| Statistica Sinica Preprint No: SS-2017-0527 | |
|---|---|
| **Title** | Multicategory Outcome Weighted Margin-based Learning for Estimating Individualized Treatment Rules |
| **Manuscript ID** | SS-2017-0527 |
| **URL** | http://www.stat.sinica.edu.tw/statistica/ |
| **DOI** | 10.5705/ss.202017.0527 |
| **Complete List of Authors** | Chong Zhang<br>Jingxiang Chen<br>Haoda Fu<br>Xuanyao He<br>Ying-Qi Zhao and<br>Yufeng Liu |
| **Corresponding Author** | Yufeng Liu |
| **E-mail** | yfliu@email.unc.edu |

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527

Statistica Sinica

# Multicategory Outcome Weighted Margin-based Learning for Estimating Individualized Treatment Rules

Chong Zhang[1], Jingxiang Chen[1], Haoda Fu[2], Xuanyao He[2], Ying-Qi Zhao[3], and Yufeng Liu[1]

[1] *University of North Carolina at Chapel Hill,* [2] *Eli Lilly and Company, and* [3] *Fred Hutchinson Cancer Research Center*

*Abstract:* Owing to the heterogeneity exhibited by many chronic diseases, precise personalized medicine, also known as precision medicine, has garnered increased attention in the scientific community. One main goal of precision medicine is to develop the most effective tailored therapy for each individual patient. To this end, one needs to incorporate individual characteristics to determine a proper individual treatment rule (ITR), which is used to make suitable decisions on treatment assignments that optimize patients' clinical outcomes. For binary treatment settings, outcome-weighted learning (OWL) and several of its variations have been proposed to estimate an ITR by optimizing the conditional expected outcome, given patients' information. However, for multiple treatment scenarios, it remains unclear how to use OWL effectively. It can be shown that some direct extensions of OWL for multiple treatments, such as the one-versus-one and one-versus-rest methods, can yield suboptimal performance. In this paper, we propose a new learning method, called multicategory outcome-weighted margin-based learning (MOML), for estimating an ITR with multiple treatments. Our

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



proposed method is very general and covers OWL as a special case. We show the Fisher consistency of the estimated ITR, and establish its convergence rate properties. Variable selection using the sparse $l_1$ penalty is also considered. Simulations and a type-2 diabetes mellitus observational study are used to demonstrate the competitive performance of the proposed method.

*Key words and phrases:* Angle-based Classifier, Large-margin, Multiple Treatments, Outcome Weighted Learning, Precision Medicine, Support Vector Machine.

## 1. Introduction

An important goal of precision medicine is to develop effective statistical methods for evaluating treatments with heterogeneous effects among patients. In particular, a treatment that works for patients with certain characteristics may not be effective for others (Simoncelli, 2014). A popular method of maximizing the overall benefits that patients receive from a recommended therapy involves identifying proper individual treatment rules (ITRs), which are functions that map patient characteristics onto the treatment space.

More recently, studies have begun building ITRs for binary treatment cases. For example, Tian et al. (2014) studied the ITR problem and con-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



ducted a subgroup analysis using a regression approach. Qian and Murphy (2011) incorporated ITR detection into an optimization problem, based on a conditional expectation that contains an indicator function. Zhao et al. (2012) used a weighted classification framework and proposed outcome-weighted learning (OWL), which replaces the indicator function with a surrogate loss. Zhou et al. (2017) proposed using the residuals from a linear regression between the outcome and the covariates to improve the finite-sample performance of the method proposed by Zhao et al. (2012). Zhang et al. (2012) proposed a robust ITR method to handle potential regression model misspecification when modeling the outcome.

Despite the successful developments in ITR estimation for binary treatments, how the idea should be adapted to multicategory treatment scenarios requires additional research. In general, some regression-based methods can be applied for this purpose under parametric assumptions, such as certain model mean structures (Robins et al., 2008). However, violating these assumptions can lead to misleading results. In this study, we develop a statistical learning framework for conducting optimal ITR detection for nominal multicategory treatment cases. For simplicity, in the remainder of the paper, we use the term multicategory to represent "nominal multicategory" when there is no confusion.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



In the classification literature, large-margin classifiers are popular and widely used in practice. Well-known examples include the support vector machine (SVM) and penalized logistic regression (PLR) (Hastie et al. (2009)). There are two main types of large-margin classifiers: soft and hard classifiers (Liu et al., 2011). The essential difference is whether obtaining the classifier requires estimating the conditional probability of each class. Soft classifiers, such as the PLR, estimate the class conditional probability, whereas hard classifiers, such as the SVM, target the classification boundary only. Liu et al. (2011) showed that the performance of soft and hard classifiers can vary for problems with different settings. In addition, they proposed the large-margin unified machine (LUM) loss family, which includes both soft and hard classifiers by using a tuning parameter, and works well for different problems.

To solve $k$-class multicategory problems, one direct approach uses sequential binary classifiers. In particular, there are two common approaches in the literature, namely, the one-versus-one and one-versus-rest approaches (Allwein et al., 2001). However, these sequential binary classifiers can be suboptimal. A common approach handling a $k$-class problem simultaneously is to estimate $k$ functions with the sum-to-zero constraint (Lee et al., 2004; Liu and Yuan, 2011; Zhang and Liu, 2013). Recently, Zhang and

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



Liu (2014) pointed out that this approach can be inefficient because one needs to add an extra sum-to-zero constraint to the optimization problem to guarantee the identifiability and desirable properties of the classifiers. In this way, an extra computational cost is incurred when solving the corresponding constrained optimization problem. To overcome this drawback, Zhang and Liu (2014) proposed an angle-based large-margin classification technique using $k-1$ functions, without the sum-to-zero constraint. This method was shown to perform well in terms of both prediction accuracy and computational efficiency.

With the success of large-margin classifiers in conducting standard classifications, it is desirable to adapt them to the OWL framework to help find an ITR for multicategory treatments. In this paper, we propose a new technique called multicategory outcome-weighted margin-based learning (MOML) to solve this problem. We start with the binary treatment scenario, and then generalize the methods to the muticategory treatment case. In particular, we use the vertices of a $k$-vertex simplex, with the origin as its center, in a $k-1$ Euclidean space to represent the $k$ treatments. Next, we construct $k-1$ functions to map the covariates of each patient onto a $k-1$-dimensional vector. Then, we define the prediction as the treatment that has the smallest angle between this vector and the

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



corresponding vertex of the simplex. Motivated by Zhao et al. (2012), we specify the objective function in the $loss + penalty$ form. The loss part is the weighted expectation of a loss function, $\ell(\cdot)$, of the angle between the $(k-1)$-dimensional function vector and the vertex of the actual treatment. The penalty term is used to control the model complexity. In this paper, we compare two penalty terms: $l_1$ and $l_2$ penalties. Note that the former can lead to sparse models and, hence, can be used for variable selection. Based on the loss term introduced, MOML detects the ITR as follows: for patients who have a good clinical outcome, the estimated optimal treatment should have a small angle with the actual treatment; on the other hand, for patients who have poor clinical results, the estimated optimal treatments should have large angles with the actual treatments.

The main contributions of this study are as follows. First, we propose using outcome-weighted margin-based learning (OML) to achieve ITR estimation for binary treatments. This learning technique produces a flexible class of decision functions that includes both soft and hard classifiers to obtain additional information and better prediction performance. Second, we propose a weighted angle-based method to adapt OML to multicategory treatment scenarios. For soft classifiers, we discuss how to obtain the estimated ratio of clinical rewards for each treatment pair in order to deter-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



mine the balance between the cost and the gain. We show the consistency properties and convergence rates of excess risks for MOML. In addition, we compare MOML with the one-versus-one and one-versus-rest extensions of OWL. Third, for the case of linear decision boundaries, we propose using an $l_1$ penalty to achieve variable sparsity. We further show that this technique leads to variable selection consistency, under certain assumptions.

