Near-optimal Individualized Treatment Recommendations

Haomiao Meng, Ying-Qi Zhao, Haoda Fu and Xingye Qiao*

Abstract

Individualized treatment recommendation (ITR) is an important analytic framework for precision medicine. The goal is to assign proper treatments to patients based on their individual characteristics. From the machine learning perspective, the solution to an ITR problem can be formulated as a weighted classification problem to maximize the average benefit that patients receive from the recommended treatments. Several methods have been proposed for ITR in both binary and multicategory treatment setups. In practice, one may prefer a more flexible recommendation with multiple treatment options. This motivates us to develop methods to obtain a set of near-optimal individualized treatment recommendations alternative to each other, called alternative individualized treatment recommendations (A-ITR). We propose two methods to estimate the optimal A-ITR within the outcome weighted learning (OWL) framework. We show the consistency of these methods and obtain an upper bound for the risk between the theoretically optimal recommendation and the estimated one. We also conduct simulation studies, and apply our methods to a real data set for Type 2 diabetic patients with injectable antidiabetic treatments. These numerical studies have shown the usefulness of the proposed A-ITR framework. We develop a R package aitr which can be found at https://github.com/menghaomiao/aitr.

Keywords: individualized treatment recommendation; set-valued classification; angle-based classification; reproducing kernel Hilbert space; statistical learning theory.

---

*Correspondence to: Xingye Qiao (e-mail: qiao@math.binghamton.edu). Haomiao Meng is a PhD student in the Department of Mathematical Sciences at Binghamton University, State University of New York, Binghamton, New York, 13902; Yingqi Zhao is an Associate Member at Public Health Sciences Division of the Fred Hutchinson Cancer Research Center, Seattle, Washington, 98109; Haoda Fu is a Senior Research Advisor at Eli Lilly and Company, Indianapolis, Indiana, 46285; Xingye Qiao is an Associate Professor in the Department of Mathematical Sciences at Binghamton University, State University of New York, Binghamton, New York, 13902.

1 Introduction

The individualized treatment recommendation (ITR) has drawn increasing attentions in recent years. Because patients respond differently to the same treatment in some diseases (Lesko, 2007; Insel, 2009), it is desirable to individualize the treatment according to patients' characteristics. Mathematically, an ITR is a map from the covariates to a treatment. The goal is to find the optimal ITR so that the average benefit that patients will receive by following such a recommendation is maximized.

In the literature, many statistical approaches have been proposed for solving the optimal ITR. For indirect modeling-based methods, one first builds a parametric or semi-parametric model to estimate the expected outcome based on a patient's characteristics, then recommends the treatment that renders the optimal outcome to the patient (Robins, 2004; Qian and Murphy, 2011; Schulte et al., 2014). However, they require correct model specification and accurate estimation to work well practically. One may also solve the optimal ITR directly. Zhao et al. (2012) proposed a classification-based method, coined as the outcome weighted learning (OWL), to estimate the optimal ITR. They transformed the ITR problem into a weighted classification problem and used support vector machine (SVM), a classification method, to solve it. Built on top of the OWL framework, there is a rapidly growing literature on different aspects of the ITR problem. Zhao et al. (2014) and Cui et al. (2017) extended the OWL framework to accommodate survival outcome. Zhou et al. (2017) and Liu et al. (2018) proposed residual weighted learning (RWL) and augmented outcome-weighted learning (AOL) respectively to reduce the variability of weight in OWL to enhance its performance. Chen et al. (2018) proposed generalized OWL (GOWL) to solve an ITR with ordinal treatments. Zhang et al. (2018a) proposed angle-based approach for the multicategory case (in which there are more than two treatments to choose from). Recently Zhao et al. (2019) and Huang et al. (2019) considered replacing the weight in OWL with a doubly-robust estimator to further improve the robustness of OWL. Methods based on other learning algorithms such as trees (Laber and Zhao, 2015; Kallus, 2016; Doubleday et al., 2018; Zhu et al., 2017) and nearest neighbors (Zhou and Kosorok, 2017; Wu et al., 2019) are also studied. Another example of direct-search methods is the work by Zhang et al. (2012), which searched for the ITR among a pre-specified class of decision rules that optimized a doubly robust augmented inverse probability weighted estimator of the overall population mean outcome.

Despite of the success of these methods in recommending a single "optimal" treatment to patients, a method that can suggest multiple "near-optimal" treatment options to a patient is not fully studied. Such options could be desirable when several treatments have comparable effects. Laber et al. (2014) and Lizotte and Laber (2016) proposed a set-valued dynamic treatment regime. In particular, if there are two treatments available (1 and -1), their set-valued rule may report $\{1\}$, $\{-1\}$, or $\{1, -1\}$. However, this approach is applicable only to cases with two competing outcomes. They would recommend the set $\{1, -1\}$ if any one treatment cannot be proven to be inferior to the other based on the two outcomes. On the other hand, they used a regression-based method to estimate the optimal set-valued rule, which may suffer an inconsistency issue if the model is mis-specified. Yuan (2015) considered a framework to allow a reject option in ITR estimation based on OWL. However, the method is restricted to the binary case (only two possible treatments).

In this paper, we propose to study the ITR problem in the setting with only one clinical outcome from a new perspective. Different from the previous ITR work, it provides a set of ITRs that are near the optimality and are alternative to each other, which we called alternative individualized treatment recommendations (A-ITR). Specifically, multiple treatments are recommended to the patient if they are expected to result in similar clinical outcomes for the patient. There are multiple reasons such alternative options are desired. Firstly, for some patients, since multiple treatments may yield the same or similar outcomes, the ranking between the top treatment options may vary due to some randomness or noise in the learning process. In the case that the expected outcomes for multiple outcomes are indistinguishable, it is morally inappropriate to withhold such important information from the patients. Secondly, such alternative options allow patients to incorporate other factors into their choice of the final treatment plan. These factors include the healthcare expense, the painfulness of the treatment, the life quality and life style, and so on. Specifically, when two treatments are expected to have similar outcomes, it is reasonable for the patient to choose an option which is covered by the insurance, that is less painful, or that does not significantly compromise the quality of life. In this sense, conventional ITR methods that only recommend one treatment to a patient may prevent patients from making informed decisions about their lives.

We will propose two methods to estimate A-ITR. Parallel to the development of the con-

ventional ITR methods, we first introduce a regression-based plug-in method to estimate the optimal A-ITR, which will serve as the baseline. Within the OWL framework, we propose two classification-based methods. The technical tool we will use is multicategory classification with reject and refine options (Zhang et al., 2018b).

The rest of the paper is organized as follows. In Section 2, we review the background of the ITR and the classification with reject and refine options problems. We then introduce the proposed A-ITR framework and discuss several estimation methods in Section 3. Discussions about the algorithm and the tuning procedure can be found in Section 4. In Section 5, we study the statistical learning theory for our proposed methods. Simulation studies and an application to Type 2 diabetes mellitus data are provided in Sections 6 and 7 respectively. Some concluding remarks are given in Section 8. All technical proofs are provided in the supplementary materials.

2 Background

In this section we briefly review the background information of both ITR and the problem of classification with reject and refine options.

2.1 Individualized Treatment Recommendation

Denote the covariates of a patient by $\boldsymbol{X} \in \mathcal{X}$. Each treatment is denoted by a random variable A, where $A \in \mathcal{A} = \{1, 2, \ldots, k\}$ (k treatments available.) After assigning a treatment to a patient, we observe an outcome $Y \in \mathbb{R}^+$. Here we assume Y is bounded and *smaller* Y is preferred. Then an individualized treatment recommendation, previously often referred to as an individualized treatment rule, is defined to be a map $d : \mathcal{X} \to \mathcal{A}$.

Let $Y^*(j)$ denote the potential outcome that would have been observed when treatment j is assigned to the patient. The actual observed outcome Y is related to the potential outcomes by $Y = \sum_{j \in \mathcal{A}} Y^*(j) \mathbb{1}[A = j]$. Define $p(A = j | \boldsymbol{X})$ as the conditional probability of treatment j given $\boldsymbol{X}$. We assume the following assumption.

Assumption 1. *For any j, $Y^*(j)$ is independent of A given $\boldsymbol{X}$; $p(A = j | \boldsymbol{X}) > 0$ almost everywhere.*

Under Assumption 1, it was shown by Qian and Murphy (2011) and Kallus (2016) that the expected outcome under ITR d is

$$\mathbb{E}^d(Y) = \mathbb{E}(Y^*(d(\boldsymbol{X}))) = \mathbb{E}[\mathbb{E}(Y^*(A)|A = d(\boldsymbol{X}), \boldsymbol{X}))] = \mathbb{E}\left[\frac{\mathbb{1}[A = d(\boldsymbol{X})]}{p(A|\boldsymbol{X})}Y\right], \qquad (1)$$

where $\mathbb{E}^d$ is the expectation under ITR d. Note that $p(A|\boldsymbol{X})$ is usually known in a randomized trial, while in an observational study $p(A|\boldsymbol{X})$ is unknown and needs to be estimated first.

Denote $\mu_j = \mathbb{E}(Y|\boldsymbol{X}, A = j)$, for $j = 1, \ldots, k$. Then the optimal ITR d^* under (1) is

$$d^* = \operatorname*{argmin}_d \mathbb{E}^d(Y) = \operatorname*{argmin}_j \mu_j, \qquad (2)$$

that is, the optimal treatment for a patient has the smallest (the best) expected outcome.

Many methods have been proposed for estimating the optimal ITR. One method is often called "regression and comparison" or Q-learning (Robins, 2004; Qian and Murphy, 2011). One first estimates the conditional mean $\mu_j(\boldsymbol{x}) = \mathbb{E}(Y \mid \boldsymbol{X} = \boldsymbol{x}, A = j)$ for each treatment j, then the optimal treatment is obtained by plugging the estimators in (2). However, this method relies on the accuracy of the regression model. If the model is mis-specified, the error could be fairly substantial. Another group of methods treat the problem as a classification problem. One example is called outcome weight learning (OWL) or O-learning (Zhao et al., 2012, 2014, 2019; Zhou et al., 2017; Zhang et al., 2018a). In the OWL framework, we rewrite the ITR solution as

$$d^* = \operatorname*{argmin}_d \mathbb{E}\left[\frac{Y}{p(A|\boldsymbol{X})}\mathbb{1}[A = d(\boldsymbol{X})]\right], \qquad (3)$$

which is closely related to a weighed classification problem with weight $Y/p(A|\boldsymbol{X})$. To overcome the non-continuity and non-convexity of the 0-1 loss, we can replace $\mathbb{1}[A = d(\boldsymbol{X})]$ by a convex surrogate loss $L(A, \boldsymbol{f}(\boldsymbol{X}))$ in the empirical counterpart and solve instead

$$\hat{\boldsymbol{f}} = \operatorname*{argmin}_{\boldsymbol{f}} \mathbb{E}_n\left[\frac{Y}{p(A|\boldsymbol{X})}L(A, \boldsymbol{f}(\boldsymbol{X}))\right], \qquad (4)$$

where $\mathbb{E}_n$ denotes the empirical expectation, and $\boldsymbol{f}$ is a multi-dimensional function defined on $\mathcal{X}$. The estimated ITR $\hat{d}$ is then obtained from $\hat{\boldsymbol{f}}$.

The relationship between $\hat{d}$ and $\hat{\boldsymbol{f}}$ depends on the loss function L and the choice of $\boldsymbol{f}$. Zhao et al. (2012) proposed to replace the 0-1 loss by hinge loss in the binary case ($k = 2$, $A \in \{1, -1\}$), that is, $L(A, f) = (1 - Af)_+$, where $x_+ = \max(x, 0)$, and f is a 1-dimensional function. In the current setting that a smaller Y is preferred, they could have used $L(A, f) = (1 + Af)_+$. They

showed that the optimal ITR can be estimated by $\hat{d} = \text{sign}\,(\hat{f})$. Zhang et al. (2018a) then extended to the multicategory case using a large-margin loss and the angle-based learning framework (Zhang and Liu, 2014). Specifically, define $\boldsymbol{f}(\boldsymbol{x}) = (f_1, \ldots, f_{k-1})^T(\boldsymbol{x}) \in \mathbb{R}^{k-1}$ and $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_k$ are vertices of a $(k-1)$-dimensional simplex, that is

$$
\boldsymbol{W}_j = \begin{cases} (k-1)^{-1/2}\mathbf{1}_{k-1}, & j = 1 \\ -(1+k^{1/2})(k-1)^{-3/2}\mathbf{1}_{k-1} + [k/(k-1)]^{1/2}\,\boldsymbol{e}_{j-1}, & 2 \leq j \leq k, \end{cases}
$$

where $\mathbf{1}_{k-1}$ is a $(k-1)$-dimensional vector with all 1 and $\boldsymbol{e}_{j-1} \in \mathbb{R}^{k-1}$ is a vector with the $(j-1)$th element 1 and 0 elsewhere. They let $L(A, \boldsymbol{f}(\boldsymbol{x})) = \ell(\langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{x})\rangle)$ where ℓ is a typical large-margin surrogate loss for binary classification (except that it is increasing instead of decreasing.) From the geometry point of view, treatment j is represented by vertex j of the simplex, and the angle between $\boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{W}_j$, $\angle(\boldsymbol{W}_j, \boldsymbol{f}(\boldsymbol{x}))$, indicates how far away $\boldsymbol{f}(\boldsymbol{x})$ is from each of these treatments. The resulting ITR was estimated by $\hat{d}(\boldsymbol{x}) = \text{argmin}_j \angle(\boldsymbol{W}_j, \hat{\boldsymbol{f}}(\boldsymbol{x})) = \text{argmax}_j \langle \boldsymbol{W}_j, \hat{\boldsymbol{f}}(\boldsymbol{x})\rangle$, that is, the treatment whose corresponding vertex is closest to $\hat{\boldsymbol{f}}(\boldsymbol{x})$.

