Learning Optimal Group-structured Individualized Treatment Rules with Many Treatments

Haixu Ma   HAIXUMA@LIVE.UNC.EDU
Department of Statistics and Operations Research
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599, USA

Donglin Zeng   DZENG@EMAIL.UNC.EDU
Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599, USA

Yufeng Liu   YFLIU@EMAIL.UNC.EDU
Department of Statistics and Operations Research
Department of Genetics
Department of Biostatistics
Carolina Center for Genome Science
Lineberger Comprehensive Cancer Center
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599, USA

Abstract

Data driven individualized decision making problems have received a lot of attentions in recent years. In particular, decision makers aim to determine the optimal Individualized Treatment Rule (ITR) so that the expected specified outcome averaging over heterogeneous patient-specific characteristics is maximized. Many existing methods deal with binary or a moderate number of treatment arms and may not take potential treatment effect structure into account. However, the effectiveness of these methods may deteriorate when the number of treatment arms becomes large. In this article, we propose GRoup Outcome Weighted Learning (GROWL) to estimate the latent structure in the treatment space and the optimal group-structured ITRs through a single optimization. In particular, for estimating group-structured ITRs, we utilize the Reinforced Angle based Multicategory Support Vector Machines (RAMSVM) to learn group-based decision rules under the weighted angle based multi-class classification framework. Fisher consistency, the excess risk bound, and the convergence rate of the value function are established to provide a theoretical guarantee for GROWL. Extensive empirical results in simulation studies and real data analysis demonstrate that GROWL enjoys better performance than several other existing methods.

Keywords: Angle-based multicategory classification, Group structure, Individualized treatment rules, Precision medicine, Support Vector Machine.

1. Introduction

A common data-driven individualized decision making problem seeks to optimize the expected value of a specified outcome, by carefully determining the Individualized Treatment Rule (ITR) based on individual characteristics and contextual information. Since the treatment effect may contain significant heterogeneity, it is necessary to tailor treatment decision rules to different subgroups of individuals. For example, using a large-scale Electronic Health Records (EHR) database, a physician may assign an optimal individualized therapy based on a patient's specific characteristics to maximize the quality of health care (Wu et al., 2020).

Machine learning based approaches for estimating an optimal ITR have been studied intensively in the literature. These methods can be usually classified into two categories. The first category consists of model-based indirect learning methods such as modeling the conditional treatment effects given the individual characteristics (Q-learning) (Watkins, 1989; Qian and Murphy, 2011), modeling the contrast between two candidate treatment effects (A-learning) (Murphy, 2003), sub-group identification methods based on a weighted loss minimization problem (Tian et al., 2014; Chen et al., 2017), and direct learning methods (D-learning) (Qi and Liu, 2018). The second category circumvents the need for modeling conditional mean functions by directly estimating the ITR that maximizes the value function based on Inverse Probability Weighting (IPW) (Zhao et al., 2012, 2015). To combine the advantages of methods in the two categories discussed above, Zhang et al. (2012), Liu et al. (2018) and Athey and Wager (2021) proposed doubly robust augmented IPW estimation to overcome model misspecification issues. In addition, extensions to more than two treatments were studied in Zhang et al. (2021) and Qi et al. (2020).

Despite great development for estimating the optimal ITRs with a moderate number of treatments in the literature as discussed above, in some clinical problems, there can be many treatment options available. For instance, Rashid et al. (2021) analyzed the Patient-Derived Xenograft (PDX) dataset which permits the evaluation of more than 20 treatments in the allowable treatment space. Another potential challenge for learning the optimal ITR is the situation with unbalanced structure of treatment propensity scores. For example, in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study (Rush et al., 2004), the ratio of the number of patients who were provided with the cognitive therapy and the number of patients who received venlafaxine is only around 1:3. Another example is that, when studying Type 2 Diabetes (T2D) treatment patterns, Montvida et al. (2018) concluded that the baseline treatments such as Metformin and Insulin would dominate other treatment options in the EHR database.

With many treatments but limited data size, model-based indirect methods are difficult to model conditional treatment effect due to the large number of interaction terms between treatments and features. In addition, it can be impractical to fit a useful regression model without enough observations for certain treatments. Therefore, the estimated optimal ITR induced by the indirect methods can be inaccurate with large variability due to the poor performance of the regression model. On the other hand, IPW-based direct learning methods utilize a plug-in approach for possible unbalanced propensities that appear in the denominator of IPW-based value function. Suffering from unbalanced structure

of propensities in the presence of many treatments, small values in propensity scores can lead to large variability of the estimated value function.

It is interesting to point out that many treatments may work similarly for patients, due to the fact that the development of drugs is often based on intervening the same disease symptoms and mechanisms. For example, for treating depression in the STAR*D study, the 7 treatment options at Level 2 are often combined with one class of treatments involving selective serotonin reuptake inhibitors (SSRI) and the other class of treatments without SSRI because the treatments within the same class have similar treatment effects (Liu et al., 2018; Pan and Zhao, 2021). Hence, it can be helpful to identify such latent structure in the treatment space. Moreover, utilizing this latent cluster treatment structure allows us to group homogeneous treatments together and helps reducing the dimension of the treatment space. This motivates us to explore specific latent structure for treatments to identify optimal treatment groups.

To the best of our knowledge, not much has been done in the literature for estimating the optimal ITR with latent structure for treatments. Rashid et al. (2021) imposed a hierarchy binary group structure based on the conditional treatment effects for the treatments in the PDX study. This estimated group structure helps producing high-quality ITRs and identifying the important genes that are known to be associated with response to treatment. In addition, several existing methods explored combining treatment decision rules for different patients when the conditional treatment effects cannot be distinguished. Specifically, Laber et al. (2014), Ertefaie et al. (2016) and Meng et al. (2020) proposed recommending a set of near-optimal individualized treatment recommendations that are alternative to each other to a patient. However, these methods are not tailored to deal with many treatment options.

In this article, we propose estimating the latent multiple group structure of treatments and associated optimal group-structured ITRs within a single optimization. Considering grouping structure, our proposed method reduces the dimension of the treatment space and automatically clusters the treatments with similar treatment effects into the same group. In particular, we define our value function associated with both treatment partition and group-based decision rules in the IPW-based direct learning framework. The optimal treatment partition and group-based decision rules are obtained by maximizing the value function. When the treatment effects employ exact homogeneous group structure, our defined optimal partition can induce the same expected homogeneous group structure. Under the estimated optimal partition for the treatment space, the estimated group-structured ITR uses a random treatment assignment strategy, determined by randomly sampling treatment based on specific strategies within the estimated optimal treatment group. Specifically, the Reinforced Angle based Multicategory Support Vector Machines (RAMSVM) based surrogate loss function (Zhang et al., 2016) is tailored for estimating the optimal treatment group decision rules robustly in the interpretable angle-based weighted multiclass classification framework (Zhang and Liu, 2014). The group-based decision functions can give linear or non-linear decision rules to deal with complicate decision boundaries. Moreover, we prove that the surrogate loss function enjoys Fisher consistency for both group structure and group-structured ITRs. Furthermore, we present comprehensive theoretical justification on the excess risk bound, finite sample regret bound and convergence rate for our method, and allow the number of the treatment groups diverge to infinity as the sample size increases.

3

Finally, we implement efficient algorithms to solve the non-convex integer programming problem to search for the optimal partition, and the coordinate descent algorithm to solve the dual problem of RAMSVM based weighted classification problem.

The main contributions of this article are summarized as follows. Our proposed method learns the optimal ITR by identifying the latent treatment group structure in a possible large treatment space. We cluster the treatments with similar treatment effects into the same group to reduce the dimension of the possible large treatment space. In contrast to existing methods (Zhao et al., 2012; Liu et al., 2018), our method avoids using weights involving the inverse of individual treatment propensity scores, which can be close to 0 when there are many treatments. Using the treatment group propensity scores, our method can obtain more stable estimate of the value function. In addition, our method simultaneously learns the optimal group-structured ITR and clusters the treatments. Different from the two-step method (Rashid et al., 2021), we combine both supervised learning (learn the optimal ITR) and unsupervised learning (cluster the treatments) through one single optimization. Moreover, we propose an effective procedure to determine the number of unknown treatment groups. This procedure is motivated by the trade-off between the benefit and the variability of the value function. It is worth noting that our theoretical contributions are different from that in the Outcome Weighted Learning (OWL) literature (Zhao et al., 2012). In particular, we establish the generalized Fisher consistency, excess risk bound, and finite sample regret bound with respect to both *treatment partition* and *group-based decision rule* under the angle-based multi-class classification framework.

The remainder of this article is organized as follows. In Section 2, we introduce the methodology and implementation details of our proposed GRoup Outcome Weighted Learning (GROWL) method. In Section 3, we provide theoretical guarantees of GROWL. In Section 4, we conduct simulation studies to evaluate the performance of GROWL. Our method is then illustrated using the data from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study in Section 5. We conclude this article and discuss some future extensions in Section 6.

2. Methodology

In this section, we first introduce the framework of estimating optimal ITRs. Then we propose our GROWL method to estimate group-structured ITRs from the IPW-based value function.

2.1 Framework of Learning Optimal ITRs

Consider the i.i.d. training data $(X_i, A_i, R_i) \sim \mathbb{P}$ for $i = 1, \ldots, n$, where $X_i \in \mathcal{X} \subseteq \mathbb{R}^d$ denotes the patient's prognostic variables, $A_i \in \mathcal{A} = \{1, 2, \ldots, M_n\} := [M_n]$ is the treatment assignment, and $R_i \in \mathbb{R}$ is the observed outcome for each patient i. Suppose that the number of treatments M_n may diverge to infinity with a certain rate as the sample size n increases since we consider the large treatment space. We assume that the larger outcome is better and R is bounded. Let $\big(R(a)\big)_{a \in \mathcal{A}} \in \mathbb{R}^{M_n}$ be the potential outcome. In addition, define the propensity function $p(a|x) := \mathbb{P}(A = a|X = x)$ and the unknown mean-outcome function $\mu(a|x) := \mathbb{E}[R(a)|X = x]$ for $a \in [M_n]$. An Individualized Treatment Rule (ITR) $D \in \mathcal{D}$ is a map from the covariate space $\mathcal{X}$ to the treatment space $\mathcal{A}$ and $\mathcal{D} \subseteq \mathcal{A}^{\mathcal{X}}$ is a prespecified

ITR class. Our goal is to find the optimal ITR $D^* \in \mathcal{D}$, that maximizes the expected outcome, known as the value function (Zhao et al., 2012). Specifically, the value of an ITR D is defined as

$$\mathcal{V}(D) = \mathbb{E}\left[\sum_{a=1}^{M_n} \mathbb{I}[D(X) = a]R(a)\right] = \mathbb{E}\left[\sum_{a=1}^{M_n} \mathbb{I}[D(X) = a]\mu(a|X)\right].$$

Next we state the following identifiability assumptions (Rubin, 1974): (1) Consistency: $R = \sum_{a\in\mathcal{A}} \mathbb{I}[A = a]R(a)$; (2) Unconfoundedness: for each $a \in [M_n], R(a) \perp\!\!\!\perp A \mid X$; (3) Positivity: $p(a|x) > 0$ for any $x \in \mathcal{X}$. If the above assumptions are satisfied, $\mathcal{V}(D)$ can be written as the following two equivalent forms:

$$\mathcal{V}(D) = \mathbb{E}\left[\sum_{a=1}^{M_n} \mathbb{I}[D(X) = a]\mathbb{E}[R|X, A = a]\right] \tag{1}$$

$$= \mathbb{E}\left[\frac{\mathbb{I}[D(X) = A]}{p(A|X)}R\right]. \tag{2}$$

Based on (1), model-based Q-learning methods (Qian and Murphy, 2011) first give an estimate for $\widehat{\mathbb{E}}[R|X, A]$ (Q-function), then the optimal ITR $D^*(x)$ is estimated from solving $\arg\max_{a\in\mathcal{A}} \widehat{\mathbb{E}}[R|X = x, A = a]$. However, due to the large number of treatment options M_n and possible unbalanced structure of the propensity score $p(A|X)$, we may not have enough observations for some specific treatments to fit the regression model. Consequently, $\widehat{\mathbb{E}}[R|X, A]$ can be inaccurate due to potential poor performance of the regression model and the estimated optimal ITR may have large variability. Another common approach is to estimate the value function based on (2) using empirical data, and then directly search for the optimal ITR D that maximizes the empirical value function $\mathbb{E}_n\left[\frac{\mathbb{I}[D(X)=A]}{p(A|X)}R\right]$ (Zhao et al., 2012). Note that the propensity score $p(A|X)$ appears in the denominator of $\mathbb{E}_n\left[\frac{\mathbb{I}[D(X)=A]}{p(A|X)}R\right]$. For the case with many treatments where insufficient data are observed for some specific treatments, it is likely to have the propensity score $p(A|X)$ close to 0 for some treatments. Hence, this can cause large variability of the empirical estimate for the value function.

2.2 GRoup Outcome Weighted Learning (GROWL)

Next we introduce our proposed GROWL method using the idea of latent group structure for the treatment space. We consider M_n treatments can be partitioned into K_n disjoint latent groups where $2 \leqslant K_n \leqslant M_n$. We allow K_n go to infinity with a certain rate as the sample size increases. Denote δ as the partition of $\mathcal{A}$, which is a map from $\mathcal{A}$ to $\{1, 2, \ldots, K_n\} := [K_n]$. Under δ, denote $G_k^\delta = \{a \mid \delta(a) = k, a \in \mathcal{A}\}$ as the k-th treatment set for $k \in [K_n]$. Intuitively, for a reasonable partition, the treatments that belong to the same treatment group should have similar treatment effects. In contrast, the treatment effects from different treatment groups should have relatively large differences. Hence, we need to first define the optimal partition δ^* that maximizes the expected outcome.

To start with, we define the following group-structured ITR class, denoted as $\mathcal{D} = \bigcup_\delta \mathcal{D}_\delta$. Specifically, associated with a partition δ, a group-structured ITR in $\mathcal{D}_\delta$ is obtained from a

random treatment assignment strategy π_δ given as

$$\pi_\delta(a|x) = \mathbb{I}[\delta(a) = D_g(x)]\frac{p(a|x)}{p(\delta(a)|x)}, \tag{3}$$

where D_g is a group-based decision rule mapping from $\mathcal{X}$ to treatment group space $[K_n]$, and $p(\delta(a)|x) := \mathbb{P}[A \in G^\delta_{\delta(a)}|X = x]$ is the propensity score for the $\delta(a)$-th treatment group under δ. Then, for a given partition δ and a group-based decision rule D_g, the value function of group-structured ITR equals to the expectation of weighted conditional treatment effects. With these notations in place, we can express the value of group-structured ITR $\mathcal{V}_1(\delta, D_g)$ as follows:

$$\begin{aligned}
\mathcal{V}_1(\delta, D_g) &= \mathbb{E}\left[\sum_{k=1}^{K_n} \mathbb{I}[D_g(X) = k] \sum_{a:\delta(a)=k} \frac{p(a|X)}{p(\delta(a)|X)} \mathbb{E}[R|X, A = a]\right] \\
&= \mathbb{E}\left[\mathbb{I}[D_g(X) = \delta(A)]\frac{R}{p(\delta(A)|X)}\right].
\end{aligned} \tag{4}$$

Remark 1 *Based on the definition in (3), group-structured ITR is obtained by first estimating the group-based decision rule $D_g(x)$, then sampling one treatment from group $G^\delta_{D_g(x)}$ with probabilities proportional to propensity scores. Note that sampling proportional to propensity scores has the following advantages and interpretations: (a) the propensity score $p(A|X)$ showing up in the denominator of the value function in (2) would be cancelled by that in the sample strategy term. Thus, only the group-based propensity $p(\delta(A)|X)$ appears in the value function $\mathcal{V}_1(\delta, D_g)$. Hence, it can give a more stable estimate for $\mathcal{V}_1(\delta, D_g)$ using empirical data especially when propensity scores are small for some treatments; (b) it makes sense to assign higher probabilities to choose prevalent treatments as they are often affordable and acceptable in practice.*

For any given δ, the optimal group-based decision rule D_g^δ is given by

$$D_g^\delta \in \underset{D_g:\mathcal{X}\to[K_n]}{\arg\max} \ \mathcal{V}_1(\delta, D_g), \tag{5}$$

and the corresponding optimal value for δ is

$$\mathcal{V}_1^*(\delta) = \mathcal{V}_1(\delta, D_g^\delta). \tag{6}$$

The optimal partition δ^* is defined as

$$\delta^* \in \arg\max_\delta \mathcal{V}_1^*(\delta) := \Delta^*, \tag{7}$$

where Δ^* is the optimal equivalent partition class and each element in this set achieves the maximum value. Observing that $\sum_{a:\delta(a)=k} \frac{p(a|X)}{p(\delta(a)|X)}\mathbb{E}[R|A = a, X] = \mathbb{E}[R|A \in G_k^\delta, X]$ for $k \in [K_n]$, the optimal value for δ in (6) can be written as

$$\begin{aligned}
\mathcal{V}_1^*(\delta) &= \max_{D_g:\mathcal{X}\to[K_n]} \mathbb{E}_X\left[\sum_{k=1}^{K_n} \mathbb{I}[D_g(X) = k]\mathbb{E}[R|A \in G_k^\delta, X]\right] \\
&= \mathbb{E}_X\left[\max_{k\in[K_n]} \mathbb{E}[R|A \in G_k^\delta, X]\right].
\end{aligned}$$

Hence, the optimal partition δ^* has the following interpretation. Averaging over the marginal distribution of X, the maximum of conditional treatment effects under the group domain, which is a mixture mean of conditional treatment effects under the individual treatment domain, is optimized under the optimal partition δ^*.

It is worth noting that, when treatment effects have homogeneous structure, our defined optimal partition δ^* in (7) would lead to this expected natural group structure. In particular, treatment effects have homogeneous structure if treatments can be partitioned into K_0 homogeneous groups $\mathcal{G}^0 = \{G_1^0, G_2^0, \ldots, G_{K_0}^0\}$ and the treatment effects are identical within each treatment group set G_k^0 for $k \in [K_0]$. For each $k \in [K_0]$ and each pair of treatments i, j within the same treatment group G_k^0, we have $\mathbb{E}[R|A = i, X] = \mathbb{E}[R|A = j, X] = \mathbb{E}[R|A \in G_k^0, X]$ a.e. in X. Meanwhile, for each pair of treatments r, s that belong to two different treatment groups respectively, $\mathbb{E}[R|A = r, X] \neq \mathbb{E}[R|A = s, X]$ holds with a positive probability in X. In this case, GROWL aims to combine treatments with identical treatment effects based on homogeneous structure $\mathcal{G}_0$ to reduce the dimension of treatment space and learn the optimal group-structured ITR. Denote δ^0 to be the partition that induces the group structure $\mathcal{G}_0$, then the following Lemma 2 holds, which demonstrates that our δ^* is properly defined for the expected homogeneous structure.

Lemma 2 *Suppose $K_n = K_0$ and the defined optimal partition δ^* is unique. Then, we have $\delta^* = \delta^0$.*

Next we illustrate how to solve $(\delta^*, D_g^*) \in \arg\max_{\delta, D_g : \mathcal{X} \to [K_n]} \mathcal{V}_1(\delta, D_g)$ in order to get an estimate for the optimal partition δ^* and the associated optimal group-based decision rule D_g^* under the IPW-based direct learning framework. Since

$$\mathcal{V}_1(\delta, D_g) + \mathcal{R}(\delta, D_g) = \mathbb{E}\left[\frac{R}{p(\delta(A)|X)}\right],$$

where the risk function $\mathcal{R}(\delta, D_g) := \mathbb{E}\left[\frac{R}{p(\delta(A)|X)}\mathbb{I}[D_g(X) \neq \delta(A)]\right]$. Hence, maximizing $\mathcal{V}_1(\delta, D_g)$ is equivalent to minimizing the generalized risk function:

$$\widetilde{\mathcal{R}}(\delta, D_g) = \mathcal{R}(\delta, D_g) - \mathbb{E}\left[\frac{R}{p(\delta(A)|X)}\right].$$

In practice, we use empirical risk minimization to approximate the generalized risk function $\widetilde{\mathcal{R}}(\delta, D_g)$ by $\mathbb{E}_n\left[\frac{R}{p(\delta(A)|X)}\mathbb{I}[D_g(X) \neq \delta(A)]\right] - \mathbb{E}_n\left[\frac{R}{p(\delta(A)|X)}\right]$, where $\mathbb{E}_n$ is the empirical average based on the training data. To alleviate the difficulty of the discontinuity and non-convexity of the 0-1 loss in the treatment group-based weighted misclassification error, for each $x \in \mathcal{X}$ and $a \in \mathcal{A}$, we propose replacing the 0-1 loss function $\mathbb{I}[D_g(x) \neq \delta(a)]$ by a angle-based loss $L_\phi(\delta(a), \boldsymbol{f}(x))$, as proposed in Zhang and Liu (2014) and Zhang et al. (2016). The group-based decision rule D_g is determined by the decision function $\boldsymbol{f}$ mapping from $\mathcal{X}$ to $\mathbb{R}^{K_n-1}$. Specifically, We encode the k-th treatment group as a vector $\mathbf{W}_k \in \mathbb{R}^{K_n-1}$ with

$$\mathbf{W}_k = \begin{cases} (K_n - 1)^{-1/2}\mathbf{1}_{K_n-1}, & k = 1, \\ -(1 + \sqrt{K_n})/(K_n - 1)^{3/2}\mathbf{1}_{K_n-1} + \left(\frac{K_n}{K_n-1}\right)^{1/2}\mathbf{e}_{k-1}, & k = 2, 3, \ldots, K_n, \end{cases}$$

where $\mathbf{1}_{K_n-1}$ is a vector of ones of length $K_n - 1$, and $\mathbf{e}_k \in \mathbb{R}^{K_n-1}$ is a vector with the k-th element equal to one, and zero elsewhere. Specifically, when $K_n = 2$, we have $\mathbf{W}_1 = 1$ and $\mathbf{W}_2 = -1$, which corresponds to the standard coding procedure in the binary classification problem. In addition, based on this coding procedure, one can check that, this treatment group simplex is symmetric with all vertices share an equal distance from each other in $\mathbb{R}^{K_n-1}$. We refer Zhang and Liu (2014) for more details about the angle-based classification method. The RAMSVM-based loss consists of a convex combination of two loss functions

$$L_\phi\big(\delta(a), \boldsymbol{f}(x)\big) := (1-\gamma)\left[\sum_{k \neq \delta(a)}(1 + \langle \mathbf{W}_k, \boldsymbol{f}(x)\rangle)^+\right] + \gamma\left[(K_n - 1 - \langle \mathbf{W}_{\delta(a)}, \boldsymbol{f}(x)\rangle)^+\right], \quad (8)$$

where $\gamma \in [0, 1]$. The final group-based decision rule is obtained from

$$D_g(x) = \underset{k \in [K_n]}{\arg\max}\langle \mathbf{W}_k, \boldsymbol{f}(x)\rangle.$$

The corresponding optimization problem is

$$\min_{\delta, \boldsymbol{f}_n \in \mathcal{F}_n}\left\{\mathbb{E}_n\left[\frac{R}{p(\delta(A)|X)}L_\phi\big(\delta(A), \boldsymbol{f}(X)\big)\right] - K_n\mathbb{E}_n\left[\frac{R}{p(\delta(A)|X)}\right] + \lambda_n\|\boldsymbol{f}\|_{\mathcal{F}_n}^2\right\}, \quad (9)$$

where $\mathcal{F}_n$ is a pre-specified function class of $\{\boldsymbol{f} : \mathcal{X} \to \mathbb{R}^{K_n-1}\}$, λ_n is a tuning parameter, and $\|\cdot\|_{\mathcal{F}_n}$ is the functional penalty associated with $\mathcal{F}_n$ to overcome overfitting.