The remainder of the paper is organized as follows. In Section 2, we review the OWL method and show how OML is introduced for the ITR estimation under the binary treatment setting. Then, we extend OML to multicategory cases, and explain how to maintain Fisher consistency by choosing a loss function. We also point out how the fitted decision functions can be connected to the ratios of the predicted clinical rewards under soft classifiers. In Sections 3 and 4, we provide six simulated examples and an application to a type-2 diabetes mellitus observational study, respectively, to evaluate the finite-sample performance of MOML. Discussions and conclusions are provided in Section 5. Several additional theories, including the excess risk convergence rate and selection consistency, and all technical details and proofs are provided in the online Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



## 2. Methodology

In this section, we first introduce the concepts and notation related to ITRs in Section 2.1, and then discuss how to use binary margin-based classifiers to find the optimal ITR for two treatments in Section 2.2. In Section 2.3, we extend the proposed method to the case of multiple treatments.

### 2.1 ITRs and OWL

Suppose we observe the training data set $\{(\boldsymbol{x}_i, a_i, r_i);\ i = 1, \ldots, n\}$ from an underlying distribution $P(\boldsymbol{X}, A, R)$, where $\boldsymbol{X} \in \mathbb{R}^p$ is a patient's covariate vector, $A \in \{1, \ldots, k\}$ is the treatment, and $R$ is the observed clinical outcome, namely, the reward. In particular, $P(\boldsymbol{x}, a, r) = f_0(\boldsymbol{x})\text{pr}(a|\boldsymbol{x})f_1(r|\boldsymbol{x}; a)$, where $f_0$ is the unknown density of $\boldsymbol{X}$, $\text{pr}(a|\boldsymbol{x})$ is the probability of receiving treatment $a$ for a patient with covariates $\boldsymbol{x}$, and $f_1$ is the unknown density of $R$, conditional on $(\boldsymbol{X}; A)$. We assume that larger values of $R$ are more desirable. In this paper, we focus on $k$-arm trials. An ITR $D$ is a mapping from the covariate space $\mathbb{R}^p$ onto the treatment set $\{1, \ldots, k\}$.

Before discussing multicategory treatments, we first introduce the binary optimal ITR, and formulate it as an outcome-weighted binary classification problem. To better understand ITRs, we use $E$ to denote the expectation with respect to $P$. For any ITR $D(\cdot)$, we let $P^D$ be the distri-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



bution of $\{\boldsymbol{X}, A, R\}$, under which the treatment $A$ is decided by $D(\boldsymbol{X})$, with $P^D(\boldsymbol{x}, a, r) = f_0(\boldsymbol{x})I(a = D(\boldsymbol{x}))f_1(r|\boldsymbol{x}; a)$, and let $E^D$ be the corresponding expectation. Therefore, $P^D$ is the distribution with the same $\boldsymbol{X}$-marginal as $P$ and, given $\boldsymbol{X} = \boldsymbol{x}$, the conditional distribution of $R$ is $P(r|\boldsymbol{X} = \boldsymbol{x}; A = D(\boldsymbol{x}))$. We assume $\text{pr}(A = a|\boldsymbol{x}) > 0$ for any $a \in \{1, \ldots, k\}$. One can verify that $P^D$ is absolutely continuous with respect to $P$, and that the Radon−Nikodym derivative is $dP^D/dP = I\{a = D(\boldsymbol{x})\}/\pi_a(\boldsymbol{x})$, where $I(\cdot)$ is the indicator function, and $\pi_a(\boldsymbol{x}) = \text{pr}(A = a|\boldsymbol{x})$. Consequently, the expected reward for a given ITR $D$ is

$$E^D(R) = \int R dP^D = \int R \frac{dP^D}{dP} dP = \int R \frac{I\{A = D(\boldsymbol{X})\}}{\pi_A(\boldsymbol{X})} dP.$$

An optimal ITR $D^*$ is defined as $D^* = \text{argmax}_D \, E^D(R) = \text{argmax}_D \, E\left[R\frac{I\{A=D(\boldsymbol{X})\}}{\pi_A(\boldsymbol{X})}\right]$. An equivalent expression of $D^*$ is that, for any $\boldsymbol{x}$, $D^*(\boldsymbol{x}) = \text{argmax}_{a\in\{1,\ldots,k\}} \, E(R|\boldsymbol{X} = \boldsymbol{x}; A = a)$. In other words, $D^*$ is an optimal ITR if for any $\boldsymbol{x}$, the expected reward that corresponds to $D^*(\boldsymbol{x})$ is larger than that of any treatment in $\{1, \ldots, k\}\backslash D^*(\boldsymbol{x})$. The optimal rule $D^*(\boldsymbol{x})$ is estimated from the observed training data from the joint distribution of $(\boldsymbol{X}, A, R)$. For a future patient with observed covariate $\boldsymbol{x}$, the optimal treatment is predicted based on the estimated $D^*(\boldsymbol{x})$ .

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



In the literature, a common approach to finding $D^*$ is to estimate $E(R \mid A = a; \boldsymbol{X} = \boldsymbol{x})$ for each treatment, using parametric or semiparametric regression models (Robins, 2004; Moodie et al., 2009; Qian and Murphy, 2011). For a new patient with covariates $\boldsymbol{x}$, the treatment recommendation is the maximum $\hat{E}\{R \mid A = a; \boldsymbol{X} = \boldsymbol{x}\}$.

When there are two treatments, we can express them as $A \in \{+1, -1\}$. Qian and Murphy (2011) showed that, in this case, finding $D^*$ can be formulated as a binary classification problem. In particular, one can verify that $D^*$ is the minimizer of

$$\int \frac{R}{\pi_A(\boldsymbol{X})} I\{A \neq D(\boldsymbol{X})\} dP. \tag{2.1}$$

Note that (2.1) can be viewed as a weighted 0–1 loss in a weighted binary classification problem. To see this, note that with the training data set $\{(\boldsymbol{x}_i, a_i, r_i);\ i = 1, \ldots, n\}$, we wish to minimize the following empirical loss that corresponds to (2.1):

$$\frac{1}{n} \sum_{i=1}^{n} \frac{r_i}{\pi_{a_i}(\boldsymbol{x}_i)} I\{a_i D(\boldsymbol{x}_i) \neq 1\}. \tag{2.2}$$

However, because the indicator function is discontinuous, solving (2.2) can be NP-hard. To overcome this difficulty, we can use a surrogate loss func-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



tion $\ell(\cdot)$ for binary margin-based classification. Zhao et al. (2012) proposed OWL, which employs a hinge loss in the SVM for the optimization. In particular, they assumed that $r_i \geq 0$ for all $i$, and used a single function $f(\boldsymbol{x})$ for classification, as is typical in binary margin-based classifiers. The treatment is assigned by $D(\boldsymbol{x}) = \text{sign}\{f(\boldsymbol{x})\}$. The corresponding optimization problem in Zhao et al. (2012) can be written as

$$\underset{f}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} \frac{r_i}{\pi_{a_i}(\boldsymbol{x}_i)} \{1 - a_i f(\boldsymbol{x}_i)\}_+ + \lambda J(f), \tag{2.3}$$

where $(1-u)_+ = \max(0, 1-u)$ is the hinge loss function, $J(f)$ is a penalty on $f$ to prevent overfitting, and $\lambda$ is the tuning parameter.

Note that Zhao et al. (2012) considered only nonnegative rewards; thus the corresponding problem remains a convex optimization. When there are negative rewards, they recommend shifting all rewards by a constant. Chen et al. (2017) showed that the performance of OWL varies with the choice of the shifting constant. To address this problem, they modified the loss to handle negative rewards directly.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



## 2.2 OML for Binary Treatments

As discussed in Section 1, there are many open problems, despite the seminal progress in Zhao et al. (2012). In particular, many choices of margin-based loss functions have not been fully studied in the literature. To investigate this problem, we propose an OML method. In Section 2.2, we focus on the case where $k = 2$ and $A \in \{+1, -1\}$, and propose the following OML optimization problem:

$$\underset{f}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \frac{r_i}{\pi_{a_i}(\boldsymbol{x}_i)} \ell\{a_i f(\boldsymbol{x}_i)\} + \lambda J(f), \tag{2.4}$$

where $\ell(\cdot)$ is a loss function in a margin-based classification. Here, $\ell(\cdot)$ denotes a classification method. For example, SVMs use the hinge loss in (2.3), and a logistic regression uses the deviance loss $\ell(u) = \log\{1 + \exp(-u)\}$. See the Supplementary Material for plots of several commonly used loss functions. We generalize our OML method to handle problems with multiple treatments in Section 2.3.

