Function

- To formulate the ITR estimation problem, we need a measure of how well a given rule will perform
- Let P denote the distribution of (X, A, R) and E denote expectation w.r.t. P
- For any given ITR $\mathcal{D}$, we let $P^{\mathcal{D}}$ denote the distribution of (X, A, R) given that $A = \mathcal{D}(X)$, with expectation denoted by $E^{\mathcal{D}}$
- Taking expectation of the reward R w.r.t. $P^{\mathcal{D}}$ gives us the value function $\mathcal{V}(\mathcal{D})$

$$E^{\mathcal{D}}(R) = \int R \frac{dP^{\mathcal{D}}}{dP} dP = E\left[\frac{I(A = \mathcal{D}(X))}{\pi(A|X)} R\right] = \mathcal{V}(\mathcal{D})$$

A Classification Problem

- An optimal ITR, $\mathcal{D}^*$, is a rule that maximizes $\mathcal{V}(\mathcal{D})$

$$\mathcal{D}^* = \underset{\mathcal{D}}{\operatorname{argmax}}\, E\left[\frac{I(A = \mathcal{D}(X))}{\pi(A|X)} R\right] \tag{1}$$

- Notice that the ITR, $\mathcal{D}^*$, which maximizes $\mathcal{V}(\mathcal{D})$, is equivalent to the $\mathcal{D}$ that minimizes

$$\begin{aligned} & E[R|A = 1] + E[R|A = -1] - \mathcal{V}(\mathcal{D}) \\ & = E\left[\frac{I(A \neq \mathcal{D}(X))}{\pi(A|X)} R\right] \end{aligned} \tag{2}$$

- This is equivalent to minimizing the classification error when predicting which treatments patients received

Heuristic motivation

- If $\mathcal{D}$ recommends the treatment that a subject did not receive, that subject is considered to be misclassified
- Consider a subpopulation where one treatment produces better outcomes than the other
  - Due to positivity, we should observe subjects on treatment $A = 1$ and $A = -1$ in this subpopulation
  - Regardless of the treatment recommended by $\mathcal{D}$, there will be misclassification error
  - Since the observations are weighted by their outcome, recommending the treatment with better (larger) outcomes will result in less misclassification error

Reframing as a Convex Problem

- When estimating (2) from data, we can replace $\mathcal{D}(X)$ with $\text{sign}(f(X))$

$$n^{-1}\sum_{i=1}^{n}\frac{R_i}{\pi(A_i|X_i)}I(A_i \neq \text{sign}(f(X_i)))$$

- The weighted classification error is not convex
- If we replace the 0-1 loss with a hinge loss the problem becomes convex

$$n^{-1}\sum_{i=1}^{n}\frac{R_i}{\pi(A_i|X_i)}(1 - A_i f(X_i))^{+} + \lambda_n\|f\|^2 \qquad (3)$$

- The estimated optimal decision function $\hat{\mathcal{D}}(X)$ minimizes Equation (3)

Weighted Support Vector Machine

- The modified classification error in (3) now looks like a weighted version of a support vector machine (SVM)
- If we assume that the decision function $f(x)$ is a linear function of x, $f(x) = \langle \beta, x \rangle + \beta_0$, then (3) is equivalent to

$$\min \quad \frac{1}{2}\|\beta\|^2 + \kappa \sum_{i=1}^{n} \frac{R_i}{\pi_i}\xi_i, \tag{4}$$

$$\text{subject to} \quad A_i(\langle \beta, X \rangle + \beta_0) \geq (1 - \xi_i), \quad \xi_i > 0$$

where $\kappa > 0$ is a tuning parameter, $\pi_i = \pi(A_i|X_i)$, and R_i/π_i is the weight applied to the i^{th} point

Nonlinear Decision Rules

- As with standard SVM, we can find the dual representation of (4) as

$$\max_{\alpha} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j A_i A_j \langle X_i, X_j \rangle$$

$$\text{subject to} \quad 0 \leq \alpha_i \leq \kappa R_i / \pi_i \text{ and } \sum_{i=1}^{n} \alpha_i A_i = 0$$