REMARK. In the conventional ITR literature, one typically assumes that larger values of the outcome Y are preferred, so that instead of minimization, d^* is the solution to the maximization of the objective (3), or equivalently, $\text{argmin}_d \mathbb{E}\left[\frac{Y}{p(A|\boldsymbol{X})}\mathbb{1}[A \neq d(\boldsymbol{X})]\right]$, which was indeed a weighed classification. In this article, we assume that smaller values of Y are preferred (due to a technical concern about computational complexity.) As consequences, $\mathbb{1}[A \neq d(\boldsymbol{X})]$ is replaced by $\mathbb{1}[A = d(\boldsymbol{X})]$ in (3); additionally, the surrogate loss function is flipped with respect to the origin so that it is an increasing function instead of a decreasing function.

2.2 Classification with Reject and Refine Options

In this article, we aim to provide set-valued recommendations that are near the optimality and are alternative to each other. We borrow the idea of multicategory classification with reject (and refine) options as a technical tool. Classification with a reject option has been widely studied. Herbei and Wegkamp (2006) formulated the problem as a minimization problem under the 0-d-1 loss. That is, the loss of a misclassified instance is 1 and the loss of a rejected instance is d, where $0 \leq d \leq 1/2$. Bartlett and Wegkamp (2008) proposed an estimation procedure under the hinge loss. Yuan and Wegkamp (2010) extended this framework to a broad class of surrogate loss

functions. Zhang et al. (2018b) generalized it to the multicategory case.

We first introduce binary classification with reject option. Let $(\boldsymbol{X}, A)$ be a pair of random variable with $\boldsymbol{X} \in \mathcal{X}$ and class label $A \in \{1, -1\}$[1], and denote $p_j(\boldsymbol{x}) = p(A = j|\boldsymbol{X} = \boldsymbol{x})$ as the conditional class probability given $\boldsymbol{X}$. The goal is to train a classifier $\phi(\boldsymbol{x})$ that produces three possible outputs: 1, -1, and 0. Here 0 stands for a "reject" option, meaning that the classifier refuses to make a prediction based on the information available. Note that the decision "0" can be viewed as a set-valued decision of $\{1, -1\}$. Chow (1970) proposed the 0-d_0-1 loss with corresponding risk function $P(\phi(\boldsymbol{X}) \neq A, \phi(\boldsymbol{X}) \neq 0) + d_0 P(\phi(\boldsymbol{X}) = 0)$ and it was shown that the Bayes rule under this risk is

$$\phi^*(\boldsymbol{x}) = \begin{cases} 1, & p_1(\boldsymbol{x}) > 1 - d_0 \\ -1, & p_1(\boldsymbol{x}) < d_0 \\ 0, & d_0 \leq p_1(\boldsymbol{x}) \leq 1 - d_0. \end{cases}$$

Here $d_0 \in [0, 1/2]$ controls the cost for refusing to make a classification. Intuitively, we produce the reject option "0" only when both p_1 and p_2 are close to $1/2$. Bartlett and Wegkamp (2008) proposed a bent hinge loss to estimate the optimal rule ϕ^*. The bent hinge loss is defined as $\ell(u) = \max(0, 1 - u, 1 - (1 - d_0)u/d_0)$, i.e., the common hinge loss with a bent slope at 0. The effect of such bent slope is to shrink f to 0 when p_1 and p_2 are close. For $f^*(\boldsymbol{x}) = \operatorname{argmin}_{f(\boldsymbol{x})} \mathbb{E}(\ell(Af)|\boldsymbol{X} = \boldsymbol{x})$, we have $\phi^* = \operatorname{sign}(f^*)$.

The situation is much more complicated for multicategory classification. Suppose there are 3 classes, that is, $\mathcal{A} = \{1, 2, 3\}$, then the possible values for the classifier $\phi(\boldsymbol{X})$ are $\{1\}$, $\{2\}$, $\{3\}$, $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, and $\{1, 2, 3\}$. In general, assuming there are k classes, $\phi(\boldsymbol{X})$ can be any element in the power set of $\{1, \ldots, k\}$ (except the empty set). In addition to the reject option, which can be written as $\{1, \ldots, k\}$, Zhang et al. (2018b) introduced the so-called refine option, in which a refined decision has a cardinality greater than 1 and less than k. It contains all those class labels which are nearly as plausible as the most plausible class. Zhang et al. (2018b) proposed to use a class of loss functions in conjunction with the angle-based learning framework Zhang and Liu (2014) to train a set-valued classifier that can render these different options. We note that both the reject option and the refine option are set-valued decisions, and they are analogous to

---

[1] Although the class label is often denoted as Y in the classification literature, the role of the class label is comparable to the role of the treatment option in the ITR setting. Hence we denote the class label as A here.

the set-valued recommendations in this article.

3 Methodology

In this section, we introduce the framework of alternative individualized treatment recommendations (A-ITR) and propose two methods to estimate the optimal A-ITR.

3.1 A-ITR Framework

There are several situations in which ITRs with additional alternative options are desirable. Even with small errors, when several treatments are near the optimality, the ranking of these treatments based on their estimated outcomes may differ from their true ranking. In this case, reporting only one treatment based on the estimated value is problematic. Secondly, when the error in the learning problem is substantially large, the so-called optimal treatment reported by conventional ITRs may lead to an outcome that is much worse than some of the other treatment options. In these situations, recommending a single treatment only adds to the distrust that patients may already have towards such black-box algorithms that they know little about. On the other hand, A-ITR provides a safety net, preventing from committing to a single treatment that is only one out of multiple treatments with similar or indistinguishable outcomes. Morally, as patients are more mindful about their financial responsibility and their quality of life, it is more appropriate to present these alternative options and have the patients themselves to make an informed decision, especially when many of these decisions are life changing.

An A-ITR is a set-valued map $\phi : \mathcal{X} \to 2^{\mathcal{A}} \backslash \emptyset$. Inspired by the idea of classification with reject and refine options, we formally define the optimal A-ITR as,

$$\phi^*(\boldsymbol{x}) = \{j; \mu_j(\boldsymbol{x})/\mu_{(1)}(\boldsymbol{x}) \leq c\}, \tag{5}$$

where $\mu_{(1)} < \cdots < \mu_{(k)}$ denote the ordered conditional mean outcomes, and $c \geq 1$ is a user-predefined number. This optimal A-ITR is a near-optimal treatment recommendation set since it contains all the treatment options with μ_j close to that of the optimal one, up to a multiplicative constant c. Note that ϕ^* may contain only one element, that is, the treatment with the smallest mean outcome, which corresponds to the conventional ITR. If it includes all the treatments, it

is a non-informative recommendation, analogous to the reject option in set-valued classification. Here we call c the near-optimal parameter.

When $c = 1$, the optimal A-ITR reduces to the optimal ITR defined in (2) since $\phi^* = \{j; \mu_j/\mu_{(1)} \leq 1\} = \{j; \mu_j = \mu_{(1)}\} = \arg\min_j \mu_j$. This means that the proposed optimal A-ITR is a generalization of the conventional optimal ITR. Figure 1 shows theoretical regions for different types of recommendations from the optimal A-ITR for 2 different values of c in a $k = 3$ case.
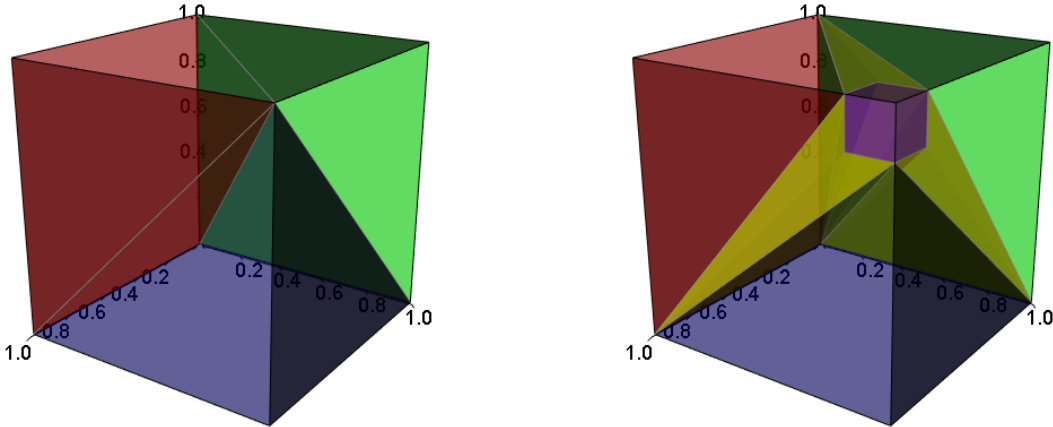
Figure 1: The optimal A-ITR when $k = 3$ for different c (left: $c = 1$; right: $c = 1.2$). Any point in the plot represents (μ_1, μ_2, μ_3) (suppose $Y^*(j) \in (0, 1)$) with the recommendation illustrated by colors. Points in the red, green, and blue regions contain only one treatment option; the yellow region contains recommendations with two options; and the purple region includes those with all three options.

3.2 Estimation

We consider two types of methods to estimate the optimal A-ITR: the regression-based methods and the classification-based methods. For regression-based methods, we can use Q-learning to first estimate the conditional mean $\mu_j(\boldsymbol{x}) = \mathbb{E}(Y|\boldsymbol{x}, A = j)$ for each treatment j, then plug into (5), i.e., $\hat{\phi}(\boldsymbol{x}) = \{j; \hat{\mu}_j(\boldsymbol{x})/\hat{\mu}_{(1)}(\boldsymbol{x}) \leq c\}$. The success of this regression-based plug-in method relies on accurate estimation of μ_j.

In contrast, the classification-based method targets on estimating the true boundary between different decision regions, bypassing the need to estimate μ_j directly. In the rest of the section, we propose two classification-based methods within the OWL framework, both of which are based on the angle-based learning approach (Zhang and Liu, 2014).

Zhang et al. (2018a) first made use of the angle-based learning approach to solve the ITR

problem, in which they denoted $\boldsymbol{W}_1 \ldots, \boldsymbol{W}_k$ as the vertices of a $(k-1)$-dimensional simplex and they chose the loss $L(A, \boldsymbol{f})$ in (4) to be a function that only depends on the inner product $\langle \boldsymbol{W}_A, \boldsymbol{f} \rangle$, namely, $L(A, \boldsymbol{f}) = \ell(\langle \boldsymbol{W}_A, \boldsymbol{f} \rangle)$. Define $\boldsymbol{f}^*$ to be the population minimizer under such loss, that is

$$\boldsymbol{f}^* = \operatorname*{argmin}_{\boldsymbol{f} \in \{\mathcal{X} \to \mathbb{R}^{k-1}\}} \mathbb{E}\left[\frac{Y}{p(A|\boldsymbol{X})}\ell(\langle \boldsymbol{W}_A, \boldsymbol{f}(\boldsymbol{X}) \rangle)\right]. \tag{6}$$

The end product of Zhang et al. (2018a) was a single-treatment ITR. In the ideal case that $\boldsymbol{f}^*$ can be obtained, their ITR was defined as $d_{\boldsymbol{f}^*}(\boldsymbol{x}) = \operatorname{argmax}_j \langle \boldsymbol{W}_j, \boldsymbol{f}^*(\boldsymbol{x}) \rangle$, and it can be shown that as long as ℓ is convex and strictly increasing, Fisher consistency holds, i.e.$d_{\boldsymbol{f}^*} = d^*$. In practice, given the training data set $\{(\boldsymbol{x}_i, a_i, y_i)\}_{i=1}^n$, $\hat{\boldsymbol{f}}$, the estimate of $\boldsymbol{f}^*$, is obtained by,

$$\hat{\boldsymbol{f}} = \operatorname*{argmin}_{\boldsymbol{f} \in \mathcal{F}} \frac{1}{n}\sum_{i=1}^n \frac{y_i}{p(a_i|\boldsymbol{x}_i)}\ell(\langle \boldsymbol{W}_{a_i}, \boldsymbol{f}(\boldsymbol{x}_i) \rangle), \text{ subject to } J(\boldsymbol{f}) \leq s, \tag{7}$$

where $\mathcal{F} \subseteq \{\boldsymbol{f}; \mathcal{X} \to \mathbb{R}^{k-1}\}$ is a class of functions, and $J(\boldsymbol{f})$ is a penalty term to prevent overfitting.

Both our proposed A-ITR methods are derived from the population minimizer $\boldsymbol{f}^*$ (6) or the empirical minimizer $\hat{\boldsymbol{f}}$ (7) obtained in this way. The differences lie in the loss function ℓ they use, and how they convert $\boldsymbol{f}^*$ or $\hat{\boldsymbol{f}}$ to the final set-valued recommendations.