To explore different soft and hard classifiers, we need to define the theoretical minimizer of a classifier. First, we assume that $r_i \geq 0$. Consequently, (2.4) is convex if $\ell(\cdot)$ and $J(f)$ are convex, in which case, it can be solved using standard optimization methods, such as those in Boyd and

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



Vandenberghe (2004). We defer the discussion of negative rewards until after Theorem 1. Define the conditional expected loss with respect to (2.4) as $S(\boldsymbol{x}) = E[\frac{R}{\pi_A(\boldsymbol{X})}\ell\{Af(\boldsymbol{X})\} \mid \boldsymbol{X} = \boldsymbol{x}]$, where the expectation is taken with respect to the marginal distribution of $(R, A)$, for a given $\boldsymbol{x}$. We define the theoretical minimizer of $S(\boldsymbol{x})$ as

$$f^*(\boldsymbol{x}) = \operatorname*{argmin}_f S(\boldsymbol{x}) = \operatorname*{argmin}_f E[\frac{R}{\pi_A(\boldsymbol{X})}\ell\{Af(\boldsymbol{X})\} \mid \boldsymbol{X} = \boldsymbol{x}].$$

Note that $f^*$ depends on the loss function $\ell$.

Next, we discuss the consistency of a classifier. In the standard margin-based classification literature, Fisher consistency (Lin, 2002; Liu, 2007), also known as classification calibration (Bartlett et al., 2006), is a fundamental requirement of classifiers. For problems that require finding optimal ITRs using classification, a method is said to be Fisher consistent if the predicted treatment based on $f^*$ leads to the best expectation of the outcome rewards (Zhao et al., 2012). In other words, for binary problems, the method is Fisher consistent if $\text{sign}\{f^*(\boldsymbol{x})\} = \text{argmax}_a R(\boldsymbol{x}, a)$, where $R(\boldsymbol{x}, a) = \int (R \mid \boldsymbol{X} = \boldsymbol{x}, A = a) dP$ is the expected reward for a given treatment $a$ at a fixed $\boldsymbol{x}$. Zhao et al. (2012) proved that the OWL method using the hinge loss is Fisher consistent for nonnegative rewards. In the next proposition, we

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



provide a more general result that can be applied to various loss functions.

**Proposition 1.** *To find optimal ITRs using binary margin-based classifiers, assume that the rewards are nonnegative. Then, the method is Fisher consistent if $\ell(\cdot)$ is differentiable at 0, and $\ell(u) < \ell(-u)$, for any $u > 0$.*

Proposition 1 shows that in ITR problems, many binary margin-based classifiers are Fisher consistent. For instance, both soft and hard classifiers in the LUM loss family (Liu et al., 2011) are Fisher consistent. Note that the LUM family uses a parameter $c$ to control whether the classification is soft ($c = 0$) or hard ($c \to \infty$). See the appendix for additional information on LUM loss functions.

In a standard margin-based classification, in addition to Fisher consistency, $f^*$ can also be used to estimate the class conditional probabilities. This approach is widely used in the literature. See, for example, Hastie et al. (2009) and Liu et al. (2011), among others. For completeness, we include a brief explanation on how to estimate probabilities using $f^*$ in the appendix. For problems that employ binary classifiers to find optimal ITRs, the next theorem shows that when we use certain loss functions, $f^*$ can be used to find the ratio between $R(\boldsymbol{x}, +1)$ and $R(\boldsymbol{x}, -1)$.

**Theorem 1.** *To find optimal ITRs using binary margin-based classifiers, assume that the rewards are nonnegative. Furthermore, assume that the*

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



*loss function* $\ell(\cdot)$ *is differentiable with* $\ell'(u) < 0$*, for all* $u$*. Then, we have that*

$$\frac{R(\boldsymbol{x}, +1)}{R(\boldsymbol{x}, -1)} = \frac{\ell'(-f^*)}{\ell'(f^*)}. \tag{2.5}$$

As a result, for any new observation $\boldsymbol{x}$, once we obtain the fitted classification function $\hat{f}(\boldsymbol{x})$, we can estimate the ratio of $R(\boldsymbol{x}, +1)$ to $R(\boldsymbol{x}, -1)$ using $\ell'\{-\hat{f}(\boldsymbol{x})\}/\ell'\{\hat{f}(\boldsymbol{x})\}$, which provides more information than the ITR itself does.

**Remark 1.** Theorem 1 shows that estimating the ratio of expected rewards in ITR problems is similar to the class conditional probability estimation in a standard margin-based classification. In particular, let $P_{+1}(\boldsymbol{x})$ and $P_{-1}(\boldsymbol{x})$ be the conditional class probabilities for classes $+1$ and $-1$, respectively, in a binary classification (see the appendix for further details). We can verify that, with similar conditions on $\ell$, we can use $\ell'(-\hat{f})/\ell'(\hat{f})$ to estimate $P_{+1}(\boldsymbol{x})/P_{-1}(\boldsymbol{x})$. For example, in a standard logistic regression, estimating $P_{+1}(\boldsymbol{x})/P_{-1}(\boldsymbol{x})$ by $\ell'(-\hat{f})/\ell'(\hat{f})$ is equivalent to using the logit link function for probability estimation. Similar discussions on class probability estimations for standard multicategory classification problems are presented in Zou et al. (2008), Zhang and Liu (2014), and Neykov et al.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



(2016).

Using Theorem 1, we can explore the difference between using soft and hard classifiers to find optimal ITRs. In particular, we plot $\log\{R(\boldsymbol{x},+1)/R(\boldsymbol{x},-1)\}$, denoted by $r_{+1-1}$, against $f^*$ for some loss functions in the LUM family in Figure 1. We can see that, with soft classifiers ($c=0$), there is a one-to-one correspondence between $r_{+1-1}$ and $f^*$. In other words, we can estimate the ratio between the expected rewards for any new patients using the estimated $\hat{f}$. This ratio information can be important in practical problems, as discussed in Section 1. As discussed in Section 3, if the underlying ratios are smooth functions, soft classifiers tend to perform better than hard classifiers in terms of accurately estimating the ratios.

For $c>0$, the flat region of $r_{+1-1}$ makes estimating this ratio more difficult. In particular, if $\hat{f}\in[-c/(1+c),c/(1+c)]$, then the method cannot provide an estimate of $r_{+1-1}$. As $c$ increases, the flat region enlarges. In the limit ($c\to\infty$), the hard classifier provides little information about $r_{+1-1}$. In other words, hard classifiers bypass the estimation of $r_{+1-1}$ and focus on the boundary (i.e., $R(\boldsymbol{x},+1)=R(\boldsymbol{x},-1)$ in binary problems) estimation only. In the Supplementary Material, we show that when the underlying ratios are close to step functions, hard classifiers outperform soft classifiers, because an accurate estimation of $r_{+1-1}$ can be very difficult.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



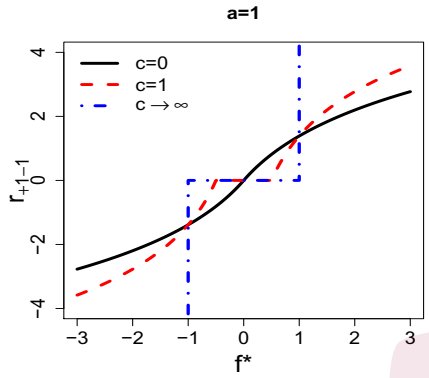


Figure 1: Plot of $\log\{R(\boldsymbol{x},+1)/R(\boldsymbol{x},-1)\}$ ($r_{+1-1}$ on the $y$ axis) against $f^*$ for some LUM loss functions. Here, $c = 0$ corresponds to the soft LUM loss, and $c \to \infty$ corresponds to the SVM hinge loss, which is a hard classifier. Note that $a$ is another parameter in the LUM family (see the appendix), and $a = 1$ and $c = 1$ correspond to the loss function in a distance-weighted discriminant analysis (Marron et al., 2007).