2.3 Implementation of GROWL

We introduce efficient algorithms to solve the optimization problem (9). To this end, we follow the procedure proposed by Liu et al. (2018) to replace R with the residual $R - s(X)$. The rational is that removing the main effect that is independent of treatment should not affect the treatment decision while using residuals can significantly reduce the variability of weights to improve algorithm performance. However, note that the residual $R - s(X)$ can take negative values, which would break the convexity of the minimization problem. In this case, we can switch the treatment group to other different treatment groups under the uniform sampling procedure. Specifically, it can be checked that, for any fixed δ, the following two optimization problems are equivalent:

$$\min_{\boldsymbol{f}} \mathbb{E}\left[\frac{R - s(X)}{p(\delta(A)|X)}L_\phi\big(\delta(A), \boldsymbol{f}(X)\big)\right] \iff \min_{\boldsymbol{f}} \mathbb{E}\left[\frac{(R - s(X))^+}{p(\delta(A)|X)}L_\phi\big(\delta(A), \boldsymbol{f}(X)\big)\right]$$
$$+ (K_n - 1)\mathbb{E}\left[\frac{(R - s(X))^-}{p(\delta(A)|X)}L_\phi\big(\widetilde{\delta}(A), \boldsymbol{f}(X)\big)\right], \quad (10)$$

where $u^+ = \max(u, 0), u^- = \max(-u, 0)$, and the conditional distribution of the random variable $\widetilde{\delta}(A)$ is determined by $Pr(\widetilde{\delta}(A) = k|\delta(A), X) = \frac{1}{K_n-1}$ for $k \neq \delta(A)$ and 0 for $k = \delta(A)$. In this way, the weight term can be easily computed by $\frac{(R-s(X))^+}{p(\delta(A)|X)} + (K_n - 1)\frac{(R-s(X))^-}{p(\delta(A)|X)}$. For simplicity of notations, in the following of this section, we use $R^\star$ to denote $(R - s(X))^+ + (K_n - 1)(R - s(X))^-$. The derivations of why the two optimization problems in (10) are equivalent can be seen in Appendix C.

Next we specify the decision function $\boldsymbol{f} : \mathcal{X} \to \mathbb{R}^{K_n-1}$ in a product Reproducing Kernel Hilbert Space (RKHS) $\mathcal{F}_n = \otimes_{k=1}^{K_n-1}\mathcal{H}_\kappa^k$. We develop efficient algorithms to solve (9) after

replacing R with $R^\star$ and switching treatments for observations with negative residuals. Our implementation consists of two steps. Step 1: under any fixed partition candidate δ, we convert the RAMSVM-based weighted classification problem (9) to a dual quadratic programming problem with box constraints. Then we solve the dual problem using coordinate descent algorithm to obtain the estimated optimal decision function under δ, denoted as $\widehat{\boldsymbol{f}}^\delta$; Step 2: Treatment partition estimation step: after plugging $(\delta, \widehat{\boldsymbol{f}}^\delta)$ back into (9) to get the value (smaller is preferred) for the candidate δ, we propose to use the genetics algorithm (Goldberg and Holland, 1988), which is a stochastic search and evolutionary algorithm to obtain the optimal $\widehat{\delta}$. Alternatively, we can also use the coordinate descent type of greedy algorithm to adjust the partition.

For step 1, we propose the following algorithm to solve the weighted classification problem when specifying $\boldsymbol{f}$ in the product linear space or product RKHS. Specifically, let $\omega_i = \frac{R_i^\star}{p(\delta(A_i)|X_i)}$ be the weight for subject $i \in [n]$. For $x \in \mathcal{X}$, denote $\boldsymbol{f}(x) = (f_1(x), \ldots, f_{K_n-1}(x))^T$ and $\mathbf{W}_{j,k}$ represents the k-th element of $\mathbf{W}_j$, where $j \in [K_n]$ and $k \in [K_n - 1]$.

For linear decision functions, we assume $f_k(x) = x^T \boldsymbol{\beta}_k$ with $k \in [K_n - 1]$, where $\boldsymbol{\beta}_k$'s are our parameters of interest. The penalty term $\|\boldsymbol{f}\|_{\mathcal{F}_n}^2 = \sum_{k=1}^{K_n-1} \boldsymbol{\beta}_k^T \boldsymbol{\beta}_k$. Note that we include the intercepts in $\mathcal{X}$ to simplify notation. After introducing slack variables for (9) and taking partial derivative of the Lagrangian function with respect to each $\boldsymbol{\beta}_k$ and slack variables, we can derive the following dual problem with respect to the Lagrangian multiplier $\alpha_{i,j}$ and obtain $(\widehat{\alpha}_{i,j})_{i\in[n];j\in[K_n]}$ by solving

$$
\begin{aligned}
\min_{(\alpha_{i,j})_{i\in[n];j\in[K_n]}} & \frac{1}{2n\lambda} \sum_{k=1}^{K_n-1} \left[\sum_{i=1}^{n} \alpha_{i,\delta(A_i)} \mathbf{W}_{\delta(A_i),k} X_i - \sum_{i=1}^{n} \sum_{j\neq\delta(A_i)} \alpha_{i,j} \mathbf{W}_{j,k} X_i \right]^T \\
& \left[\sum_{i=1}^{n} \alpha_{i,\delta(A_i)} \mathbf{W}_{\delta(A_i),k} X_i - \sum_{i=1}^{n} \sum_{j\neq\delta(A_i)} \alpha_{i,j} \mathbf{W}_{j,k} X_i \right] \\
& - \sum_{i=1}^{n} \alpha_{i,\delta(A_i)}(K_n - 1) - \sum_{i=1}^{n} \sum_{j\neq\delta(A_i)} \alpha_{i,j}, \\
\text{s.t. } & 0 \leqslant \alpha_{i,j} \leqslant \omega_i \left(\gamma \mathbb{I}[j = \delta(A_i)] + (1-\gamma)\mathbb{I}[j \neq \delta(A_i)]\right) \ (i \in [n]; j \in [K_n]).
\end{aligned}
\tag{11}
$$

Moreover, we can calculate

$$
\widehat{\boldsymbol{\beta}}_k = \frac{1}{n\lambda} \left[\sum_{i=1}^{n} \widehat{\alpha}_{i,\delta(A_i)} \mathbf{W}_{\delta(A_i),k} X_i - \sum_{i=1}^{n} \sum_{j\neq\delta(A_i)} \widehat{\alpha}_{i,j} \mathbf{W}_{j,k} X_i \right].
$$

Note that one can verify that the quadratic optimization function in (11) is strictly convex with respect to each $\alpha_{i,j}$. The constraints in (11) are box constraints. Therefore, (11) can be solved efficiently by the well-known coordinate descent algorithm. Compared with standard Quadratic Programming (QP) algorithms for solving the dual problem, the coordinate descent algorithm can enjoy a faster computational speed and obtain more accurate solutions (Zhang et al., 2016). The final estimated group-based ITR is obtained by

$\hat{D}_g(x) = \arg\max_{j \in [K_n]} \langle \mathbf{W}_j, \hat{\boldsymbol{f}}(x) \rangle$, where $\hat{\boldsymbol{f}}(x) = (\hat{f}_1(x), \ldots, \hat{f}_{K_n-1}(x))^T$ and $\hat{f}_k(x) = x^T \hat{\boldsymbol{\beta}}_k$ for $k \in [K_n - 1]$.

To deal with more complicated functions, we generalize the linear approach to obtain a nonlinear decision function in RKHS. To begin with, denote κ to be the corresponding kernel function and $\mathbf{G} = \left(\kappa(X_i, X_{i'}) \right)_{i,i' \in [n]}$ to be the gram matrix. We assume $\mathbf{G}$ is invertible. Denote $\mathbf{G}_i$ to be the i-th column of $\mathbf{G}$. By using the L_2 norm in $\otimes_{k=1}^{K_n-1} \mathcal{H}_\kappa^k$ for the penalty term, i.e., $\|\boldsymbol{f}\|_{\mathcal{F}_n}^2 = \sum_{k=1}^{K_n-1} \boldsymbol{\theta}_k^T \mathbf{G} \boldsymbol{\theta}_k$, we can represent the decision function as $f_k(x) = \theta_{k,0} + \sum_{i=1}^n \theta_{k,i} \kappa(X_i, x)$ for $k \in [K_n - 1]$. Here, $\boldsymbol{\theta}_k = (\theta_{k,1}, \ldots, \theta_{k,n})^T$ is our kernel product coefficient vector for $k \in [K_n - 1]$. Similar to the steps in linear case, $(\hat{\alpha}_{i,j})_{i \in [n]; j \in [K_n]}$ can be obtained by solving the following dual problem

$$
\begin{aligned}
\min_{(\alpha_{i,j})_{i \in [n]; j \in [K_n]}} & \frac{1}{2n\lambda} \sum_{k=1}^{K_n-1} \left[\sum_{i=1}^n \alpha_{i,\delta(A_i)} \mathbf{W}_{\delta(A_i),k} \mathbf{G}_i - \sum_{i=1}^n \sum_{j \neq \delta(A_i)} \alpha_{i,j} \mathbf{W}_{j,k} \mathbf{G}_i \right]^T \\
& \mathbf{G}^{-1} \left[\sum_{i=1}^n \alpha_{i,\delta(A_i)} \mathbf{W}_{\delta(A_i),k} \mathbf{G}_i - \sum_{i=1}^n \sum_{j \neq \delta(A_i)} \alpha_{i,j} \mathbf{W}_{j,k} \mathbf{G}_i \right] \\
& + \frac{1}{2n\lambda} \sum_{k=1}^{K_n-1} \left[\sum_{i=1}^n \alpha_{i,\delta(A_i)} \mathbf{W}_{\delta(A_i),k} - \sum_{i=1}^n \sum_{j \neq \delta(A_i)} \alpha_{i,j} \mathbf{W}_{j,k} \right]^2 \\
& - \sum_{i=1}^n \alpha_{i,\delta(A_i)}(K_n - 1) - \sum_{i=1}^n \sum_{j \neq \delta(A_i)} \alpha_{i,j}, \\
\text{s.t. } & 0 \leqslant \alpha_{i,j} \leqslant \omega_i \left(\gamma \mathbb{I}[j = \delta(A_i)] + (1-\gamma) \mathbb{I}[j \neq \delta(A_i)] \right) \ (i \in [n]; j \in [K_n]).
\end{aligned}
\tag{12}
$$

Furthermore, we can obtain

$$
\hat{\boldsymbol{\theta}}_k = \frac{1}{n\lambda} \mathbf{G}^{-1} \left[\sum_{i=1}^n \hat{\alpha}_{i,\delta(A_i)} \mathbf{W}_{\delta(A_i),k} \mathbf{G}_i - \sum_{i=1}^n \sum_{j \neq \delta(A_i)} \hat{\alpha}_{i,j} \mathbf{W}_{j,k} \mathbf{G}_i \right]
$$

$$
\hat{\theta}_{k,0} = \frac{1}{n\lambda} \left[\sum_{i=1}^n \hat{\alpha}_{i,\delta(A_i)} \mathbf{W}_{\delta(A_i),k} - \sum_{i=1}^n \sum_{j \neq \delta(A_i)} \hat{\alpha}_{i,j} \mathbf{W}_{j,k} \right].
$$

One can check that (12) can be solved in an analogous manner as (11). The final decision function is obtained from $\hat{f}_k(x) = \hat{\theta}_{k,0} + \sum_{i=1}^n \hat{\theta}_{k,i} \kappa(X_i, x)$ for $k \in [K_n - 1]$. More details about how the original problem (9) is transformed to the dual problems (11) and (12) in step 1 are provided in Appendix C.

For step 2, after we plug $(\delta, \hat{\boldsymbol{f}}^\delta)$ back to (9), we get the value for the candidate partition δ. We formulate the partition space as the discrete problem of partitioning M_n numbers into K_n groups. To solve this non-convex integer programming problem, when M_n and K_n are relatively small, we can implement the genetics algorithm using the R package called `GA` introduced in Scrucca (2013). Furthermore, if $K_n \ll M_n$ and both K_n and M_n are large, then the total number of partitions can be very large. Consequently, the genetics

algorithm can be time consuming. Hence, to deal with this case, we propose a coordinate descent type of greedy algorithm to search for the optimal partition iteratively. Specifically, at each iteration, we minimize (9) by successively adjusting the group assignment for one specific treatment while holding the assignment of other treatments fixed. We go through each treatment in a cyclic fashion until convergence. The initial partition can be obtained via clustering the fitted conditional expected outcome for each treatment. The conditional expected outcome can be roughly estimated by L_1 penalized regression (Qian and Murphy, 2011), random forest or latent supervised clustering using the pairwise fusion penalty (Chen et al., 2021).

2.4 Selection of Treatment Group Number

Our analysis so far treats the group number K_n as given. However, K_n is typically unknown in practice. We propose the following effective procedure to determine K_n. We first randomly split the observed data $\{(X_i, A_i, R_i)\}_{i=1}^n$ into two folds. For each group number $1 \leqslant K \leqslant M_n$, denote the $\widehat{\delta}_K$ and $\widehat{D}_{g,K}$ as the estimated optimal partition and associated group-based decision rule learned from one fold of the training data based on the implementations discussed in Section 2.3. Then, we calculate the estimated value function $\widehat{\mathcal{V}}_1(\widehat{\delta}_K, \widehat{D}_{g,K})$ for each K using

$$\widehat{\mathcal{V}}_1(\widehat{\delta}_K, \widehat{D}_{g,K}) = \frac{\mathbb{E}_n\big[R\mathbb{I}[\widehat{D}_{g,K}(X) = \widehat{\delta}_K(A)]/p(\widehat{\delta}_K(A)|X)\big]}{\mathbb{E}_n\big[\mathbb{I}[\widehat{D}_{g,K}(X) = \widehat{\delta}_K(A)]/p(\widehat{\delta}_K(A)|X)\big]}, \tag{13}$$

where $\mathbb{E}_n$ denotes the empirical mean of the other fold of observed data. Note that when $K = 1$, all the treatments are grouped together and the associated $\widehat{\mathcal{V}}_1(\widehat{\delta}_1, \widehat{D}_{g,1})$ corresponds to the value when we randomly recommend treatments. Thus, we can obtain the estimated benefit function of the estimated group-structured ITR for each K with

$$\widehat{B}_{en}(K) = \widehat{\mathcal{V}}_1(\widehat{\delta}_K, \widehat{D}_{g,K}) - \widehat{\mathcal{V}}_1(\widehat{\delta}_1, \widehat{D}_{g,1})$$
$$= \widehat{\mathcal{V}}_1(\widehat{\delta}_K, \widehat{D}_{g,K}) - \mathbb{E}_n[R].$$

We replicate the above process T times. For each replication $t = 1, 2, \ldots, T$, denote the benefit function as $\widehat{B}_{en}^{(t)}(K)$. We propose the following procedure that can be interpreted as the trade-off between the benefit and the variability of the estimated group-structured ITR to determine the optimal K_n:

$$\widehat{K}_n = \underset{1 < K \leqslant M_n}{\arg\max} \left\{ \underset{t}{\texttt{mean}} \left[\{\widehat{B}_{en}^{(t)}(K)\}_{t=1}^T \right] \right\}. \tag{14}$$

One can also replace R with $R - \widehat{s}(X)$ in (13) to remove variability coming from estimating the main effect.

The group number estimator (14) can be interpreted as follows. Denote δ_K^* as the optimal partition when the group number is K. For $1 \leqslant K \leqslant M_n$, let $B_{en}(K) := \mathbb{E}_X\big[\max_{k \in [K]} \mathbb{E}[R|A \in G_k^{\delta_K^*}, X] - R\big]$ be the maximum benefit when the group number is specified as K and $B_{en}^* := \mathbb{E}_X\big[\max_{a \in \mathcal{A}} \mathbb{E}[R|X, A = a] - R\big] = B_{en}(M_n)$ be the optimal benefit. The optimal benefit B_{en}^* corresponds to the case that we do not consider any

group structure in the treatment space. We first consider the case that the treatments have homogeneous group structure $\mathcal{G}_0$ discussed in Section 2.2 and the true value of group number equals to K_0. Then, similar to the proof of Lemma 2 shown in Appendix C, one can check that if setting $K > K_0$, then the optimal partition defined in (7) would result in over identified group structures. These over identified optimal group structures can be any refinement of $\mathcal{G}_0$. In particular, these refined optimal partitions δ_K^*'s of $\delta_{K_0}^*$ all lead to the same optimal benefit $B_{en}(K) = B_{en}^*$ when $K_0 \leqslant K \leqslant M_n$. In this case, the bias of the value function is 0. However, the stochastic error bound and the convergence rate of the estimated value function shown in Theorem 6 in Section 3.3 increases with a polynomial rate $\mathcal{O}(K_n^2)$ as K_n becomes larger. This demonstrates that as K_n increases, the variability of the group-structured ITR becomes larger. Based on Xia et al. (2009), this variability is involved in (14) by using T times of sample splitting. Therefore, the group number selection procedure (14) incorporates the penalization of variability to avoid the over identified group structures when $K \geqslant K_0$ for the homogeneous case.

When the homogeneous case does not hold, then our optimal benefit $B_{en}(K)$ may be strictly less than the optimal benefit B_{en}^* for all $1 \leqslant K < M_n$. One can check that $B_{en}(K)$ is a non-decreasing function as K increases by induction, and finally equals to B_{en}^* when $K = M_n$. For the non-homogeneous case, together with the same analysis of the increasing variability as K increases, the selection of group number can be interpreted as a trade-off between the benefit and the variability of the group-structured ITR.

Remark 3 *In practice, when the estimated optimal $\widehat{K}_n$ is relatively small, we conclude that the gain of smaller variability would dominate the potential loss of the benefit for the group-structured ITR. Hence, a homogeneous or nearly-homogeneous group structure of the treatment effects is expected. In contrast, when $\widehat{K}_n$ is close to M_n, the gain of higher benefit would dominate the negative effect of the large variability. For this scenario, we expect that the treatments may behave very differently from each other on the same patient. When $\widehat{K}_n = M_n$, GROWL would learn the optimal ITR under the individual treatment domain without grouping. Its performance becomes similar to the traditional methods that do not consider the treatment structure.*

3. Theoretical Properties

In this section, we establish Fisher consistency of the optimal partition and associated group-based ITR for GROWL. We further obtain an excess risk bound and derive the convergence rate for the value function with the diverging group number K_n.

3.1 Fisher Consistency

Denote $(\delta_\phi^*, \boldsymbol{f}_\phi^*)$ as the optimal partition and associated optimal decision function under the generalized L_ϕ risk function $\widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f})$:

$$(\delta_\phi^*, \boldsymbol{f}_\phi^*) \in \underset{\delta, \boldsymbol{f}:\mathcal{X} \to \mathbb{R}^{K_n-1}}{\arg\min} \left\{ \widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}) := \mathcal{R}_\phi(\delta, \boldsymbol{f}) - K_n \mathbb{E}\left[\frac{R}{p(\delta(A)|X)}\right] \right\}, \qquad (15)$$

where the L_ϕ risk function $\mathcal{R}_\phi(\delta, \boldsymbol{f})$ is defined as

$$\mathcal{R}_\phi(\delta, \boldsymbol{f}) := \mathbb{E}\left[\frac{R}{p(\delta(A)|X)} L_\phi\big(\delta(A), \boldsymbol{f}(X)\big)\right]. \tag{16}$$

Let $\boldsymbol{f}_\phi^\delta$ be the optimal decision function for any fixed δ: $\boldsymbol{f}_\phi^\delta = \arg\min_{\boldsymbol{f}:\mathcal{X}\to\mathbb{R}^{K_n-1}} \widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f})$. Under the angle-based weighted classification framework, a classifier is said to be Fisher consistent if for each partition δ and $x \in \mathcal{X}$, the predicted treatment group has the maximum conditional group treatment effect under δ:

$$\arg\max_{k\in[K_n]}\langle \mathbf{W}_k, \boldsymbol{f}_\phi^\delta(x)\rangle = \arg\max_{k\in[K_n]}\mathbb{E}[R|A \in G_k^\delta, X = x].$$

For our problem, we establish the generalized Fisher consistency results for both partition δ and the decision rule under the group domain if we choose L_ϕ to be the surrogate loss function, i.e., the derived optimal decision rule is the same as the one using the 0-1 loss. In particular, the following generalized Fisher consistency holds:

Theorem 4 *Let δ^* and Δ^* be defined in (7) and $(\delta_\phi^*, \boldsymbol{f}_\phi^*)$ be defined in (15). Denote $D_g^*(x)$ to be the optimal deterministic group-based ITR under $\delta^* \in \Delta^*$, which leads to the Bayesian rule $\arg\max_{k\in[K_n]}\mathbb{E}[R|A \in G_k^{\delta^*}, X = x]$. If $\gamma \in [0, \frac{1}{2}]$, then we have $\delta_\phi^* \in \Delta^*$ and $\arg\max_{k\in[K_n]}\langle\mathbf{W}_k, \boldsymbol{f}_\phi^*(x)\rangle = D_g^*(x)$.*

3.2 Excess Risk

Next we establish the excess risk of 0-1 loss can be upper bounded by that of RAMSVM loss. To start with, we introduce the following notations. For any group-based ITR D_g, there exists a decision function $\boldsymbol{f} : \mathcal{X} \to \mathbb{R}^{K_n-1}$ such that $D_g(x) = \arg\max_{k\in[K_n]}\langle\mathbf{W}_k, \boldsymbol{f}(x)\rangle$. Similar to the definition of $\mathcal{R}(\delta, D_g)$, we define

$$\mathcal{R}(\delta, \boldsymbol{f}) = \mathbb{E}\left[\mathbb{I}[\arg\max_{k\in[K_n]}\langle\mathbf{W}_k, \boldsymbol{f}(X)\rangle \neq \delta(A)]\frac{R}{p(\delta(A)|X)}\right],$$

and denote $\widetilde{\mathcal{R}}(\delta, \boldsymbol{f}) = \mathcal{R}(\delta, \boldsymbol{f}) - \mathbb{E}[R/p(\delta(A)|X)]$. Then the generalized Bayesian risk is denoted as $\widetilde{\mathcal{R}}^* = \inf_{\delta,\boldsymbol{f}}\left\{\widetilde{\mathcal{R}}(\delta, \boldsymbol{f})|\delta, \boldsymbol{f} : \mathcal{X} \to \mathbb{R}^{K_n-1}\right\}$. In terms of the value function $\mathcal{V}_1(\delta, D_g)$ defined in (4), we observe that $\mathcal{V}_1(\delta^*, D_g^*) - \mathcal{V}_1(\delta, D_g) = \widetilde{\mathcal{R}}(\delta, \boldsymbol{f}) - \widetilde{\mathcal{R}}^*$. Note that in GROWL, we replace the 0-1 loss with the L_ϕ loss. Recall we have defined the generalized L_ϕ risk function $\widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f})$ in (15) and (16). Similarly, the infimum of generalized L_ϕ risk function is defined as $\widetilde{\mathcal{R}}_\phi^* = \inf_{\delta,\boldsymbol{f}}\left\{\widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f})|\delta, \boldsymbol{f} : \mathcal{X} \to \mathbb{R}^{K_n-1}\right\}$. In addition, under any fixed δ, let $\boldsymbol{f}^\delta = \arg\min_{\boldsymbol{f}:\mathcal{X}\to\mathbb{R}^{K_n-1}} \mathcal{R}(\delta, \boldsymbol{f})$ be the optimal decision function under the group domain. Denote $\boldsymbol{f}^* = \arg\min_{\boldsymbol{f}:\mathcal{X}\to\mathbb{R}^{K_n-1}} \mathcal{R}(\delta^*, \boldsymbol{f})$.