3.2.1 Two-step OWL Method

For the two-step method, we use a convex, differentiable, and increasing loss function ℓ_D. To calculate the A-ITR, we have to first order the vertices, for any $\boldsymbol{f}$ (which may be $\boldsymbol{f}^*$ or $\hat{\boldsymbol{f}}$), in the manner of order statistics, i.e., $\langle \boldsymbol{W}_{(1)}, \boldsymbol{f} \rangle > \cdots > \langle \boldsymbol{W}_{(k)}, \boldsymbol{f} \rangle$. The resultant two-step estimator of the optimal A-ITR is then defined as

$$\phi_{\boldsymbol{f}}^D(\boldsymbol{x}) = \left\{ j; \frac{\ell'_D(\langle \boldsymbol{W}_{(1)}, \boldsymbol{f}(\boldsymbol{x}) \rangle)}{\ell'_D(\langle \boldsymbol{W}_j, \boldsymbol{f}(\boldsymbol{x}) \rangle)} \leq c \right\}. \tag{8}$$

Here ℓ'_D is the first derivative of ℓ_D, and the superscript "D" indicates that $\boldsymbol{f}$ is the population solution to (6) or the empirical solution to (7), based on a *differentiable* loss function. Our estimator is partially justified and motivated by the following result.

Proposition 1 (Zhang et al., 2018a). *Let $\boldsymbol{f}^*$ be the population minimizer in (6) in which ℓ is a convex and differentiable function ℓ_D with $\ell'_D(u) > 0$ for all u. For any $i \neq j \in \{1, \ldots, k\}$, we have*

$$\frac{\mu_j}{\mu_i} = \frac{\ell'_D(\langle \boldsymbol{W}_i, \boldsymbol{f}^* \rangle)}{\ell'_D(\langle \boldsymbol{W}_j, \boldsymbol{f}^* \rangle)}.$$

9

Proposition 1 implies the following Fisher-consistent-like result for our proposed A-ITR estimator.

Proposition 2. *Let $\boldsymbol{f}^*$ be the population minimizer in (6) in which ℓ is a convex and differentiable function ℓ_D with $\ell'_D(u) > 0$ for all u. The A-ITR $\phi^D_{\boldsymbol{f}^*}$ defined in (8) based on $\boldsymbol{f}^*$ coincides with the optimal A-ITR ϕ^* (5).*

This method is a two-step procedure because it first estimates the ratios of conditional means μ_j/μ_i using $\ell'_D(\langle \boldsymbol{W}_i, \hat{\boldsymbol{f}} \rangle)/\ell'_D(\langle \boldsymbol{W}_j, \hat{\boldsymbol{f}} \rangle)$ then plugs it back into (5). Note that it does not estimate each conditional mean individually, but their ratios. The issue remains that if $\boldsymbol{f}^*$ cannot be accurately estimated, the ratio of the conditional means cannot be accurately estimated.

3.2.2 One-step OWL Method

The one-step method aims to directly obtain a set-valued recommendation without calculating $\ell'_D(\langle \boldsymbol{W}_i, \hat{\boldsymbol{f}} \rangle)/\ell'_D(\langle \boldsymbol{W}_j, \hat{\boldsymbol{f}} \rangle)$. The crucial difference here is the use of a bent loss function, defined as

$$\ell_B(\langle \boldsymbol{W}_j, \boldsymbol{f} \rangle) = \ell_1(\langle \boldsymbol{W}_j, \boldsymbol{f} \rangle) + \ell_2(\langle \boldsymbol{W}_j, \boldsymbol{f} \rangle),$$

where $\ell_1 > 0$ is a convex and increasing function with $\ell'_1(u) = 1$ for all $u \geq 0$, and $\ell_2(u) = (c-1)u_+$ with $c \geq 1$. Such a loss function is bent at 0, since $\ell'_B(0-) = 1$ and $\ell'_B(0+) = c$. An example of bent loss is the bent hinge loss, $\ell_B(u) = (1+u)_+ + (c-1)u_+$ (see Figure 2.) The bent loss has been a critical tool that helps to achieve reject (and refine) options in the classification literature (Bartlett and Wegkamp, 2008; Zhang et al., 2018b).

The main effect of the bent loss is to shrink the so-called angle margins for class (or treatment) j, defined as $\langle \boldsymbol{W}_j, \boldsymbol{f} \rangle$, towards 0, similar to the shrinkage effect of the lasso penalty. Likewise, the additional slope $c \geq 1$, the near-optimal parameter, for positive u is analogous to a larger penalty parameter in lasso regression, which would encourage a sparser model. Note that here such a shrinkage effect is applied to the classes (treatments) with nonnegative angle margins only. Specifically, the angle margin for class j will tend to be shrunken to 0 if $1 < \mu_j/\mu_{(1)} \leq c$. Formally, $\boldsymbol{f}^*$, the solution to (6) with the bent loss ℓ_B, enjoys the following theoretical property. Proposition 3 is derived from Proposition 1 in Zhang et al. (2018b)

Proposition 3. *For the sequence $\mu_{(1)} < \cdots < \mu_{(k)}$, if there exists an integer $r \in \{1, \ldots, k-1\}$ such that $\mu_{(j)}/\mu_{(1)} < c$ for $j = 1, \ldots, r$ and $\mu_{(j)}/\mu_{(1)} > c$ for $j = r+1, \ldots, k$, then $\boldsymbol{f}^*$ as defined in*
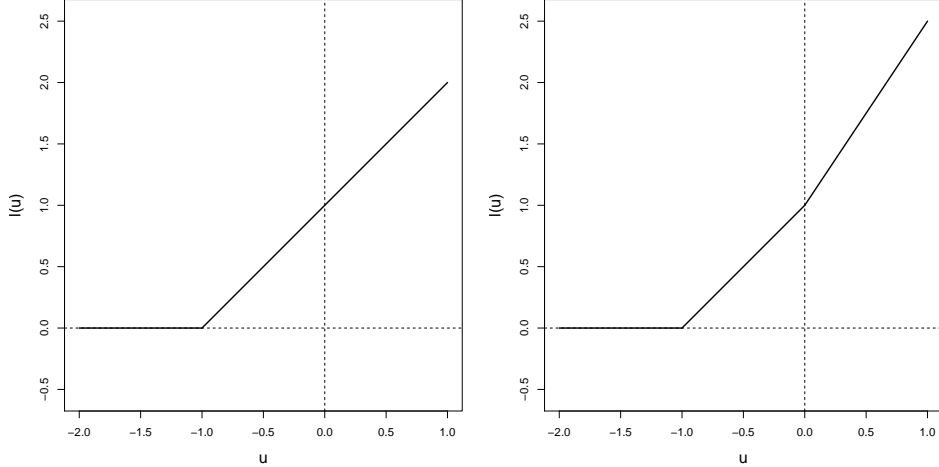
Figure 2: Hinge loss (left panel) and bent hinge loss with $c = 1.5$ (right panel).

(6) in which ℓ is a convex and increasing function ℓ_B with $\ell'_B(0-) = 1$ and $\ell'_B(0+) = c \geq 1$ satisfies that $\langle \boldsymbol{W}_{(1)}, \boldsymbol{f}^* \rangle > 0$, $\langle \boldsymbol{W}_{(2)}, \boldsymbol{f}^* \rangle = \cdots = \langle \boldsymbol{W}_{(r)}, \boldsymbol{f}^* \rangle = 0$, and $\langle \boldsymbol{W}_{(r+1)}, \boldsymbol{f}^* \rangle = \cdots = \langle \boldsymbol{W}_{(k)}, \boldsymbol{f}^* \rangle < 0$; otherwise, $\langle \boldsymbol{W}_{(j)}, \boldsymbol{f}^* \rangle = 0$ for $j = 1, \ldots, k$.

Proposition 3 shows that when a bent loss is used, the population minimizer $\boldsymbol{f}^*$ (6) has the nice property that all those treatments within the near-optimal net (defined as $\mu_{(j)}/\mu_{(1)} \leq c$) have non-negative angle margins; all those treatments outside the near-optimal net have negative angle margins. This naturally leads to the following set-valued recommendation,

$$\phi^B_{\boldsymbol{f}}(\boldsymbol{x}) = \{j; \langle \boldsymbol{W}_j, \boldsymbol{f}(\boldsymbol{x}) \rangle \geq 0\}. \tag{9}$$

The superscript "B" means that $\boldsymbol{f}$ is the population minimizer $\boldsymbol{f}^*$ (6) or the empirical minimizer $\hat{\boldsymbol{f}}$ (7), with the loss ℓ being a *bent* loss ℓ_B, as opposed to a differentiable loss function in the two-step method.

Proposition 3 implies that $\{j; \mu_j(\boldsymbol{x})/\mu_{(1)}(\boldsymbol{x}) < c\} \subseteq \phi^B_{\boldsymbol{f}^*}(\boldsymbol{x}) \subseteq \{j; \mu_j(\boldsymbol{x})/\mu_{(1)}(\boldsymbol{x}) \leq c\}$. The following assumption is necessary to resolve the identifiable issue of (9) and to show its optimality.

Assumption 2. *For any positive c_0, $P(\mu_j(\boldsymbol{X}) = c_0 \mu_i(\boldsymbol{X})) = 0$ for $\forall i \neq j \in \{1, \ldots, k\}$ in which $\mu_j(\boldsymbol{X}) = \mathbb{E}(Y|\boldsymbol{X}, A = j)$ is the conditional mean outcome for treatment j.*

Assumption 2 guarantees the sets $\{\boldsymbol{x} \in \mathcal{X}; \mu_j(\boldsymbol{x})/\mu_i(\boldsymbol{x}) = 1\}$ and $\{\boldsymbol{x} \in \mathcal{X}; \mu_j(\boldsymbol{x})/\mu_i(\boldsymbol{x}) = c\}$ (where c is the near-optimal parameter in (5)) have measure 0 for any $i \neq j$ so that $\boldsymbol{f}^*$ is identifiable almost everywhere. Under Assumption 2, we have the following proposition, analogous to Fisher

consistency in classification.

Proposition 4. *Suppose Assumption 2 holds. Let $\boldsymbol{f}^*$ be the population minimizer in (6) in which ℓ is a convex and increasing function ℓ_B with $\ell'_B(0-) = 1$ and $\ell'_B(0+) = c \geq 1$. The A-ITR $\phi^B_{\boldsymbol{f}^*}$ defined in (9) based on $\boldsymbol{f}^*$ coincides with the optimal A-ITR ϕ^* (5) with the near-parameter parameter c.*

Note that for both classification-based methods, a single-valued ITR can be easily defined by recommending the treatment option with the largest angle margin, that is, $\operatorname{argmax}_j \langle \boldsymbol{W}_j, \boldsymbol{f}(\boldsymbol{x}) \rangle$.

Unlike the regression-based method, the two classification-based methods do not estimate the conditional mean outcome. The success of the regression-based method relies on accurate estimation of μ_j at every $\boldsymbol{x}$ of interest, while reasonable performance is expected for the classification-based methods as long as the estimation is accurate around the "boundaries". However, the two-step method and the one-step method seem to have different focuses. Both methods start with finding a discriminant function to minimize outcome-weighted classification error for the purpose of minimizing the expected outcome. As a consequence, both methods have "good" performances near boundaries that distinguish the optimal treatment from the non-optimal treatments for each patient. The one-step method, additionally, uses a bent loss with a shrinkage effect that is capable of determining whether a treatment is *close enough to*, not whether it is *equal to*, the optimal treatment (more precisely speaking, whether the ratio between the conditional outcomes $\mu_j(\boldsymbol{x})/\mu_{(1)}(\boldsymbol{x})$ is less than c or not). This is theoretically justified by Proposition 3. Hence, the one-step method also has "good" performance near such new notions of boundaries (that is, boundaries between the top few "equally good" treatment options, and the others.)