Next, we discuss how to address negative rewards using our OML method. Recall that when all $r_i \geq 0$, we can use a surrogate loss function $\ell$ that is a convex upper bound of the 0−1 loss, as from (2.2) to (2.4). When $r_i < 0$, the corresponding 0−1 loss is equivalent to $-|r_i|I\{a_i D(\boldsymbol{x}_i) \neq 1\}$, which can be regarded as a *−1−0 loss* (Chen et al., 2017). In this case, because the reward is negative, it is desirable to consider the other treatment, rather than $a_i$. Based on these observations, we propose the following

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



optimization of binary problems for both positive and negative rewards:

$$\underset{f}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \frac{|r_i|}{\pi_{a_i}(\boldsymbol{x}_i)} \ell_{r_i}\{a_i f(\boldsymbol{x}_i)\} + \lambda J(f), \tag{2.6}$$

where $\ell_{r_i}(u) = \ell(u)$ if $r_i \geq 0$, and $\ell_{r_i}(u) = \ell(-u)$ if $r_i < 0$ (the inverted loss). Note that $\ell(-u) - 1$ is the tight convex upper bound of the $-1-0$ loss as long as $\ell$ is convex, and minimizing $\ell\{-a_i f(\boldsymbol{x}_i)\} - 1$ and $\ell\{-a_i f(\boldsymbol{x}_i)\}$ with respect to $f$ are equivalent. The treatment recommendation rule for negative rewards is still $D(\boldsymbol{x}) = \operatorname{sign}\{f(\boldsymbol{x})\}$.

The next theorem shows that our binary OML method with negative rewards also enjoys Fisher consistency, with mild conditions on the loss function.

**Theorem 2.** *When finding optimal ITRs using binary OML classifiers (2.6), a method is Fisher consistent if $\ell(\cdot)$ is differentiable at zero, and $\ell(u) < \ell(-u)$ for any $u > 0$.*

From Theorem 2, by including the inverted loss functions for negative rewards, our OML method can still be asymptotically consistent. In contrast, the estimation of the rewards ratio becomes more complicated if $R$ can be negative. The next theorem shows that our OML method is able to provide an upper or lower bound for the corresponding rewards ratios, under some

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



mild assumptions.

**Theorem 3.** *To find optimal ITRs using binary margin-based classifiers, assume that the expected rewards satisfy $R(\boldsymbol{x}, a) > 0$, for all $\boldsymbol{x}$ and $a$. Furthermore, assume that the loss function $\ell(\cdot)$ is differentiable, with $\ell'(u) < 0$ for all $u$. Then, we have that*

$$\begin{cases} \frac{R(\boldsymbol{x},+1)}{R(\boldsymbol{x},-1)} \geq \frac{\ell'(-f^*)}{\ell'(f^*)}, & \text{if } R(\boldsymbol{x},+1) > R(\boldsymbol{x},-1), \\ \frac{R(\boldsymbol{x},+1)}{R(\boldsymbol{x},-1)} \leq \frac{\ell'(-f^*)}{\ell'(f^*)}, & \text{if } R(\boldsymbol{x},+1) < R(\boldsymbol{x},-1). \end{cases} \tag{2.7}$$

Theorem 3 shows that $\ell'(-\hat{f})/\ell'(\hat{f})$ can be used as a lower bound for the rewards ratio when treatment $+1$ is better, and an upper bound if $-1$ is better. The condition that $R(\boldsymbol{x}, a) > 0$ for all $\boldsymbol{x}$ and $a$ can be satisfied, for example, when patients with no treatments have zero expected rewards, and all treatments under study have preliminary results to show that they are effective overall. Note that when there are negative rewards, our OML method cannot provide an accurate estimation of the rewards ratio, but can provide a bound (see the proof of Theorem 3 in the Supplementary Material for further details), yet the method is still Fisher consistent. Hence, we can see that in ITR problems, calculating a rewards estimation can be more difficult than a treatment recommendation. This is analogous to a standard classification, in which a probability estimation can be more difficult than

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



making a label prediction.

In the next section, we generalize our OML method to handle problems with multiple treatments.

## 2.3 MOML

To find $D^*$ in a practical problem with $k > 2$ treatments, we can employ sequential binary classifiers, such as the one-versus-one and one-versus-rest approaches. However, these can lead to inconsistent ITR estimators (see the Supplementary Material for a proof of the inconsistency of the one-versus-rest SVM approach). As discussed in Section 1, it can be desirable to have a multicategory classifier that considers all $k$ treatments simultaneously in one optimization problem.

In the literature, many commonly used simultaneous multicategory margin-based classifiers employ $k$ classification functions for the $k$ classes. Furthermore, they impose a sum-to-zero constraint on the $k$ functions to reduce the parameter space and to ensure certain theoretical properties, such as Fisher consistency. Recently, Zhang and Liu (2014) showed that this approach can be redundant and suboptimal in terms of computational speed and classification accuracy. To overcome these difficulties, Zhang and Liu (2014) proposed an angle-based classification method. In this pa-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



per, we propose identifying optimal ITRs with multiple treatments in an angle-based classification framework.

The standard angle-based classification can be summarized as follows. Let $\{(\boldsymbol{x}_i, y_i);\ i = 1, \ldots, n\}$ be the training data set, where $y$ represents the class label. Define a simplex $\boldsymbol{W}$ with $k$ vertices $\{\boldsymbol{W}_1, \ldots, \boldsymbol{W}_k\}$ in a $(k-1)$-dimensional space, such that

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

where $\mathbf{1}_{k-1}$ is a vector of ones of length $k-1$, and $\boldsymbol{e}_j \in \mathbb{R}^{k-1}$ is a vector with the $j$th element equal to one, and zero elsewhere. This simplex is symmetric with all vertices an equal distance from each other. The angle-based classifier uses a $(k-1)$-dimensional classification function vector $\boldsymbol{f} = (f_1, \ldots, f_{k-1})^T$, which maps $\boldsymbol{x}$ to $\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^{k-1}$. Note that $\boldsymbol{f}$ introduces $k$ angles with respect to $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_k$, namely, $\angle(\boldsymbol{f}, W_j);\ j = 1, \ldots, k$. The prediction rule is based on which angle is the smallest. In particular, $\hat{y}(\boldsymbol{x}) = \text{argmin}_{j \in \{1,\ldots,k\}} \angle(\boldsymbol{f}, W_j)$, where $\hat{y}(\boldsymbol{x})$ is the predicted label for $\boldsymbol{x}$. Figure 2 illustrates how to make predictions using this angle-based classification when $k = 2, 3$, and 4. When $k = 3$, for example, the mapped observation $\hat{f}$ is predicted as the class corresponding to $\boldsymbol{W}_1$, because $\theta_1$ is the

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



smallest angle. Based on the observation that $\text{argmin}_{j\in\{1,\ldots,k\}} \angle(\boldsymbol{f}, W_j) = \text{argmax}_{j\in\{1,\ldots,k\}} \langle \boldsymbol{f}, \boldsymbol{W}_j \rangle$, Zhang and Liu (2014) proposed the following optimization problem for the angle-based classifier:

$$\underset{f}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} \ell\{\langle \boldsymbol{W}_{y_i}, \boldsymbol{f}(\boldsymbol{x}_i)\rangle\} + \lambda J(\boldsymbol{f}), \tag{2.8}$$

where $\ell(\cdot)$ is a binary margin-based surrogate loss function, which is typically nonnegative and satisfies $\ell(u) < \ell(-u)$ for any $u > 0$, $J(\boldsymbol{f})$ is a penalty on $\boldsymbol{f}$ to prevent overfitting, and $\lambda$ is a tuning parameter to balance the goodness of fit and the model complexity. One advantage of the angle-based classifier is that it is free of the sum-to-zero constraint and, thus, leaning more efficient for large data sets.