The following theorem shows the relationship between the generalized excess 0-1 risk $\widetilde{\mathcal{R}}(\delta, \boldsymbol{f}) - \widetilde{\mathcal{R}}^*$ and generalized excess L_ϕ risk $\widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}) - \widetilde{\mathcal{R}}_\phi^*$ under some bounded restrictions for $\boldsymbol{f}$.

Theorem 5 (Bound for excess risk) *For any partition δ, any measurable function $\boldsymbol{f}$: $\mathcal{X} \to \mathbb{R}^{K_n-1}$ such that $\langle \mathbf{W}_k, \boldsymbol{f}(x) \rangle \in [-1, K_n - 1]$ holds for $\forall x \in \mathcal{X}$ and $\forall k \in [K_n]$, and any probability distribution for (X, A, R), we have*

$$\widetilde{\mathcal{R}}(\delta, \boldsymbol{f}) - \widetilde{\mathcal{R}}^* \leqslant \widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}) - \widetilde{\mathcal{R}}_\phi^*.$$

Note that Theorem 5 is different from Theorem 3.2 in Zhao et al. (2012) in the sense that we consider multiple treatments, and dealing with both partition δ and the decision function.

3.3 Convergence Rate

Define the estimated optimal partition $\widehat{\delta}_n$ and group-based decision function $\widehat{\boldsymbol{f}}_n$ as

$$(\widehat{\delta}_n, \widehat{\boldsymbol{f}}_n) \in \underset{\delta, \boldsymbol{f} \in \otimes_{k=1}^{K_n-1} \mathcal{H}_\kappa^k}{\arg\min} \left\{ \mathbb{E}_n \left[\frac{R}{p(\delta(A)|X)} L_\phi(\delta(A), \boldsymbol{f}(X)) - K_n \frac{R}{p(\delta(A)|X)} \right] + \lambda_n \|\boldsymbol{f}\|_{\mathcal{F}_n}^2 \right\}. \quad (17)$$

For a fixed partition δ, denote the optimal estimated group-based decision function as

$$\widehat{\boldsymbol{f}}_n^\delta = \underset{\boldsymbol{f} \in \otimes_{k=1}^{K_n-1} \mathcal{H}_\kappa^k}{\arg\min} \left\{ \mathbb{E}_n \left[\frac{R}{p(\delta(A)|X)} L_\phi(\delta(A), \boldsymbol{f}(X)) - K_n \frac{R}{p(\delta(A)|X)} \right] + \lambda_n \|\boldsymbol{f}\|_{\mathcal{F}_n}^2 \right\}. \quad (18)$$

Specifically, for the decision function class, we restrict our consideration to the product RKHS associated with Radial Basis Function (RBF) kernels:

$$\kappa(x, x') = \exp(-\sigma_n^2 \|x - x\|^2), \ x, x' \in \mathcal{X},$$

where $\sigma_n > 0$ is a bandwidth parameter varying with n. For theoretical convenience, we assume $\widehat{\boldsymbol{f}}_n$ satisfies the extra bounded constraint $\langle \mathbf{W}_k, \widehat{\boldsymbol{f}}_n(x) \rangle \in [-1, K_n - 1]$ for $\forall x \in \mathcal{X}$ and $\forall k \in [K_n]$. This constraint does not show up in the algorithm discussed in Section 2.3 because it makes the computation algorithm in Section 2.3 become more complicate and inefficient. Our numerical experience suggests that removing the constraint for $\widehat{\boldsymbol{f}}_n$ can yield better classification performance than including it.

Next we show that $\widetilde{\mathcal{R}}(\widehat{\delta}_n, \widehat{\boldsymbol{f}}_n)$ converges to $\widetilde{\mathcal{R}}^*$ and equivalently, the value function $\mathcal{V}_1(\widehat{\delta}_n, \widehat{D}_{g,n})$ converges to $\mathcal{V}_1(\delta^*, D_g^*)$ where the estimated group-based ITR $\widehat{D}_{g,n}(x) :=$ $\arg\max_{k \in [K_n]} \langle \mathbf{W}_k, \widehat{\boldsymbol{f}}_n(x) \rangle$. We start with introducing the following quantity:

$$\mathcal{A}_{\sigma_n}^\delta(\lambda_n) = \inf_{\boldsymbol{f} \in \otimes_{k=1}^{K_n-1} \mathcal{H}_\kappa^k} \left\{ \lambda_n \|\boldsymbol{f}\|_\kappa^2 + \widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}) - \widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}^\delta) \right\}.$$

For a fixed δ, the term $\mathcal{A}_{\sigma_n}^\delta(\lambda_n)$ describes how well the regularized RAMSVM-risk approximates the optimal RAMSVM-risk in the RKHS. This quantity is often referred as the approximation error term (Steinwart and Scovel, 2007). Specifically, when $K_n = 2$, Steinwart and Scovel (2007) proposed a geometric noise assumption to upper bound $\mathcal{A}_{\sigma_n}^\delta(\lambda_n)$ in the context of hinge loss based SVM classification problem under any fixed δ. In this paper, we generalize the geometric noise assumption so that we can upper bound $\mathcal{A}_{\sigma_n}^\delta(\lambda_n)$ for the multicategory group-based ITR problem under the RAMSVM-based loss. Under each δ, denote the difference of two group treatment effects as

$$\eta_{i,j}^\delta(x) = \mathbb{E}[R|A \in G_i^\delta, X = x] - \mathbb{E}[R|A \in G_j^\delta, X = x], \ i \neq j \ \text{and} \ i, j \in [K_n].$$

Define the decision regions for each pair of treatment groups $i, j \in [K_n]$ to be $\mathcal{O}_{ij}^{\delta,-} = \{x \in \mathcal{X} \mid \eta_{i,j}^{\delta}(x) < 0\}$ and $\mathcal{O}_{ij}^{\delta,+} = \{x \in \mathcal{X} \mid \eta_{ij}^{\delta}(x) > 0\}$. Then let $\mathcal{X}_i^{\delta} = \bigcap_{j \neq i} \mathcal{O}_{ij}^{\delta,+}$ for $i \in [K_n]$ be the subset of $\mathcal{X}$ where the treatment effect of group i dominates any other treatment groups under partition δ. Denote the function $\eta^{\delta}(x) = \sum_{i=1}^{K_n} \mathbb{I}[x \in \mathcal{X}_i^{\delta}] \sup_{j \neq i} |\eta_{i,j}^{\delta}(x)|$ as the maximum difference of the group treatment effects for each region $\mathcal{X}_i^{\delta}$. Furthermore, denote the following distance function to the decision boundary as $\Delta^{\delta}(x) = \sum_{i=1}^{K_n} \mathbb{I}[x \in \mathcal{X}_i^{\delta}] \inf_{j \neq i} \operatorname{dist}(x, \mathcal{X}_j^{\delta})$, where $\operatorname{dist}(x, \mathcal{O})$ is the distance between a point x and a set $\mathcal{O}$. Then we define the following generalized geometric noise assumption:

Assumption 1 (Generalized geometric noise assumption) *For a fixed δ and $i \in [K_n]$, there exists a constant $U > 0$ such that for any $t > 0$, we have*

$$\mathbb{E}_X \left[exp\left(-\frac{(\Delta^{\delta}(X))^2}{t} \right) |\eta^{\delta}(X)| \right] \leqslant U t^{qd/2},$$

where d is the dimension of $\mathcal{X}$ and q serves as the geometric noise exponent.

One can check that when $K_n = 2$, Assumption 1 is consistent with Definition 2.3 in Steinwart and Scovel (2007) and Definition 3.8 in Zhou et al. (2017). In some sense, this geometric noise exponent q describes the concentration of the measure $|\eta^{\delta}(x)| dP_X$ near the decision boundary. In the case of complete separation, i.e., $\eta^{\delta}(x) > \varphi_0 > 0$ for some constant φ_0, q can be as large as possible.

Let Γ_n be the total number of partitions. Recall the definition of $\mathcal{V}_1^*(\delta)$ and Δ^* in (6) and (7). Consider K_n can diverge to infinity as the sample size n increases. Then, for any n, there exist a positive gap $\Psi_n := \frac{1}{2} \inf_{\delta \notin \Delta^*} \{\mathcal{V}_1^*(\delta^*) - \mathcal{V}_1^*(\delta)\} > 0$, such that $\widetilde{\mathcal{R}}(\bar{\delta}, \boldsymbol{f}^{\bar{\delta}}) - \widetilde{\mathcal{R}}^* = \mathcal{V}_1^*(\delta^*) - \mathcal{V}_1^*(\bar{\delta}) > \Psi_n$ holds for any non-optimal partition $\bar{\delta} \notin \Delta^*$. Here, Ψ_n can be interpreted as the signal to characterize the minimum distance of the value function between the optimal partitions and any other non-optimal partitions. Intuitively, we need the signal Ψ_n to be large enough so that we can distinguish the optimal partitions from non-optimal partitions. Now we are ready to present the main theorem for the convergence rate of GROWL.

Theorem 6 (Convergence rate of $\widehat{\delta}_n$ and value function for diverging M_n, K_n)
Suppose the generalized geometric noise assumption holds for an optimal partition $\delta^* \in \Delta^*$ with exponent $0 < q < \infty$ and constant U. Further assume $|\frac{R}{p(\delta^*(A)|X)}| \leqslant Z_n$ and $\theta(\mathcal{X}) \leqslant M_1$, where $\{Z_n\}$ is a sequence, and $\theta(\mathcal{X})$ is denoted as the volume of $\mathcal{X}$. Let $\epsilon_n = K_n^{1/2} Z_n^{1/2} n^{-1/2} \lambda_n^{1/2} + K_n Z_n n^{-1/2} + K_n Z_n \left((\lambda_n)^{-\frac{2}{2+v} + \frac{(2-v)(1+\theta)}{(2+v)(1+q)}} n^{-\frac{2}{2+v}} + n^{-1} \lambda_n^{-1} + \lambda_n^{\frac{q}{q+1}} \right)$. Then, if $M_n K_n^2 / \log n \to 0, n\lambda_n \to \infty, \Psi_n/\epsilon_n \to \infty$ hold for any $\theta > 0, 0 < v < 2$, and take $\sigma_n = \lambda_n^{-1/(q+1)d}$, the following results hold:

(I) *For any $\theta > 0, 0 < v < 2$, we have*

$$Pr(\widehat{\delta}_n \in \Delta^*) = 1 - \mathcal{O}(\Gamma_n \epsilon_n);$$

(II) $\widetilde{\mathcal{R}}(\widehat{\delta}_n, \widehat{\boldsymbol{f}}_n) - \widetilde{\mathcal{R}}^* = \mathcal{O}_p(\Gamma_n \epsilon_n).$

15

For Theorem 6, we can choose $\lambda_n = n^{-\frac{q+1}{3q+1}}$, and let (θ, v) be sufficiently small. When data are well separated under one of the optimal partitions, q can be sufficiently large. Note that if the group-based propensity score $p(\delta^*(A)|X)$ has balanced structure under δ^*, then as $K_n \to \infty$, $p(\delta^*(A)|X)$ would decay uniformly. Thus, we have $|\frac{R}{p(\delta^*(A)|X)}| \leqslant Z_n = \mathcal{O}(K_n)$. In this case, the convergence rate for the value function can achieve $\mathcal{V}_1(\delta^*, D_g^*) - \mathcal{V}_1(\widehat{\delta}_n, \widehat{D}_{g,n}) = \widetilde{\mathcal{R}}(\widehat{\delta}_n, \widehat{f}_n) - \widetilde{\mathcal{R}}^* = \mathcal{O}_p(\Gamma_n K_n^2 n^{-\frac{1}{3}})$.

Note that ϵ_n in Theorem 6 describes the rate of the stochastic error bound for the regret. Thus, the assumption $\Psi_n/\epsilon_n \to \infty$ implies that, the signal Ψ_n should be large enough and dominates the noise ϵ_n so that $\widehat{\delta}_n$ can finally belong to Δ^*. To better under illustrate this assumption, we consider the following simple example. Suppose the number of treatment groups is fixed. Then, under the homogeneous case, the best partition among the partition set without the optimal partitions corresponds to the case that only one treatment is misclustered. In this case, one can check that the defined signal term $\Psi_n = \mathcal{O}(1/M_n)$ based on the proof of Lemma 2. Hence, due to $M_n = o(\log n)$, the assumption $\Psi_n/\epsilon_n \to \infty$ is satisfied since ϵ_n decays at a polynomial rate of n.

The pipeline for proving Theorem 6 is stated as the follow two steps: First, for any optimal partition $\delta^* \in \Delta^*$, we establish a finite sample bound for the difference between the expected outcome using the estimated group-based decision function $\widehat{f}_n^{\delta^*}$ based on the training data and that of the optimal group-based decision function f^{δ^*} under δ^*; Second, due to $\Psi_n/\epsilon_n \to \infty$, as n goes to infinity, the stochastic error ϵ_n of $\widetilde{\mathcal{R}}(\bar{\delta}, f^{\bar{\delta}})$ arising from using $\mathbb{E}_n$ to estimate $\mathbb{E}$ in (17) would be dominated by the gap Ψ_n for any $\bar{\delta} \notin \Delta^*$. Hence, $\widehat{\delta}_n$ in (17) would finally belong to Δ^* when n is sufficient large since $\delta^* \in \Delta^*$ maximizes the value function $\mathcal{V}_1$. Then, the convergence rate is determined by the rate of $Pr(\widehat{\delta}_n \in \Delta^*) \to 1$ and the convergence rate of the first step when treating the partition is fixed as $\delta^* \in \Delta^*$. Specifically, the novelty of our technical proof arises from bounding the approximation bias term $\mathcal{A}_{\sigma_n}^{\delta}(\lambda_n)$ with order $\mathcal{O}(K_n^2 \lambda_n^{q/(q+1)})$ and deriving the finite value reduction bound in a multicategory setting. The intermediate results deriving from the first step generalize Lemma 3.9 in Zhou et al. (2017) and Theorem 3.4 in Zhao et al. (2012) from binary treatments to multiple treatment groups K_n that may diverge to infinity as the sample size increases. More details are provided in Appendix B.

For the case that the number of treatments M_n and treatment groups K_n are fixed, it is straightforward to derive the following Corollary 7 from Theorem 6. Note that for the fixed group number case, $\Psi_n/\epsilon_n \to \infty$ is trivially satisfied since Ψ_n is a constant.

Corollary 7 (Convergence rate of $\widehat{\delta}_n$ and value function for fixed M_n and K_n)
Suppose the generalized geometric noise assumption holds for an optimal partition $\delta^* \in \Delta^*$ with exponent $0 < q < \infty$ and constant U. Further assume $|\frac{R}{p(\delta^*(A)|X)}| \leqslant Z_0$ and $\theta(\mathcal{X}) \leqslant M_1$ where Z_0 and M_1 are constants. Then, if $n\lambda_n \to \infty$, and take $\sigma_n = \lambda_n^{-1/(q+1)d}$, the following results hold:

(I) For any $\theta > 0, 0 < v < 2$, we have

$$Pr(\widehat{\delta}_n \in \Delta^*) = 1 - \mathcal{O}\left((n\lambda_n)^{-1/2} + (\lambda_n)^{-\frac{2}{2+v} + \frac{(2-v)(1+\theta)}{(2+v)(1+q)}} n^{-\frac{2}{2+v}} + \lambda_n^{\frac{q}{q+1}}\right);$$

$$(II)\ \widetilde{\mathcal{R}}(\widehat{\delta}_n, \widehat{\mathbf{f}}_n) - \widetilde{\mathcal{R}}^* = \mathcal{O}_p\left((n\lambda_n)^{-1/2} + (\lambda_n)^{-\frac{2}{2+v} + \frac{(2-v)(1+\theta)}{(2+v)(1+q)}} n^{-\frac{2}{2+v}} + \lambda_n^{\frac{q}{q+1}}\right).$$

4. Simulation Studies

We evaluate the finite-sample performance of our proposed method using several simulation studies.

4.1 Homogeneous Settings

In this simulation study, we consider the setting where the treatment responses for the treatments in the same group are equivalent, but differ for the treatments across different groups. We generate 10-dimensional independent prognostic variables $X_1, \ldots, X_{10}$, following $U[-1,1]$. The outcome R is normally distributed with $\mathbb{E}[R|A,X] = 1 + 2X_1 + X_2 + 0.5X_3 + T_0(X,A)$ and standard deviation 1, where $T_0(X,A)$ reflects the interaction between the treatment and the prognostic variables. In addition, we assume that the treatment effects have the homogeneous grouping structure $\mathcal{G}^0$ induced by δ^0 discussed in Section 2.2. Specifically, we consider the following three scenarios:

Scenario 1. $M_n = 10, K_n = 2, \mathcal{G}^0 = \{\{1,2,3,4,5\}, \{6,7,8,9,10\}\}$ and $T_0(X,A) = 1.8(0.2 - X_1 - X_2)(\mathbb{I}[A \in \{1,2,3,4,5\}] \times (-1) + \mathbb{I}[A \in \{6,7,8,9,10\}] \times 1)$;

Scenario 2. $M_n = 10, K_n = 2, \mathcal{G}^0 = \{\{1,2,3,4,5\}, \{6,7,8,9,10\}\}$ and $T_0(X,A) = 3.5(0.8 - X_1^2 - X_2^2)(\mathbb{I}[A \in \{1,2,3,4,5\}] \times (-1) + \mathbb{I}[A \in \{6,7,8,9,10\}] \times 1)$;

Scenario 3. $M_n = 15, K_n = 3, \mathcal{G}^0 = \{\{1,2,3,4,5\}, \{6,7,8,9,10\}, \{11,12,13,14,15\}\}$ and $T_0(X,A) = 5((-0.2 + X_1 + 2X_2)\mathbb{I}[A \in \{1,2,3,4,5\}] + (0.3 + 2X_1 + X_2)\mathbb{I}[A \in \{6,7,8,9,10\}] + (-0.2 + 3X_1)\mathbb{I}[A \in \{11,12,13,14,15\}])$.

Scenario 1 corresponds to 10 treatment arms belonging to two treatment groups with underlying linear decision boundaries whereas Scenario 2 considers the circle decision boundary. Scenario 3 includes 15 treatments compared with the first two and deals with three treatment groups with linear decision boundary. Since our studies are especially interested in the case that the propensity score of some specific treatments may be very small, we perform the following two designs varying from balanced to unbalanced designs within each scenario:

(a) Balanced Design: $p(a|x) = \frac{1}{M_n}$ for each $a \in \mathcal{A} = [M_n]$ and each $x \in \mathcal{X}$;

(b) Unbalanced Design: The value of $p(a|x)$ for some specific treatments can be very small compared with other treatments.

Under the unbalanced design, for each $x \in \mathcal{X}$, the propensity scores for the first two scenarios are set to be $((\frac{1}{20}, \frac{1}{20}, \frac{1}{20}, \frac{3}{20}, \frac{1}{5}), (\frac{1}{20}, \frac{1}{20}, \frac{1}{10}, \frac{1}{10}, \frac{1}{5}))$ while the propensity scores equal to $((\frac{1}{20}, \frac{1}{20}, \frac{1}{20}, \frac{11}{120}, \frac{11}{120}), (\frac{1}{20}, \frac{1}{20}, \frac{1}{20}, \frac{11}{120}, \frac{11}{120}), (\frac{1}{20}, \frac{1}{20}, \frac{7}{90}, \frac{7}{90}, \frac{7}{90}))$ for Scenario 3. In addition, we conduct more simulation settings when the propensity scores are more unbalanced and may depend on the covariates. These additional results are shown in Appendix D. For GROWL, we use the linear kernel for Scenarios 1 and 3 and utilize the Gaussian kernel for Scenario 2 corresponding to different shapes of the decision boundary. The tuning parameter λ_n is chosen to maximize the empirical value function $\mathbb{E}_n[R\mathbb{I}[\widehat{D}(X) = A]/p(A|X)]/\mathbb{E}_n[\mathbb{I}[\widehat{D}(X) = A]/p(A|X)]$ by 10-fold cross-validation among $\{\frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 16\}$. For the Gaussian kernel, we fix the inverse bandwidth of the kernel σ_n with $1/(2\widehat{\tau}^2)$, where $\widehat{\tau}$ is the median of

the pairwise Euclidean distance of the covariates (Wu and Liu, 2007). The treatment group number is determined by the trade-off procedure (14) with $T = 50$ shown in Section 2.4.

The following four methods are compared under each scenario:

(a) SL: Super Learner based Q-learning method to estimate $\mathbb{E}[R|X, A]$ (Polley and Van Der Laan, 2010);

(b) AD: Multi-armed Angle-based Direct learning using linear terms for Scenarios 1 and 3 and polynomial terms for Scenario 2 (Qi et al., 2020);

(c) PLS: L_1-Penalized Least Squares method (Qian and Murphy, 2011), which estimates $\mathbb{E}[R|X, A]$ using the basis sets $(1, X, A, XA)$ for Scenarios 1 and 3, and $(1, X^2, A, X^2A)$ for Scenario 2;

(d) GROWL: Our proposed method.

SL aims to find the optimal combination of multiple estimated Q-functions by minimizing the cross-validated risk. For the collection of learning algorithms, we include ridge regression (Hoerl and Kennard, 1970), elastic net (Zou and Hastie, 2005), random forest (Breiman, 2001), XGBoosting (Chen and Guestrin, 2016), and neural network (Venables and Ripley, 2013). We refer Polley and Van Der Laan (2010) for more implementation details about SL. In addition, the tuning parameters in AD and PLS are selected to maximize $\mathbb{E}_n\big[R\mathbb{I}[\widehat{D}(X) = A]/p(A|X)\big]/\mathbb{E}_n\big[\mathbb{I}[\widehat{D}(X) = A]/p(A|X)\big]$ by 10-fold cross-validation.

We evaluate the above methods using the empirical value function and the group-based misclassification rate between the estimated group decision rules and the true group decision rules on an independently generated testing data of size $10,000$. The empirical value function is calculated by the mean of treatment effects under the empirical distribution of X based on the estimated decision rule. Note that under the homogeneous setting, the maximum group-based treatment effect equals to the maximum individual treatment based effect. Hence, the misclassification rate under the group domain is equivalent to the misclassification rate under the individual treatment domain. For each scenario, the training sample sizes vary from 200, 400 to 600 and we replicate the simulations for 200 times.

