Lecture 8: Outcome Weighted Learning for Static Treatment Strategies
Motivation
The goal of the Nefazodone-CBASP clinical trial (Keller et al., 2000) is to determine the best treatment choice among

- Pharmacotherapy (nefazodone).
- Psychotherapy (cognitive behavioral-analysis system of psychotherapy (CBASP)).
- Combination of both.

- 681 patients, with 50 prognostic variables measured on each patient.

Further Goal
Can we reduce depression by creating individualized treatment rules based on prognostic data?
Challenges

- Identify the optimal individualized treatment rule using training data where optimal treatment is unknown.
- High-dimensional predictors; arbitrary order nonparametric interactions.
Learning Framework
Observe independently and identically distributed training data $(X_i, A_i, R_i), i = 1, \ldots, n$.
- $X$: baseline variables, $X \in \mathbb{R}^d$,
- $A$: binary treatment options, $A \in \{-1, 1\}$,
- $R$: outcome (larger is better), $R \in \mathbb{R}^+$, $R$ is bounded.

Randomized study with known randomization probability of the treatment.

Construct individualized treatment rule (ITR)

$$\mathcal{D}(X) : \mathbb{R}^d \rightarrow \{-1, 1\}.$$ 

Goal

Maximize the expected outcome if the ITR is implemented in the future.
1. Let $P$ denote the distribution of $(X, A, R)$, where treatments are randomized, and $P^D$ denoted the distribution of $(X, A, R)$, where treatments are chosen according to $D$. The value function of $D$ (Qian&Murphy, 2011) is

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

2. **Optimal Individualized Treatment Rule:**

   $$D^* \in \arg\max_D V(D).$$

   $$E(R|X, A = 1) > E(R|X, A = -1) \Rightarrow D^*(X) = 1$$
   $$E(R|X, A = 1) < E(R|X, A = -1) \Rightarrow D^*(X) = -1$$
Optimal Individualized Treatment Rule Discovery

Traditional approach: regression-based

\[
\text{(X, A, R)} \quad \overset{\text{Minimize}}{\rightarrow} \quad \text{Predict} \quad \hat{E}(R|A, X) \quad \overset{\text{argmax}_{A \in \{-1, 1\}}}{\rightarrow} \quad \text{Optimal ITR}
\]

**Problem**: mismatch between minimizing the prediction error and maximizing the value function.

**Our approach**

\[
\text{(X, A, R)} \quad \overset{\text{Maximize } \mathcal{V}(D)}{\rightarrow} \quad \text{Optimal ITR}
\]

Can we directly maximize the value function?
Intuition: Classification

Given a new observation $X^{\text{new}}$, predict the class label $D^{*,\text{new}}$.

- No direct information on the true class labels, $D^*$.  
- Can we assign the right treatment based on the observed information?
Learning Method
Optimal Individualized Treatment Rule $\mathcal{D}^*$

Maximize the value

$$E \left[ \frac{I(A = \mathcal{D}(X))}{P(A|X)} R \right]$$

Minimize the risk

$$E \left[ \frac{I(A \neq \mathcal{D}(X))}{P(A|X)} R \right]$$

- For any rule $\mathcal{D}$, $\mathcal{D}(X) = \text{sign}(f(X))$ for some function $f$.
- Empirical approximation to the risk function:
  $$n^{-1} \sum_{i=1}^{n} \frac{R_i}{P(A_i|X_i)} I(A_i \neq \text{sign}(f(X_i)))$$.
- **Computation challenges**: non-convexity and discontinuity of 0-1 loss.
Hinge Loss: \( \phi(Af(X)) = (1 - Af(X))^+ \), where \( x^+ = \max(x, 0) \)
Objective Function: Regularization Framework

\[
\min_f \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{R_i}{P(A_i|X_i)} \phi(A_i f(X_i)) + \lambda_n \|f\|^2 \right\}.
\]

- \(\|f\|\) is some norm for \(f\), and \(\lambda_n\) controls the severity of the penalty on the functions.
- A linear decision rule: \(f(X) = X^T \beta + \beta_0\), with \(\|f\|\) as the Euclidean norm of \(\beta\).
- Estimated individualized treatment rule:

\[
\hat{D}_n = \text{sign}(\hat{f}_n(X)),
\]

where \(\hat{f}_n\) is the solution.
The dual problem is a convex optimization problem.
Quadratic programming; Karush-Kuhn-Tucker conditions.
Linear decision rules may be insufficient.
Kernel trick, \( k : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \).
Nonlinear decision rule with \( f(x) = \beta k(\cdot, x) + \beta_0 \).
Reproducing kernel Hilbert space (RKHS) \( \mathcal{H}_k \) with norm denoted by \( \| \cdot \|_k \):
\[
\mathcal{H}_k = \left\{ g(x) = \sum_{i=1}^{m} \alpha_i k(x_i, x) \right\}.
\]
A linear kernel yields a linear decision rule.
Theoretical Results
Risk and Surrogate $\phi$-Risk