To generalize our OML method from the binary setting to handle multicategory problems, we propose the following optimization:

$$\underset{f}{\text{argmin}} \frac{1}{n} \sum_{i=1}^{n} \frac{|r_i|}{\pi_{a_i}(\boldsymbol{x}_i)} \ell_{r_i}\{\langle \boldsymbol{W}_{a_i}, \boldsymbol{f}(\boldsymbol{x}_i)\rangle\} + \lambda J(\boldsymbol{f}), \tag{2.9}$$

where $\ell_{r_i}$ is defined as in (2.6). For the penalty term $J(f)$, we discuss two options: $l_2$ and $l_1$ penalties. When applying the $l_1$ penalty, we can remove covariates that have zero coefficient estimates in all $k-1$ components of the fitted $\boldsymbol{f}$. We show in Section 4 that such a sparse penalty can exhibit selec-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



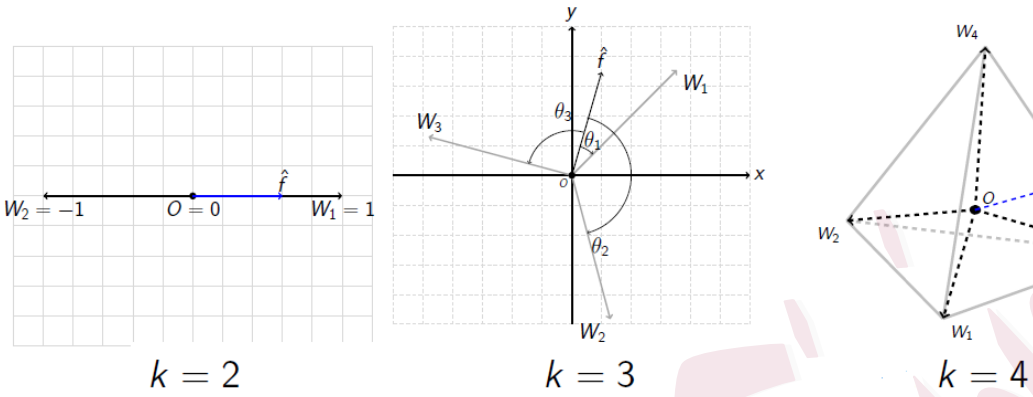


Figure 2: Illustration of the angle-based classification with $k = 2, 3$, and 4. For example, when $k = 3$ (as the plot in the middle shows), the mapped observation $\hat{f}$ is predicted as the class corresponding to $\boldsymbol{W}_1$, because $\theta_1 < \theta_3 < \theta_2$.

tion consistency under linear learning. For a new patient with the covariate vector $\boldsymbol{x}$, once the fitted classification function vector $\hat{\boldsymbol{f}}$ is obtained, the corresponding treatment recommendation is $\text{argmax}_{a\in\{1,\ldots,k\}}\langle \boldsymbol{W}_a, \hat{\boldsymbol{f}}(\boldsymbol{x})\rangle$. We can verify that when $k = 2$, (2.9) reduces to (2.6). Hence, for the statistical learning theory (see the Supplementary Material), we focus on multicategory classification, and the results can be applied to binary cases directly.

Next, we study the Fisher consistency of MOML for multicategory treatments. In the literature on standard margin-based classification, Fisher consistency is more complicated in multicategory problems than it is in binary settings. For example, it is known that the binary SVM is Fisher consistent (Lin, 2002). However, its direct generalization to the multicate-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



gory classifier is inconsistent, both in the framework with $k$ functions and a sum-to-zero constraint (Liu, 2007), and in the framework of angle-based classification (Zhang and Liu, 2014). To overcome these challenges, many new multicategory SVMs have been proposed. See, for example, Lee et al. (2004) and Liu and Yuan (2011), among others. To find optimal ITRs, we have the following result for the Fisher consistency of our MOML method in multicategory treatment problems.

Before presenting our main result, we introduce an important assumption. First, recall that the expected reward for a given treatment $j$ at $\boldsymbol{x}$ is $R(\boldsymbol{x}, a) = \int (R \mid \boldsymbol{X} = \boldsymbol{x}, A = a) dP$. Define the positive part of a conditional reward as $R_j^+(\boldsymbol{x}) = \int (R \mid \boldsymbol{X} = \boldsymbol{x}, A = j) I(R > 0) dP$, and the negative part as $R_j^-(\boldsymbol{x}) = \int (R \mid \boldsymbol{X} = \boldsymbol{x}, A = j) I(R < 0) dP$. We can verify that $R(\boldsymbol{x}, j) = R_j^+(\boldsymbol{x}) + R_j^-(\boldsymbol{x})$. Here, $R_j^-(\boldsymbol{x})$ can be used to measure the possibility and severity of adverse effects for treatment $j$ on patients with the covariate vector $\boldsymbol{x}$. The next assumption requires that $R^-(\boldsymbol{x})$ of the best treatment for a given patient should not be small.

**Assumption 1.** *For a patient with the covariate vector $\boldsymbol{x}$, denote the best treatment by $j$ (i.e., $R(\boldsymbol{x}, j) > R(\boldsymbol{x}, i)$, for any $i \neq j$). Then, $R_j^-(\boldsymbol{x}) \geq R_i^-(\boldsymbol{x})$, for any $i \neq j$.*

Assumption 1 is desirable, and often necessary for practical problems.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



In particular, for any patient, we should expect that the best treatment does not have a large probability of adverse effects, and that its adverse effects are relatively mild. Assumption 1 can be satisfied, for example, when the rewards are all positive, or when the marginal distributions of the rewards for different patients and treatments are the same, except for a constant shift (e.g., normal distributions with a common variance). With Assumption 1, we are ready to present the theorem for the Fisher consistency of our MOML method.

**Theorem 4.** *To find optimal ITRs using MOML classifiers (2.9), suppose Assumption 1 is valid. Then the method is Fisher consistent if $\ell(\cdot)$ is convex and strictly decreasing. Moreover, MOML with a hinge loss is not Fisher consistent.*

Note that Theorem 4 provides a sufficient condition for the MOML classifier to be Fisher consistent. In the literature, some classifiers have loss functions that do not satisfy the condition in Theorem 4, yet we can still verify that the corresponding MOML method is Fisher consistent. For example, we can use a similar approach to that in the proof of Theorem 4 to show that our MOML method using the proximal SVM loss is Fisher consistent. On the other hand, our MOML SVM (i.e., using the standard hinge loss) is not Fisher consistent. To overcome this challenge, we propose

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



using the LUM loss function with a large, but finite $c$. This loss function is very close to the SVM hinge loss, which corresponds to $c \to \infty$, and can preserve Fisher consistency. Note that a similar approach was previously used in Zhang and Liu (2014) to obtain a Fisher consistent angle-based classifier.

To estimate the ratio of the expected rewards for different treatments, we have the following theorem.

**Theorem 5.** *Suppose the loss function $\ell(u)$ is convex and differentiable, with $\ell'(u) < 0$ for all $u$. If the random reward satisfies $R \geq 0$, then for any $i \neq j \in \{1, \ldots, k\}$, we have*

$$\frac{R(\boldsymbol{x}, i)}{R(\boldsymbol{x}, j)} = \frac{\ell'(\langle \boldsymbol{f}^*, \boldsymbol{W}_j \rangle)}{\ell'(\langle \boldsymbol{f}^*, \boldsymbol{W}_i \rangle)}.$$

From Theorem 5, once $\hat{\boldsymbol{f}}(\boldsymbol{x})$ is obtained for a new patient with $\boldsymbol{x}$, we can estimate the rewards ratio between the $i$th and $j$th treatments as $\ell'\{\langle \hat{\boldsymbol{f}}(\boldsymbol{x}), \boldsymbol{W}_j \rangle\} / \ell'\{\langle \hat{\boldsymbol{f}}(\boldsymbol{x}), \boldsymbol{W}_i \rangle\}$. Additional discussions on soft and hard classifiers are provided in the Supplementary Material.

We also develop additional theoretical results for MOML such as the convergence rate of excess risks. In addition, we show that MOML enjoys variable selection consistency under linear ITRs with $J(\boldsymbol{f})$ as the $l_1$ penalty.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



Additional information is included in the additional statistical learning theory section of the Supplementary Material.