We present the empirical value function of each scenario under different designs using boxplots in Figure 1. Results of group-based misclassification rates are included in Figure 5 of Appendix D. We also report the square root of Mean Square Error (MSE) of the empirical value function and the misclassification rate in Table 1. Based on Lemma 2 and Theorem 4, we have shown that δ^0 should equal to the optimal partition δ^* defined in (7), and δ^* should equal to the optimal partition δ^*_ϕ derived from (15). Accordingly, the Ratio column in Table 1 reports the ratio of our estimated $\widehat{\delta}_n$ exactly being δ^0 among the 200 replications. It can be seen that $\widehat{\delta}_n$, the estimation of δ^*_ϕ, converges to δ^0 with a high ratio as n becomes larger, which confirms part (I) of Theorem 6. In general, as the trial design becomes more unbalanced, all these methods perform worse, in the sense that MSE becomes larger for each scenario. Without considering the group structure for the treatments, SL, PLS and AD suffer from the inaccurate estimation of functions related to individual-treatment effects because of the small amount of observations for some specific treatments. However, GROWL estimates the group-structured ITR, which reduces the dimension of the treatment space and clusters the treatments that employ similar treatment effects into the same group. In addition, since GROWL estimates the ITR in the treatment group domain, the variance of the value function shrinks quicker than other methods as the training sample size increases. As is demonstrated in Figure 1 and Table 1, our method outperforms other methods in most

cases with higher empirical value functions, smaller misclassfication rates, and especially lower variabilities for both evaluation criteria.

Table 1: Results for Ratio of finding the optimal partition δ^0 and square root of **MSE** of Empirical Value Function and Misclassification Rate evaluated on the independent test data under the **homogeneous** settings. The best values are in bold.

	$n = 200$			$n = 400$			$n = 600$		
	Ratio(%)	Value	Misclassification	Ratio(%)	Value	Misclassification	Ratio(%)	Value	Misclassification
Balanced Design									
Scenario **1**									
SL	—	0.109	0.112	—	0.046	0.069	—	0.028	0.052
AD	—	0.132	0.126	—	0.056	0.098	—	0.044	0.086
PLS	—	0.070	0.092	—	0.041	0.065	—	0.030	0.057
GROWL	97.0	**0.046**	**0.068**	99.5	**0.014**	**0.044**	100	**0.008**	**0.034**
Scenario **2**									
SL	—	0.181	0.167	—	0.068	0.097	—	0.052	0.078
AD	—	0.106	0.133	—	0.073	0.111	—	0.069	0.107
PLS	—	0.086	0.112	—	0.047	0.079	—	0.041	0.069
GROWL	92.5	**0.079**	**0.101**	98.5	**0.036**	**0.069**	99.0	**0.027**	**0.060**
Scenario **3**									
SL	—	0.567	0.260	—	0.169	0.154	—	0.119	0.108
AD	—	1.310	0.463	—	0.752	0.369	—	0.565	0.342
PLS	—	**0.401**	**0.226**	—	0.199	0.177	—	0.191	0.170
GROWL	67.5	0.511	0.269	92.0	**0.125**	**0.119**	97.5	**0.080**	**0.089**
Unbalanced Design									
Scenario **1**									
SL	—	0.258	0.164	—	0.073	0.088	—	0.046	0.068
AD	—	0.416	0.216	—	0.270	0.145	—	0.075	0.101
PLS	—	0.146	0.129	—	0.045	0.067	—	0.032	0.058
GROWL	91.5	**0.085**	**0.078**	99.0	**0.022**	**0.047**	99.0	**0.009**	**0.035**
Scenario **2**									
SL	—	0.255	0.199	—	0.127	0.139	—	0.082	0.102
AD	—	0.246	0.197	—	0.123	0.138	—	0.089	0.120
PLS	—	0.166	0.148	—	0.072	0.093	—	0.047	0.075
GROWL	83.5	**0.080**	**0.103**	97.5	**0.037**	**0.070**	98.0	**0.029**	**0.061**
Scenario **3**									
SL	—	0.675	0.288	—	0.183	0.162	—	0.121	0.112
AD	—	1.484	0.481	—	0.814	0.386	—	0.664	0.355
PLS	—	**0.423**	**0.236**	—	0.231	0.184	—	0.198	0.174
GROWL	56.0	0.570	0.281	83.5	**0.132**	**0.124**	98.0	**0.083**	**0.090**

4.2 Non-homogeneous Settings

In many cases, it is possible that the treatment effects do not have exactly homogeneous grouping structure $\mathcal{G}^0$ assumed in Section 4.1. In this section, we perform nearly homogeneous and nonhomogeneous scenarios to examine our method. Specifically, we generalize Scenario 1 in Section 4.1 with the following Scenario 4 indexed by a parameter $\theta > 0$:

Scenario 4: $M_n = 10$, $T_0(X, A) = 1.8\big(0.2 - X_1 - X_2\big)\big(\mathbb{I}[A \in \{1,2,3,4,5\}] \times (-1 - \frac{A}{\theta}) + \mathbb{I}[A \in \{6,7,8,9,10\}] \times (1 + \frac{A-5}{\theta})\big)$.

The parameter θ determines the level of heterogeneity of treatment effects. As θ becomes smaller, the treatment effects become more diverse and thus the group structure
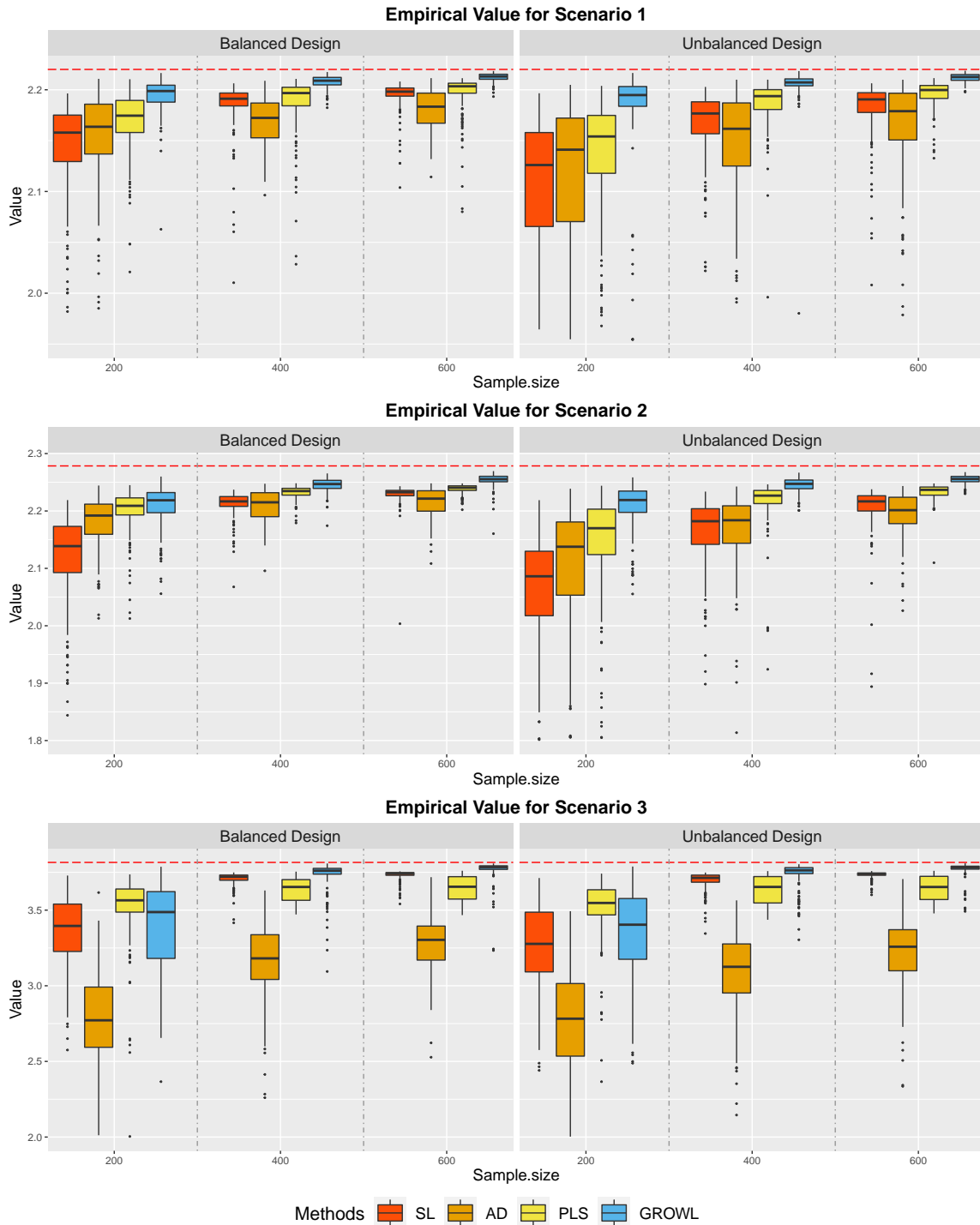
Figure 1: Boxplots of Empirical Value Function evaluated on the independent test data under the **homogeneous** settings. Red horizontal dashed lines show oracle values for each scenario.

tends to disappear. When $\theta = +\infty$, Scenario 4 has the exact homogeneous structure $\mathcal{G}^0 = \big\{\{1,2,3,4,5\},\{6,7,8,9,10\}\big\}$ shown in Scenario 1. We vary θ from $40, 20, 10$ to 5. In particular, the simulation scenarios in this section only focus on the unbalanced design. For GROWL, the treatment group number is determined by the trade-off procedure in (14). Other simulation settings and comparison methods are the same as those in Section 4.1.

Similar to Section 4.1, we provide boxplots of the empirical value function in Figure 2, and MSE of the empirical value function in Table 2. When θ decreases from 40 to 5, the treatment effects vary from nearly homogeneous structure to nonhomogeneous structure. For nearly homogeneous cases with $\theta = 40$ and 20, our method still outperforms other methods while for nonhomogeneous cases with $\theta = 10$ and 5, PLS performs the best. The results are consistent with Remark 3. For each $1 \leqslant K_n < 10$, the maximum benefit of group-structured ITR $B_{en}(K_n)$ is strictly less than the optimal benefit B_{en}^* because the conditional treatment effects within the same group are different under δ_K^* in nonhomogeneous cases. When $\theta = +\infty$, these two values are equal. As θ decreases, the gap between these two values increases. Based on our simulation results, for each θ and training sample size, over 95% of the replications suggest $\widehat{K}_n = 2$ based on the trade-off procedure (14), and over 90% of the estimated partition still equals to the same two-group structure $\mathcal{G}^0 = \big\{\{1,2,3,4,5\},\{6,7,8,9,10\}\big\}$ as the homogeneous setting. Hence, in Scenario 4, GROWL tends to sacrifice the benefit while retain small variability of the value function, and the gain of small variability continues to dominate the loss of benefit when θ decreases from 40 to 5. In Figure 2, one can observe that the empirical value of our method would not converge to the optimal value (shown by dashed lines) and a positive gap would exist. In addition, when sample size increases from 200 to 600, the relative improvement ratio in terms of the MSE for GROWL decreases when θ has smaller values. However, due to the usage of the group structure, the variability of our method is very small compared with others shown in Figure 2. Therefore, the trade-off between the benefit and variability of the value function for the group-structured ITR estimated by GROWL is clear. From Table 2, GROWL is still competitive in nonhomogeneous settings in terms of the MSE criterion.

5. Application to the STAR*D Study

In this section, we apply our proposed GROWL to analyze the data from the STAR*D study (Rush et al., 2004). The STAR*D study performed research on outpatients with nonpsychotic major depressive disorder. The goal of the study was to compare various treatment options for the patients who failed to obtain a satisfactory response with citalopram (CIT), an initial antidepressant treatment. The primary outcome was measured by the Quick Inventory of Depressive Symptomatology (QIDS) score ranging from 0 to 27, where higher scores indicate more severe depression.

The STAR*D data consist of four levels. In our analysis, we focus on the 1407 eligible patients who received treatments at Level 2. In particular, at Level 2, patients were asked to indicate their preference of either switching to one of the 4 different treatments, i.e., bupropion (BUP), cognitive therapy (CT), sertraline (SER), and venlafaxine (VEN), or augmenting their existing CIT with 3 options, i.e., CIT+BUP, CIT+buspirone (BUS), and CIT+CT. If a patient indicated no preference, then he/she was assigned to any of the above 7 treatments. To encourage the active collaboration and shared decision-making
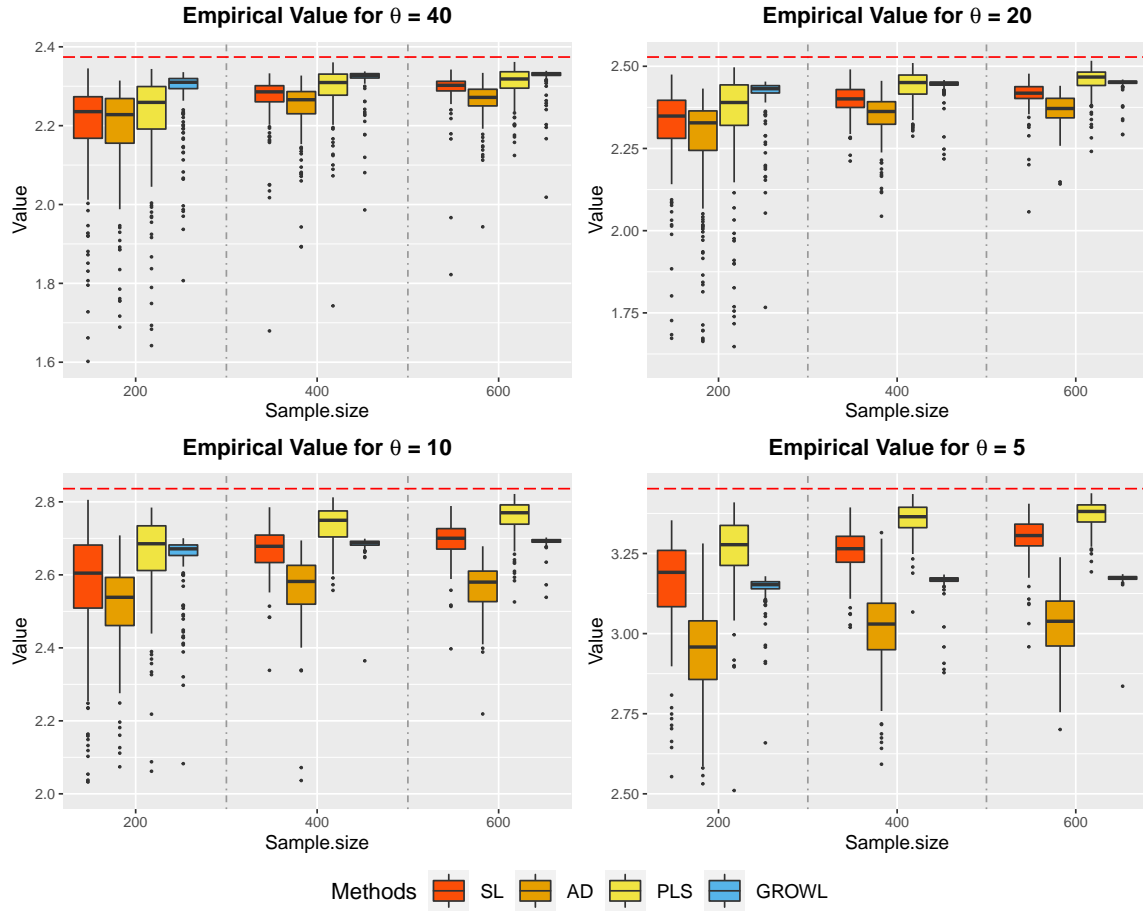
Figure 2: Boxplots of Empirical Value Function evaluated on the independent test data under the **nonhomogeneous** settings and unbalanced design. Red horizontal dashed lines show oracle values for each scenario.

Table 2: Results for square root of **MSE** of Empirical Value Function evaluated on independent test data under the **nonhomogeneous** settings and unbalanced designs. The best values are in bold.

	$n = 200$	$n = 400$	$n = 600$
	$\theta = 40$		
SL	0.265	0.136	0.091
AD	0.424	0.174	0.132
PLS	0.212	0.102	0.076
GROWL	**0.119**	**0.069**	**0.058**
	$\theta = 20$		
SL	0.266	0.137	0.123
AD	0.480	0.223	0.167
PLS	0.245	0.100	0.084
GROWL	**0.141**	**0.091**	**0.079**
	$\theta = 10$		
SL	0.306	0.178	0.151
AD	0.546	0.298	0.294
PLS	0.214	**0.112**	**0.091**
GROWL	**0.206**	0.154	0.146
	$\theta = 5$		
SL	0.339	0.208	0.165
AD	0.664	0.454	0.442
PLS	**0.233**	**0.125**	**0.092**
GROWL	0.315	0.293	0.282

with patients, we consider the patients' preference as part of the intervenable treatment options and assume the future patients' preference can be intervened when recommending treatments. Hence, we have a total of $3+4+7 = 14$ treatment options and these preference-related treatments are often called patient-centered medications in the literature (Robinson et al., 2008). Figure 3 demonstrates the distribution of the observation numbers for the 14 patient-centered treatments. Due to the relatively large treatment space and the unbalanced structure of propensities of treatments, it can be seen that only a few observations were obtained for many treatment options, especially for the treatments in the "no-preference" (Nop-) group.

We apply four methods (SL, PLS, AD, GROWL) discussed in Section 4 to estimate the optimal treatment rules among the 14 treatment options for patients. Specifically, the reward R in our study is calculated by the reduction of QIDS score from the start to the end of Level 2. Hence, a higher value of R is preferred. Feature variables X include QIDS score at the start of Level 2, reduction of QIDS score during Level 1, and other demographic variables such as gender, race, age, education level, employment status, and marital status, etc. The estimated propensities $\widehat{p}(A|X)$ are obtained from fitting the multinomial logistic regression model. For PLS, we use terms $(1, X, X^2, A, XA, X^2A)$ to fit the L_1 penalized linear regression to estimate conditional treatment effects; For AD, polynomial terms are also included as the basis set of decision functions; For GROWL, we implement the Gaussian kernel for decision functions. Comparisons of all these methods were based on 200 repetitions of three-fold cross-validation, where two folds are used to train the model. For our proposed GROWL, we
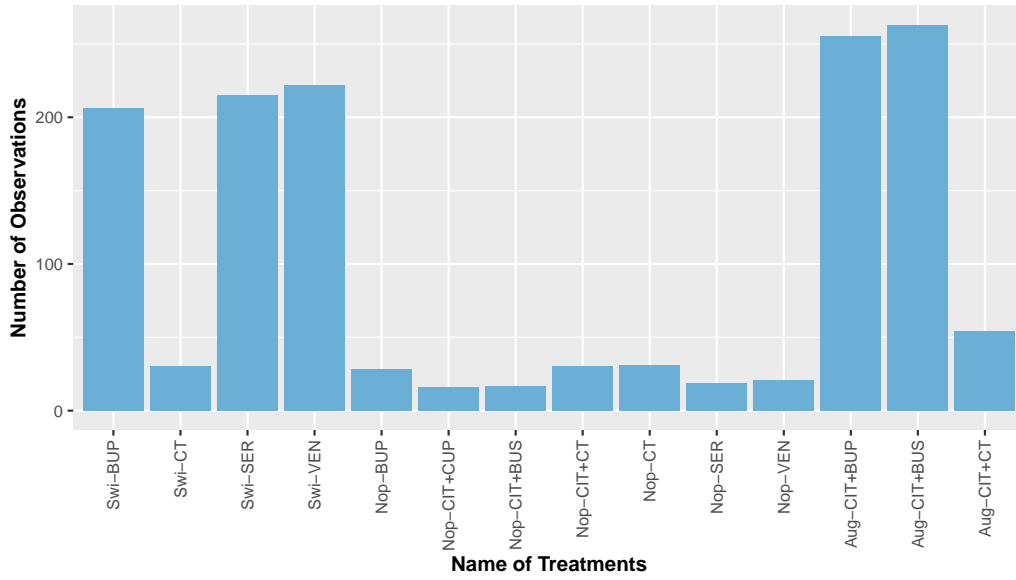
Figure 3: Distribution of observations for the 14 treatment options in the STAR*D dataset. "Swi", "Aug" and "Nop" correspond to patients that switch to other treatments, augment existing treatments, and have no preference for two previous options.

follow equation (14) discussed in Section 2.4 to determine the number of groups with training data. We evaluate the four methods on the remaining one fold of testing data based on the empirical value function $\mathbb{E}_n\big[R\mathbb{I}[\widehat{D}(X) = A]/\widehat{p}(A|X)\big]/\mathbb{E}_n\big[\mathbb{I}[\widehat{D}(X) = A]/\widehat{p}(A|X)\big]$, where $\mathbb{E}_n$ denotes the empirical average of testing data.
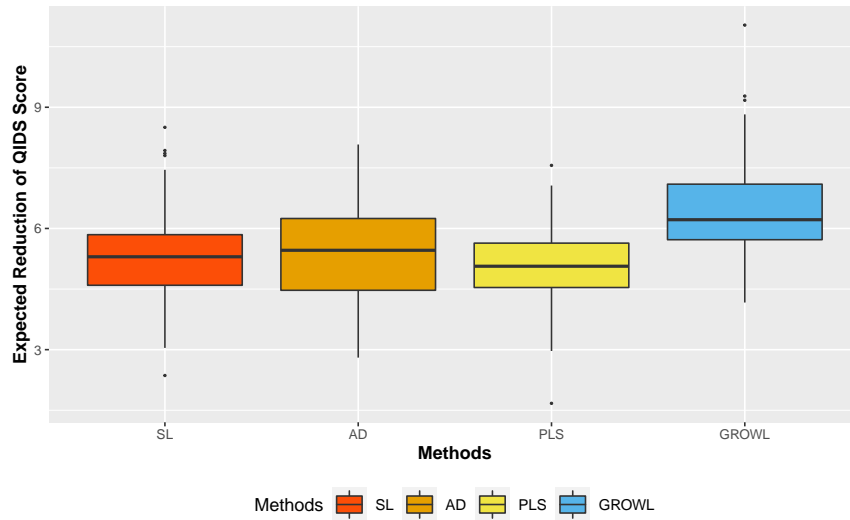
Figure 4: Boxplots of Expected Reduction of QIDS score during Level 2 for patients in testing data based on 200 repetitions for the STAR*D study (higher value is better).

The testing results are shown in Figure 4. The means of expected reduction of QIDS score during Level 2 by using GROWL is 6.44, which outperform the mean value of SL (5.28), AD (5.32) and PLS (5.04). Thus, compared with methods without considering the treatment partition, GROWL substantially improves the performance of the optimal ITR estimation. The estimated group numbers are 5 or 6 for most of the repetitions. Among the 200 estimated partitions of the treatment space, the patient-centered treatments containing SER, CIT+BUP, CIT+BUS, and CIT+CT are often combined within one group and the treatments containing BUP, CT and VEN are integrated with another group with high frequency. It is interesting to point out that, the treatments SER, CIT+BUP, CIT+BUS, and CIT+CT are often considered as one class of treatments including Selective Serotonin Reuptake Inhibitors (SSRI) while the treatments BUP, CT and VEN are non-SSRI treatments because the treatments within the same group have similar treatment effects (Liu et al., 2018; Pan and Zhao, 2021). In addition, the patient-centered treatments with the same patients' preference are often clustered into the same group. With the overall dataset, we implement the GROWL with the Gaussian kernel and obtain the final estimated group structure with 5 treatment groups:

$$
\widehat{\mathcal{G}}_0 = \Big\{ \{\text{Swi-BUP, Swi-VEN, Nop-VEN}\}, \{\text{Swi-SER}\},
$$
$$
\{\text{Aug-CIT+BUP, Aug-CIT+BUS, Aug-CIT+CT}\},
$$
$$
\{\text{Swi-CT, Nop-CIT+CT, Nop-CT, Nop-VEN}\},
$$
$$
\{\text{Nop-CIT+BUP, Nop-CIT+BUS, Nop-SER}\} \Big\}.
$$

It can be seen that these patient-centered treatments with the same preference work similarly within the SSRI groups and the non-SSRI groups respectively.