To illustrate the additional strength of the one-step method, we show the boundaries between recommendations for a toy example (the details of which will be revisited in the numerical studies) in Figure 3, in which the top row shows the single-valued ITR and the second row the set-valued A-ITR, by the Bayes rule, the two-step method and the one-step method respectively. Both classification-based methods give good approximations to the Bayes ITR boundaries, shown in the top row. However, the two-step method seems to have a bigger error in terms of the A-ITR boundaries when compared to the Bayes rule (shown in the bottom row), than the one-step method does. This is probably due to the fact that the optimization for the two-step method is not designed to capture this subtle pattern, at least not for a finite sample.
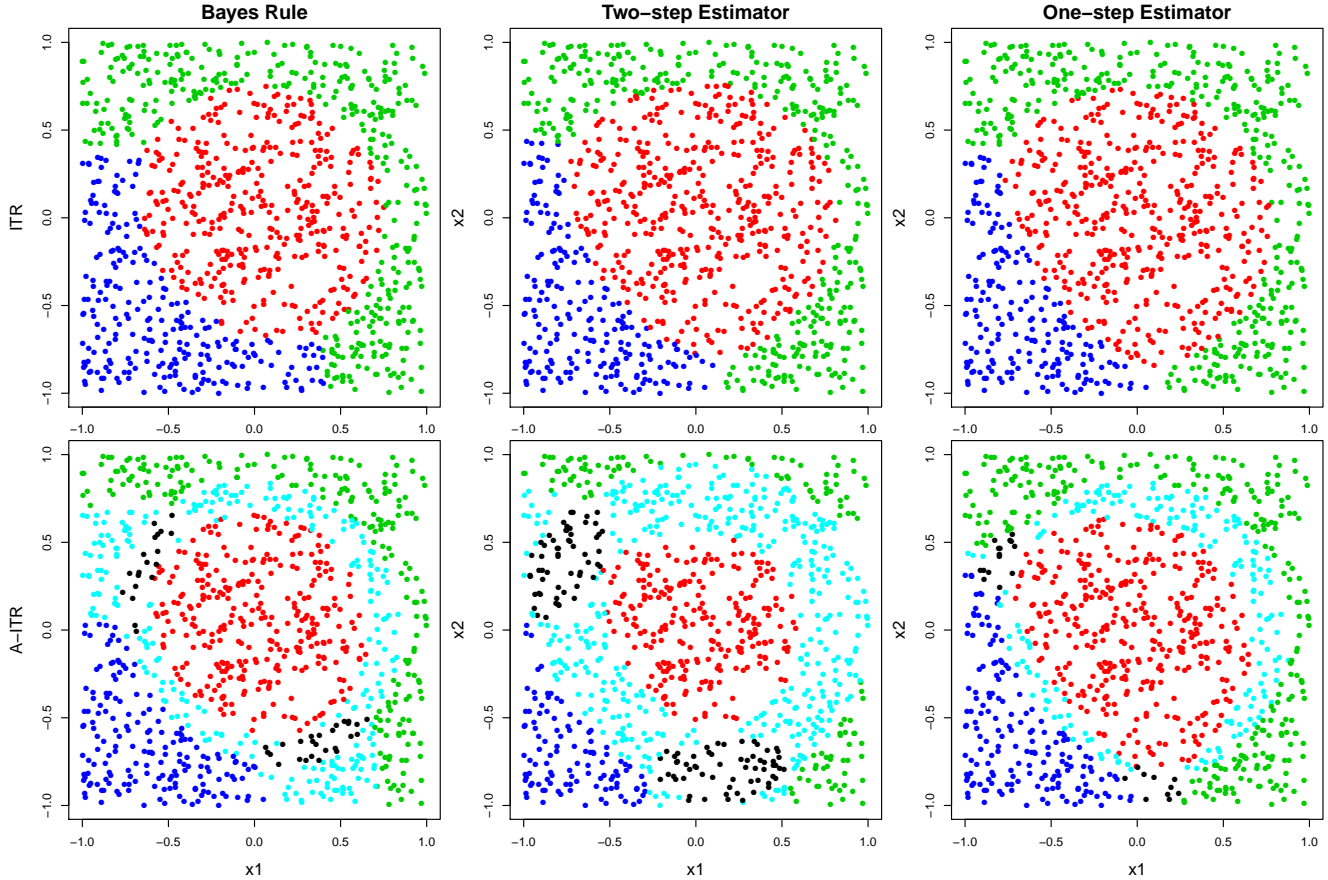
12

Figure 3: ITR (top) and A-ITR (bottom) for a toy example given by Bayes rule, the two-step estimator and one-step estimator. Cyan color indicates recommendations with two options, and black color indicates three options. Both estimators give similar results to the Bayes rule in terms of the single-valued ITR. The two-step method does not provide as good A-ITR results as the one-step method.

A potential drawback of the one-step method is that it relies on an additional Assumption 2 to be identifiable. In general such an assumption is very weak, though it can be nontrivial in cases where the outcome can only take finite values. For example, if the possible outcomes are integers between 1 to 10, then Assumption 2 may not hold for some c.

4 Implementations

In this section, we discuss various aspects of the implementations for the proposed methods, including the optimization, the normalization of the predictive function, and the parameter tuning.

4.1 Algorithm

In this section, we introduce the optimization procedure to estimate $\boldsymbol{f}^*$ defined in (6). Instead of the constrained problem (7), we solve the regularized problem:

$$\min_{\boldsymbol{f} \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p(a_i|\boldsymbol{x}_i)} \ell(\langle \boldsymbol{W}_{a_i}, \boldsymbol{f}(\boldsymbol{x}_i) \rangle) + \lambda J(\boldsymbol{f}), \tag{10}$$

where λ is a tuning parameter. It is a weighted classification problem with weight $w_i = y_i/p(a_i|\boldsymbol{x}_i)$.

In terms of the function class $\mathcal{F}$, there are linear learning and kernel learning (Steinwart et al., 2007; Hofmann et al., 2008; Hastie et al., 2009). Let $\boldsymbol{f} = (f_1, \dots, f_{k-1})^T \in \mathcal{F}$, and for simplicity, we add a constant term to $\boldsymbol{x}$. Then for linear learning, we have $f_j(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{\beta}_j$, and the corresponding penalty $J(\boldsymbol{f}) = \sum_{j=1}^{k-1} ||\boldsymbol{\beta}_j||^2 = \sum_{j=1}^{k-1} \boldsymbol{\beta}_j^T \boldsymbol{\beta}_j$. For kernel learning, $f_j(\boldsymbol{x}) = \sum_{i=1}^{n} K(\boldsymbol{x}_i, \boldsymbol{x}) \alpha_{ij} + \alpha_{0j}$, where $K(\cdot, \cdot)$ is a kernel function. The penalty term becomes $J(\boldsymbol{f}) = \sum_{j=1}^{k-1} \boldsymbol{\alpha}_j^T P \boldsymbol{\alpha}_j + \sum_{j=1}^{k-1} \alpha_{0j}^2$, where P is the gram matrix. Note that we include the intercept term into $J(\boldsymbol{f})$ and a benefit by doing this is the reduction of the complexity of the algorithm. Zhang et al. (2016) shows theoretically that it can achieve the same convergence rate as the case without the intercept term.

We proposed the two-step method and the one-step method. The two-step method is based on a differentiable loss ℓ_D, while the one-step method is based on a bent loss $\ell_B(u) = \ell_1(u) + \ell_2(u)$, where ℓ_1 is convex and $\ell_2(u) = (c-1)u_+$. Since ℓ_D is similar to a special case of ℓ_B with $c = 1$, here we only need to focus on the algorithm for the bent loss ℓ_B. In the rest of this section, we use linear learning to demonstrate our algorithm, and have deferred the details about kernel learning to the supplementary materials.

We first consider the case when ℓ_1 is differentiable. In this case, we use the ADMM (Boyd et al., 2011) algorithm to solve (10). The ADMM algorithm is used when the objective function can be written as a sum of two convex functions, which, in our case, are ℓ_1 and ℓ_2.

To start with, we denote the coefficient matrix as $B_{p \times (k-1)} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{k-1}]$. Then we create another copy of the coefficients $G_{p \times (k-1)} = [\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_{k-1}]$, and let $Z_{p \times (k-1)} = [\boldsymbol{z}_1, \dots, \boldsymbol{z}_{k-1}]$. Recall that $w_i = y_i/p(a_i|\boldsymbol{x}_i)$, then we minimize the augmented Lagrangian

$$L_\rho(B, G, Z) = \sum_{i=1}^{n} w_i \ell_1(\langle \boldsymbol{W}_{a_i}, B^T \boldsymbol{x}_i \rangle) + \sum_{i=1}^{n} w_i \ell_2(\langle \boldsymbol{W}_{a_i}, G^T \boldsymbol{x}_i \rangle) + \frac{n\lambda}{2} \sum_{j=1}^{k-1} \boldsymbol{\beta}_j^T \boldsymbol{\beta}_j$$

$$+ \sum_{j=1}^{k-1} \boldsymbol{z}_j^T (\boldsymbol{\beta}_j - \boldsymbol{\gamma}_j) + \frac{\rho}{2} \sum_{j=1}^{k-1} (\boldsymbol{\beta}_j - \boldsymbol{\gamma}_j)^T (\boldsymbol{\beta}_j - \boldsymbol{\gamma}_j),$$

where $\rho > 0$ controls the step size.

At step t, for each $j = 1, \ldots, k-1$ we can update B^t, G^t and Z^t as

$$\boldsymbol{\beta}_j^t = \underset{\boldsymbol{\beta}_j}{\operatorname{argmin}}\, L_\rho([\boldsymbol{\beta}_1^t, \ldots, \boldsymbol{\beta}_j, \ldots, \boldsymbol{\beta}_{k-1}^{t-1}], G^{t-1}, Z^{t-1}),$$

$$\boldsymbol{\gamma}_j^t = \underset{\boldsymbol{\gamma}_j}{\operatorname{argmin}}\, L_\rho(B^t, [\boldsymbol{\gamma}_1^t, \ldots, \boldsymbol{\gamma}_j, \ldots, \boldsymbol{\gamma}_{k-1}^{t-1}], Z^{t-1}),$$

$$\boldsymbol{z}_j^t = \boldsymbol{z}_j^{t-1} + \rho(\boldsymbol{\beta}_j^t - \boldsymbol{\gamma}_j^t)$$

until matrix B converges. Note that in the two-step method where $c = 1$, we have $\ell_2(u) = 0$. In this case, we can force $B = G$ and only update $\boldsymbol{\beta}_j^t$'s until they converge.

Next we consider the case when ℓ_1 is not differentiable. In the literature of classification, a non-differentiable loss that has been commonly used is hinge loss. Note that in our case, since we prefer smaller outcomes, we define the hinge loss as $\ell_1(u) = (1 + u)_+$ (see Figure 2). That is, we flip the traditional hinge loss around the y-axis to make it an increasing function. A typical approach to a problem with the hinge loss is to transform it into a quadratic programming (QP) problem in its duality (Fung and Mangasarian, 2005; Hastie et al., 2009; Zhang et al., 2018b). Specifically, the dual problem of (10) can be written as

$$\min_{\alpha_j, \gamma_j} \frac{n\lambda}{2} \sum_{j=1}^{k-1} \boldsymbol{\beta}_j^T \boldsymbol{\beta}_j - \sum_{i=1}^{n} \alpha_i$$

$$\text{s.t. } 0 \leq \alpha_i \leq w_i, \ 0 \leq \gamma_i \leq w_i, \ i = 1, \ldots, n,$$

where $\boldsymbol{\beta}_j = -\frac{1}{n\lambda} \sum_{i=1}^{n} (\alpha_i + (c-1)\gamma_i) W_{a_i,j} \boldsymbol{x}_i$, and $W_{a_i,j}$ is the jth component of $\boldsymbol{W}_{a_i}$. Note that the weight $w_i = y_i/p(a_i|\boldsymbol{x}_i)$ serves as the upper bound of the box constraints. Because the objective function is quadratic in α_i and γ_i, it has explicit solution at each iteration. Thus it converges very fast by using algorithms such as coordinate decent (Zhang et al., 2018b).

In practice there may be numerical errors to the solution. Moreover, due to different choices of the tuning parameter λ, the scale of the resulting angle margins may vary much between different tuning trials. We propose the following normalization procedure for the one-step A-ITR $\phi_{\boldsymbol{f}}^B$ (9) to boost the empirical performance. The idea is that instead of recommending all treatments with angle margins greater than or equal to 0, we change the threshold to a small number varying around 0. Such a threshold is a fixed constant δ multiplied by a measure of the scale, chosen to

15

be the magnitude of the smallest angle margin. The normalized two-step A-ITR is then

$$\phi_{\hat{\boldsymbol{f}}}^{B}(\boldsymbol{x}) = \{j; \langle \boldsymbol{W}_j, \hat{\boldsymbol{f}}(\boldsymbol{x}) \rangle \geq \delta M(\boldsymbol{x})\}, \tag{11}$$

where δ is a tuning parameter around 0 and $M(\boldsymbol{x}) = |\langle \boldsymbol{W}_{(k)}, \hat{\boldsymbol{f}}(\boldsymbol{x}) \rangle|$ is the magnitude of the smallest angle margin (note that $\langle \boldsymbol{W}_{(k)}, \hat{\boldsymbol{f}}(\boldsymbol{x}) \rangle$ is negative).

4.2 Tuning Procedure

In this paper, the estimation procedure involves two tuning parameters. The first one is the regularization parameter λ in (10) which appears in both the two-step and one-step methods. The second one is the normalization parameter δ in (11) for the one-step method only. We will tune these two parameters differently in two steps.

The first step is to tune λ. For each λ, the estimated solution is $\hat{\boldsymbol{f}}$. Then we define the corresponding single-treatment ITR as $d_{\hat{\boldsymbol{f}}} = \text{argmax}_j \langle \boldsymbol{W}_j, \hat{\boldsymbol{f}} \rangle$ and calculate its empirical average of the expected outcome (1), which is given by

$$\frac{\sum_{i=1}^n \left(\mathbb{1}\left[a_i = d_{\hat{\boldsymbol{f}}}(\boldsymbol{x}_i)\right] y_i / p(a_i | \boldsymbol{x}_i) \right)}{\sum_{i=1}^n \left(\mathbb{1}\left[a_i = d_{\hat{\boldsymbol{f}}}(\boldsymbol{x}_i)\right] / p(a_i | \boldsymbol{x}_i) \right)}$$

(Zhao et al., 2012; Zhang et al., 2018a). We choose the λ that yields the smallest empirical risk for the resulting ITR, even if our ultimate goal is to obtain a set-valued A-ITR. This can substantially simplify the tuning process. We found that other more complicated tuning procedures have led to a similar performance.

For the one-step method, we need to continue to tune δ. For the same λ (same resulting ITR), because different δ's may lead to slightly different set-valued A-ITRs and recommendations with different carnalities, we must actually compare the resulting A-ITRs to choose the best δ, instead of using the ITR as a proxy. However, there are some difficulties in evaluating the performance of the estimated A-ITR. Compared to the conventional ITR, the challenge here is that when the recommendation includes two or more treatment options, there are multiple potential outcomes and it is difficult to quantify the "overall" benefit for such a recommendation.