## 3. Numerical Studies

In this section, we use six simulation studies with both linear and nonlinear ITR boundaries to assess the finite-sample performance of the proposed MOML method. For all examples, we fit MOML using the $l_2$ penalty, and compare it to the standard OWL (Zhao et al. (2012)) with extensions of one-versus-rest (OWL-1) and one-versus-one (OWL-2). Furthermore, to evaluate the performance of the variable selection, as discussed in Section 3.2, we implement MOML using the $l_1$ penalty (MOML-$l_1$) for all linear ITR boundary examples. When fitting OWL, we replace the hinge loss with the modified loss in (7) to improve its performance for a fair comparison. For the one-versus-rest extension, we conduct sequential one-versus-rest binary optimal treatment estimations (i.e., 1 vs. others, 2 vs. others, $\cdots$, $k$ vs. others), and then pick the treatment recommended by the classifier $\hat{\boldsymbol{f}}_j$ with the largest magnitude among $j = 1, \cdots, k$. For the one-versus-one extension, we first estimate the decision function $\hat{\boldsymbol{f}}_l$, for $l = 1, \cdots, k(k-1)/2$, based on each pair of treatments (i.e., 1 vs. 2, 1 vs. 3, $\cdots$, $k-1$ vs. $k$), and then pick the treatment suggested by $\hat{\boldsymbol{f}}_l$ with the largest magnitude.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



Note that the one-versus-one extension uses only a subset of the data to fit each $\hat{\boldsymbol{f}}_l$. For a meaningful comparison, we restrict $\boldsymbol{f}$ to be linear functions of $\boldsymbol{x}$ for all of the models in the linear ITR boundary examples, and apply Gaussian kernel learning to fit $\boldsymbol{f}$ in nonlinear ITR boundary examples.

When we generate the data sets, we first simulate a training set, which is used to fit the model. We also generate an independent and equal-size tuning set to find the best combination of tuning parameters, as well as a much larger testing set to evaluate the model performance (10 times as big as the training set). For the tuning parameter range, we choose $a$ from $\{0.1, 1, 10\}$, let $c$ vary in $\{0, 1, 10, 100, 1000\}$, and let $\lambda$ vary in $\{0.001, 0.01, 0.1, 1, 10\}$. We report the averages and standard deviations of the misclassification rates and the empirical value functions of the testing sets as the criteria for model assessment. The empirical value function is defined as $\mathbb{P}_n^*[I(A = D(\boldsymbol{X}))R/\pi_A(\boldsymbol{X})]/\mathbb{P}_n^*[I(A = D(\boldsymbol{X}))/\pi_A(\boldsymbol{X})]$, where $\mathbb{P}_n^*$ denotes the empirical average of the testing data set (Zhao et al., 2012). The value function is treated as a more comprehensive measure of how close the estimated ITR is to the true optimal ITR. We repeat the simulations 50 times in each example.

In the first four examples, we generate the data sets in which the optimal treatment boundaries are linear functions of the covariates. We add

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



additional covariates as random noise in Examples 3 and 4. In the last two examples, we discuss nonlinear ITR scenarios, and perform Gaussian kernel learning classifiers. We let the dimensions of the covariates $\boldsymbol{x}$ vary in $p \in \{10, 50\}$ for all examples. The kernel bandwidth $\tau$ is fixed as $1/(2\hat{\sigma}^2)$, where $\hat{\sigma}$ is the median of the pairwise Euclidean distance of the simulated covariates (Wu and Liu, 2007). The details of each setting are presented below.

**Example 1** We consider three points $(\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3)$ that are equal distances from the $p$-dimensional space to represent the cluster centroids of the true optimal treatments. For each $\boldsymbol{c}_j$, where $j = 1, 2, 3$, we generate its covariate $X_i$ from a multivariate normal distribution $N(\boldsymbol{c}_j, I_p)$, where $I_p$ is a $p$-dimensional identity matrix. The actually assigned $A_i$ follows a discrete uniform distribution $U\{1, 2, 3\}$. The reward $R_i$ follows a Gaussian distribution $N(\mu(X_i, A_i, d_i), 1)$, where $\mu(X_i, A_i, d_i) = X_i^T\boldsymbol{\beta} + 5 \cdot I(A_i = d_i)$, $\boldsymbol{\beta}^T = (\mathbf{1}_{p/2}^T, -\mathbf{1}_{p/2}^T)$, and $d_i$ is the optimal treatment for $X_i$, as determined by the cluster centroids. The training data set is of size 300.

**Example 2** We define a five-treatment scenario in which the five centroids $(\boldsymbol{c}_1, \cdots, \boldsymbol{c}_5)$ form a simplex in $\mathbb{R}^4$. The marginal distribution $X_i|c_j$ follows a normal distribution with mean $\boldsymbol{c}_j$ and covariate matrix $0.1I_p$. The treatment $A_i$ follows a discrete uniform $U\{1, \cdots, 5\}$. The reward

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



$R_i \sim N(\mu(X_i, A_i, d_i), 0.1)$, where $\mu(X_i, A_i, d_i) = X_i^T\boldsymbol{\beta} + 3 \cdot I(A_i = d_i) + 1$ and $\boldsymbol{\beta}^T = 0.1 \times (\mathbf{1}_{p/2}^T, -\mathbf{1}_{p/2}^T)$. The training data set is of size 500.

**Example 3** This example includes 10 treatments, and the optimal ITR boundary depends on the first two covariates, that is, $(X_1, X_2)$. The 10 corresponding centroids $(\boldsymbol{c}_1, \cdots, \boldsymbol{c}_{10})$ are spread out evenly on the unit circle $X_1^2 + X_2^2 = 1$, and the marginal distribution of $(X_1, X_2)^T$ is a normal distribution with mean $\boldsymbol{c}_j$ and covariate matrix $0.03I_2$. Similarly to Example 2, $A_i \sim U\{1, \cdots, 10\}$ and $R_i \sim N(\mu(X_i, A_i, d_i), 1)$, where $\mu(X_i, A_i, d_i) = X_i^T\boldsymbol{\beta} + 5 \cdot I(A_i = d_i) - 2$ and $\boldsymbol{\beta}^T = (\mathbf{1}_5^T, -\mathbf{1}_5^T, \mathbf{0}_{p-10}^T)$. The training data set is of size 600.

**Example 4** All settings are the same as Example 2, except that $\boldsymbol{\beta}^T = 0.1 \times (1, 1, -1, -1, \mathbf{0}_{p-4}^T)$.

**Example 5** This is a three-class example, with each centroid $c_j$, for $j = 1, 2, 3$, distributed on two mess points with equal probabilities. The marginal distribution of $(X_1, X_2)^T$ is a mixture of two normal distributions $0.5N[(\cos(j\pi/3), \sin(j\pi/3))^T, 0.08I_2] + 0.5N[(\cos(\pi + j\pi/3), \sin(\pi + j\pi/3))^T, 0.08I_2]$. The treatment $A_i \sim U\{1, 2, 3\}$ and the reward $R_i \sim N(\mu(X_i, A_i, d_i), 1)$, where $\mu(X_i, A_i, d_i) = X_i^T\boldsymbol{\beta} + 5 \cdot I(A_i = d_i) - 1$ and $\boldsymbol{\beta}^T = (\mathbf{1}_{p/2}^T, -\mathbf{1}_{p/2}^T)$. The training data set is of size 300.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



**Example 6** In this example, the optimal treatment $d_i$ for each $X_i$ is determined with probability 95% by the signs of two underlying nonlinear functions, $f_1(X) = X_1^2 + X_2^2 + \exp\{0.5X_3\}$ and $f_2(X) = X_4^2 - X_5^3 - X_6$. A random noise is added to $d_i$ with probability 5% to create a positive Bayes error. In particular, we have $d_i$ defined as

$$d_i = d(X_i) = \begin{cases} 1 + [\text{sign}(f_1(X_i) - m_1)]_+ + 2 \times [\text{sign}(f_2(X_i) - m_2)]_+ & \text{with prob. } 0.95 \\ U_i & \text{with prob. } 0.05 \end{cases},$$

where $m_1$ and $m_2$ are the medians of $f_1$ and $f_2$, respectively, and $U_i$ follows a discrete $U\{1, 2, 3, 4\}$, which is independent of $(A_i, X_i)$. The covariate $X_i$ follows a continuous uniform distribution $U(0, 1)$, $A_i \sim U\{1, \cdots, 4\}$, and $R_i \sim N(\mu(X_i, A_i, d_i), 1)$, where $\mu(X_i, A_i, d_i) = X_i^T\boldsymbol{\beta} + 5 \cdot I(A_i = d_i) - 1$ and $\boldsymbol{\beta}^T = (\mathbf{1}_{p/2}^T, -\mathbf{1}_{p/2}^T)$. The training data set is of size 500.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