To better interpret the decision rule and examine the effects of the feature variables, we implement our GROWL with linear kernel based on $\widehat{\mathcal{G}}_0$. For the STAR*D dataset, the mean of expected reduction of QIDS score for GROWL with linear kernel is 6.11, demonstrating that GROWL with the linear kernel still outperforms other methods. Note that we have 13 feature variables (including the intercept) and 5 treatment groups. Therefore, we obtain a 4×13 estimated coefficient matrix $\widehat{B}$ for the linear decision function. The k-th column of Table 3 in Appendix D demonstrates $\widehat{B}^T \mathbf{W}_k$ for the k-th treatment group where $k = 1, 2, \ldots, 5$. We can see that nearly all feature variables play an important role in the estimated optimal ITR. In particular, for the important biomarker, the QIDS score, the patients with higher QIDS score reduction during Level 1 are suggested with augmenting the current CIT treatment implemented at Level 1, while patients with low QIDS reduction are recommended with switching to other treatments at Level 2.

6. Discussion

In this article, we propose a new method called GROWL to cluster treatment options and at the same time, estimate the optimal group-structured ITR within one RAMSVM-based objective function. Other comparison methods estimate the ITR under the individual treatment domain while GROWL focuses on the group treatment domain. When homogeneous

or nearly-homogeneous treatment group structure is satisfied for the treatments, GROWL is able to find the expected partition with high accuracy and the superior performance of GROWL is demonstrated in our numerical studies. Under heterogeneous settings when θ is small shown in Section 4.2, our method tends to sacrifice the benefit while reduce the variability significantly. In this case, our method is still superior to other methods when the sample size is small. Another advantage of GROWL is that it combines both supervised and unsupervised learning through one single optimization.

From a broad perspective, our method is not limited to ITR problems. It can be viewed as a multicategory classification technique. In particular, consider using observed data $(x_1, y_1), \ldots, (x_n, y_n)$ to classify the covariate $x \in \mathcal{X}$ as a specific class y in a large class space $\mathcal{Y}$. Due to the large number of labels, insufficient data are observed for some specific labels. Consequently, standard classification methods can become ineffective. On the other hand, the conditional probability $Pr[Y = y|X]$ may employ possible similar patterns for some classes $y \in \mathcal{Y}$. Then our method tends to estimate this pattern with group-based structure to reduce the dimension of the classification label space and classify the observations in the group domain. Although our paper mainly focuses on estimating the optimal ITR in the decision making framework, the essence is similar because the conditional treatment effects $\mathbb{E}[R|A = a, X]$ play a similar role as $Pr[Y = y|X]$ in classification problems.

Several possible extensions can be explored for future studies. First, most scenarios considered in the paper is that the group structure of the treatment effects is independent with the marginal distribution of feature variables X. However, consider the case that homogeneous group structure is completely different with different values of $x \in \mathcal{X}$, the optimal partition in (7) tends to sacrifice some subgroups of individuals. Consequently, more value would be lost because the optimal partition is obtained via averaging the marginal distribution of X. For these more complex scenarios, it will be interesting to estimate different partitions targeting subgroups of individuals. Secondly, our method can be extended to learn group structures for multi-stage Dynamic Treatment Regimes (DTR) (Murphy, 2003; Zhao et al., 2015). This can be an interesting direction for future research.

Acknowledgments

Appendix

The appendix contains detailed proofs of the main results, additional figures and tables, and more implementation details. In Appendix A, we introduce some useful lemmas to prove the theorems in the main paper. In Appendix B, we give the detailed proofs of theorems in the main paper. In Appendix C, we prove the lemmas in the main paper and the lemmas in Appendix A. We also provide proofs of equation (10) and how the dual problems (11) and (12) are derived. Additional results of more simulations and real data analysis are given in Appendix D. Finally, implementation details are summarized in Appendix E.

Appendix A. Useful Lemmas

Lemma 8 (Bound of excess risk for any fixed δ) *For any partition δ, any measurable function $\boldsymbol{f} : \mathcal{X} \to \mathbb{R}^{K_n-1}$ such that $\langle \mathbf{W}_k, \boldsymbol{f}(x) \rangle \in [-1, K_n - 1]$ holds for $\forall x \in \mathcal{X}$ and $\forall k \in [K_n]$, we have*

$$\mathcal{R}(\delta, \boldsymbol{f}) - \mathcal{R}(\delta, \boldsymbol{f}^\delta) \leqslant \mathcal{R}_\phi(\delta, \boldsymbol{f}) - \mathcal{R}_\phi(\delta, \boldsymbol{f}^\delta).$$

Note that Lemma 8 generalizes the excess risk bound in Theorem 3.2 in Zhao et al. (2012) from the binary setting to the multi-class setting for any fixed δ.

Lemma 9 (Bound for approximation error) *Fix a partition δ. Suppose the generalized geometric noise assumption holds for δ with noise exponent $0 < q < \infty$ and constant U. Set $\sigma_n = \lambda_n^{-1/(q+1)d}$. Then for the Gaussian kernel κ, there is a constant $c > 0$ depending only on the dimension d, the geometric noise exponent q and constant U, such that for all $\lambda_n > 0$, we have*

$$\mathcal{A}_{\sigma_n}^\delta(\lambda_n) \leqslant cK_n^2 \lambda_n^{q/(q+1)}.$$

Lemma 10 (Finite sample regret bound for optimal δ^*) *Suppose the generalized geometric noise assumption holds for an optimal partition $\delta^* \in \Delta^*$ with exponent $0 < q < \infty$ and constant U. Assume $|\frac{R}{p(\delta^*(A)|X)}| \leqslant Z_n$. Then for the Gaussian kernel κ and any $\theta > 0, 0 < v < 2$, there exists a constant C (depending on θ, v, d), such that for all $\tau \geqslant 1$ and $\sigma_n = \lambda_n^{-1/(q+1)d}$,*

$$Pr^* \left(\widetilde{\mathcal{R}}(\delta^*, \widehat{\boldsymbol{f}}_n^{\delta^*}) \leqslant \widetilde{\mathcal{R}}^* + \epsilon \right) \geqslant 1 - e^{-\tau},$$

where Pr^* denotes the outer probability for possibly nonmeasurable sets, and

$$\epsilon = CK_n Z_n \left[(\lambda_n)^{-\frac{2}{2+v} + \frac{(2-v)(1+\theta)}{(2+v)(1+q)}} n^{-\frac{2}{2+v}} + \frac{\tau}{n\lambda_n} + \lambda_n^{\frac{q}{q+1}} \right].$$

Note that Lemma 10 is a generalization of Theorem 3.4 in Zhao et al. (2012) to deal with the multiclass ITR problem under the L_ϕ-based loss.

Appendix B. Proofs of Theorems

Proof of Theorem 4:

Based on the definition of L_ϕ loss, $\mathcal{R}_\phi(\delta, \boldsymbol{f})$ equals to

$$\mathbb{E}_{X,A} \left[\frac{\mathbb{E}[R|A,X]}{p(\delta(A)|X)} \times [(1-\gamma) \sum_{j \neq \delta(A)} (1 + \langle \mathbf{W}_j, \boldsymbol{f}(X) \rangle)^+ + \gamma(K_n - 1 - \langle \mathbf{W}_{\delta(A)}, \boldsymbol{f}(X) \rangle)^+] \right]$$

$$= \mathbb{E}_X \left[\sum_{k=1}^{K_n} \sum_{a:\delta(a)=k} \frac{\mathbb{E}[R|A=a,X]p(a|X)}{p(\delta(A)|X)} \times [(1-\gamma) \sum_{j \neq k} (1 + \langle \mathbf{W}_j, \boldsymbol{f}(X) \rangle)^+ + \gamma(K_n - 1 - \langle \mathbf{W}_k, \boldsymbol{f}(X) \rangle)^+] \right]$$

$$= \mathbb{E}_X \left[\sum_{k=1}^{K_n} \mathbb{E}[R|A \in G_k^\delta, X] \, [(1-\gamma) \sum_{j \neq k} (1 + \langle \mathbf{W}_j, \boldsymbol{f}(X) \rangle)^+ + \gamma(K_n - 1 - \langle \mathbf{W}_k, \boldsymbol{f}(X) \rangle)^+] \right].$$

For each fixed δ and fixed $x \in \mathcal{X}$, denote $g^\delta(x) = \arg\max_{k \in [K_n]} \mathbb{E}[R|A \in G_k^\delta, X = x]$. Then based on the proof of Theorem 1 in Zhang et al. (2016), when $\gamma \in [0, 1/2]$, we have $\langle \mathbf{W}_{g^\delta(x)}, \boldsymbol{f}_\phi^\delta(x) \rangle = K_n - 1$ and $\langle \mathbf{W}_j, \boldsymbol{f}_\phi^\delta(x) \rangle = -1$ for $j \neq g^\delta(x)$. Hence, classical Fisher Consistency holds for group-based decision rule for each fixed δ. By plugging $\boldsymbol{f}_\phi^\delta$ into above equation, we have

$$\inf_{\boldsymbol{f}:\mathcal{X} \to \mathbb{R}^{K_n-1}} \mathcal{R}_\phi(\delta, \boldsymbol{f}) = \mathcal{R}_\phi(\delta, \boldsymbol{f}_\phi^\delta)$$

$$= \mathbb{E}_X \left[\sum_{k \neq g^\delta(X)}^{K_n} \mathbb{E}[R|A \in G_k^\delta, X] \, [(1-\gamma)K_n + \gamma K_n] \right]$$

$$= \mathbb{E}_X \left[K_n \sum_{k=1}^{K_n} \mathbb{E}[R|A \in G_k^\delta, X] \, - K_n \max_{k \in [K_n]} \mathbb{E}[R|A \in G_k^\delta, X] \right]$$

$$= \mathbb{E} \left[K_n \frac{R}{p(\delta(A)|X)} \right] - K_n \mathbb{E}_X \left[\max_{k \in [K_n]} \mathbb{E}[R|A \in G_k^\delta, X] \right].$$

Hence, we have

$$\delta_\phi^* \in \arg\min_\delta \left\{ \widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}_\phi^\delta) = \mathcal{R}_\phi(\delta, \boldsymbol{f}_\phi^\delta) - K_n \mathbb{E} \left[\frac{R}{p(\delta(A)|X)} \right] \right\}$$

$$= \arg\max_\delta \mathbb{E}_X \left[\max_{k \in [K_n]} \mathbb{E}[R|A \in G_k^\delta, X] \right] = \Delta^*.$$

Then $\arg\max_{k \in [K_n]} \langle \mathbf{W}_k, \boldsymbol{f}_\phi^*(x) \rangle = \arg\max_{k \in [K_n]} \mathbb{E}[R|A \in G_k^{\delta^*}, X = x]$ follows straightforward. ■

Proof of Theorem 5:

To prove the excess risk bound, we notice the following decomposition:

$$\widetilde{\mathcal{R}}(\delta, \boldsymbol{f}) - \widetilde{\mathcal{R}}^* = \left(\widetilde{\mathcal{R}}(\delta, \boldsymbol{f}) - \widetilde{\mathcal{R}}(\delta, \boldsymbol{f}^\delta) \right) + \left(\widetilde{\mathcal{R}}(\delta, \boldsymbol{f}^\delta) - \widetilde{\mathcal{R}}(\delta^*, \boldsymbol{f}^*) \right)$$

$$= \left(\mathcal{R}(\delta, \boldsymbol{f}) - \mathcal{R}(\delta, \boldsymbol{f}^\delta) \right) + \left(\widetilde{\mathcal{R}}(\delta, \boldsymbol{f}^\delta) - \widetilde{\mathcal{R}}(\delta^*, \boldsymbol{f}^*) \right)$$

$$\leq \left(\mathcal{R}_\phi(\delta, \boldsymbol{f}) - \mathcal{R}_\phi(\delta, \boldsymbol{f}^\delta) \right) + \left(\widetilde{\mathcal{R}}(\delta, \boldsymbol{f}^\delta) - \widetilde{\mathcal{R}}(\delta^*, \boldsymbol{f}^*) \right)$$

$$\leq \left(\mathcal{R}_\phi(\delta, \boldsymbol{f}) - \mathcal{R}_\phi(\delta, \boldsymbol{f}^\delta) \right) + \left(\widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}^\delta) - \widetilde{\mathcal{R}}_\phi^* \right)$$

$$= \widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}) - \widetilde{\mathcal{R}}_\phi^*.$$

The first inequality follows from Lemma 8. For the second inequality, notice that

$$\widetilde{\mathcal{R}}(\delta, \boldsymbol{f}^\delta) - \widetilde{\mathcal{R}}(\delta^*, \boldsymbol{f}^*) \leqslant K_n \left(\widetilde{\mathcal{R}}(\delta, \boldsymbol{f}^\delta) - \widetilde{\mathcal{R}}(\delta^*, \boldsymbol{f}^*) \right) = \widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}^\delta) - \widetilde{\mathcal{R}}_\phi^*.$$

Then the proof is complete. ∎

Proof of Theorem 6:

For each partition δ, take the following notations. Define the random variable $S^\delta = \frac{R}{p(\delta(A)|X)}$. Denote $V_\phi^\delta(\boldsymbol{f}) = S^\delta L_\phi(\delta, \boldsymbol{f})$ as the weighted loss. Similarly, under δ, let $\mathbb{P}^\delta$ be the population measure of $\left(X, \delta(A), R/p(\delta(A)|X)\right)$ and $\mathbb{P}_n^\delta$ be the associated empirical measure. Recall that $\Psi_n = \frac{1}{2}\min_{\delta \notin \Delta^*} \mathcal{V}_1^*(\delta^*) - \mathcal{V}_1^*(\delta) > 0$. Then, for any $\delta \notin \Delta^*$, we have $\inf_{\boldsymbol{f}} \widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}) - \widetilde{\mathcal{R}}_\phi^* > \Psi_n$. To prove part (I) in this theorem, we need to control the probability of event $\left\{\widehat{\delta}_n \notin \Delta^*\right\}$. Take an optimal $\delta^* \in \Delta^*$. Note that based on the definition of $\widehat{\delta}_n$, the following two events are equivalent:

$$\left\{\widehat{\delta}_n \in \Delta^*\right\} \Leftrightarrow \bigcap_{\delta \notin \Delta^*} \left\{ \mathbb{P}_n^{\delta^*}(V_\phi^{\delta^*}(\widehat{\boldsymbol{f}}_n)) + \lambda_n \left\|\widehat{\boldsymbol{f}}_n\right\|_{\mathcal{F}_n}^2 - K_n \mathbb{P}_n^{\delta^*}(S^{\delta^*}) \leqslant \mathbb{P}_n^\delta(V_\phi^\delta(\widehat{\boldsymbol{f}}_n)) + \lambda_n \left\|\widehat{\boldsymbol{f}}_n\right\|_{\mathcal{F}_n}^2 - K_n \mathbb{P}_n^\delta(S^\delta) \right\}.$$

Hence, $Pr\left[\widehat{\delta}_n \in \Delta^*\right]$

$$= 1 - Pr\left[\bigcup_{\delta \notin \Delta^*} \left\{ \mathbb{P}_n^{\delta^*}(V_\phi^{\delta^*}(\widehat{\boldsymbol{f}}_n)) + \lambda_n \left\|\widehat{\boldsymbol{f}}_n\right\|_{\mathcal{F}_n}^2 - K_n \mathbb{P}_n^{\delta^*}(S^{\delta^*}) > \mathbb{P}_n^\delta(V_\phi^\delta(\widehat{\boldsymbol{f}}_n)) + \lambda_n \left\|\widehat{\boldsymbol{f}}_n\right\|_{\mathcal{F}_n}^2 - K_n \mathbb{P}_n^\delta(S^\delta) \right\} \right]$$

$$\geqslant 1 - \sum_{\delta \notin \Delta^*} Pr\left[\mathbb{P}_n^{\delta^*}(V_\phi^{\delta^*}(\widehat{\boldsymbol{f}}_n)) + \lambda_n \left\|\widehat{\boldsymbol{f}}_n\right\|_{\mathcal{F}_n}^2 - K_n \mathbb{P}_n^{\delta^*}(S^{\delta^*}) > \mathbb{P}_n^\delta(V_\phi^\delta(\widehat{\boldsymbol{f}}_n)) + \lambda_n \left\|\widehat{\boldsymbol{f}}_n\right\|_{\mathcal{F}_n}^2 - K_n \mathbb{P}_n^\delta(S^\delta) \right]$$

$$= 1 - \sum_{\delta \notin \Delta^*} \left(1 - Pr\left[\mathbb{P}_n^{\delta^*}(V_\phi^{\delta^*}(\widehat{\boldsymbol{f}}_n)) + \lambda_n \left\|\widehat{\boldsymbol{f}}_n\right\|_{\mathcal{F}_n}^2 - K_n \mathbb{P}_n^{\delta^*}(S^{\delta^*}) \leqslant \mathbb{P}_n^\delta(V_\phi^\delta(\widehat{\boldsymbol{f}}_n)) + \lambda_n \left\|\widehat{\boldsymbol{f}}_n\right\|_{\mathcal{F}_n}^2 - K_n \mathbb{P}_n^\delta(S^\delta) \right] \right)$$

After rearranging both sides, we can only control the following probability for each fixed $\delta \notin \Delta^*$ in order to lower bound $Pr\left[\widehat{\delta}_n \in \Delta^*\right]$:

$$Pr\left[\mathbb{P}_n^{\delta^*}(V_\phi^{\delta^*}(\widehat{\boldsymbol{f}}_n)) - \mathbb{P}_n^\delta(V_\phi^\delta(\widehat{\boldsymbol{f}}_n)) - K_n(\mathbb{P}_n^{\delta^*}(S^{\delta^*}) - \mathbb{P}_n^\delta(S^\delta)) \leqslant 0 \right]$$

$$= Pr\left[\mathbb{P}_n^{\delta^*}(V_\phi^{\delta^*}(\widehat{\boldsymbol{f}}_n)) - \mathbb{P}_n^\delta(V_\phi^\delta(\widehat{\boldsymbol{f}}_n)) - K_n(\mathbb{P}_n^{\delta^*}(S^{\delta^*}) - \mathbb{P}_n^\delta(S^\delta)) - (\widetilde{\mathcal{R}}_\phi^* - \widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}^\delta)) \leqslant \widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}^\delta) - \widetilde{\mathcal{R}}_\phi^* \right]$$

$$\geqslant Pr\left[\mathbb{P}_n^{\delta^*}(V_\phi^{\delta^*}(\widehat{\boldsymbol{f}}_n)) - \mathbb{P}_n^\delta(V_\phi^\delta(\widehat{\boldsymbol{f}}_n)) - K_n(\mathbb{P}_n^{\delta^*}(S^{\delta^*}) - \mathbb{P}_n^\delta(S^\delta)) - (\widetilde{\mathcal{R}}_\phi^* - \widetilde{\mathcal{R}}_\phi(\delta, \boldsymbol{f}^\delta)) \leqslant \Psi_n \right]$$

$$= Pr\left[\mathbf{I} + \mathbf{II} + \mathbf{III} - \mathbf{IV} \leqslant \Psi_n \right]$$

$$\geqslant Pr\left[\mathbf{I} + \mathbf{II} + \mathbf{III} \leqslant \Psi_n \right],$$

where

$$\mathbf{I} = (\mathbb{P}_n^{\delta^*} - \mathbb{P}^{\delta^*})(V_\phi^{\delta^*}(\widehat{\boldsymbol{f}}_n)) - (\mathbb{P}_n^\delta - \mathbb{P}^\delta)(V_\phi^\delta(\widehat{\boldsymbol{f}}_n)),$$

$$\mathbf{II} = K_n(\mathbb{P}^{\delta^*} - \mathbb{P}_n^{\delta^*})(S^{\delta^*}) - K_n(\mathbb{P}^\delta - \mathbb{P}_n^\delta)(S^\delta),$$

$$\mathbf{III} = \mathbb{P}^{\delta^*}(V_\phi^{\delta^*}(\widehat{\boldsymbol{f}}_n - \boldsymbol{f}^{\delta^*})),$$

$$\mathbf{IV} = \mathbb{P}^\delta(V_\phi^\delta(\widehat{\boldsymbol{f}}_n - \boldsymbol{f}^\delta)).$$

Note that the last inequality above is due to $\mathbf{IV} \geqslant 0$. Here, after removing $\mathbf{IV}$, the advantage is that we only need the generalized geometric noise condition holds for δ^* to bound $\mathbf{III}$ following from Lemma 10.

Now, we would bound the above three terms respectively using concentration techniques. First, to bound **I**, we start with obtaining a bound for $\left\|\widehat{\boldsymbol{f}}_n\right\|_{\mathcal{F}_n}^2$. Since $\mathbb{P}_n^{\widehat{\delta}_n}(V_\phi^{\widehat{\delta}_n}(\widehat{\boldsymbol{f}}_n) - K_n S^{\widehat{\delta}_n}) + \lambda_n \left\|\widehat{\boldsymbol{f}}_n\right\|_{\mathcal{F}_n}^2 \leqslant \mathbb{P}_n^\delta(V_\phi^\delta(\boldsymbol{f}) - K_n S^\delta) + \lambda_n \|\boldsymbol{f}\|_{\mathcal{F}_n}^2$ holds for any δ and $\boldsymbol{f} \in \otimes_{k=1}^{K_n-1}\mathcal{H}_\kappa^k$, we can select $\delta = \widehat{\delta}_n$ and $\boldsymbol{f} = \boldsymbol{0}$ to get

$$\left\|\widehat{\boldsymbol{f}}_n\right\|_{\mathcal{F}_n}^2 \leqslant \frac{1}{\lambda_n}\frac{1}{n}\sum_{i=1}^n S_i^\delta L_\phi(\widehat{\delta}_n, \boldsymbol{0}) \leqslant \frac{2(K_n-1)}{\lambda_n}Z_n,$$

so that only the constrained class $\mathcal{F}_n = \left\{\sqrt{\lambda_n}\boldsymbol{f} \in \otimes_{k=1}^{K_n-1}\mathcal{H}_\kappa^k : \left\|\sqrt{\lambda_n}\boldsymbol{f}\right\|_{\mathcal{F}_n} \leqslant \sqrt{2(K_n-1)Z_n}\right\}$ would be considered. Note that RAMSVM loss L_ϕ is Lipschitz continuous with respect to $\boldsymbol{f}$ with Lipschitz constant $K_n Z_n$. Then, for each δ, McDiarmid's inequality (Bartlett and Mendelson, 2002) implies that with probability at least $1 - e^{-\tau}$,

$$\sup_{\boldsymbol{f}\in\mathcal{F}_n}(\mathbb{P}_n^\delta - \mathbb{P}^\delta)V_\phi^\delta(\boldsymbol{f}) \leqslant \mathbb{E}\left[\sup_{\boldsymbol{f}\in\mathcal{F}_n}(\mathbb{P}_n^\delta - \mathbb{P}^\delta)V_\phi^\delta(\boldsymbol{f})\right] + 2K_n Z_n\sqrt{\frac{2\tau}{n}}.$$

Next we define the Rademacher complexity over $\mathcal{F}_n$ as $\mathrm{R}_n^\delta := \mathbb{E}\sup_{\boldsymbol{f}\in\mathcal{F}_n}\mathbb{P}_n(\varepsilon V_\phi^\delta(\boldsymbol{f}))$, where ε is the Rademacher variable independent of $(X, \delta(A), S^\delta)$ under $\mathbb{P}^\delta$. Then by standard symmetrization arguments and Lemma 22 in Bartlett and Mendelson (2002), we have

$$\mathbb{E}\left[\sup_{\boldsymbol{f}\in\mathcal{F}_n}(\mathbb{P}_n^\delta - \mathbb{P}^\delta)V_\phi^\delta(\boldsymbol{f})\right] \leqslant 2\mathrm{R}_n^\delta \leqslant 2\sqrt{\frac{2(K_n-1)Z_n}{n\lambda_n}}.$$

Similar proof holds for $\mathbb{P}^\delta - \mathbb{P}_n^\delta$. Therefore, we have $Pr\left[\mathbf{I} \leqslant \epsilon_1\right] \geqslant 1 - 2e^{-\tau}$, where

$$\epsilon_1 = 4\sqrt{\frac{2(K_n-1)Z_n}{n\lambda_n}} + 4K_n Z_n\sqrt{\frac{2\tau}{n}}.$$

Next we bound the second term **II**. Note that by assumption, for each δ, $(K_n-1)|S^\delta| \leqslant (K_n-1)Z_n$. By Hoeffding inequality (Steinwart and Christmann, 2008), we have

$$Pr\left[\mathbf{II} \leqslant \epsilon_2\right] \geqslant 1 - e^{-\tau},$$

where

$$\epsilon_2 = (K_n-1)Z_n\sqrt{\frac{\tau}{2n}}.$$

Finally, bounding **III** with Lemma 10, we have for any $\theta > 0, v \in [0, 2]$, such that for all $\tau \geqslant 1$ and $\sigma_n = \lambda_n^{-1/(q+1)d}$, we have

$$Pr^*[\mathcal{R}_\phi(\delta^*, \widehat{\boldsymbol{f}}_n^{\delta^*}) \leqslant \mathcal{R}_\phi^* + \epsilon_3] = Pr[\mathbf{III} \leqslant \epsilon_3] \geqslant 1 - e^{-\tau},$$

and

$$\epsilon_3 = CK_n Z_n\left[(\lambda_n)^{-\frac{2}{2+v} + \frac{(2-v)(1+\theta)}{(2+v)(1+q)}}n^{-\frac{2}{2+v}} + \frac{\tau}{n\lambda_n} + \lambda_n^{\frac{q}{q+1}}\right],$$

where C is a constant depending on v, θ and d.