Although the proposed optimal A-ITR ϕ^* defined in (5) is not a Bayes rule under any loss

function, we can consider a closely related loss function, whose risk function is given by

$$\mathbb{E}\left[\frac{Y\mathbb{1}[A \in \phi(\boldsymbol{X})]}{p(A|\boldsymbol{X})(1 + (|\phi(\boldsymbol{X})| - 1)c)}\right], \tag{12}$$

where $\phi : \mathcal{X} \to 2^{\mathcal{A}} \backslash \emptyset$ is a set-valued predictor and $|\cdot|$ denotes the cardinality of a set. Compared to the expected outcome $\mathbb{E}^d(Y)$ defined in (1), this quantity is a weighted outcome with weight $1/(1 + (|\phi| - 1)c)$ under ϕ. If we force $|\phi| \equiv 1$, it reduces to $\mathbb{E}^d(Y)$. More importantly, it can be shown that the minimizer of (12), denoted by ϕ^+, is

$$\phi^+(\boldsymbol{x}) = \underset{\phi(\boldsymbol{x}) \in 2^{\mathcal{A}}}{\operatorname{argmin}} \mu^\phi(\boldsymbol{x}), \quad \text{where } \mu^\phi(\boldsymbol{x}) \triangleq \frac{1}{1 + (|\phi(\boldsymbol{x})| - 1)c} \sum_{j \in \phi(\boldsymbol{x})} \mu_j(\boldsymbol{x}).$$

Here μ^ϕ defines a new criterion that generalizes the expected outcomes under a set-valued treatment recommendation ϕ. To see that, note that for $\phi(\boldsymbol{x}) = \{1\}$, $\mu^\phi(\boldsymbol{x}) = \mu_1(\boldsymbol{x})$, while for $\phi(\boldsymbol{x}) = \{1, 2\}$, $\mu^\phi(\boldsymbol{x}) = (\mu_1(\boldsymbol{x}) + \mu_2(\boldsymbol{x}))/(1 + c)$, which is smaller than the simple average $(\mu_1(\boldsymbol{x}) + \mu_2(\boldsymbol{x}))/2$ when $c > 1$. Suppose treatment 1 is better than treatment 2 ($\mu_1(\boldsymbol{x}) < \mu_2(\boldsymbol{x})$). We can show that $\phi_2(\boldsymbol{x}) \triangleq \{1, 2\}$ is as good as $\phi_1(\boldsymbol{x}) \triangleq \{1\}$ under this new criterion if and only if $\mu_2(\boldsymbol{x})/\mu_1(\boldsymbol{x}) \leq c$, which is exactly the near-optimal recommendation set defined in (5).

Intuitively, $\phi^+(\boldsymbol{x})$ is an optimal set of treatments selected to minimize the "average" clinical outcome with a penalty on the cardinality of the recommendation set. Note that when $c = 1$, ϕ^+ is the same as the optimal ITR d^*. Moreover, when $k = 2$, ϕ^+ is the same as the optimal A-ITR ϕ^*, as shown above. When $k \geq 3$, ϕ^+ and ϕ^* are different but are nested within each other in the following way: if we let $S_t^* = \{\boldsymbol{x} \in \mathcal{X}; |\phi^*(\boldsymbol{x})| \leq t\}$ and $S_t^+ = \{\boldsymbol{x} \in \mathcal{X}; |\phi^+(\boldsymbol{x})| \leq t\}$, then we have $S_1^* = S_1^+$, and $S_t^* \subseteq S_t^+$ for $t = 2, \ldots, k - 1$. Figure 4 demonstrates their relationship when $k = 3$.

From Figure 4, we observe that the regions with only one treatment are the same ($S_1^* = S_1^+$), while the regions containing two or three treatments are slightly different. In general, the boundaries between the size-1 decisions and their complements are the same for the two rules ϕ^+ and ϕ^*. They only differ in the boundaries between recommendations with different cardinalities (for example, the boundary between size-2 decisions and size-3 decisions). Due to the similarity between ϕ^+ and ϕ^*, although ϕ^* does not minimize the weighted outcome defined in (12), it is close. This justifies the use of the weighted outcome (12) as a new criterion for the tuning parameter selection. Specifically, we choose the δ value that can yield the smallest value of the

17

following empirical counterpart of (12),

$$\frac{\sum_{i=1}^{n}\left(\mathbb{1}\left[a_i \in \phi_{\hat{f}}(\boldsymbol{x}_i)\right]y_i / \left[p(a_i|\boldsymbol{x}_i)(1 + (|\phi_{\hat{f}}(\boldsymbol{x}_i)| - 1)c)\right]\right)}{\sum_{i=1}^{n}\left(\mathbb{1}\left[a_i \in \phi_{\hat{f}}(\boldsymbol{x}_i)\right] / \left[p(a_i|\boldsymbol{x}_i)|\phi_{\hat{f}}(\boldsymbol{x}_i)|\right]\right)}. \tag{13}$$

In addition to the tuning parameter selection, we may also use this criterion to select different methods for conducting A-ITRs. In the real data analysis, we will use this criterion to select between the two proposed classification-based methods.
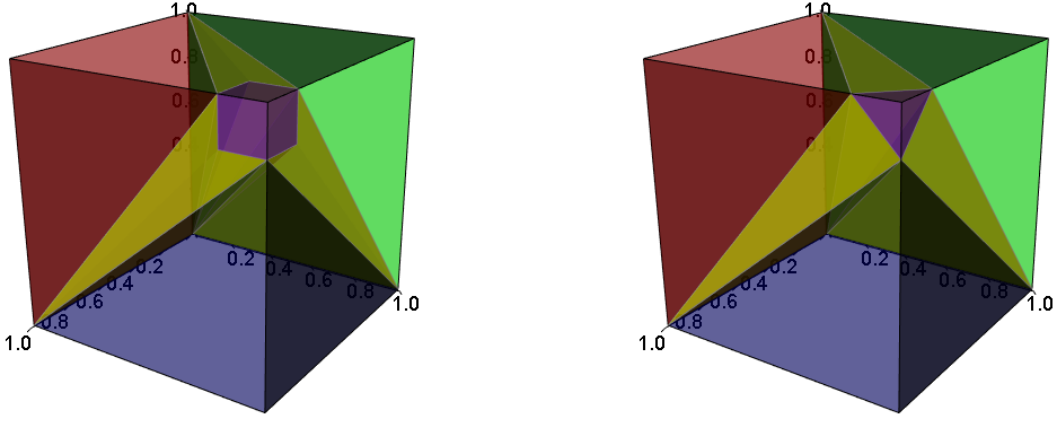
Figure 4: Comparison between ϕ^* (left panel) and ϕ^+ (right panel) when $k = 3$ and $c = 1.2$. Any point in the plot represents (μ_1, μ_2, μ_3) (suppose $Y^*(j) \in (0,1)$) with the recommendation illustrated by colors. As in Figure 1, the red, green, and blue regions contain only one treatment; the yellow region contains two treatments; and the purple region includes all three treatments. $S_1^* = S_1^+$ (unions of red, blue and green regions), and $S_2^* \subseteq S_2^+$ (all but the purple regions).

5   Statistical Learning Theory

In this section, we study the convergence rate of the excess ℓ-risk in both linear learning and kernel learning settings. We assume the random vector $\boldsymbol{Z} = (\boldsymbol{X}, A, Y)$ follows a certain distribution P that satisfies Assumption 1. Furthermore, we make an additional assumption.

Assumption 3. *There is a constant $C > 0$ such that $|Y/p(A|\boldsymbol{X})| \leq C$ holds. For simplicity, we set $C = 1$ through out this section.*

For $\boldsymbol{f}$ and $\boldsymbol{f}'$, two $(k-1)$-dimensional functions, and ℓ, an increasing, convex and Lipchitz loss

function, denote

$$e_\ell(\boldsymbol{f}, \boldsymbol{f}') = \mathbb{E}\left[\frac{Y}{p(A|\boldsymbol{X})}\ell(\langle \boldsymbol{W}_A, \boldsymbol{f}\rangle)\right] - \mathbb{E}\left[\frac{Y}{p(A|\boldsymbol{X})}\ell(\langle \boldsymbol{W}_A, \boldsymbol{f}'\rangle)\right].$$

We call $e_\ell(\boldsymbol{f}, \boldsymbol{f}^*)$ the excess ℓ-risk of $\boldsymbol{f}$ if $\boldsymbol{f}^*$ is optimal within a certain function space $\mathcal{F}$.

5.1 Linear Learning

We first consider the linear function space, that is, we assume $\boldsymbol{f} = (f_1, \dots, f_{k-1})^T$ with $f_j(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}_j$ for $j = 1, \dots, k-1$. For simplicity, we assume each covariate is bounded by $[0, 1]$.

Assumption 4. $\boldsymbol{X} \in \mathcal{X} = [0, 1]^p$.

Now consider the following function space,

$$\mathcal{F}(p, s) = \{\boldsymbol{f} = (f_1, \dots, f_{k-1})^T; f_j(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}_j, j = 1, \dots, k-1, J(\boldsymbol{f}) \leq s\},$$

where $J(\boldsymbol{f}) = \sum_{j=1}^{k-1}||\boldsymbol{\beta}_j||_2^2 = \sum_{j=1}^{k-1}\sum_{l=1}^{p}\beta_{lj}^2$. Let $\mathcal{F}(p) = \cup_{0 \leq s < \infty}\mathcal{F}(p, s)$. Define

$$\boldsymbol{f}^{(p)} = \underset{\boldsymbol{f} \in \mathcal{F}(p)}{\operatorname{argmin}}\, \mathbb{E}\left[\frac{Y}{p(A|\boldsymbol{X})}\ell(\langle \boldsymbol{W}_A, \boldsymbol{f}\rangle)\right],$$

$$\boldsymbol{f}^{(p,s)} = \underset{\boldsymbol{f} \in \mathcal{F}(p,s)}{\operatorname{argmin}}\, \mathbb{E}\left[\frac{Y}{p(A|\boldsymbol{X})}\ell(\langle \boldsymbol{W}_A, \boldsymbol{f}\rangle)\right], \quad \text{and}$$

$$\hat{\boldsymbol{f}} = \underset{\boldsymbol{f} \in \mathcal{F}(p,s)}{\operatorname{argmin}}\, \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p(a_i|\boldsymbol{x}_i)}\ell(\langle \boldsymbol{W}_{a_i}, \boldsymbol{f}(\boldsymbol{x}_i)\rangle).$$

Theorem 1 gives the convergence rate for the excess ℓ-risk $e_\ell(\hat{\boldsymbol{f}}, \boldsymbol{f}^{(p)})$, where $p = p_n$, $s = s_n$ can grow with n as $n \to \infty$.

Theorem 1. *Let* $\tau_n = (n^{-1}\ln p_n)^{1/2} \to 0$ *as* $n \to \infty$. *For linear learning, suppose Assumptions 1, 2, 3, and 4 hold. We have*

$$e_\ell(\hat{\boldsymbol{f}}, \boldsymbol{f}^{(p_n)}) = O\left(\max(c(p_n s_n)^{1/2}\tau_n \ln \tau_n^{-1}, \delta_n)\right),$$

almost surely under P, *where* $\delta_n = e_\ell(\boldsymbol{f}^{(p_n, s_n)}, \boldsymbol{f}^{(p_n)})$.

In Theorem 1, δ_n stands for the approximation error between the optimal $\boldsymbol{f}$ in $\mathcal{F}(p_n, s_n)$ and the optimal $\boldsymbol{f}$ in $\mathcal{F}(p_n)$. So if $s_n \to \infty$, δ_n converges to 0. On the other hand, the first term $O\left(c(p_n s_n)^{1/2}\tau_n \ln \tau_n^{-1}\right)$ is the estimation error between $\hat{\boldsymbol{f}}$ and $\boldsymbol{f}^{(p_n, s_n)}$, and as we increase s_n, it becomes larger. The optimal tuning parameter s_n is then chosen such that $c(p_n s_n)^{1/2}\tau_n \ln \tau_n^{-1} \sim \delta_n$.

In Theorem 1 we may allow $s_n \to \infty$ with an appropriately chosen rate. The reason is that when we include diverging number of covariates, i.e., $p \to \infty$, $\boldsymbol{f}^{(p)}$ can become more complicated, and thus we need a larger s_n to accommodate this change. However, in practice it may not be necessary since the true model usually depends on a finite number of covariates. So we could simplify Theorem 1 if we make the assumption that there is a finite s^* such that $\boldsymbol{f}^{(p)} \in \mathcal{F}(p, s^*)$ for all p. For example, suppose $f_j^*(\boldsymbol{x}) = f_j^{(p)}(\boldsymbol{x}) = \sum_{l=1}^m x_l \beta_{lj}^*$ for $j = 1, \ldots, k-1$ and all $p = m, \ldots, \infty$. Then we can choose $s^* = \sum_{j=1}^{k-1} \sum_{l=1}^m |\beta_{lj}^*|^2$.

Corollary 1. *Suppose $\boldsymbol{f}^*$ defined in (6) only depends on finite many covariates, and that Assumptions 1, 2, 3, and 4 hold. We have*

$$e_\ell(\hat{\boldsymbol{f}}, \boldsymbol{f}^*) = O\left(c p_n^{1/2} \tau_n \ln \tau_n^{-1}\right) = O\left(c(n^{-1} p_n \ln p_n)^{1/2} \ln(n(\ln p_n)^{-1})\right),$$

almost surely under P.