- By replacing $\langle X_i, X_j \rangle$ with a kernel, $\mathcal{K}(X_i, X_j)$, we can fit non-linear decision boundaries

Risk

- For any ITR $\mathcal{D}(x) = \text{sign}(f(x))$ associated with decision function $f(x)$, we define the risk

$$\mathcal{R}(f) = E\left[\frac{R}{\pi(A_i|X_i)} I(A \neq \text{sign}(f(X)))\right]$$

- We want to find the function f which minimizes the risk $\mathcal{R}(f)$, but OWL instead finds the function which minimizes the ϕ-risk

$$\mathcal{R}_\phi(f) = E\left[\frac{R}{\pi(A_i|X_i)} \phi(Af(X))\right]$$

where $\phi(t) = (1 - t)^+$ is the hinge loss

Fisher consistency

- Let $\mathcal{R}^*(f) = \inf_f\{\mathcal{R}(f) | f : \mathcal{X} \mapsto \mathbb{R}\}$ be the minimal (Bayes) risk
- Similarly, let $\mathcal{R}^*_\phi(f)$ be the minimal ϕ-risk
- Is there any connection between $\mathcal{R}^*(f)$ and $\mathcal{R}^*_\phi(f)$?
- Zhao et al. were able to show that the function f which minimizes $\mathcal{R}_\phi(f)$ is the same f which minimizes $\mathcal{R}(f)$
- Despite using the surrogate loss ϕ, the decision rule found by OWL is the same decision rule that gives the optimal ITR in (1)

Limitations

- OWL is nonparametric, so it is not necessarily the most efficient approach
  - When we have knowledge of the specific parametric or semiparametric form of the mean outcome, a likelihood-based method may be more efficient
- Note that the optimal ITR from (1) is not affected by a shift in the reward R

$$\underset{\mathcal{D}}{\operatorname{argmax}}\, E\left[\frac{I(A=\mathcal{D}(X))}{\pi(A|X)}R\right] \equiv \underset{\mathcal{D}}{\operatorname{argmax}}\, E\left[\frac{I(A=\mathcal{D}(X))}{\pi(A|X)}(R+c)\right]$$

  - Simulations have shown that OWL is sensitive to shifts in the outcome

Limitations, cont.

- The misclassification error targeted by OWL is based on the differences between the estimated and true treatment assignments
- In high dimensional covariate space or with a flexible enough decision function, the misclassification error can be reduced by having the decision function match the treatment assignment
- This can result in suboptimal decision rules since the treatments are randomly assigned in the trial, and it is unlikely that the majority of subjects are assigned optimal treatments

Outline

Outcome Weighted Learning

Efficient Augmentation and Relaxation Learning (EARL)

Efficient Augmentation and Relaxation Learning (EARL)

- Efficient Augmentation and Relaxation Learning (EARL) is an extension of OWL which combines use of outcome weighting, double robustness, and the use of a surrogate
  - Y-Q Zhao, EB Laber, Y Ning, S Saha, and BE Sands (2019). Efficient augmentation and relaxation learning for individualized treatment rules using observational data. *Journal of Machine Leanring Research* 20(48):1-23.
- The double-robustness is incorporated through the outcome weighting
  - A regression model for the outcome is estimated
  - A model for the propensity score is also estimated

EARL

- The data and assumptions for EARL are similar to OWL
  - The data consist of the triple (X, A, Y) (Y replaces R in OWL)
  - The data are collected from observational data (but could also come from a randomized trial)
  - Clinical outcome Y need not be bounded nor positive but we require higher values to be better for the patient
  - The usual causal assumptions are required, including no-unmeasured confounders, strict positivity ($> \tau > 0$), etc.

