Published in final edited form as:
J Am Stat Assoc. 2012 September 1; 107(449): 1106–1118. doi:10.1080/01621459.2012.695674.

Estimating Individualized Treatment Rules Using Outcome Weighted Learning

Yingqi Zhao, Ph.D. [Candidate],
Department of Biostatistics, University of North Carolina at Chapel Hill, NC 27599

Donglin Zeng [Associate professor],
Department of Biostatistics, University of North Carolina at Chapel Hill, NC 27599

A. John Rush [Professor and Vice-Dean], and
Office of Clinical Sciences, Duke-National University of Singapore Graduate Medical School, Singapore 169857

Michael R. Kosorok [Professor and Chair]
Department of Biostatistics, and Professor, Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Yingqi Zhao: yqzhao@live.unc.edu; Donglin Zeng: dzeng@email.unc.edu; A. John Rush: john.rush@duke-nus.edu.sg; Michael R. Kosorok: kosorok@unc.edu

Abstract

There is increasing interest in discovering individualized treatment rules for patients who have heterogeneous responses to treatment. In particular, one aims to find an optimal individualized treatment rule which is a deterministic function of patient specific characteristics maximizing expected clinical outcome. In this paper, we first show that estimating such an optimal treatment rule is equivalent to a classification problem where each subject is weighted proportional to his or her clinical outcome. We then propose an outcome weighted learning approach based on the support vector machine framework. We show that the resulting estimator of the treatment rule is consistent. We further obtain a finite sample bound for the difference between the expected outcome using the estimated individualized treatment rule and that of the optimal treatment rule. The performance of the proposed approach is demonstrated via simulation studies and an analysis of chronic depression data.

Keywords

Dynamic Treatment Regime; Individualized Treatment Rule; Weighted Support Vector Machine; RKHS; Risk Bound; Bayes Classifier; Cross Validation

1. INTRODUCTION

In many different diseases, patients can show significant heterogeneity in response to treatments. In some cases, a drug that works for a majority of individuals may not work for a subset of patients with certain characteristics. For example, molecularly targeted cancer drugs are only effective for patients with tumors expressing targets (Grünwald & Hidalgo 2003; Buzdar 2009), and significant heterogeneity exists in responses among patients with different levels of psychiatric symptoms (Piper et al. 1995; Crits-Christoph et al. 1999). Thus significant improvements in public health could potentially result from judiciously treating individuals based on his or her prognostic or genomic data rather than a “one size fits all” approach. Treatments and clinical trials tailored for patients have enjoyed recent popularity in clinical practice and medical research, and, in some cases, have provided high

quality recommendations accounting for individual heterogeneity (Sargent et al. 2005; Flume et al. 2007; Insel 2009). These proposals have focused on smaller, specific and well-defined subgroups, sought to provide guidance in clinical decision making based on individual differences, and have attempted to achieve better risk minimization and benefit maximization.

One statistical approach for developing individual-adaptive interventions is to classify subjects into different risk levels estimated by a parametric or semiparametric regression model using prognostic factors, and then to assign therapy according to risk level (Eagle et al. 2004; Marlowe et al. 2007; Cai et al. 2010). However, the parametric or semiparametric model assumptions may not be valid due to the complexity of the disease mechanism and individual heterogeneity. Moreover, these approaches require preknowledge in allocating the optimal treatment to each risk category. There is also a significant literature examining discovery and development of personalized treatment relying on predicting patient responses to optional regimens (Rosenwald et al. 2002; van't Veer & Bernards 2008), where the optimal decision leads to the best predicted outcome. One recent paper by Qian & Murphy (2011) applies a two-step procedure which first estimates a conditional mean for the response and then estimates the rule maximizing this conditional mean. A rich linear model is used to sufficiently approximate the conditional mean, with the estimated rule derived via l_1 penalized least squares (l_1-PLS). The method includes variable selection to facilitate parsimony and ease of interpretation. The conditional mean approximation requires estimating a prediction model of the relationship between pretreatment prognostic variables, treatments and clinical outcome using a prediction model. Reduction in the mean response is related to the excess prediction error, through which an upper bound can be constructed for the mean reduction of the associated treatment rule. However, by inverting the model to find the optimal treatment rule, this method emphasizes prediction accuracy of the clinical response model instead of directly optimizing the decision rule.