The convergence of excess ℓ-risk $e_\ell(\hat{\boldsymbol{f}}, \boldsymbol{f}^*)$ in Corollary 1 requires that $p_n = o(n)$. Particularly, when p_n grows no faster than n^{1-r}, where $0 < r < 1$, it can be verified that the error rate is at an order of no greater than $n^{-r/2}(\ln n)^{3/2}$. This result is consistent with most of the classical asymptotic theory that the dimension of covariates should not be greater than the number of observations. Furthermore, we observe that if $p_n = O(1)$, then $e_\ell(\hat{\boldsymbol{f}}, \boldsymbol{f}^*) = O(n^{-1/2} \ln n)$, which is almost $O(n^{-1/2})$.

5.2   Kernel Learning

Next we discuss the convergence rate of excess ℓ-risk for kernel learning. We denote $\boldsymbol{f} = (f_1, \ldots, f_{k-1})^T$ to be a function in a reproducing kernel Hilbert space (RKHS) H with kernel function $K(\cdot, \cdot)$. Then by the RKHS theory, we can write $f_j(\boldsymbol{x}) = \sum_{i=1}^n K(\boldsymbol{x}_i, \boldsymbol{x}) \alpha_{ij} + \alpha_{0j}$ for $j = 1, \ldots, k-1$. To develop the theory for the proposed methods, we still need one more assumption.

Assumption 5. *Suppose H is a separable RKHS equipped with kernel function $K(\cdot, \cdot)$. There exists a positive number B, such that $K(\boldsymbol{x}, \boldsymbol{x}') \leq B$ for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$.*

Assumption 5 states that the RKHS is separable and the kernel function is bounded. This is true for many commonly used kernel functions. For example, for the Gaussian kernel, we may

20

take $B = 1$. We define the function space as

$$\mathcal{F}(n, s) = \{\boldsymbol{f} = (f_1, \ldots, f_{k-1})^T; f_j(\boldsymbol{x}) = \sum_{i=1}^{n} K(\boldsymbol{x}_i, \boldsymbol{x})\alpha_{ij} + \alpha_{0j}, j = 1, \ldots, k-1, J(\boldsymbol{f}) \leq s\},$$

where $J(\boldsymbol{f}) = \sum_{j=1}^{k-1} \boldsymbol{\alpha}_j^T K \boldsymbol{\alpha}_j + \sum_{j=1}^{k-1} \alpha_{0j}^2$ and K is the gram matrix. Recall we have included intercepts in the penalty for simplicity. Let $\mathcal{F}(\infty) = \lim_{n \to \infty} \cup_{0 \leq s < \infty} \mathcal{F}(n, s)$, and define

$$\boldsymbol{f}^{(\infty)} = \operatorname*{argmin}_{\boldsymbol{f} \in \mathcal{F}(\infty)} \mathbb{E}\left[\frac{Y}{p(A|\boldsymbol{X})}\ell(\langle \boldsymbol{W}_A, \boldsymbol{f}\rangle)\right],$$

$$\boldsymbol{f}^{(n,s)} = \operatorname*{argmin}_{\boldsymbol{f} \in \mathcal{F}(n,s)} \mathbb{E}\left[\frac{Y}{p(A|\boldsymbol{X})}\ell(\langle \boldsymbol{W}_A, \boldsymbol{f}\rangle)\right], \quad \text{and}$$

$$\hat{\boldsymbol{f}} = \operatorname*{argmin}_{\boldsymbol{f} \in \mathcal{F}(n,s)} \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p(a_i|\boldsymbol{x}_i)}\ell(\langle \boldsymbol{W}_{a_i}, \boldsymbol{f}(\boldsymbol{x}_i)\rangle),$$

The following theorem gives the convergence rate of $e_\ell(\hat{\boldsymbol{f}}, \boldsymbol{f}^{(\infty)})$ when $s = s_n$ grows with n.

Theorem 2. *For RKHS learning, suppose Assumptions 1, 2, 3, and 5 hold. We have*

$$e_\ell(\hat{\boldsymbol{f}}, \boldsymbol{f}^{(\infty)}) = O\left(\max(cB(s_n/n)^{1/2}\ln n, \delta_n)\right),$$

almost surely under P, where $\delta_n = e_\ell(\boldsymbol{f}^{(n,s_n)}, \boldsymbol{f}^{(\infty)})$.

Similar to the linear case, there is a trade-off between the approximation error δ_n and the estimation error $O(cB(s_n/n)^{1/2}\ln n)$ in Theorem 2, and the optimal tuning parameter s_n is determined roughly when $cB(s_n/n)^{1/2}\ln n \sim \delta_n$.

Compared to Theorem 1, the excess ℓ-risk for RKHS learning seems to yield a faster rate. However, this is not always truely the case due to Assumption 5, which requires a bounded kernel function, and implies a restriction on the number of covariates p. For example, for linear kernel we have $K(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{x}^T\boldsymbol{x}' \leq p$ under Assumption 4. For Assumption 5 to be true, we have to let $p = O(1)$. In this case both convergence rates are $e_\ell(\hat{\boldsymbol{f}}, \boldsymbol{f}^{(p)}) = O((s_n/n)^{1/2}\ln n)$; that of the kernel learning is no faster than that of the linear learning. In general, to obtain a faster rate than that of the linear learning, we need a kernel function that does not increase in p, such as the Gaussian kernel.

Note that the approximation error δ_n converges to 0 as n increases, and both the convergence rate of δ_n and that of the resulting $e_\ell(\hat{\boldsymbol{f}}, \boldsymbol{f}^*)$ depend on the choice of the kernel. To illustrate the magnitude of δ_n and its impact on the excess risk, consider a binary example where $X \sim \text{Unif}(0, 1)$ and $f^*(x) = (1 + x)^2$. With the polynomial kernel of degree 2 we have $f^{(\infty)} = f^*$ and

$B = \max_{x,x'}(1 + xx')^2 = 2$. Given a training set $\{x_1, \ldots, x_n\}$, let $x_{(n)}$ be the largest order statistic and define $f_{(n)}(x) = \left(1 + xx_{(n)}\right)^2$. It can be shown that for any $s_n \geq 1$, $f_{(n)} \subseteq \mathcal{F}(n, s_n)$ thus $\delta_n = e_\ell(f^{(n,s_n)}, f^{(\infty)}) \leq e_\ell(f_{(n)}, f^{(\infty)}) \leq c\mathbb{E}||f_{(n)} - f^{(\infty)}||_2$. Note that the difference between $f_{(n)}$ and $f^{(\infty)}$ is maximized at 1, so $\delta_n \leq c\mathbb{E}|f_{(n)}(1) - f^{(\infty)}(1)| = c\mathbb{E}\left(3 - 2x_{(n)} - x_{(n)}^2\right)$. Because the density function of $x_{(n)}$ is $nx^{n-1}\mathbf{1}_{(0,1)}$, we have $\delta_n \leq c\int_0^1 (3 - 2x - x^2)nx^{n-1}dx = \frac{2c(2n+3)}{(n+1)(n+2)}$. Hence in this example, the order of δ_n is at most $O(n^{-1})$, thus $e_\ell(\hat{f}, f^*) = e_\ell(\hat{f}, f^{(\infty)}) = O(n^{-1/2}\ln n)$.

6 Simulation Studies

In this section, we study the numerical performance of the proposed methods.

6.1 Comparing Set-valued Recommendations

For two ITRs d_1 and d_2, we can compare them by evaluating the expected outcome defined in (1). However, for two A-ITRs ϕ_1 and ϕ_2, it is difficult to quantify which one is better due to the fact that a measure for the overall benefit is not well defined when multiple treatments are recommended. Although in Section 4.2 we have proposed the weighted expected outcome (12) for evaluating two A-ITRs, the optimal A-ITR ϕ^+ under this new criterion is still different from the desired near-optimal recommendation set ϕ^*. So in the simulation studies, in addition to the empirical weighted outcome (13), we consider another means to compare different A-ITRs, using the expected outcome of the best and the worst treatments among the treatments that are recommended, averaged over a set of observations. We conduct such an evaluation for different types of recommendations separately to see how the A-ITR performs differently on them. Based on the size of the true optimal A-ITR ϕ^*, we split the covariate space $\mathcal{X}$ into three regions corresponding to three kinds of recommendations:

$$R_1 = \{\text{only one treatment is suggested}\} = \{\boldsymbol{x} \in \mathcal{X}; |\phi^*(\boldsymbol{x})| = 1\},$$

$$R_2 = \{\text{more than one treatment but not all of them are suggested}\}$$

$$= \{\boldsymbol{x} \in \mathcal{X}; 1 < |\phi^*(\boldsymbol{x})| < k\},$$

$$R_3 = \{\text{all treatments are suggested}\} = \{\boldsymbol{x} \in \mathcal{X}; |\phi^*(\boldsymbol{x})| = k\}.$$

Note that R_1, R_2 and R_3 are disjoint and $\mathcal{X} = R_1 \cup R_2 \cup R_3$. When $c = 1$, ϕ^* is the optimal ITR d^* and $\mathcal{X} = R_1$. When $c > 1$, we may have non-empty regions R_2 and R_3.

For two A-ITRs ϕ_1 and ϕ_2, we will compare them separately on R_1, R_2 and R_3. In each region, since multiple treatments may be suggested, we can compare the expected minimal outcome and the expected maximal outcome that they may lead to. Recall $Y^*(j)$ is the potential outcome by taking treatment j. Mathematically, we consider a performance interval,

$$\left(\mathbb{E} \left[\mathbb{E} \left(\min_{j \in \phi(\mathbf{X})} Y^*(j) | \mathbf{X} \right) \right], \mathbb{E} \left[\mathbb{E} \left(\max_{j \in \phi(\mathbf{X})} Y^*(j) | \mathbf{X} \right) \right] \right),$$

where the first quantity indicates the expected outcome if one can always use the best treatment within the recommended set $\phi(\mathbf{x})$ and the second quantity represents the worst situation, i.e., how bad it can be if one always chooses the worst treatment among the recommended options. Note that on R_1, the two quantities are the same under ϕ^* since only one treatment is recommended. As we increase c, we expect that this interval becomes wider on R_2 and R_3 since the diversity of the recommended options increases. From the definition of this interval, we claim that ϕ_1 is better than ϕ_2 if both the lower and the upper limits of this interval under ϕ_1 are smaller than their counterparts under ϕ_2, on each region.

6.2 Results

To compare the performance of our proposed methods, we consider three simulation examples. For each example, we consider two different dimensions with $\mathbf{X}$ uniformly sampled from $\mathcal{X} = [0, 1]^5$ and $\mathcal{X} = [0, 1]^{10}$. For simplicity, we assume $A \perp \mathbf{X}$ and $p(A|\mathbf{X}) = 1/k$, and let $Y = \mu_A(\mathbf{X}) + \epsilon$ where $\epsilon \sim N(0, 1/2)$. In each case, we first generate a training sample with sample size 2000 to fit the model, then use a test set with sample size 1000 to evaluate the performance. We compare three methods, namely, the regression-based method, the two-step classification-based method with squared loss, and the one-step classification-based method with the bent hinge loss. For each method, we output both ITR and A-ITR with $c = 1.2$. Finally, we repeat each simulation for 100 times and report the averages.

Example 1: This is an example with three treatments, where two conditional mean outcome functions are polynomial and the other is linear. Specifically, we have $\mu_1(\mathbf{X}) = 1 + 3X_1^2 + 3X_2^2$, $\mu_2(\mathbf{X}) = 3 - 0.5X_1^2 + 0.5X_2^2$, and $\mu_3(\mathbf{X}) = 3 + X_1 + X_2$. The upper panel in Figure 5 shows the