Therefore, combining the bound for $\mathbf{I}, \mathbf{II}, \mathbf{III}$, we have

$$Pr^*[\mathbf{I} + \mathbf{II} + \mathbf{III} \leqslant \epsilon_1 + \epsilon_2 + \epsilon_3] \geqslant 1 - 4e^{-\tau},$$

where ϵ_1, ϵ_2 and ϵ_3 depend on n and go to 0 as $n \to \infty$. Note that $\epsilon_1 = \mathcal{O}(K_n^{1/2} Z_n^{1/2} n^{-1/2} \lambda_n^{-1/2} + K_n Z_n n^{-1/2})$, $\epsilon_2 = \mathcal{O}(K_n Z_n n^{-1/2})$ and $\epsilon_3 = \mathcal{O}\Big(K_n Z_n (\lambda_n)^{-\frac{2}{2+v} + \frac{(2-v)(1+\theta)}{(2+v)(1+q)}} n^{-\frac{2}{2+v}} + K_n Z_n (n\lambda_n)^{-1} + K_n Z_n \lambda_n^{\frac{q}{q+1}} \Big)$. Based on $\Psi_n/(\epsilon_1 + \epsilon_2 + \epsilon_3) \to \infty$, we can find sufficient large n such that $\epsilon_1 + \epsilon_2 + \epsilon_3 < \Psi_n$, then we have

$$Pr^*[\widehat{\delta}_n \in \Delta^*] = 1 - \mathcal{O}(\Gamma_n \epsilon_n).$$

To prove (II), for each fixed $\tau > 1, \theta > 0, 0 < v < 2$, consider

$$\begin{aligned}
&Pr\left[\widetilde{\mathcal{R}}_\phi(\widehat{\delta}_n, \widehat{\boldsymbol{f}}_n) - \widetilde{\mathcal{R}}_\phi^* \geqslant 2\epsilon_3 \right] \\
=&Pr\left[\left(\widetilde{\mathcal{R}}_\phi(\widehat{\delta}_n, \widehat{\boldsymbol{f}}_n) - \widetilde{\mathcal{R}}_\phi(\delta^*, \widehat{\boldsymbol{f}}_n) \right) + \left(\widetilde{\mathcal{R}}_\phi(\delta^*, \widehat{\boldsymbol{f}}_n) - \widetilde{\mathcal{R}}_\phi^* \right) \geqslant 2\epsilon_3 \right] \\
\leqslant&Pr\left[\widetilde{\mathcal{R}}_\phi(\widehat{\delta}_n, \widehat{\boldsymbol{f}}_n) - \widetilde{\mathcal{R}}_\phi(\delta^*, \widehat{\boldsymbol{f}}_n) \geqslant \epsilon_3 \right] + Pr\left[\widetilde{\mathcal{R}}_\phi(\delta^*, \widehat{\boldsymbol{f}}_n) - \widetilde{\mathcal{R}}_\phi^* \geqslant \epsilon_3 \right].
\end{aligned}$$

Next we bound the above two terms respectively. For the first term, take n sufficient large such that $\epsilon_1 + \epsilon_2 + \epsilon_3 < \Psi_n$, then

$$\begin{aligned}
&Pr\left[\widetilde{\mathcal{R}}_\phi(\widehat{\delta}_n, \widehat{\boldsymbol{f}}_n) - \widetilde{\mathcal{R}}_\phi(\delta^*, \widehat{\boldsymbol{f}}_n) \geqslant \epsilon_3 \right] \\
=&Pr\left[\widetilde{\mathcal{R}}_\phi(\widehat{\delta}_n, \widehat{\boldsymbol{f}}_n) - \widetilde{\mathcal{R}}_\phi(\delta^*, \widehat{\boldsymbol{f}}_n) \geqslant \epsilon_3 \Big| \widehat{\delta}_n \in \Delta^* \right] Pr\left[\widehat{\delta}_n \in \Delta^* \right] \\
&+ Pr\left[\widetilde{\mathcal{R}}_\phi(\widehat{\delta}_n, \widehat{\boldsymbol{f}}_n) - \widetilde{\mathcal{R}}_\phi(\delta^*, \widehat{\boldsymbol{f}}_n) \geqslant \epsilon_3 \Big| \widehat{\delta}_n \notin \Delta^* \right] Pr\left[\widehat{\delta}_n \notin \Delta^* \right] \\
\leqslant&0 + Pr\left[\widehat{\delta}_n \notin \Delta^* \right] \\
\leqslant&4e^{-\tau}.
\end{aligned}$$

For the second term, Lemma 10 implies that $Pr\left[\widetilde{\mathcal{R}}_\phi(\delta^*, \widehat{\boldsymbol{f}}_n) - \widetilde{\mathcal{R}}_\phi^* \geqslant \epsilon_3 \right] \leqslant e^{-\tau}$. Combining both parts, when $\epsilon_1 + \epsilon_2 + \epsilon_3 < \Psi_n$, we have for all $\tau \geqslant 1$,

$$Pr\left[\widetilde{\mathcal{R}}_\phi(\widehat{\delta}_n, \widehat{\boldsymbol{f}}_n) - \widetilde{\mathcal{R}}_\phi^* \leqslant \mathcal{O}(\epsilon_1 + \epsilon_2 + \epsilon_3) \right] \geqslant 1 - 5e^{-\tau}.$$

Therefore, we have

$$\widetilde{\mathcal{R}}(\widehat{\delta}_n, \widehat{\boldsymbol{f}}_n) - \widetilde{\mathcal{R}}^* = \mathcal{O}_p(\Gamma_n \epsilon_n).$$

Note that when R is replaced with $R^\star = (R - s(X))^+ + (K_n - 1)(R - s(X))^-$ discussed in Section 2.3, the final convergence rate would also include the rate coming from estimating $s(X)$. More details can be found in Liu et al. (2018). $\blacksquare$

Appendix C. Proofs of Lemmas, Equation (10), and Dual Problems (11) and (12)

Proof of Lemma 2:

For any fixed $\delta : \mathcal{A} \to [K_n]$,

$$
\begin{aligned}
\mathcal{V}_1^*(\delta) &= \max_{D_g:\mathcal{X} \to [K_n]} \mathbb{E}\left[\mathbb{I}[D_g(X) = \delta(A)]\frac{R}{p(\delta(A)|X)}\right] \\
&= \max_{D_g:\mathcal{X} \to [K_n]} \mathbb{E}_X\left[\sum_{a \in \mathcal{A}} \mathbb{I}[D_g(X) = \delta(a)]\frac{p(a|X)}{p(\delta(a)|X)}\mathbb{E}[R|X, A = a]\right] \\
&= \max_{D_g:\mathcal{X} \to [K_n]} \mathbb{E}_X\left[\sum_{k=1}^{K_n} \mathbb{I}[D_g(X) = k]\sum_{a:\delta(a)=k}\frac{p(a|X)}{p(\delta(a)|X)}\mathbb{E}[R|X, A = a]\right] \\
&= \mathbb{E}_X\left[\max_{k \in [K_n]}\sum_{a:\delta(a)=k}\frac{p(a|X)}{p(\delta(a)|X)}\mathbb{E}[R|X, A = a]\right] \\
&\leqslant \mathbb{E}_X\left[\max_{a \in \mathcal{A}}\mathbb{E}[R|X, A = a]\right] \\
&= \mathbb{E}_X\left[\max_{k \in [K_n]}\mathbb{E}[R|X, A \in G_k^0]\right] = \mathcal{V}_1^*(\delta^0),
\end{aligned}
$$

where the last equation is because of the homogeneous group structure. So, taking supremum over all of the partitions gives $\sup_{\delta:\mathcal{A}\to[K_n]}\mathcal{V}_1^*(\delta) \leqslant \mathcal{V}_1^*(\delta^0)$. Therefore, based on the definition of δ^*, we have $\delta^* = \delta^0$.

Proof of Lemma 8:

First, for the excess risk of 0-1 loss,

$$
\begin{aligned}
&\mathcal{R}(\delta, \boldsymbol{f}) - \mathcal{R}(\delta, \boldsymbol{f}^\delta) \\
=&\mathbb{E}\left[\frac{R}{p(\delta(A)|X)}\left(\mathbb{I}[\arg\max_j\langle\mathbf{W}_j, \boldsymbol{f}(X)\rangle \neq \delta(A)] - \mathbb{I}[\arg\max_j\langle\mathbf{W}_j, \boldsymbol{f}^\delta(X)\rangle \neq \delta(A)]\right)\right] \\
=&\mathbb{E}\left[\sum_{k=1}^{K_n}\mathbb{E}[R|A \in G_k^\delta, X]\left(\mathbb{I}[\arg\max_j\langle\mathbf{W}_j, \boldsymbol{f}(X)\rangle \neq \delta(A)] - \mathbb{I}[\arg\max_j\langle\mathbf{W}_j, \boldsymbol{f}^\delta(X)\rangle \neq \delta(A)]\right)\right]
\end{aligned}
$$

Then, for the excess risk of RAMSVM loss,

$$
\begin{aligned}
&\mathcal{R}_\phi(\delta, \boldsymbol{f}) - \mathcal{R}_\phi(\delta, \boldsymbol{f}^\delta) \\
=&\mathbb{E}\left[\frac{R}{p(\delta(A)|X)}\left((1-\gamma)\sum_{k \neq \delta(A)}(1 + \langle\mathbf{W}_k, \boldsymbol{f}(X)\rangle)^+ + \gamma(K_n - 1 - \langle\mathbf{W}_{\delta(A)}, \boldsymbol{f}(X)\rangle)^+\right)\right] \\
&- \mathbb{E}\left[\frac{R}{p(\delta(A)|X)}\left((1-\gamma)\sum_{k \neq \delta(A)}(1 + \langle\mathbf{W}_k, \boldsymbol{f}^\delta(X)\rangle)^+ + \gamma(K_n - 1 - \langle\mathbf{W}_{\delta(A)}, \boldsymbol{f}^\delta(X)\rangle)^+\right)\right] \\
=&\mathbb{E}\left[\sum_{k=1}^{K_n}\mathbb{E}[R|A \in G_k^\delta, X]\left((1-\gamma)\sum_{k \neq \delta(A)}(1 + \langle\mathbf{W}_k, \boldsymbol{f}(X)\rangle)^+ + \gamma(K_n - 1 - \langle\mathbf{W}_{\delta(A)}, \boldsymbol{f}(X)\rangle)^+\right)\right]
\end{aligned}
$$

$$-\mathbb{E}\left[\sum_{k=1}^{K_n}\mathbb{E}[R|A\in G_k^\delta,X]\left((1-\gamma)\sum_{k\neq\delta(A)}(1+\langle\mathbf{W}_k,\boldsymbol{f}^\delta(X)\rangle)^+ +\gamma(K_n-1-\langle\mathbf{W}_{\delta(A)},\boldsymbol{f}^\delta(X)\rangle)^+)\right)\right]$$

$$\geqslant\mathbb{E}\left[\sum_{k=1}^{K_n}\mathbb{E}[R|A\in G_k^\delta,X]\left((1-\gamma)\sum_{k\neq\delta(A)}(1+\langle\mathbf{W}_k,\boldsymbol{f}(X)\rangle)+\gamma(K_n-1-\langle\mathbf{W}_{\delta(A)},\boldsymbol{f}(X)\rangle))\right)\right]$$

$$-\mathbb{E}\left[\sum_{k=1}^{K_n}\mathbb{E}[R|A\in G_k^\delta,X]\left((1-\gamma)\sum_{k\neq\delta(A)}(1+\langle\mathbf{W}_k,\boldsymbol{f}^\delta(X)\rangle)+\gamma(K_n-1-\langle\mathbf{W}_{\delta(A)},\boldsymbol{f}^\delta(X)\rangle))\right)\right]$$

$$=\mathbb{E}\left[\sum_{k=1}^{K_n}\mathbb{E}[R|A\in G_k^\delta,X]\left((K_n-1-\langle\mathbf{W}_k,\boldsymbol{f}(X)\rangle)-(K_n-1-\langle\mathbf{W}_k,\boldsymbol{f}^\delta(X)\rangle)\right)\right]$$

$$=\mathbb{E}\left[\sum_{k=1}^{K_n}\mathbb{E}[R|A\in G_k^\delta,X]\langle\mathbf{W}_k,\boldsymbol{f}^\delta(X)-\boldsymbol{f}(X)\rangle\right].$$

The inequality is because, taking out the positive operator would make the first term smaller while leave the second term unchanged due to $\langle\mathbf{W}_k,\boldsymbol{f}^\delta(x)\rangle\in[-1,K_n-1]$ for every $x\in\mathcal{X}$. Next, we can only prove, for each $x\in\mathcal{X}$,

$$\mathbb{E}\left[\sum_{k=1}^{K_n}\mathbb{E}[R|A\in G_k^\delta,X=x]\langle\mathbf{W}_k,\boldsymbol{f}^\delta(x)-\boldsymbol{f}(x)\rangle\right]$$

$$\geqslant\mathbb{E}\left[\sum_{k=1}^{K_n}\mathbb{E}[R|A\in G_k^\delta,X=x]\left(\mathbb{I}[\arg\max_j\langle\mathbf{W}_j,\boldsymbol{f}(x)\rangle\neq\delta(A)]-\mathbb{I}[\arg\max_j\langle\mathbf{W}_j,\boldsymbol{f}^\delta(x)\rangle\neq\delta(A)]\right)\right].$$
(19)

Suppose $\mathbb{E}[R|A\in G_s^\delta,X=x]>\mathbb{E}[R|A\in G_k^\delta,X=x]$ for every $k\neq s$. Then based on Fisher Consistency for group-based decision rule, we have $\arg\max_{j\in[K_n]}\langle\mathbf{W}_j,\boldsymbol{f}^\delta(x)\rangle=s$. Suppose $\arg\max_j\langle\mathbf{W}_j,\boldsymbol{f}(x)\rangle=r$ where $r\neq s$. The right hand side of (19) becomes

$$\sum_{k\neq r,s}^{K_n}\mathbb{E}[R|A\in G_k^\delta,X=x](1-1)+\mathbb{E}[R|A\in G_r^\delta,X=x](0-1)+\mathbb{E}[R|A\in G_s^\delta,X=x](1-0)$$

$$=\mathbb{E}[R|A\in G_s^\delta,X=x]-\mathbb{E}[R|A\in G_r^\delta,X=x].$$

Since $\sum_{k=1}^{K_n}\mathbf{W}_k=\mathbf{0}$, the left hand side of (19) equals to

$$\sum_{k\neq s}^{K_n}\mathbb{E}[R|A\in G_s^\delta,X=x]\langle-\mathbf{W}_k,\boldsymbol{f}^\delta(x)-\boldsymbol{f}(x)\rangle+\sum_{k\neq s}^{K_n}\mathbb{E}[R|A\in G_k^\delta,X=x]\langle\mathbf{W}_k,\boldsymbol{f}^\delta(x)-\boldsymbol{f}(x)\rangle$$

$$=\sum_{k\neq s}^{K_n}\left[\mathbb{E}[R|A\in G_k^\delta,X=x]-\mathbb{E}[R|A\in G_s^\delta,X=x]\right]\langle\mathbf{W}_k,\boldsymbol{f}^\delta(x)-\boldsymbol{f}(x)\rangle$$

$$=\sum_{k\neq r,s}^{K_n}\left[\mathbb{E}[R|A\in G_s^\delta,X=x]-\mathbb{E}[R|A\in G_k^\delta,X=x]\right](1+\langle\mathbf{W}_k,\boldsymbol{f}(x)\rangle)$$

$$+\left[\mathbb{E}[R|A\in G_s^\delta,X=x]-\mathbb{E}[R|A\in G_r^\delta,X=x]\right](1+\langle\mathbf{W}_r,\boldsymbol{f}(x)\rangle).$$

The last equation is based on $\langle\mathbf{W}_k,\boldsymbol{f}^\delta(x)\rangle=-1$ for each $k\neq s$ (Theorem 1 in Zhang et al. (2016)). Then, by the definition of $\boldsymbol{f}$ and the constraint of bound, we have $\langle\mathbf{W}_k,\boldsymbol{f}(x)\rangle\geqslant-1$

for $k \neq r$ and $\langle \mathbf{W}_r, \boldsymbol{f}(x) \rangle \geqslant 0$. Hence, (19) holds and taking expectation for X in both sides of (19) gives the result. ∎

Proof of Lemma 9:

Let $\mathcal{H}_\kappa(\mathbb{R}^d)$ be the RKHS of the RBF kernel with parameter σ_n. Then based on Steinwart and Scovel (2007), the following linear operator $V_{\sigma_n} : \otimes_{k=1}^{K_n-1} L_2^k(\mathbb{R}^d) \to \otimes_{k=1}^{K_n-1} \mathcal{H}_\kappa^k(\mathbb{R}^d)$ defined by

$$V_{\sigma_n}\mathbf{g}(x) = \frac{(2\sigma_n)^{d/2}}{\pi^{d/4}} \int_{\mathbb{R}^d} e^{-2\sigma_n^2 \|x-y\|^2} \mathbf{g}(y) \; dy, \quad \mathbf{g} \in \otimes_{k=1}^{K_n-1} L_2^k(\mathbb{R}^d), x \in \mathbb{R}^d,$$

is an isometric isomorphism. Accordingly, for any fixed δ, we obtain

$$\mathcal{A}_{\sigma_n}^\delta(\lambda_n) \leqslant \inf_{\langle V_{\sigma_n}\mathbf{g}, \mathbf{W}_k \rangle \in [-1, K_n-1], \forall k \in [K_n]} \lambda_n \|\mathbf{g}\|_{L_2}^2 + \mathcal{R}_\phi(\delta, V_{\sigma_n}\mathbf{g}) - \mathcal{R}_\phi(\delta, \boldsymbol{f}^\delta).$$

Using the notation of Lemma 4.1 of Steinwart and Scovel (2007), define $\acute{\mathcal{X}} = 3\mathcal{X}$. Let $\acute{\mathcal{X}}_i^\delta$ be similarly defined as $\mathcal{X}_i^\delta$ on the domain of $\acute{\mathcal{X}}$. We fix a specific measurable function $\acute{\boldsymbol{f}}^\delta$ such that for each $i \in [K_n]$ and $x \in \acute{\mathcal{X}}_i^\delta$, $\acute{\boldsymbol{f}}^\delta$ satisfies $\langle \mathbf{W}_i, \acute{\boldsymbol{f}}^\delta(x) \rangle = K_n - 1$ and $\langle \mathbf{W}_j, \acute{\boldsymbol{f}}^\delta(x) \rangle = -1$ for $j \neq i$. For $x \in \acute{\mathcal{X}}$, it can be checked that the above property is equivalent with $\acute{\boldsymbol{f}}^\delta(x) = (K_n - 1) \sum_{i=1}^{K_n} \mathbf{W}_i \mathbb{I}[x \in \acute{\mathcal{X}}_i^\delta]$. Besides $\acute{\mathcal{X}}$, define $\acute{\boldsymbol{f}}^\delta(x) = \mathbf{0}$. By similar approach of Lemma 4.1 in Steinwart and Scovel (2007), we can make sure the ball $B(x, \Delta^\delta(x)) \subset \acute{\mathcal{X}}_i^\delta$ for every $x \in \mathcal{X}_i^\delta$ ($i \in [K_n]$) on this enlarged support for $\mathcal{X}$. Choose $\mathbf{g} = (\sigma_n^2/\pi)^{d/4} \acute{\boldsymbol{f}}^\delta$ and note that $\|\mathbf{W}_i\|_2 = 1$, we immediately obtain the following with Jensen inequality and assumption of bounded volume of $\mathcal{X}$,

$$\|\mathbf{g}\|_{L_2}^2 \leqslant (K_n - 1)(\frac{81\sigma_n^2}{\pi})^{d/4}\theta(\mathcal{X}) \leqslant (K_n - 1)(\frac{81\sigma_n^2}{\pi})^{d/4}M_1.$$

Similar to the proof of Lemma 2 in our paper,

$$\mathcal{R}_\phi(\delta, V_{\sigma_n}\mathbf{g}) - \mathcal{R}_\phi(\delta, \boldsymbol{f}^\delta)$$
$$\leqslant \sum_{k=1}^{K_n} \mathbb{E}\left[\mathbb{I}[X \in \mathcal{X}_k^\delta] \sum_{j \neq k}^{K_n} |\mathbb{E}[R|A \in G_j^\delta, X] - \mathbb{E}[R|A \in G_k^\delta, X]| \cdot |\langle \mathbf{W}_j, V_{\sigma_n}\mathbf{g}(X) - \boldsymbol{f}^\delta(X) \rangle| \right]$$
$$= \sum_{k=1}^{K_n} \mathbb{E}\left[\mathbb{I}[X \in \mathcal{X}_k^\delta] \sum_{j \neq k}^{K_n} |\eta_{j,k}^\delta(X)| \cdot |\langle \mathbf{W}_j, V_{\sigma_n}\mathbf{g}(X) - \boldsymbol{f}^\delta(X) \rangle| \right].$$