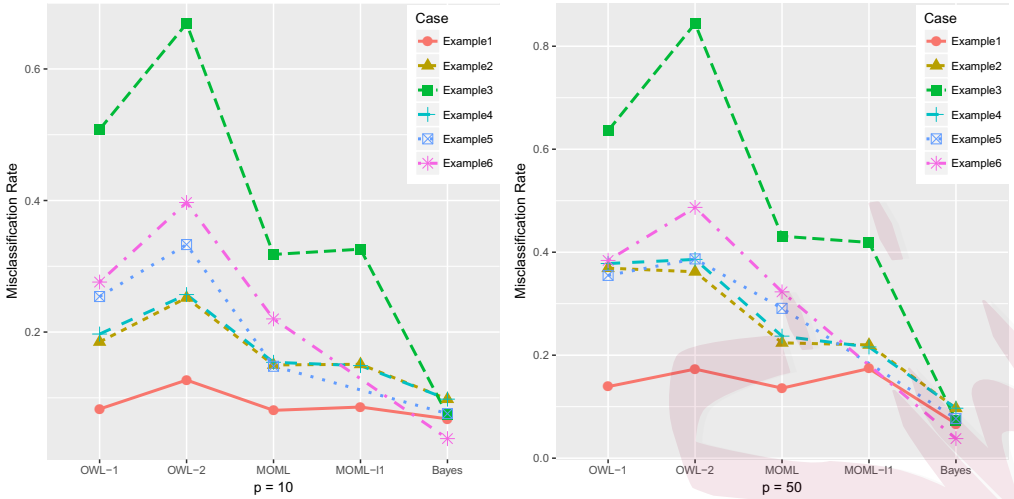


Figure 3: Plots of misclassification rates of simulation studies. OWL-1 and OWL-2 represent extensions of OWL (one-versus-rest and one-versus-one), MOML and MOML-$l_1$ represent outcome weighted margin-based learning with $l_2$ and $l_1$ penalties, respectively, and Bayes represents the empirical Bayes error.

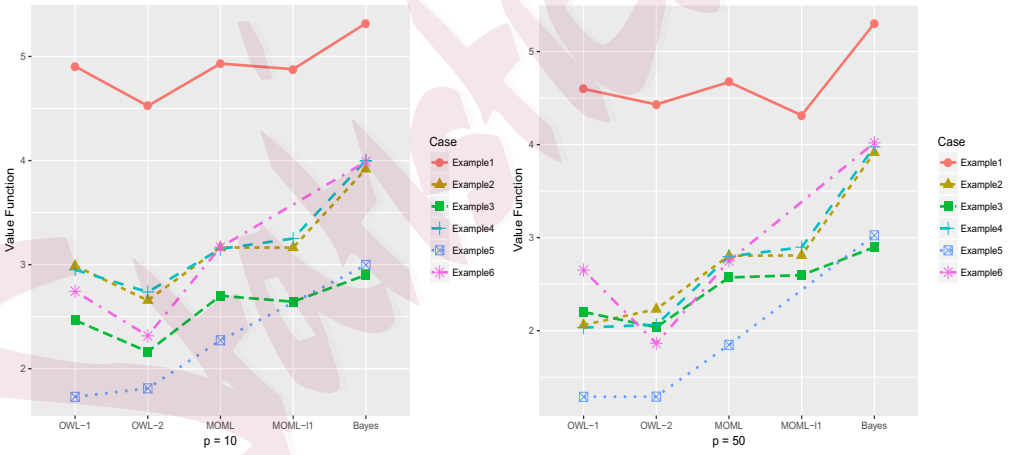


Figure 4: Plots of value functions of simulation studies. OWL-1 and OWL-2 represent extensions of OWL (one-versus-rest and one-versus-one), MOML and MOML-$l_1$ represent outcome weighted margin-based learning with $l_2$ and $l_1$ penalties, respectively, and Bayes represents the empirical Bayes error.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



Figures 3 and 4 plot the sample means of the misclassification rates and the empirical value functions produced by the models. The numerical results, with standard deviations, are reported in tables in the Supplementary Material. From the results, MOML with the $l_2$ penalty, MOML with the $l_1$ penalty, and OWL-1 (with one-versus-rest extension) perform equivalently when the underlying ITR is not very complicated and the treatment effect is sufficiently strong, as Example 1 shows when $p = 10$. Example 2 represents situations when the linear ITR becomes more complicated and the treatment effect is intermediate. Here, MOML produces significantly larger empirical value function results than the two simple OWL extensions do. Example 4 has a similar setting to Example 2, with noise variables added to the covariate set. Under this scenario, MOML with the $l_1$ penalty outperforms MOML with the $l_2$ penalty, because it is able to remove many unnecessary noise variables. This improvement in prediction accuracy becomes clearer when there are higher covariate dimensions, that is, $p = 50$. For the selection result, when $p = 10$, MOML-$l_1$ removes 64.6% of the noise, on average, while keeping all useful variables; when $p = 50$, about 57.6% of the noise is removed, and all useful variables are kept. Example 3 represents a difficult ITR detection scenario, with a large number of treatments ($k = 10$). In this case, the two MOML methods have much smaller mis-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



classification rates than those of the two OWL extensions, implying that MOML can produce stable estimation results. The variable selection results show that MOML-$l_1$ succeeded in removing 68.8% and 60.2% of the noise under $p = 10$ and $p = 50$, respectively. All true variables are kept under both cases. Examples 5 and 6 are nonlinear ITRs. In Example 5, MOML maintains a low misclassification rate when the covariate dimension is not large (i.e., $p = 10$). As more variables are added to the covariate space, all of the methods produce significantly worse prediction performance, although MOML still outperforms the two OWL extensions. As such, we recommend reducing the covariate dimension before applying nonlinear MOML in practice. In Example 6, we intentionally include outliers in the samples to assess the models' robustness. All of the methods are affected, although MOML still produces better prediction results than those of the other methods.

Finally, we explore the advantages of soft and hard classifiers using Examples 1 and 6. We try different values of $c$, and show that a properly tuned classifier performs very well. The details are provided in the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



## 4. Application to a Type-2 Diabetes Mellitus Study

In this section, we apply the proposed method to a type-2 diabetes mellitus (T2DM) observational study to assess its performance in real-life data applications. The study includes people with T2DM during the period 2012−2013, with data provided by the Clinical Practice Research Datalink (CPRD) (Herrett et al. (2015)). Four anti-diabetic therapies are considered in this study: glucagon-like peptide-1 (GLP-1) receptor agonist, long-acting insulin only, intermediate-acting insulin only, and a regime including short-acting insulin. The primary target variable is the change in HbA1c before and after the treatment. Seven clinical factors are used: age, gender, ethnicity, body mass index, high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), and smoking status. In total, 634 patients satisfy the aforementioned requirements, and around 5% have complete observations. Considering the large missing proportion, we perform the following steps. First, all factors that have a missing rate larger than 70% are removed. Second, a standard $t$ test is implemented for each remaining factor to check whether its missing indicator affects the response. If the test result is statistically significant, we keep the variable, while removing all of its missing observations. Otherwise, we delete the variable. We have 230 observations left after this cleaning process.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



| | Training | Validation |
|---|---|---|
| OWL-1-Linear | 2.712 (0.329) | 2.371 (0.483) |
| OWL-2-Linear | 2.487 (0.233) | 2.221 (0.561) |
| OWL-1-Gaussian | 4.118 (0.401) | 3.285 (0.490) |
| OWL-2-Gaussian | 4.003 (0.374) | 3.221 (0.468) |
| MOML-Linear | 2.610 (0.130) | 2.440 (0.320) |
| MOML-$l_1$-Linear | 2.813 (0.138) | 2.533 (0.182) |
| MOML-Gaussian | 4.105 (0.221) | **3.612** (0.328) |

Table 1: Analysis Results for the T2DM Data Set. Estimated averages and standard deviations (in parentheses) of the value function are reported using five-fold cross-validation with 50 replications. OWL-1 and OWL-2 represent two extensions of OWL (one-versus-rest and one-versus-one, respectively), and MOML and MOML-$l_1$ represent outcome weighted margin-based learning with $l_2$ and $l_1$ penalties, respectively. The observed average reward for the cleaned data set is 2.246.