Goal

- Minimize the risk:

$$\mathcal{L}(f) = E \left[ \frac{R}{P(A|X)} I(A \neq \text{sign}(f(X))) \right].$$

- The minimal risk $\mathcal{L}^* = \inf_f \{ \mathcal{L}(f) | f : \mathbb{R}^d \rightarrow \mathbb{R} \}$.

Surrogate

- Minimize the $\phi$-risk, where $\phi(Af(X)) = (1 - Af(X))^+$:

$$\mathcal{L}_\phi(f) = E \left[ \frac{R}{P(A|X)} \phi(Af(X)) \right].$$

- The minimal $\phi$-risk $\mathcal{L}^*_\phi = \inf_f \{ \mathcal{L}_\phi(f) | f : \mathbb{R}^d \rightarrow \mathbb{R} \}$.
Properties of Weighted Hinge Loss Function

Fisher consistency
- If \( \tilde{f} \) minimizes \( L_\phi(f) \), then \( D^*(x) = \text{sign}(\tilde{f}(x)) \), where \( D^* \) is the Bayes decision rule.
- Aim at the optimal individualized treatment rule directly.

Quantified relationship between excess risks
- \( L(f) - L^* \leq L_\phi(f) - L^*_\phi \) (Bartlett et al., 2006).
- Minimization of \( L_\phi(f) \) is a reasonable surrogate for minimization of \( L(f) \).
- The excess value is upper bounded by the excess \( \phi \)-risk.
Outcome weighted learning estimator:

\[ \hat{f}_n = \arg\min_f \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{R_i}{P(A_i|X_i)} \phi(A_if(X_i)) + \lambda_n \|f\|^2 \right\}. \]

**Theorem** Assume that we choose a sequence \( \lambda_n > 0 \) such that \( \lambda_n \rightarrow 0 \) and \( \lambda_n n \rightarrow \infty \), then for all distributions \( P \), we have

\[ \lim_{n \rightarrow \infty} \mathcal{L}_\phi(\hat{f}_n) = \inf_{f \in \bar{\mathcal{H}}_k} \mathcal{L}_\phi(f). \]

Thus, if \( f^* \) belongs to the closure of \( \mathcal{H}_k \), we have \( \lim_{n \rightarrow \infty} \mathcal{L}_\phi(\hat{f}_n) = \mathcal{L}^* \). It then follows that \( \lim_{n \rightarrow \infty} \mathcal{L}(\hat{f}_n) = \mathcal{L}^* \).

- **Sequential convergence of values to the optimal value.**
Understand the accuracy of OWL procedure.
Consider reproducing kernel Hilbert space associated with Gaussian Radial Basis Function kernels:

$$k(x, x') = \exp(-\sigma^2 \|x - x'\|^2), x, x' \in \mathbb{R}^d,$$

where $\sigma$ is the inverse bandwidth of the kernel.

Precise risk bound under certain regularity conditions. For any $\delta > 0$, $0 < \nu < 2$, if $\sigma_n = \lambda_n^{-1/(q+1)d}$, then the optimal rate for the risk is

$$\mathcal{L}(\hat{f}_n) - \mathcal{L}^* = O_p \left( n^{-\frac{2q}{(4+\nu)q+2+(2-\nu)(1+\delta)}} \right),$$

with the optimal choice of $\lambda_n$ balancing bias and variance.
What if $E(R|X = x, A = 1)$ is not too close to $E(R|X = x, A = -1)$?

**Theorem**

Under certain margin conditions, for $0 < \nu < 1$, let

$$
\lambda_n = O\left(n^{-1/(\nu+1)}\right) \quad \text{and} \quad \sigma_n = \lambda_n^{-1/(q+1)d},
$$

we have

$$
\mathcal{L}(\hat{f}_n) - \mathcal{L}^* = O_p \left(n^{\frac{-1}{\nu+1} \frac{q}{q+1}}\right).
$$

- The value converges surprisingly fast to the optimal, as fast as $n^{-1}$.
- Similar to rate results in SVM literature (Tsybakov, 2004).
Simulation Study and Application
Simulation Study

- **OWL with Gaussian kernel**: two tuning parameters
  - $\lambda_n$: the parameter for penalty.
  - $\sigma_n$: the inverse bandwidth of the kernel.