The Value Function

- Recall the value function:

$$V(d) = E\left[\frac{Y}{\pi(A;X)} I\{A = d(X)\},\right]$$

- The augmented estimator for $V(d)$ is

$$\begin{aligned}\hat{V}^{AIPWE} \;=\; \mathbb{P}_n \Bigg[& \frac{YI\{A = d(X)\}}{\hat{\pi}\{d(X);X\}} \\ & - \frac{I\{A = d(X)\} - \hat{\pi}\{d(X);X\}}{\hat{\pi}\{d(X);X\}} \hat{Q}\{X, d(X)\}\Bigg],\end{aligned}$$

where $\hat{\pi}$ is consistent for π and $\hat{Q}$ is consistent for $Q(X, A) = E[Y|X, A]$.

Finding the optimal d

- Define the estimated weight

$$\hat{W}_a = \frac{YI\{A=a\}}{\hat{\pi}\{a;X\}} - \frac{I\{A=a\} - \hat{\pi}\{a;X)\}}{\hat{\pi}\{a;X\}}\hat{Q}\{X;a\},$$

for $a \in \{-1,1\}$, and let

- $\hat{f}_n =$

$$\arg\inf_{f\in\mathcal{M}} \mathbb{P}_n \Big[|\hat{W}_1| I\{\text{sgn}(\hat{W}_1)f(X) < 0\} + |\hat{W}_{-1}| I\{-\text{sgn}(\hat{W}_{-1})f(X) < 0\}\Big],$$

where $\mathcal{M}$ is a suitable class of measureable functions.

Finding the optimal d

- The needed optimization computation for $\hat{f}_n$ is a hard non-convex problem.
- Solution: use a surrogate loss ϕ, chosen from $\phi(t) = (1-t)_+$ (hinge), $= e^{-t}$ (exponential), $= \log(1+e^{-t})$ (logistic), $= t^2$ (squared), or $= \{(1-t)_+\}^2$ (squared hinge), and compute

$$\tilde{f}_n = \arg\inf_{f\in\mathcal{M}} \mathbb{P}_n \Big[|\hat{W}_1|\phi\{\text{sgn}(\hat{W}_1)f(X)\} + |\hat{W}_{-1}|\phi\{-\text{sgn}(\hat{W}_{-1})f(X)\}\Big] + \lambda_n \|f\|^2,$$

  where $\lambda_n \geq 0$ is a (possibly) data dependent tuning parameter.

Finding the optimal d

- We actually use a modification of $\tilde{f}_n$ based on sample splitting (a kind of internal cross-validation) which helps to remove the dependence between the observations used to estimate the target function from the observations used in the estimators $\hat{\pi}$ and $\hat{Q}$, resulting in an estimator $\hat{f}_{n,k}$.
- Our decision rule then becomes $\hat{d}_{n,k} = \text{sgn}(\hat{f}_{n,k})$.
- Let $f^* \in \mathcal{M}$ satisfy
  - $d^*(x) = \text{sgn}\{f^*(X)\}$ and
  - $V(f^*) = V^* \equiv \sup_{f \in \mathcal{M}} V(f)$.

Theoretical results

- Assume that at least one of $\hat{\pi}$ and $\hat{Q}$ is correctly specified.
- Assume other reasonable regularity conditions.
- Then $\hat{d}_{n,k}$ is consistent for the d^* with a rate of convergence that dependends on the choice of surrogate ϕ.
- The hinge loss achieves the fastest convergence rates of the ones considered.
- The rates can be further improved if both $\hat{\pi}$ and $\hat{Q}$ are correctly specified.

Discussion

- Simulation studies show that EARL performs very well and compares favorably to other methods.
- The authors present a very interesting application of their method on the Ocean State Crohn's and Colitis Area Registry (OSCCAR):
  - Patients have ulcerative colitis (UC), Crohn's desease (CD), or inderterminate colitis (IC).
  - Treatment choice is "step-up" (begin gradually) or "top-down" (more aggressive), both have been established to be effective.
  - Primary outcome is disease activity score at two years (higher is worse).
  - The optimal decision rule yielded a disease activity score of 1.75 versus 2.24 from the observed data.
  - Two patient features are statistically significant: gender and body mass index (BMI).