We apply the same methods with linear and Gaussian kernels to the cleaned T2DM data set as those in the simulation analysis. We use the negative HbA1c change as the reward, because the treatment goal is to decrease HbA1c. The prosperity score $\pi_A(\boldsymbol{X})$ is calculated based on a fitted multinomial logistic regression between the assigned treatment and all covariates. We use five-fold cross-validation to choose the best tuning parameter over 50 replications. Specifically, we randomly divide the clean data into five equal-sized subsets, train the model based on every fourth set (training sets), and make a prediction using the remaining set (validation sets). The means and standard deviations of the empirical value functions for the training and validation sets are presented in Table 1.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



Table 1 shows that the proposed MOML with the Gaussian kernel gives the best predicted value function, with a smaller standard deviation than that of OWL with the Gaussian kernel. MOML-$l_1$ suggests keeping all of the variables over the 50 replicates, which indicates that the covariates remaining in the clean data may all be important when a linear function is chosen to fit the ITR. In terms of the estimated optimal treatment assignment results, the one-versus-rest extension of OWL with a Gaussian kernel (OWL-1-Gaussian) assigns around 32% of the patients to the short-acting insulin group, and the rest to the other three treatment groups in a relatively even way. MOML with the Gaussian kernel recommends that approximately 40% of the patients take the short-acting insulin, around 25% and 23% patients take intermediate and long-acting insulin, respectively, and less than 12% take the GLP-1. This conclusion is consistent with the findings of some studies on short-acting insulins, which shows the benefit of reducing HbA1c (Holman et al., 2007). On the other hand, prandial insulins can also increase the risk of hypo and weight gain. As a result, it may be worth treating some composite metric as the outcome that combines HbA1c change, hypo events, and weight gain information, to find the corresponding optimal treatment rules.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



## 5. Conclusion

In this paper, we propose a margin-based loss function to solve the optimal individual treatment estimation problem for binary treatments, and then extend it to include multicategory treatment scenarios. For binary treatments, we develop a loss based on the LUM family, such that the proposed method includes a wide range of ITRs, varying from soft to hard classifiers. The standard OWL is a special case of the proposed margin-based learning methods because the LUM family loss becomes the hinge loss when $c \to \infty$ and $a = 1$. For multiple treatments, we formulate the loss as a weighted sum of the angles between the estimated decision function $\boldsymbol{f}$ and the actual treatment $A$. We show that MOML enjoys desirable theoretical properties, and has a higher prediction accuracy than that of the other methods under both linear and nonlinear treatment assignment boundaries. Our method produces straightforward ITR results with a clear geometric interpretation. Moreover, the optimization problem of MOML is unconstrained and, hence, can be more efficient to compute compared with other multicategory methods with the sum-to-zero constraint. We also showed that the proposed MOML exhibits selection consistency using the $l_1$ penalty for the case with linear decision boundaries. This idea can be extended to nonlinear boundaries as well. One possibility is to use the idea of weighed kernels, and to

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



impose a weight vector $\boldsymbol{w}$ in front of the covariate $\boldsymbol{x}$ in the standard kernel definition (Chen et al., 2017).

## Acknowledgments

The authors would like to thank the editor, associate editor, and reviewers, for their helpful comments and suggestions. This research was supported in part by NSF grants IIS1632951 and DMS-1821231, and NIH grant R01GM126550.

## Supplementary Material

Additional theoretical results, numerical examples, and all technical proofs are provided in the online Supplementary Material.

## References

Allwein, E. L., Schapire, R. E., and Singer, Y. (2001). Reducing Multiclass to Binary: a Unifying Approach for Margin Classifiers. *Journal of Machine Learning Research* **1,** 113–141.

Bartlett, P. L., Jordan, M. I., and McAuliffe, J. D. (2006). Convexity, Classification, and Risk Bounds. *Journal of the American Statistical Association* **101,** 138–156.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge.

Chen, J., Fu, H., He, X., Kosorok, M. R., and Liu, Y. (2017). Estimating Individualized Treatment Rules for Ordinal Treatments. *arXiv:1702.04755 [stat]* arXiv: 1702.04755.

Chen, J., Zhang, C., Kosorok, M. R., and Liu, Y. (2017). Double Sparsity Kernel Learning with Automatic Variable Selection and Data Extraction. *arXiv:1706.01426 [stat]* arXiv: 1706.01426.

Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. (2009). *The Elements of Statistical Learning*. New York: Springer, 2nd edition.

Herrett, E., Gallagher, A. M., Bhaskaran, K., Forbes, H., Mathur, R., van Staa, T., and Smeeth, L. (2015). Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology* **44,** 827–836.

Holman, R. R., Thorne, K. I., Farmer, A. J., Davies, M. J., Keenan, J. F., Paul, S., Levy, J. C., and 4-T Study Group (2007). Addition of biphasic, prandial, or basal insulin to oral therapy in type 2 diabetes. *The New England Journal of Medicine* **357,** 1716–1730.

Lee, Y., Lin, Y., and Wahba, G. (2004). Multicategory Support Vector

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



Machines, Theory, and Application to the Classification of Microarray Data and Satellite Radiance Data. *Journal of the American Statistical Association* **99,** 67–81.

Lin, Y. (2002). Support Vector Machines and the Bayes Rule in Classification. *Data Mining and Knowledge Discovery* **6,** 259–275.

Liu, Y. (2007). Fisher Consistency of Multicategory Support Vector Machines. In *Eleventh International Conference on Artificial Intelligence and Statistics*, pages 289–296.

Liu, Y. and Yuan, M. (2011). Reinforced Multicategory Support Vector Machines. *Journal of Computational and Graphical Statistics* **20,** 901–919.

Liu, Y., Zhang, H. H., and Wu, Y. (2011). Soft or Hard Classification? Large Margin Unified Machines. *Journal of the American Statistical Association* **106,** 166–177.

Marron, J. S., Todd, M., and Ahn, J. (2007). Distance Weighted Discrimination. *Journal of the American Statistical Association* **102,** 1267–1271.

Moodie, E. E. M., Platt, R. W., and Kramer, M. S. (2009). Estimat-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



ing response-maximized decision rules with applications to breastfeeding. *Journal of the American Statistical Association* **104,** 155–165.

Neykov, M., Liu, J. S., and Cai, T. (2016). On the Characterization of a Class of Fisher-Consistent Loss Functions and its Application to Boosting. *Journal of Machine Learning Research* **17,** 1–32.

Qian, M. and Murphy, S. A. (2011). Performance Guarantees for Individualized Treatment Rules. *Annals of statistics* **39,** 1180–1210.

Robins, J., Orellana, L., and Rotnitzky, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine* **27,** 4678–4721.

Robins, J. M. (2004). Optimal Structural Nested Models for Optimal Sequential Decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*, pages 189–326. Springer.

Simoncelli, T. (2014). Paving the Way for Personalized Medicine: FDA's Role in a New Era of Medical Product Development. Technical report, Federal Drug Administration (FDA).

Tian, L., Alizadeh, A. A., Gentles, A. J., and Tibshirani, R. (2014). A Simple Method for Estimating Interactions between a Treatment and a Large

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



Number of Covariates. *Journal of the American Statistical Association* **109,** 1517–1532.

Wu, Y. and Liu, Y. (2007). Robust Truncated Hinge Loss Support Vector Machines. *Journal of the American Statistical Association* **102,** 974–983.

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. and Liu, Y. (2013). Multicategory Large-margin Unified Machines. *Journal of Machine Learning Research* **14,** 1349–1386.

Zhang, C. and Liu, Y. (2014). Multicategory Angle-based Large-margin Classification. *Biometrika* **101,** 625–640.

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.

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.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0527



Zou, H., Zhu, J., and Hastie, T. (2008). New Multicategory Boosting Algorithms Based on Multicategory Fisher-Consistent Losses. *The Annals of Applied Statistics* **2,** 1290–1306.

Department of Statistics and Operations Research, University of North Carolina at Chapel Hill

E-mail: zhangchong101@gmail.com

Department of Biostatistics, University of North Carolina at Chapel Hill

E-mail: jgxchen@email.unc.edu

Eli Lilly and Company

E-mail: fu_haoda@lilly.com, he_xuanyao@lilly.com

Public Health Sciences Division, Fred Hutchinson Cancer Research Center

E-mail: yqzhao@fredhutch.org

Department of Statistics and Operations Research, Department of Genetics, Department of Biostatistics, Carolina Center for Genome Sciences, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill

E-mail: yfliu@email.unc.edu