23

true boundaries for the three treatments. We report the results using polynomial kernel for both the two-step and one-step methods. The tuning parameter λ is chosen from 5^{-6} to 5^2.
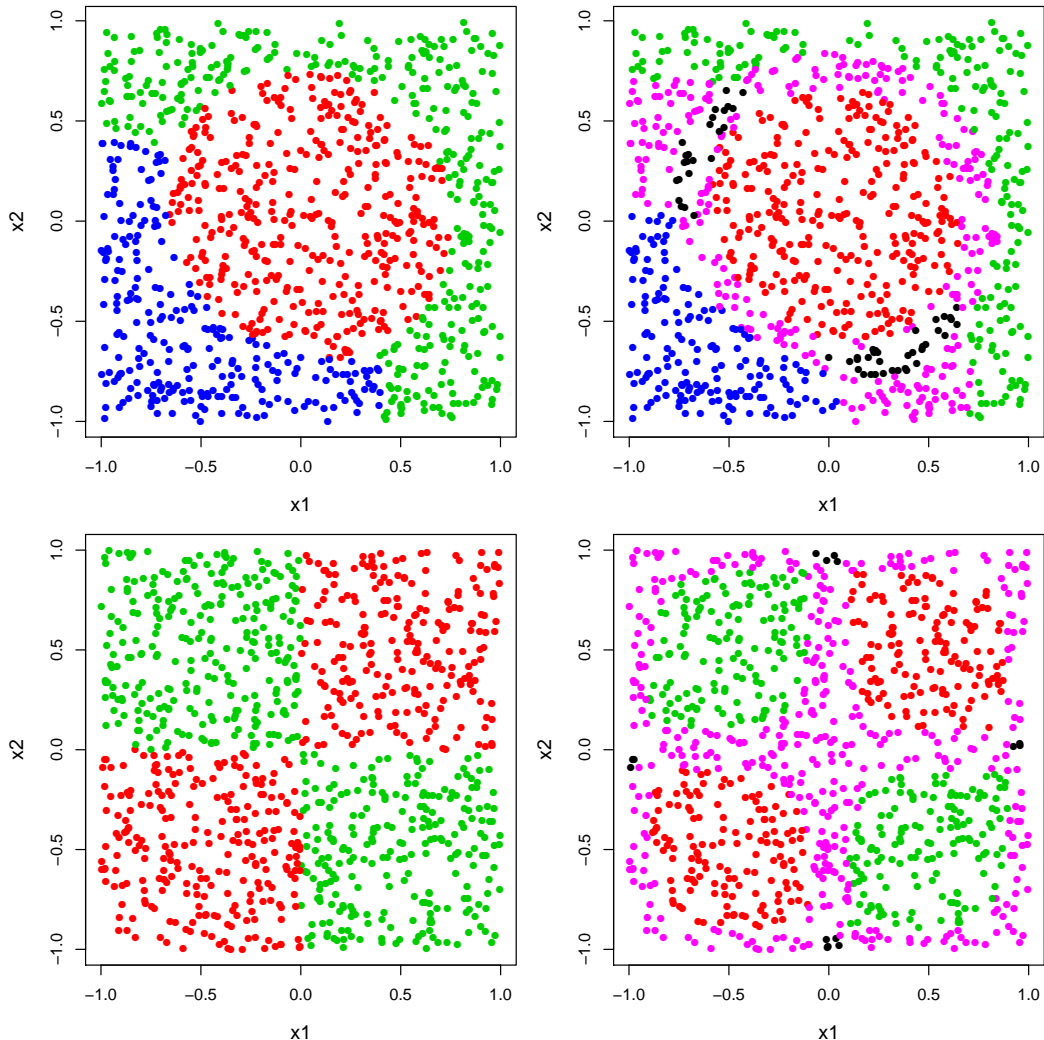

Figure 5: The true ITR and A-ITR for observations in Examples 1 and 2 with recommendation types shown in colors. The red, green and blue regions indicate that only one single treatment is suggested, which by definition is R_1, while the pink region is R_2 and the black region is R_3. The upper panel is Example 1, the lower panel Example 2. The left panel is ITR, and the right panel is A-ITR ($c = 1.2$).

Example 2: This is an example with four treatments, where all the conditional mean outcome functions are non-linear, $\mu_A(\boldsymbol{X}) = 2 + \text{sign}(A - 2.5)\cos\left(0.5\pi(X_1 + (-1)^A X_2)\right)$. Specially, treatments 2 and 4 are dominated by treatments 1 and 3 and the optimal ITR should only output either 1 or 3. However, in certain regions treatments 2 and 4 still produce fairly good outcome which can only be captured by A-ITR. The lower panel in Figure 5 shows the true boundaries.

For the two-step method, we report the results using Gaussian kernel. For the one-step method, we report the results with polynomial kernel. The tuning parameter λ is chosen from 5^{-9} to 5^{-1}.

Example 3: This is an example with three treatments and the true conditional mean outcome functions are determined by four covariates. Specifically, $\mu_1(\boldsymbol{X}) = 3 + 3X_1^2 + 3X_2^2 - 0.5\exp(0.5X_3^2 + X_4)$, $\mu_2(\boldsymbol{X}) = 3 - 2X_1^2 + \exp(X_3 + X_4^2)$, and $\mu_3(\boldsymbol{X}) = 3 - X_2^3 - 2X_3^2 + 0.5\left(\exp(X_1 + X_4) - 1\right)^2$. Similar to Example 1, we report the results using polynomial kernel for both the two-step and one-step methods. The tuning parameter λ is chosen from 5^{-7} to 1.

Table 1 collects the results of the three examples with $p = 5$ and 10. In Table 1, the results of A-ITR are in the form of intervals while the results of ITR are single numbers. We also compute the empirical weighted outcome ("All" column in Table 1) defined in (13) as an indicator for the overall performance for each method.

We note that the performance intervals for A-ITR always cover the expected outcomes of the single-valued ITR. This implies that by applying our proposed A-ITR framework, patients will potentially get a much better outcome as long as they are willing to consider other equally effective options identified by the A-ITR. Even if the patient does not choose the best option within the recommended set, the worst case is not too bad and the ratio of its outcome to that of the best option is about c if the A-ITR is accurately estimated.

We compare different methods by inspecting the length and location of the A-ITR performance interval. Recall that the A-ITR with the shortest interval, the smallest lower limit, and the smallest upper limit on each region is the best A-ITR. However, since R_3 is the region where all treatments are near the optimality, different recommendations are expected to perform similarly. Hence we focus on regions R_1 and R_2 for the purpose of comparison.

From Table 1, we note that the regression-based A-ITR, though has the smallest lower limit in some cases, always yields the longest interval, suggesting that the treatment could either go really well or really badly. This implies that the regression-based A-ITR method tends to include ineffective treatments into the near-optimal set. Part of the reason may be that the regression-based method has not accurately estimated each of the three or four potential outcome functions.

For the classification-based A-ITRs, the lower limits are roughly the same between the one-step method and the two-step method; however, the one-step method has shorter interval in most cases. This means that the one-step method is better at excluding ineffective treatment options

Example 1		$p = 5$				$p = 10$			
		$R_1(70.05\%)$	$R_2(24.62\%)$	$R_3(5.33\%)$	All	$R_1(70.05\%)$	$R_2(24.62\%)$	$R_3(5.33\%)$	All
Reg.	ITR	2.47	2.52	2.61	2.49	2.47	2.53	2.62	2.50
	A-ITR	$(1.98, 3.09)$	$(2.42, 2.99)$	$(2.47, 2.96)$	2.37	$(1.99, 3.07)$	$(2.42, 2.99)$	$(2.48, 2.95)$	2.38
2-step	ITR	2.17	2.63	2.68	2.31	2.26	2.66	2.69	2.39
	A-ITR	$(2.03, 2.49)$	$(2.50, 2.83)$	$(2.57, 2.82)$	2.28	$(2.09, 2.52)$	$(2.54, 2.82)$	$(2.60, 2.78)$	2.32
1-step	ITR	2.07	2.59	2.67	2.23	2.21	2.65	2.68	2.34
	A-ITR	$(1.99, 2.23)$	$(2.51, 2.70)$	$(2.61, 2.74)$	**2.19**	$(2.08, 2.42)$	$(2.56, 2.76)$	$(2.62, 2.75)$	**2.30**
Bayes	ITR	1.92	2.44	2.56	2.08				
	A-ITR	1.92	$(2.38, 2.74)$	$(2.46, 2.99)$	2.05				

Example 2		$p = 5$				$p = 10$			
		$R_1(56.80\%)$	$R_2(41.84\%)$	$R_3(1.36\%)$	All	$R_1(56.80\%)$	$R_2(41.84\%)$	$R_3(1.36\%)$	All
Reg.	ITR	1.56	1.62	1.92	1.60	1.56	1.62	1.93	1.60
	A-ITR	$(1.13, 1.99)$	$(1.21, 2.04)$	$(1.78, 2.09)$	1.45	$(1.14, 1.99)$	$(1.21, 2.03)$	$(1.80, 2.09)$	1.46
2-step	ITR	1.24	1.30	1.93	1.28	1.36	1.36	1.93	1.37
	A-ITR	$(1.15, 1.48)$	$(1.22, 1.46)$	$(1.83, 2.06)$	1.26	$(1.23, 1.57)$	$(1.27, 1.54)$	$(1.87, 2.03)$	1.35
1-step	ITR	1.24	1.30	1.93	1.28	1.32	1.34	1.92	1.34
	A-ITR	$(1.16, 1.41)$	$(1.23, 1.39)$	$(1.85, 2.01)$	**1.25**	$(1.20, 1.50)$	$(1.26, 1.45)$	$(1.87, 1.99)$	**1.31**
Bayes	ITR	1.13	1.25	1.87	1.19				
	A-ITR	1.13	$(1.15, 1.46)$	$(1.71, 2.30)$	1.16				

Example 3		$p = 5$				$p = 10$			
		$R_1(58.79\%)$	$R_2(29.93\%)$	$R_3(11.28\%)$	All	$R_1(58.79\%)$	$R_2(29.93\%)$	$R_3(11.28\%)$	All
Reg.	ITR	3.35	3.57	3.85	3.47	3.35	3.59	3.86	3.49
	A-ITR	$(3.01, 3.85)$	$(3.41, 3.82)$	$(3.72, 4.01)$	3.35	$(3.01, 3.87)$	$(3.43, 3.83)$	$(3.72, 4.01)$	3.37
2-step	ITR	3.09	3.53	3.82	3.30	3.23	3.59	3.84	3.41
	A-ITR	$(2.89, 3.51)$	$(3.37, 3.80)$	$(3.68, 4.01)$	3.21	$(2.99, 3.61)$	$(3.44, 3.81)$	$(3.74, 3.96)$	**3.32**
1-step	ITR	3.05	3.50	3.82	3.26	3.23	3.59	3.85	3.41
	A-ITR	$(2.87, 3.43)$	$(3.37, 3.66)$	$(3.73, 3.92)$	**3.19**	$(2.96, 3.74)$	$(3.43, 3.80)$	$(3.75, 3.95)$	3.33
Bayes	ITR	2.72	3.28	3.65	2.99				
	A-ITR	2.72	$(3.23, 3.66)$	$(3.58, 4.19)$	2.92				

Table 1: Results of the simulation studies. In each region, the expected outcome for ITR and the outcome interval for A-ITR ($c = 1.2$) are reported. The empirical weighted outcome defined in (13) is shown in the "All" column. Each number is averaged over 100 replications. In each case, the best performing method is marked in bold.

from the recommendation than the two-step method. In addition, the one-step method also has the smallest expected weighted outcome (shown in the "All" column).

7 Real Data Analysis

In this section, we apply our proposed A-ITR framework to a Type 2 diabetes mellitus (T2DM) observational study. The data set contains 1139 patients. Every patient was assigned with one out of four diabetes treatments, which are GLP-1 receptor agonists alone, long-acting insulin alone, intermediate-acting insulin alone, and insulin regimens including a short-acting insulin. The end point is the change of hemoglobin A1c level before and after the treatment, which is denoted by $\Delta HbA1c$. In practice, if the treatment works, this value is usually negative (meaning that the hemoglobin A1c level decreases). The smaller $\Delta HbA1c$ is, the more effective the treatment is.

We first preprocess the original data set. Among the 19 covariates, we exclude those with large proportion of missing values and with extremely imbalanced categories. We then impute the rest of them using the predictive mean matching method (Van Buuren, 2018). There are 10 covariates left after the preprocessing: gender, diabetic retinopathy, diabetic neuropathy, age, weight, body mass index (BMI), baseline hemoglobin A1c level, baseline high-density lipoprotein cholesterol (HDL), baseline low-density lipoprotein cholesterol (LDL), and heart disease.

For the outcome variable $\Delta HbA1c$, we can reduce its variability by subtracting an estimate of its conditional mean $\mathbb{E}(\Delta HbA1c|\boldsymbol{x})$ to make the estimation of $\hat{\boldsymbol{f}}$ more robust (Liu et al., 2016; Zhou et al., 2017). Here we use the ordinary least square regression to estimate $\mathbb{E}(\Delta HbA1c|\boldsymbol{x})$. Denote the estimated mean function fitted by regression as $\hat{m}(\boldsymbol{x})$, we then observe that $\Delta HbA1c - \hat{m}(\boldsymbol{x})$ can be positive or negative. We perform an exponential transformation to make it positive, which also justifies the use of ratio $\mu_j/\mu_{(1)}$ to determine the near-optimal recommendation set. Specifically, we let $Y = \exp\left((\Delta HbA1c - \hat{m}(\boldsymbol{x}))/5\right)$. If we further assume conditional normality for $\Delta HbA1c$ given $\boldsymbol{X}$ and treatment j, with mean $\nu_j(\boldsymbol{x}) \equiv \mathbb{E}(\Delta HbA1c|\boldsymbol{X} = \boldsymbol{x}, A = j)$ and equal variance across treatments, then $Y|(\boldsymbol{x}, j)$ follows a log-normal distribution with mean proportional

to $\exp(\nu_j(\boldsymbol{x})/5)$. Then the optimal A-ITR is,

$$\phi^*(\boldsymbol{x}) = \left\{ j; \frac{\mu_j(\boldsymbol{x})}{\mu_{(1)}(\boldsymbol{x})} = \frac{\mathbb{E}(Y|\boldsymbol{X} = \boldsymbol{x}, A = j)}{\min_i \mathbb{E}(Y|\boldsymbol{X} = \boldsymbol{x}, A = i)} = \frac{\exp(\nu_j(\boldsymbol{x})/5)}{\exp(\nu_{(1)}(\boldsymbol{x})/5)} \leq c \right\}$$
$$= \{ j; \nu_j(\boldsymbol{x}) - \nu_{(1)}(\boldsymbol{x}) \leq 5 \log c \}.$$

In this study, we choose the near-optimal parameter $c = 1.2$, so that $5 \log c \approx 0.9$. This implies that the near-optimal recommendation set is constructed by including all treatments with conditional means $\Delta HbA1c$ within 0.9 of the optimal treatment.