Now for $x \in \mathcal{X}_k^\delta$, we would bound $|\langle \mathbf{W}_j, V_{\sigma_n}\mathbf{g}(x) - \boldsymbol{f}^\delta(x)\rangle|$ when $j \neq k$. We observe

$$
\begin{aligned}
-1 &\leqslant \langle \mathbf{W}_j, V_{\sigma_n}\mathbf{g}(x)\rangle \\
&= \langle \mathbf{W}_j, (\frac{2\sigma_n^2}{\pi})^{d/2} \int_{\mathbb{R}^d} e^{-2\sigma_n^2\|x-y\|^2} \acute{\boldsymbol{f}}^\delta(y) \ dy\rangle \\
&= (\frac{2\sigma_n^2}{\pi})^{d/2} \int_{\mathbb{R}^d} e^{-2\sigma_n^2\|x-y\|^2} \langle \mathbf{W}_j, \acute{\boldsymbol{f}}^\delta(y)\rangle \ dy \\
&= (\frac{2\sigma_n^2}{\pi})^{d/2} \int_{\mathbb{R}^d} e^{-2\sigma_n^2\|x-y\|^2} \langle \mathbf{W}_j, \acute{\boldsymbol{f}}^\delta(y) - (K_n-1)\mathbf{W}_j\rangle \ dy + (K_n - 1) \\
&\leqslant (\frac{2\sigma_n^2}{\pi})^{d/2} \int_{B(x,\Delta^\delta(x))} e^{-2\sigma_n^2\|x-y\|^2} \langle \mathbf{W}_j, \acute{\boldsymbol{f}}^\delta(y) - (K_n-1)\mathbf{W}_j\rangle \ dy + (K_n - 1) \\
&= (\frac{2\sigma_n^2}{\pi})^{d/2} \int_{B(x,\Delta^\delta(x))} -K_n e^{-2\sigma_n^2\|x-y\|^2} \ dy + (K_n - 1) \\
&= -1 + K_n \mathbb{P}_{\gamma_{\sigma_n}}(|u| \geqslant \Delta^\delta(x)),
\end{aligned}
$$

where $\gamma_{\sigma_n} = (2\sigma_n^2/\pi)^{d/2} e^{-2\sigma_n^2|u|^2} \ du$ is a spherical Gaussian in $\mathbb{R}^d$. According to Ledoux and Talagrand (2013), we have $\mathbb{P}_{\gamma_{\sigma_n}}(|u| \geqslant \Delta^\delta(x)) \leqslant 4e^{-\sigma_n^2\Delta^\delta(x)^2/2d}$ and consequently,

$$
-1 \leqslant \langle \mathbf{W}_j, V_{\sigma_n}\mathbf{g}(x)\rangle \leqslant -1 + 4K_n e^{-\sigma_n^2\Delta^\delta(x)^2/2d}.
$$

Note that $\langle \mathbf{W}_j, \boldsymbol{f}^\delta(x)\rangle = -1$ for $x \in \mathcal{X}_k^\delta$ and $j \neq k$. Observe that above derivation holds for all $k \in [K_n]$. Hence, we conclude

$$
|\langle \mathbf{W}_j, V_{\sigma_n}\mathbf{g}(x) - \boldsymbol{f}(x)\rangle| \leqslant 4K_n e^{-\sigma_n^2\Delta_k^\delta(x)^2/2d}
$$

for all $x \in \bigcup_{k=1}^{K_n} \mathcal{X}_k^\delta$. Then generalized geometric noise assumption for $t = \frac{2d}{\sigma_n^2}$ yields

$$
\mathcal{R}_\phi(\delta, V_{\sigma_n}\mathbf{g}) - \mathcal{R}_\phi(\delta, \boldsymbol{f}^\delta) \leqslant 4K_n(K_n - 1)U(2d)^{qd/2}\sigma_n^{-qd}.
$$

Therefore, combining both part, we conclude that

$$
\mathcal{A}_{\sigma_n}^\delta(\lambda_n) \leqslant cK_n^2 \lambda_n^{q/(q+1)},
$$

where c depends on the dimension of covariates d, geometric noise component q and associated constant U, and upper bound of volume of covariate space M_1 when we set $\sigma_n = \lambda_n^{-1/(q+1)d}$.

$\blacksquare$

Proof of Lemma 10:

To begin with, according to 1, it suffices to prove the result for $\widetilde{\mathcal{R}}_\phi(\delta^*, \widehat{\boldsymbol{f}}_n^{\delta^*}) - \widetilde{\mathcal{R}}_\phi^*$. Then, similar to the proof idea in Theorem 3.4 in Zhao et al. (2012), we observe the following decomposition:

$$
\begin{aligned}
\widetilde{\mathcal{R}}_\phi(\delta^*, \widehat{\boldsymbol{f}}_n^{\delta^*}) - \widetilde{\mathcal{R}}_\phi^* \leqslant &\left[\lambda_n \left\|\widehat{\boldsymbol{f}}_n^{\delta^*}\right\|_{\mathcal{F}_n}^2 + \widetilde{\mathcal{R}}_\phi(\delta^*, \widehat{\boldsymbol{f}}_n^{\delta^*}) - \inf_{\boldsymbol{f}\in\otimes_{k=1}^{K_n-1}\mathcal{H}_\kappa^k} \left(\lambda_n \|\boldsymbol{f}\|_{\mathcal{F}_n}^2 + \widetilde{\mathcal{R}}_\phi(\delta^*, \boldsymbol{f})\right)\right] \\
&+ \left[\inf_{\boldsymbol{f}\in\otimes_{k=1}^{K_n-1}\mathcal{H}_\kappa^k} \left(\lambda_n \|\boldsymbol{f}\|_{\mathcal{F}_n}^2 + \widetilde{\mathcal{R}}_\phi(\delta^*, \boldsymbol{f})\right) - \widetilde{\mathcal{R}}_\phi^*\right],
\end{aligned}
$$

where the first term is referred as stochastic error and the second term is called the approximation bias term. We will bound each term separately in the following. First, the approximation bias term has been bounded by $\mathcal{O}(K_n^2 \lambda_n^{q/(q+1)})$ in Lemma 9.

Next, to bound the stochastic error term, we follow the proof of Theorem 3.4 in Zhao et al. (2012). The proof of their theorem is basically derived by verifying the conditions of Theorem 5.6 in Steinwart and Scovel (2007). To achieve that, we first point out that, with Cauchy-Schwarz inequality, RAMSVM loss $(1 - \gamma) \sum_{k \neq \delta*(A)} (1 + \langle \mathbf{W}_j, \boldsymbol{f}(X) \rangle)^+ + \gamma(K_n - 1 - \langle \mathbf{W}_{\delta*(A)}, \boldsymbol{f}(X) \rangle)^+$ is Lipschitz continuous with respect to $\boldsymbol{f}$ with Lipschitz constant Z_n. In addition, let $\mathbb{P}^{\delta*}$ be the population measure of $(X, \delta*(A), R/p(\delta*(A)|X))$ and $\mathbb{P}_n^{\delta*}$ be the associated empirical measure. Next, by the definition of $\widehat{\boldsymbol{f}}_n^{\delta*}$, we have

$$\mathbb{P}_n^{\delta*} \big(\frac{R}{p(\delta*(A)|X)} L_\phi(\delta*, \widehat{\boldsymbol{f}}_n^{\delta*}) \big) + \lambda_n \big\| \widehat{\boldsymbol{f}}_n^{\delta*} \big\|_{\mathcal{F}_n}^2 \leqslant \mathbb{P}_n^{\delta*} \big(\frac{R}{p(\delta*(A)|X)} L_\phi(\delta*, \boldsymbol{f}) \big) + \lambda_n \|\boldsymbol{f}\|_{\mathcal{F}_n}^2 ,$$

holds for any $\boldsymbol{f} \in \otimes_{k=1}^{K_n-1} \mathcal{H}_\kappa^k$. Hence, we can select $\boldsymbol{f} = \boldsymbol{0}$ to get

$$\big\| \widehat{\boldsymbol{f}}_n^{\delta*} \big\|_{\mathcal{F}_n}^2 \leqslant \frac{1}{\lambda_n} \frac{1}{n} \sum_{i=1}^n \frac{R_i}{p(\delta*(A_i)|X_i)} L_\phi(\delta*, \boldsymbol{0}) \leqslant \frac{2}{\lambda_n}(K_n - 1)Z_n.$$

Therefore, it suffices to consider a ball with radius $r = \sqrt{2(K_n - 1)Z_n/\lambda_n}$ in the product RKHS $\otimes_{k=1}^{K_n-1} \mathcal{H}_\kappa^k$, denoted by $B_{\otimes_{k=1}^{K_n-1} \mathcal{H}_\kappa^k}(r)$. We define the function class

$$\mathcal{G}_{\lambda_n} = \left\{ \frac{R}{p(\delta*(A)|X)} L_\phi(\delta*, \boldsymbol{f}) + \lambda_n \|\boldsymbol{f}\|_{\mathcal{F}_n}^2 - \frac{R}{p(\delta*(A)|X)} L_\phi(\delta*, \boldsymbol{f}_{\lambda_n}^*) - \lambda_n \|\boldsymbol{f}_{\lambda_n}^*\|_{\mathcal{F}_n}^2 : \boldsymbol{f} \in B_{\otimes_{k=1}^{K_n-1} \mathcal{H}_\kappa^k}(r) \right\},$$

where $\boldsymbol{f}_{\lambda_n}^* = \arg\min_{\boldsymbol{f}} \left\{ \mathcal{R}_\phi(\delta*, \boldsymbol{f}) + \lambda_n \|\boldsymbol{f}\|_{\mathcal{F}_n}^2 : \boldsymbol{f} \in B_{\otimes_{k=1}^{K_n-1} \mathcal{H}_\kappa^k}(r) \right\}$. Then similar to the proof of Theorem 3.4 in Zhao et al. (2012), we verify the following three conditions: (i) By the Lipschitz continuity of L_ϕ, there exists constant B such that $\sup_{g \in \mathcal{G}_{\lambda_n}} \|g\|_\infty \leqslant B$ where $B = \mathcal{O}(\lambda_n^{-1/2})$; (ii) Based on the convexity of L_ϕ, there exists constant c such that $\mathbb{E}[g^2] \leqslant c\mathbb{E}[g]$ for all $g \in \mathcal{G}_{\lambda_n}$ where $c = \mathcal{O}(\lambda^{-1})$; (iii) Theorem 2.1 in Steinwart and Scovel (2007) gives that, for all $\sigma_n > 0, 0 < v < 2, \theta > 0, \epsilon > 0$, $\sup_{\mathbb{P}_n^{\delta*}} \log \mathcal{N}(B^{-1} \mathcal{G}_{\lambda_n}, \epsilon, L_2(\mathbb{P}_n^{\delta*})) \leqslant c_2 \sigma_n^{(1-v/2)(1+\theta)d} \epsilon^{-v}$. Then the result follows from Theorem 5.6 in Steinwart and Scovel (2007).

Combining the bound for stochastic error term and approximation bias term, we complete the proof. ∎

Proof of Equation (10):

We firstly derive the equivalence of the two optimization problems based on the 0-1 loss. For any fixed δ,

$$\min_{D_g} \mathbb{E}\left[\frac{(R-s(X))^+}{p(\delta(A)|X)}\mathbb{I}[D_g(X) \neq \delta(A)]\right] + (K_n - 1)\mathbb{E}\left[\frac{(R-s(X))^-}{p(\delta(A)|X)}\mathbb{I}[D_g(X) \neq \widetilde{\delta}(A)]\right]$$

$$= \min_{D_g} \mathbb{E}_{\delta(A),X}\left[\frac{\mathbb{E}[(R-s(X))^+|\delta(A),X]}{p(\delta(A)|X)}\mathbb{I}[D_g(X) \neq \delta(A)]\right]$$

$$+ (K_n - 1)\mathbb{E}_{\delta(A),X}\left[\frac{\mathbb{E}[(R-s(X))^-|\delta(A),X]}{p(\delta(A)|X)}\sum_{j \neq \delta(A)}\frac{1}{K_n - 1}\mathbb{I}[D_g(X) \neq j]\right]$$

$$= \min_{D_g} \mathbb{E}_{\delta(A),X}\left[\frac{\mathbb{E}[(R-s(X))^+|\delta(A),X]}{p(\delta(A)|X)}\mathbb{I}[D_g(X) \neq \delta(A)]\right]$$

$$+ \mathbb{E}_{\delta(A),X}\left[\frac{\mathbb{E}[(R-s(X))^-|\delta(A),X]}{p(\delta(A)|X)}\sum_{j \neq \delta(A)}\mathbb{I}[D_g(X) \neq j]\right]$$

$$= \min_{D_g} \mathbb{E}_X\left[\sum_{i=1}^{K_n}\mathbb{E}[(R-s(X))^+|\delta(A) = i, X]\mathbb{I}[D_g(X) \neq i]\right]$$

$$+ \mathbb{E}_X\left[\sum_{i=1}^{K_n}\mathbb{E}[(R-s(X))^-|\delta(A) = i, X]\sum_{j \neq i}\mathbb{I}[D_g(X) \neq j]\right]$$

$$= \min_{D_g} \mathbb{E}_X\left[\sum_{i=1}^{K_n}\mathbb{E}[(R-s(X))^+|\delta(A) = i, X]\mathbb{I}[D_g(X) \neq i]\right]$$

$$+ \mathbb{E}_X\left[\sum_{i=1}^{K_n}\mathbb{E}[(R-s(X))^-|\delta(A) = i, X]\left(\sum_{j=1}^{K_n}\mathbb{I}[D_g(X) \neq j] - \mathbb{I}[D_g(X) \neq i]\right)\right]$$

$$= \min_{D_g} \mathbb{E}_X\left[\sum_{i=1}^{K_n}\mathbb{E}[(R-s(X))^+|\delta(A) = i, X]\mathbb{I}[D_g(X) \neq i]\right]$$

$$+ \mathbb{E}_X\left[\sum_{i=1}^{K_n}\mathbb{E}[(R-s(X))^-|\delta(A) = i, X]\left(K_n - 1 - \mathbb{I}[D_g(X) \neq i]\right)\right]$$

$$= \min_{D_g} \mathbb{E}_X\left[\sum_{i=1}^{K_n}\mathbb{E}[R - s(X)|\delta(A) = i, X]\mathbb{I}[D_g(X) \neq i]\right]$$

$$+ (K_n - 1)\mathbb{E}_X\left[\sum_{i=1}^{K_n}\mathbb{E}[(R-s(X))^-|\delta(A) = i, X]\right]$$

$$= \min_{D_g} \mathbb{E}\left[\frac{R - s(X)}{p(\delta(A)|X)}\mathbb{I}[D_g(X) \neq \delta(A)]\right]$$

$$+ (K_n - 1)\mathbb{E}_X\left[\sum_{i=1}^{K_n}\mathbb{E}[(R-s(X))^-|\delta(A) = i, X]\right]$$

Note that the second term in the last equation is not related to D_g. Hence, the problem is equivalent with minimizing the first term. This finishes the proof for the equivalence based on the 0-1 loss. The equivalence based on the RAMSVM loss can be directly guaranteed by Fisher consistency from Theorem 4. ∎

Derivations from Problem (9) to Dual Problems (11) and (12):

For the linear kernel, we solve (9) with its dual form. After introducing new slack variables $(\xi_{i,j})_{i\in[n];j\in[K_n]}$, the problem with linear learning can be written as

$$\min_{\boldsymbol{\beta}_k,\xi_{i,j}} \frac{n\lambda}{2}\sum_{k=1}^{K_n-1}\boldsymbol{\beta}_k^T\boldsymbol{\beta}_k + \sum_{i=1}^{n}\omega_i\left[\gamma\xi_{i,\delta(A_i)} + (1-\gamma)\sum_{j\neq\delta(A_i)}\xi_{i,j}\right],$$

$$\text{s.t. } \xi_{i,j} \geqslant 0 \ (i\in[n], j\in[K_n]);$$
$$\xi_{i,\delta(A_i)} + \langle \boldsymbol{f}(X_i), \mathbf{W}_{\delta(A_i)}\rangle - (K_n-1) \geqslant 0 \ (i\in[n]);$$
$$\xi_{i,j} - \langle \boldsymbol{f}(X_i), \mathbf{W}_j\rangle - 1 \geqslant 0 \ (i\in[n], j\neq\delta(A_i)).$$

The corresponding Lagrangian function L_a is defined as

$$L_a = \frac{n\lambda}{2}\sum_{k=1}^{K_n-1}\boldsymbol{\beta}_k^T\boldsymbol{\beta}_k + \sum_{i=1}^{n}\omega_i\left[\gamma\xi_{i,\delta(A_i)} + (1-\gamma)\sum_{j\neq\delta(A_i)}\xi_{i,j}\right]$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{K_n}\rho_{i,j}\xi_{i,j} - \sum_{i=1}^{n}\alpha_{i,\delta(A_i)}[\xi_{i,\delta(A_i)} + \langle\boldsymbol{f}(X_i),\mathbf{W}_{\delta(A_i)}\rangle - (K_n-1)]$$

$$-\sum_{i=1}^{n}\sum_{j\neq\delta(A_i)}\alpha_{i,j}[\xi_{i,j} - \langle\boldsymbol{f}(X_i),\mathbf{W}_j\rangle - 1],$$

where $(\alpha_{i,j})_{i\in[n];j\in[K_n]}$ and $(\rho_{i,j})_{i\in[n];j\in[K_n]}$ are the Lagrangian multipliers. Furthermore, we can rewrite L_a with

$$L_a = \frac{n\lambda}{2}\sum_{k=1}^{K_n-1}\boldsymbol{\beta}_k^T\boldsymbol{\beta}_k + \sum_{i=1}^{n}\sum_{j=1}^{K_n}\left[\omega_i\big(\gamma\mathbb{I}[j=\delta(A_i)] + (1-\gamma)\mathbb{I}[j\neq\delta(A_i)]\big) - \rho_{i,j} - \alpha_{i,j}\right]\xi_{i,j}$$

$$+\sum_{i=1}^{n}\alpha_{i,\delta(A_i)}(K_n-1) + \sum_{i=1}^{n}\sum_{j\neq\delta(A_i)}\alpha_{i,j}$$

$$-\sum_{i=1}^{n}\alpha_{i,\delta(A_i)}\langle\boldsymbol{f}(X_i),\mathbf{W}_{\delta(A_i)}\rangle + \sum_{i=1}^{n}\sum_{j\neq\delta(A_i)}\alpha_{i,j}\langle\boldsymbol{f}(X_i),\mathbf{W}_j\rangle.$$

Next we take partial derivative of L_a with respect to $(\xi_{i,j})_{i\in[n];j\in[K_n]}$ and $(\boldsymbol{\beta}_k)_{k\in[K_n-1]}$, and let them be 0. We have

$$\frac{\partial L_a}{\partial \xi_{i,j}} = \omega_i\big(\gamma\mathbb{I}[j=\delta(A_i)] + (1-\gamma)\mathbb{I}[j\neq\delta(A_i)]\big) - \rho_{i,j} - \alpha_{i,j} = 0 \ (i\in[n], j\in[K_n]),$$

and

$$\frac{\partial L_a}{\partial \boldsymbol{\beta}_k} = n\lambda\boldsymbol{\beta}_k - \sum_{i=1}^{n}\alpha_{i,\delta(A_i)}\mathbf{W}_{\delta(A_i),k}X_i + \sum_{i=1}^{n}\sum_{j\neq\delta(A_i)}\alpha_{i,j}\mathbf{W}_{j,k}X_i = 0 \ (k\in[K_n-1]).$$

When the partial derivatives for $(\boldsymbol{\beta}_k)_{k\in[K_n-1]}$ equal to 0, we have

$$\boldsymbol{\beta}_k = \frac{1}{n\lambda}\left[\sum_{i=1}^{n}\alpha_{i,\delta(A_i)}\mathbf{W}_{\delta(A_i),k}X_i - \sum_{i=1}^{n}\sum_{j\neq\delta(A_i)}\alpha_{i,j}\mathbf{W}_{j,k}X_i\right].$$

After plugging the above characterizations of $\boldsymbol{\beta}$ into L_a, we can simplify L_a with

$$L_a = -\frac{n\lambda}{2}\sum_{k=1}^{K_n-1}\boldsymbol{\beta}_k^T\boldsymbol{\beta}_k + \sum_{i=1}^n \alpha_{i,\delta(A_i)}(K_n-1) + \sum_{i=1}^n\sum_{j\neq\delta(A_i)}\alpha_{i,j}.$$

Note that maximizing L_a with respect to $\alpha_{i,j}$ is equivalent to minimizing $-L_a$, which solves the dual problem (11). The constraints of problem (11) come from $(\alpha_{i,j})_{i\in[n];j\in[K_n]} \geqslant 0$, $(\rho_{i,j})_{i\in[n];j\in[K_n]} \geqslant 0$, and $(\partial L_a/\partial\xi_{i,j})_{i\in[n];j\in[K_n]} = 0$.