- **Methods for comparison**:
  - **OWL with Linear kernel**.
  - Regression based methods:
    - $l_1$ penalized least squares ($l_1$-PLS) (Qian & Murphy, 2011) with basis function $(1, X, A, XA)$.
    - Ordinary Least Squares (OLS) with basis function $(1, X, A, XA)$.

- **Evaluation of values in terms of mean squared error (MSE)**.
  - 1000 replications; each training data set is of size 100, 200, 400 or 800.
  - Independent validation set of size 10000.

Donglin Zeng, Department of Biostatistics, University of North Carolina
Data Generation

- $X = (X_1, \ldots, X_{50}) \sim U[-1, 1]^{50}$.
- $A \in \{-1, 1\}$, $P(A = 1) = P(A = -1) = 0.5$.
- The response $R \sim N(Q_0, 1)$, where

$$Q_0 = 1 + 2X_1 + X_2 + 0.5X_3 + T_0(X, A).$$

1. $T_0(X, A) = 0.442(1 - X_1 - X_2)A$.
2. $T_0(X, A) = (0.5 - X_1^2 - X_2^2) (X_1^2 + X_2^2 - 0.3) A$. 

Donglin Zeng, Department of Biostatistics, University of North Carolina
Simulation Results

Optimal Decision Boundary

MSE for Values

Scenario 1: \( T_0(X, A) = 0.442(1 - X_1 - X_2)A \)

Donglin Zeng, Department of Biostatistics, University of North Carolina
Simulation Results

Optimal Decision Boundary

MSE for Values

Scenario 2: $T_0(X, A) = (0.5 - X_1^2 - X_2^2)(X_1^2 + X_2^2 - 0.3) A$
Nefazodone-CBASP clinical trial (Keller et al., 2000)

- 681 patients with non-psychotic chronic major depressive disorder (MDD).
- Randomized in a 1:1:1 ratio to either nefazodone, cognitive behavioral-analysis system of psychotherapy (CBASP) or the combination of nefazodone and psychotherapy.
- Primary outcome: score on the 24-item Hamilton Rating Scale for Depression (HRSD); the lower the better.
- 50 baseline variables: demographics, psychological problem diagnostics etc.
Pairwise Comparison:

- OWL: gaussian kernel.
  $l_1$-PLS and OLS: $(1, X, A, XA)$.
- Value calculated with a 5-fold cross validation type analysis.

Table: Mean HRSD (Lower is Better) from Cross Validation Procedure with Different Methods

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>$l_1$-PLS</th>
<th>OWL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nefazodone vs CBASP</td>
<td>15.87</td>
<td>15.95</td>
<td>15.74</td>
</tr>
<tr>
<td>Combination vs Nefazodone</td>
<td>11.75</td>
<td>11.28</td>
<td>10.71</td>
</tr>
<tr>
<td>Combination vs CBASP</td>
<td>12.22</td>
<td>10.97</td>
<td>10.86</td>
</tr>
</tbody>
</table>
Concluding Remarks
Conclusions

Outcome Weighted Learning procedure

- Discovers an optimal individualized therapy to improve expected (long term) outcome.
- Nonparametric approach sidesteps the inversion of the predicted model.
- Aims directly at maximizing the value function.
- Casts OWL into a non-standard classification problem without knowledge of true class labels.
- Achieves consistency and fast convergence rates, similar to the rates of SVM with the same type of assumptions on the separations.
Improved OWL
**Improved learning algorithm**

- For single-stage OWL, note $\mathcal{V}(D)$ is equivalent to

$$
E \left( \frac{|R - s(H)| I(Asign(R - s(H)) = D(H))}{P(A|H)} \right) 
+ E[s(H)] - E[(R - s(H))^+] .
$$

- It implies that OWL can be implemented using a new weighted supervised learning:

  **Class label:** Asign($R - s(H)$); **feature variables:** $H$; **weight:** $|R - s(H)|$.

- The most important feature is to use a residual variable, $R - s(H)$, so reduce weighting variability to improve learning efficiency.
See arxiv link for paper

Robust Hybrid Learning for Estimating Personalized Dynamic Treatment Regimens

by Ying Liu, Yuanjia Wang, Michael R. Kosorok, Yingqi Zhao, Donglin Zeng

https://arxiv.org/abs/1611.02314