We compare performance of the regression-based method, the two-step method, and the one-step method. For both classification-based methods, we estimate the propensity score $p(A|\boldsymbol{X})$ using logistic regression. Each method leads to a single-valued ITR and a set-valued A-ITR, and we compare the different recommendations using the 5-fold cross-validated empirical weighted outcome defined in (13), shown in Table 2.

	ITR	A-ITR
Regression	1.071	0.988
	(0.010)	(0.006)
Two-step: Linear	0.995	0.975
	(0.007)	(0.006)
Two-step: Gaussian	0.959	0.947
	(0.006)	(0.005)
One-step: Linear	1.150	1.033
	(0.008)	(0.006)
One-step: Gaussian	**0.939**	**0.935**
	(0.006)	(0.007)

Table 2: The mean 5-fold cross validated weighted outcome and its standard error (in the parenthesis) over 100 replications for T2DM data. The method that yields the best result is marked in bold.

From Table 2, we observe that the one-step method with Gaussian kernel has the best weighted outcome. To illustrate the resultant A-ITR, we split the data into training set (70%) and test set (30%). We fit the training set using one-step method with Gaussian kernel and then construct the recommendation set for patients in the test set. In our analysis, no patient is recommended to take the intermediate-acting insulin and the majority of patients are recommended to choose between the short-acting insulin and GLP-1. Specifically, 55% of patients are recommended the

short-acting insulin only, 8% are recommended GLP-1 only, and 24% are recommended to take either one of the two. For the rest 13% of patients, they are all recommended to take the long-acting insulin, including 1% who are suggested to take either long-acting insulin or GLP-1, 5% who are suggested to take either long-acting insulin or short-acting insulin, and 7% whose only option is long-acting insulin. We visualize the predicted treatments in Figure 6.
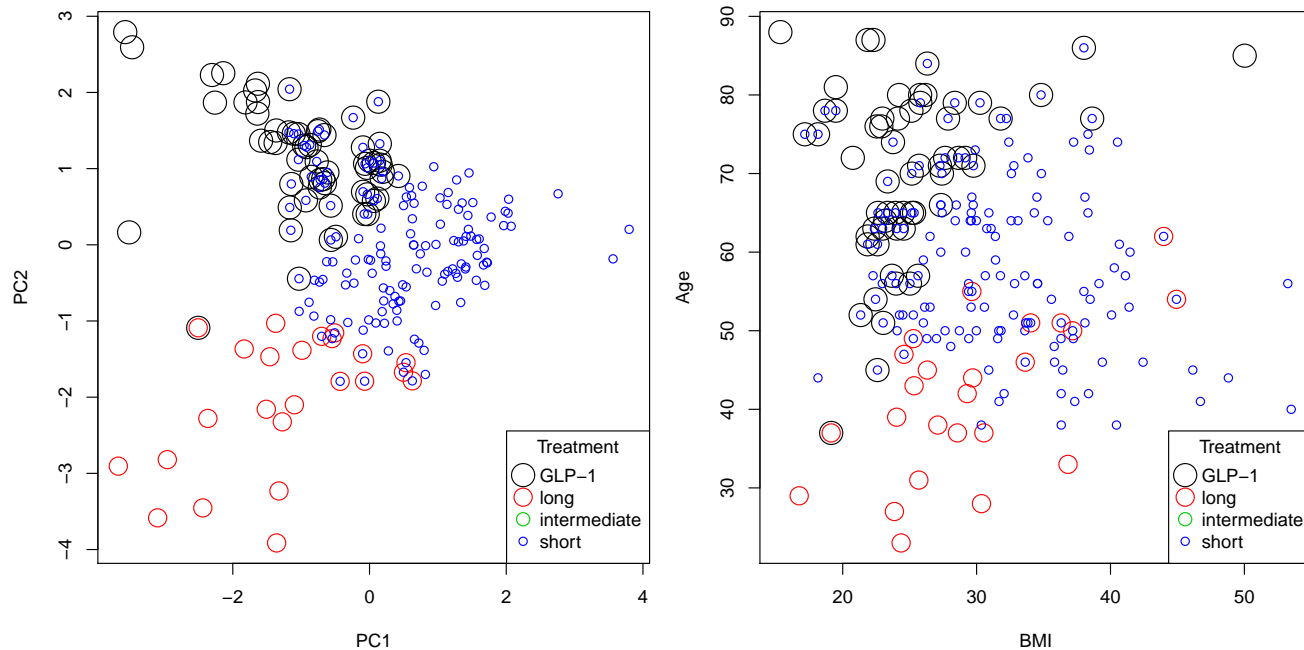

Figure 6: Predicted treatment(s) for patients with recommendations given by different colors. The data set is projected on the first two principle components (left panel), and two particular covariates, age and BMI (right panel). Concentric circles indicate multiple treatments recommended to the same patient.

From Figure 6, we can see that age and BMI are two useful biomarkers in constructing the near-optimal recommendation set. In fact, by comparing the left panel and the right panel of Figure 6, we observe that BMI behaves like the first principle component (PC1) while age behaves like the second principle component (PC2). Figure 6 suggests that for patients without obesity (BMI less than 30), younger patients should take the long-acting insulin while older patients should take GLP-1. The short-acting insulin, on the other hand, serves as an "universal" treatment that many patients can take as an alternative, and is especially effective for overweighted patients.

8 Conclusion

In this work, we propose a new individualized treatment recommendation framework, named A-ITR, that has the capacity to recommend to patients near-optimal treatment options in terms of their clinical outcomes. By adopting the A-ITR, patients have the opportunity to choose the treatment options tailored for their different financial situations, personal preference and life style choices. To estimate the optimal A-ITR, we proposed two classification-based methods based on the OWL framework. We also provide a new evaluation criterion suitable for A-ITRs, namely the weighted expected outcome, defined in (12). The simulation study shows the usefulness of this new criterion in parameter tuning and model selection.

There are several possible directions for future works. Firstly, the current A-ITR estimation is based on OWL framework, which may be sensitive to the estimated propensity score. In this case, one may consider applying the doubly-robust OWL framework (Zhao et al., 2019; Huang et al., 2019) to improve the efficiency of the estimated A-ITR. Secondly, we can consider A-ITRs with an additional competing outcome as a secondary end point (Laber et al., 2014), or A-ITR with additional safety end points formulated as constraints (Wang et al., 2018). Thirdly, we can consider other learning algorithms to estimate $\hat{\boldsymbol{f}}$ within the A-ITR framework. Finally, we can consider nontrivial extensions to other types of outcome such as survival outcome (Zhao et al., 2014; Qi et al., 2019) or dichotomous outcome (Qi et al., 2019; Klausch et al., 2018).

References

Bartlett, P. L. and Wegkamp, M. H. (2008), "Classification with a reject option using a hinge loss," *Journal of Machine Learning Research*, 9, 1823–1840.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. (2011), "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, 3, 1–122.

Chen, J., Fu, H., He, X., Kosorok, M. R., and Liu, Y. (2018), "Estimating individualized treatment rules for ordinal treatments," *Biometrics*, 74, 924–933.

Chow, C. (1970), "On optimum recognition error and reject tradeoff," *IEEE Transactions on information theory*, 16, 41–46.

Cui, Y., Zhu, R., Kosorok, M., et al. (2017), "Tree based weighted learning for estimating individualized treatment rules with censored data," *Electronic journal of statistics*, 11, 3927–3953.

Doubleday, K., Zhou, H., Fu, H., and Zhou, J. (2018), "An Algorithm for Generating Individualized Treatment Decision Trees and Random Forests," *Journal of Computational and Graphical Statistics*, 27, 849–860.

Fung, G. M. and Mangasarian, O. L. (2005), "Multicategory proximal support vector machine classifiers," *Machine learning*, 59, 77–97.

Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer series in statistics, Springer.

Herbei, R. and Wegkamp, M. H. (2006), "Classification with reject option," *Canadian Journal of Statistics*, 34, 709–721.

Hofmann, T., Schölkopf, B., and Smola, A. J. (2008), "Kernel methods in machine learning," *The annals of statistics*, 1171–1220.

Huang, X., Goldberg, Y., and Xu, J. (2019), "Multicategory individualized treatment regime using outcome weighted learning," *Biometrics*, 75, 1216–1227.

Insel, T. R. (2009), "Translating scientific opportunity into public health impact: a strategic plan for research on mental illness," *Archives of general psychiatry*, 66, 128–133.

Kallus, N. (2016), "Learning to personalize from observational data," *arXiv preprint arXiv:1608.08925*.

Klausch, T., van de Ven, P., van de Brug, T., Brakenhoff, R. H., van de Wiel, M. A., and Berkhof, J. (2018), "Estimating Bayesian Optimal Treatment Regimes for Dichotomous Outcomes using Observational Data," *arXiv preprint arXiv:1809.06679*.

Laber, E. and Zhao, Y. (2015), "Tree-based methods for individualized treatment regimes," *Biometrika*, 102, 501–514.

Laber, E. B., Lizotte, D. J., and Ferguson, B. (2014), "Set-valued dynamic treatment regimes for competing outcomes," *Biometrics*, 70, 53–61.

Lesko, L. (2007), "Personalized medicine: elusive dream or imminent reality?" *Clinical Pharmacology & Therapeutics*, 81, 807–816.

Liu, Y., Wang, Y., Kosorok, M. R., Zhao, Y., and Zeng, D. (2016), "Robust hybrid learning for estimating personalized dynamic treatment regimens," *arXiv preprint arXiv:1611.02314*.

— (2018), "Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens," *Statistics in medicine*, 37, 3776–3788.

Lizotte, D. J. and Laber, E. B. (2016), "Multi-objective Markov decision processes for data-driven decision support," *The Journal of Machine Learning Research*, 17, 7378–7405.

Qi, Z., Liu, D., Fu, H., and Liu, Y. (2019), "Multi-Armed Angle-Based Direct Learning for Estimating Optimal Individualized Treatment Rules With Various Outcomes," *Journal of the American Statistical Association*, 1–33.

Qian, M. and Murphy, S. A. (2011), "Performance guarantees for individualized treatment rules," *Annals of statistics*, 39, 1180.

Robins, J. M. (2004), "Optimal structural nested models for optimal sequential decisions," in *Proceedings of the second seattle Symposium in Biostatistics*, Springer, pp. 189–326.

Schulte, P. J., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2014), "Q-and A-learning methods for estimating optimal dynamic treatment regimens," *Statistical science: a review journal of the Institute of Mathematical Statistics*, 29, 640.

Steinwart, I., Scovel, C., et al. (2007), "Fast rates for support vector machines using Gaussian kernels," *The Annals of Statistics*, 35, 575–607.

Van Buuren, S. (2018), *Flexible imputation of missing data*, Chapman and Hall/CRC.

Wang, Y., Fu, H., and Zeng, D. (2018), "Learning optimal personalized treatment rules in consideration of benefit and risk: with an application to treating type 2 diabetes patients with insulin therapies," *Journal of the American Statistical Association*, 113, 1–13.

Wu, P., Zeng, D., and Wang, Y. (2019), "Matched Learning for Optimizing Individualized Treatment Strategies Using Electronic Health Records," *Journal of the American Statistical Association*, 1–23.

Yuan, M. (2015), "Outcome weighted learning with a reject option," in *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine*, SIAM, chap. 14, pp. 239–248.

Yuan, M. and Wegkamp, M. (2010), "Classification methods with reject option based on convex risk minimization," *Journal of Machine Learning Research*, 11, 111–130.

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

Zhang, C., Chen, J., Fu, H., He, X., Zhao, Y., and Liu, Y. (2018a), "Multicategory Outcome Weighted Margin-based Learning for Estimating Individualized Treatment Rules," *Statistica sinica*.

Zhang, C. and Liu, Y. (2014), "Multicategory angle-based large-margin classification," *Biometrika*, 101, 625–640.

Zhang, C., Liu, Y., and Wu, Y. (2016), "On quantile regression in reproducing kernel Hilbert spaces with the data sparsity constraint," *The Journal of Machine Learning Research*, 17, 1374–1418.

Zhang, C., Wang, W., and Qiao, X. (2018b), "On Reject and Refine Options in Multicategory Classification," *Journal of the American Statistical Association*, 113, 730–745.

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, 1106–1118.

Zhao, Y.-Q., Laber, E. B., Ning, Y., Saha, S., and Sands, B. E. (2019), "Efficient augmentation and relaxation learning for individualized treatment rules using observational data." *Journal of Machine Learning Research*, 20, 1–23.

Zhao, Y.-Q., Zeng, D., Laber, E. B., Song, R., Yuan, M., and Kosorok, M. R. (2014), "Doubly robust learning for estimating individualized treatment with censored data," *Biometrika*, 102, 151–168.

Zhou, X. and Kosorok, M. R. (2017), "Causal nearest neighbor rules for optimal treatment regimes," *arXiv preprint arXiv:1711.08451.*

Zhou, X., Mayer-Hamblett, N., Khan, U., and Kosorok, M. R. (2017), "Residual weighted learning for estimating individualized treatment rules," *Journal of the American Statistical Association*, 112, 169–187.

Zhu, R., Zhao, Y.-Q., Chen, G., Ma, S., and Zhao, H. (2017), "Greedy outcome weighted tree learning of optimal personalized treatment rules," *Biometrics*, 73, 391–400.