For the general kernel, we similarly introduce the slack variables $(\xi_{i,j})_{i\in[n];j\in[K_n]}$. If the intercepts $(\theta_{k,0})_{k\in[K_n-1]}$ are included in the penalty term, then (9) is equivalent to

$$\min_{\boldsymbol{\theta}_k,\theta_{k,0},\xi_{i,j}}\frac{n\lambda}{2}\sum_{k=1}^{K_n-1}\boldsymbol{\theta}_k^T\mathbf{G}\boldsymbol{\theta}_k + \frac{n\lambda}{2}\sum_{k=1}^{K_n-1}\theta_{k,0}^2 + \sum_{i=1}^n\omega_i\left[\gamma\xi_{i,\delta(A_i)} + (1-\gamma)\sum_{j\neq\delta(A_i)}\xi_{i,j}\right],$$

$$\text{s.t. } \xi_{i,j} \geqslant 0 \ (i\in[n], j\in[K_n]);$$
$$\xi_{i,\delta(A_i)} + \langle\boldsymbol{f}(X_i),\mathbf{W}_{\delta(A_i)}\rangle - (K_n-1) \geqslant 0 \ (i\in[n]);$$
$$\xi_{i,j} - \langle\boldsymbol{f}(X_i),\mathbf{W}_j\rangle - 1 \geqslant 0 \ (i\in[n], j\neq\delta(A_i)).$$

Similar to the linear case, we introduce the Lagrangian multipliers $(\rho_{i,j})_{i\in[n];j\in[K_n]}$ and $(\alpha_{i,j})_{i\in[n];j\in[K_n]}$, calculate partial derivative with respect to $(\boldsymbol{\theta}_q)_{q\in[K_n-1]}, (\theta_{q,0})_{q\in[K_n-1]}$, and $(\xi_{i,j})_{i\in[n];j\in[K_n]}$, and set the derivatives to 0. Then we have that

$$\boldsymbol{\theta}_k = \frac{1}{n\lambda}\mathbf{G}^{-1}\left[\sum_{i=1}^n\alpha_{i,\delta(A_i)}\mathbf{W}_{\delta(A_i),k}\mathbf{G}_i - \sum_{i=1}^n\sum_{j\neq\delta(A_i)}\alpha_{i,j}\mathbf{W}_{j,k}\mathbf{G}_i\right],$$

and

$$\theta_{k,0} = \frac{1}{n\lambda}\left[\sum_{i=1}^n\alpha_{i,\delta(A_i)}\mathbf{W}_{\delta(A_i),k} - \sum_{i=1}^n\sum_{j\neq\delta(A_i)}\alpha_{i,j}\mathbf{W}_{j,k}\right],$$

where $\mathbf{G}_i$ is the i-th column of $\mathbf{G}$. After plugging the above characterizations of $\boldsymbol{\theta}$ into the Lagrangian function and rewriting it, we get (12). $\blacksquare$

Appendix D. Additional Simulation Results

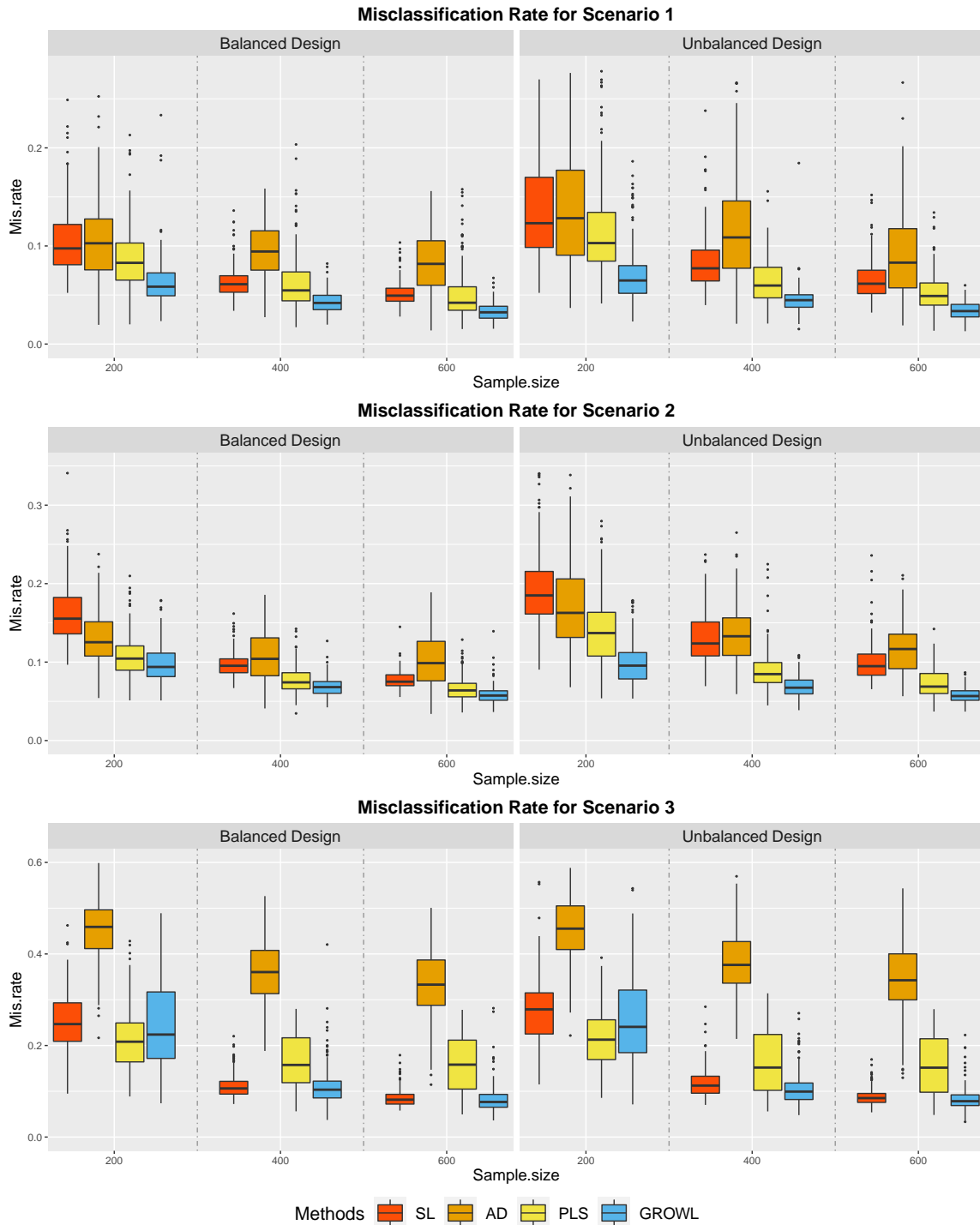### D.1. Misclassification Rate Under the Homogeneous Settings.

Figure 5: Boxplots of Misclassificaiton Rate evaluated on the independent test data under the **homogeneous** settings.

D.2. Simulations Under More Unbalanced Designs.

To better demonstrate the performance of GROWL, we conduct several other simulations under more unbalanced designs for Scenario 3 in homogeneous setting. For the unbalanced design (II), the propensity scores of 15 treatments are set to be $\big((0.035, 0.035, 0.035, 0.114, 0.114), (0.035, 0.035, 0.035, 0.114, 0.114), (0.035, 0.035, 0.088, 0.088, 0.088)\big)$, and for the unbalanced design (III), the propensity scores are set to be $\big((0.020, 0.020, 0.020, 0.137, 0.137), (0.020, 0.020, 0.020, 0.137, 0.137), (0.020, 0.020, 0.098, 0.098, 0.098)\big)$. Varying from balanced design, unbalanced design (I) (same setting as shown in Section 4.1), unbalanced design (II), and unbalanced design (III), the treatment propensities become more and more unbalanced. In addition, we conduct another simulation setting where the propensity scores of one of the three treatment groups is extremely small. Specifically, for this extreme case, the propensity scores are $\big((0.070, 0.070, 0.087, 0.087, 0.087), (0.100, 0.100, 0.100, 0.100, 0.100), (0.020, 0.020, 0.020, 0.020, 0.020)\big)$. Overall, GROWL still performs the best in most cases shown in Figures 6 and 7. As treatment propensities become more unbalanced, GROWL may need more training data in order to learn the true partition δ^0 correctly. In addition, a roughly correct $\widehat{\delta}_n$ is still helpful in terms of the performance of the value function. Compared with other methods that do not consider the treatment structure, GROWL is able to combine the similar treatments into the treatment groups. The decision rules learned from GROWL perform better because they are estimated under the treatment groups that have more observations than the individual treatments.
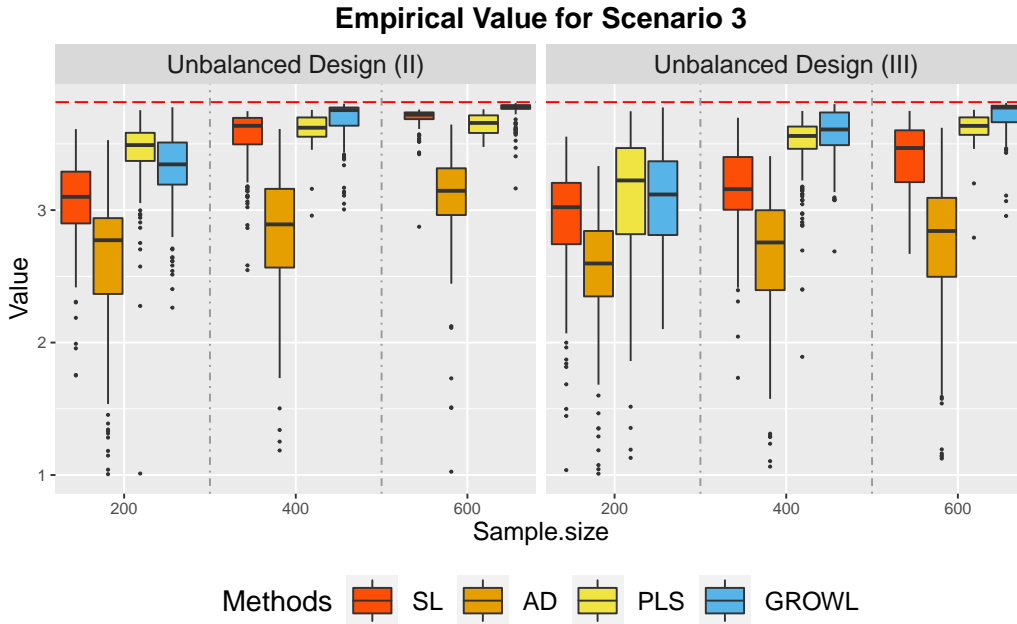

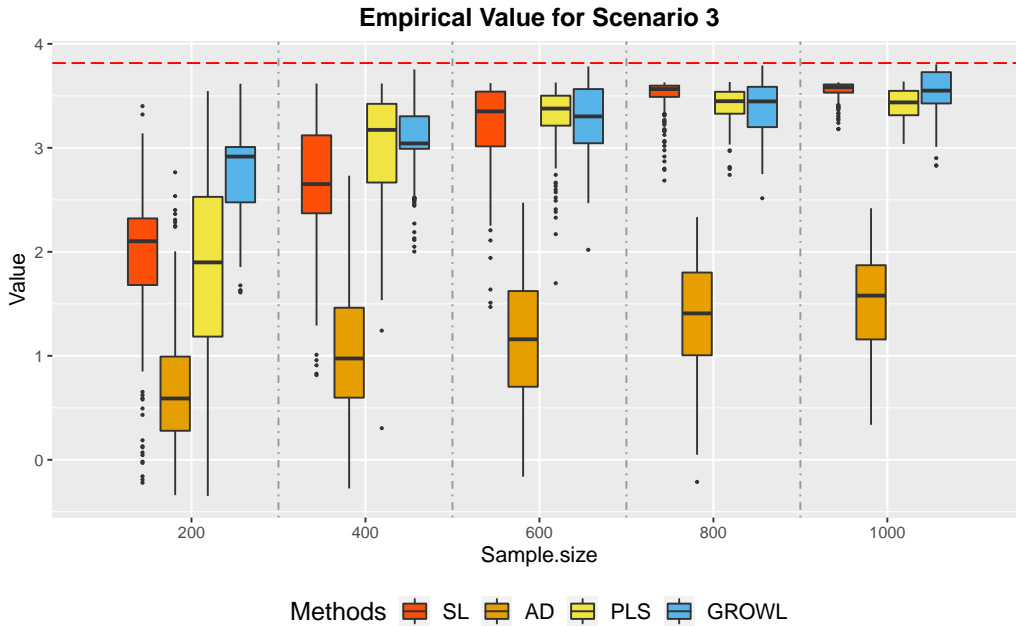Figure 6: Boxplots of Empirical Value Function under more **unbalanced** designs for the homogeneous case.

Figure 7: Boxplots of Empirical Value Function under **extreme** designs for the homogeneous case.

D.3. Simulations When Propensity Scores Depend on the Covariates.

We conduct more general simulation settings when the treatment propensity scores depend on the covariates. Specifically, the treatment propensity score $p(a|x)$ is provided with the following multinomial model:

$$\log \frac{p(a|X)}{p(1|X)} = \beta_{a,0} + \beta_{a,1}X_1 + \beta_{a,2}X_2 + \cdots + \beta_{a,10}X_{10},$$

for $a = 2, \ldots, M$, where $\beta_{i,j}$'s are all generated independently from $U[-0.1, 0.1]$. For the homogeneous case, we take Scenario 1 as an example. The value shown in Figure 8 demonstrates that GROWL still has superior performance over other methods. For the non-homogeneous case in Scenario 4, as shown in Figure 9, the overall trend of GROWL is similar to that of the unbalanced design in Figure 2. GROWL is still competitive, and especially has smaller variance than other methods.
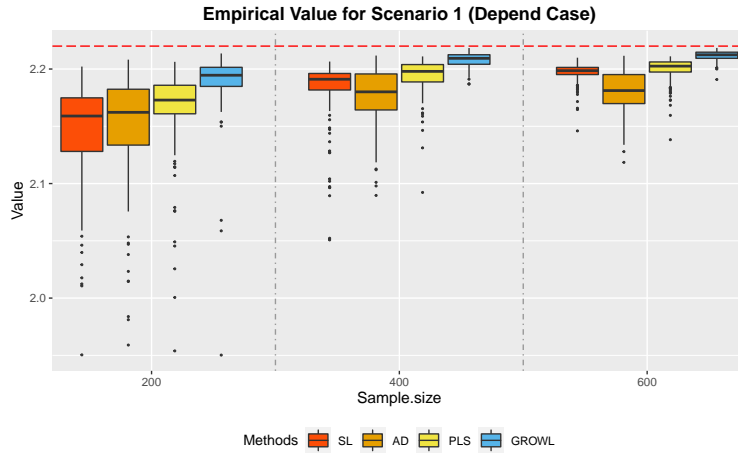
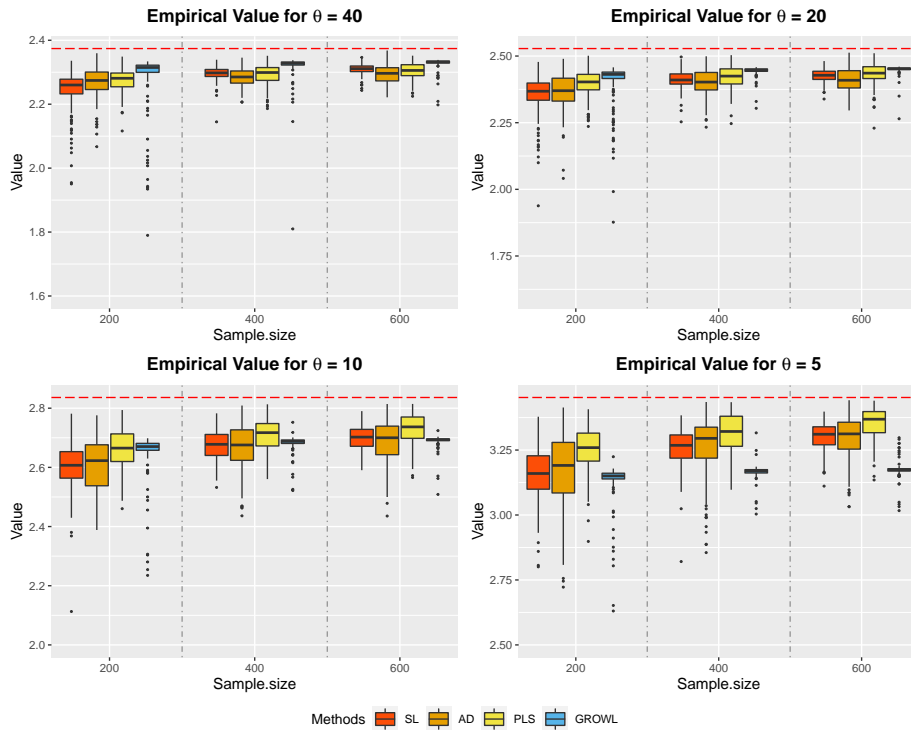Figure 8: Boxplots of Empirical Value Function under the **dependent** design for the homogeneous case.

Figure 9: Boxplots of Empirical Value Function under the **dependent** design for the non-homogeneous case.

D.4. Estimated Coefficients in STAR*D Analysis.

Table 3: Estimated coefficients of linear comparison function for GROWL on the STAR*D dataset. Larger coefficients encourage better reward.

Variable Name	Group 1	Group 2	Group 3	Group 4	Group 5
intercept	0.713	0.168	0.658	0.191	-1.730
gender (female)	-0.839	0.249	1.635	-0.444	-0.601
ethnic (white)	-0.878	-0.180	-0.764	0.221	1.601
age	-0.959	-0.319	0.579	-0.746	1.445
depression history (yes)	-1.427	0.021	-0.368	0.528	1.245
marital	1.049	0.371	-1.636	0.282	-0.066
school years	1.421	-0.151	-0.033	0.153	-1.390
education	0.526	-1.230	1.166	-0.841	0.379
student (yes)	-0.813	1.704	-0.541	-0.377	0.026
employment	0.178	-0.867	-1.085	1.369	0.405
volunteer work	0.066	-0.750	-0.056	-0.886	1.626
QIDS change during Level 1	-1.574	0.136	1.206	0.228	0.004
QIDS at the start of Level 2	0.743	0.576	-0.952	0.851	-1.218

Appendix E. Implementation Details

Algorithm 1: GROWL using the Genetic Algorithm

Initialize.

a. Fit (penalized) linear regression $R \sim X$ with training data and get residuals r;

b. Input the estimated initial partition set Δ_0.

For each partition δ in Δ_0, **do**

a. Fit treatment A into group $\delta(A)$ based on δ

b. **If** $r > 0$, Stay with the same assigned treatment $\delta(A)$, and set $\bar{\delta}(A) = \delta(A)$;

   Else Uniformly switch $\delta(A)$ to arbitrary unassigned treatment $\widetilde{\delta}(A)$, and set $\bar{\delta}(A) = \widetilde{\delta}(A)$;

c. Use $(R, \bar{\delta}(A), X)$ to fit RAMSVM with weights $w_G = \frac{|r|}{p(\delta(A)|X)}$;

d. Get fitted decision function $\widehat{\boldsymbol{f}}^\delta$ based on RAMSVM;

e. Plug $(\delta, \widehat{\boldsymbol{f}}^\delta)$ back into the empirical average of the risk function $\widetilde{\mathcal{R}}_\phi$;

f. Get the risk value for δ.

Updated the set Δ using the **Genetic Algorithm** until convergence;

Obtain the optimal $\widehat{\delta}$ and corresponding $\widehat{D}_g$ under group domain;

Sample one treatment from group $\widehat{D}_g(X)$ based on π_δ, and finally get the ITR $D(X)$.

Algorithm 2: GROWL with the Greedy Adjustment

Initialize.

a. Fit (penalized) linear regression $R \sim X$ with training data and get residuals r;

b. Input the estimated initial partition set Δ_0.

Update δ, **do**

    For $j = 1, 2, \ldots, M_n, 1, 2, \ldots, M_n, \ldots$, **do**

a. Adjust the assignment of the j-th treatment and hold the assignment for others fixed;

b. Get the risk value for the adjusted δ in the same way shown in Algorithm 1;

c. Obtain the locally best δ in cyclic fashion.

Until convergence.

Obtain the optimal $\widehat{\delta}$ and corresponding $\widehat{D}_g$ under group domain;

Sample one treatment from group $\widehat{D}_g(X)$ based on π_δ, and finally get the ITR $D(X)$.

References

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89 (1):133–161, 2021.

Peter L Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3:463–482, 2002.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

Jingxiang Chen, Quoc Tran-Dinh, Michael R Kosorok, and Yufeng Liu. Identifying heterogeneous effect using latent supervised clustering with adaptive fusion. *Journal of Computational and Graphical Statistics*, 30(1):43–54, 2021.

Shuai Chen, Lu Tian, Tianxi Cai, and Menggang Yu. A general statistical framework for subgroup identification and comparative treatment scoring. *Biometrics*, 73(4):1199–1209, 2017.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

Ashkan Ertefaie, Tianshuang Wu, Kevin G Lynch, and Inbal Nahum-Shani. Identifying a set that contains the best dynamic treatment regimes. *Biostatistics*, 17(1):135–148, 2016.

David E Goldberg and John Henry Holland. *Genetic Algorithms and Machine Learning*. Kluwer Academic Publishers, 1988.

Arthur E Hoerl and Robert W Kennard. Ridge regression: applications to nonorthogonal problems. *Technometrics*, 12(1):69–82, 1970.

Eric B Laber, Daniel J Lizotte, and Bradley Ferguson. Set-valued dynamic treatment regimes for competing outcomes. *Biometrics*, 70(1):53–61, 2014.

Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Springer Science & Business Media, 2013.

Ying Liu, Yuanjia Wang, Michael R Kosorok, Yingqi Zhao, and Donglin Zeng. Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens. *Statistics in Medicine*, 37(26):3776–3788, 2018.

Haomiao Meng, Yingqi Zhao, Haoda Fu, and Xingye Qiao. Near-optimal individualized treatment recommendations. *Journal of Machine Learning Research*, 21(183):1–28, 2020.

Olga Montvida, Jonathan Shaw, John J Atherton, Frances Stringer, and Sanjoy K Paul. Long-term trends in antidiabetes drug usage in the us: real-world evidence in patients newly diagnosed with type 2 diabetes. *Diabetes Care*, 41(1):69–78, 2018.

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

Yinghao Pan and Yingqi Zhao. Improved doubly robust estimation in learning optimal individualized treatment rules. *Journal of the American Statistical Association*, 116(533): 283–294, 2021.

Eric C Polley and Mark J Van Der Laan. Super learner in prediction. *U.C. Berkeley Division of Biostatistics Working Paper Series*, 2010.

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

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

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

Naim U Rashid, Daniel J Luckett, Jingxiang Chen, Michael T Lawson, Longshaokan Wang, Yunshu Zhang, Eric B Laber, Yufeng Liu, Jen Jen Yeh, Donglin Zeng, and Michael R Kosorok. High-dimensional precision medicine from patient-derived xenografts. *Journal of the American Statistical Association*, 116(535):1140–1154, 2021.

Janice H Robinson, Lynn C Callister, Judith A Berry, and Karen A Dearing. Patient-centered care and adherence: Definitions and applications to improve outcomes. *Journal of the American Academy of Nurse Practitioners*, 20(12):600–607, 2008.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

A John Rush, Maurizio Fava, Stephen R Wisniewski, Philip W Lavori, Madhukar H Trivedi, Harold A Sackeim, Michael E Thase, Andrew A Nierenberg, Frederic M Quitkin, T Michael Kashner, David J Kupfer, Jerrold F Rosenbaum, Jonathan Alpert, Jonathan W Stewart, Patrick J McGrath, Melanie M Biggs, Kathy Shores-Wilson, Barry D Lebowitz, Louise Ritz, and George Niederehe. Sequenced treatment alternatives to relieve depression (star* d): rationale and design. *Controlled Clinical Trials*, 25(1):119–142, 2004.

Luca Scrucca. Ga: A package for genetic algorithms in r. *Journal of Statistical Software*, 53(4):1–37, 2013.

Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer Science & Business Media, 2008.

Ingo Steinwart and Clint Scovel. Fast rates for support vector machines using gaussian kernels. *The Annals of Statistics*, 35(2):575–607, 2007.

Lu Tian, Ash A Alizadeh, Andrew J Gentles, and Robert Tibshirani. A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association*, 109(508):1517–1532, 2014.

William N Venables and Brian D Ripley. *Modern Applied Statistics with S-PLUS*. Springer Science & Business Media, 2013.

Christopher John Cornish Hellaby Watkins. *Learning From Delayed Rewards*. PhD thesis, University of Cambridge, UK, 1989.

Peng Wu, Donglin Zeng, and Yuanjia Wang. Matched learning for optimizing individualized treatment strategies using electronic health records. *Journal of the American Statistical Association*, 115(529):380–392, 2020.

Yichao Wu and Yufeng Liu. Robust truncated hinge loss support vector machines. *Journal of the American Statistical Association*, 102(479):974–983, 2007.

Yingcun Xia, Howell Tong, Wai Keung Li, and Li-Xing Zhu. An adaptive estimation of dimension reduction space. In *Exploration of A Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics*, pages 299–346. World Scientific, 2009.

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

Chong Zhang and Yufeng Liu. Multicategory angle-based large-margin classification. *Biometrika*, 101(3):625–640, 2014.

Chong Zhang, Yufeng Liu, Junhui Wang, and Hongtu Zhu. Reinforced angle-based multicategory support vector machines. *Journal of Computational and Graphical Statistics*, 25(3):806–825, 2016.

Chong Zhang, Jingxiang Chen, Haoda Fu, Xuanyao He, Yingqi Zhao, and Yufeng Liu. Multicategory outcome weighted margin-based learning for estimating individualized treatment rules. *Statistica Sinica*, 30(4):1857–1879, 2021.

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

Yingqi Zhao, Donglin Zeng, Eric B Laber, and Michael R Kosorok. New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, 110(510):583–598, 2015.

Xin Zhou, Nicole Mayer-Hamblett, Umer Khan, and Michael R Kosorok. Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association*, 112(517):169–187, 2017.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.