In this paper, we proposed a new method for solving this problem which circumvents the need for conditional mean modeling followed by inversion by directly estimating the decision rule which maximizes clinical response. Specifically, we demonstrate that the optimal treatment rule can be estimated within a weighted classification framework, where the weights are determined from the clinical outcomes. We then alleviate the computational problem by substituting the 0–1 loss in the classification with a convex surrogate loss as is done with the support vector machine (SVM) via the hinge loss (Cortes & Vapnik 1995). The directness of this outcome weighted learning (OWL) approach enables us to better select targeted therapy while making full use of available information.

The remainder of the paper is organized as follows. In Section 2, we provide the mathematical concepts and framework for individualized treatment rules, and then formulate the problem as OWL. The proposed weighted SVM approach for constructing the optimal ITR is then developed in detail. In Section 3, consistency and risk bound results are established for the estimated rules. Faster convergence rates can be achieved with additional marginal assumptions on the data generating distribution. We present simulation studies to evaluate performance of the proposed method in Section 4. The method is then illustrated on the Nefazodone-CBASP data (Keller et al. 2000) in Section 5. In Section 6, we discuss future work. The proofs of theoretical results are given in the Appendix.

2. METHODOLOGY

2.1 Individualized Treatment Rule (ITR)

We assume the data are collected from a two-arm randomized trial. That is, treatment assignments, denoted by $A \in \mathcal{A} = \{-1, 1\}$, are independent of any patient's prognostic

variables, which are denoted as a d-dimensional vector $X = (X_1, \ldots, X_d)^T \in \mathcal{X}$. We let R be the observed clinical outcome, also called the "reward," and assume that R is bounded, with larger values of R being more desirable. Thus an individualized treatment rule (ITR) is a map from the space of prognostic variables, $\mathcal{X}$, to the space of treatments, $\mathcal{A}$. An optimal ITR is a rule that maximizes the expected reward if implemented.

Mathematically, we can quantify the optimal ITR in terms of the relationship among (X, A, R). To see this, denote the distribution of (X, A, R) by P and expectation with respect to the P is denoted by E. For any given ITR $\mathcal{D}$, we let $P^{\mathcal{D}}$ denote the distribution of (X, A, R) given that $A = \mathcal{D}(X)$, i.e., the treatments are chosen according to the rule $\mathcal{D}$; correspondingly, the expectation with respect to $P^{\mathcal{D}}$ is denoted by $E^{\mathcal{D}}$. Then under the assumption that $P(A = a) > 0$ for $a = 1$ and -1, it is clear that $P^{\mathcal{D}}$ is absolutely continuous with respect to P and $dP^{\mathcal{D}}/dP = I(a = \mathcal{D}(x))/P(A = a)$, where $I(\cdot)$ is the indicator function. Thus, the expected reward under the ITR $\mathcal{D}$ is given as

$$E^{\mathcal{D}}(R) = \int R dP^{\mathcal{D}} = \int R \frac{dP^{\mathcal{D}}}{dP} dP = E\left[\frac{I(A=\mathcal{D}(X))}{A\pi+(1-A)/2}R\right],$$

where $\pi = P(A = 1)$. This expectation is called the value function associated with $\mathcal{D}$ and is denoted $\mathcal{V}(\mathcal{D})$. Consequently, an optimal ITR, $\mathcal{D}^*$, is a rule that maximizes $\mathcal{V}(\mathcal{D})$, i.e.,

$$\mathcal{D}^* \in \underset{\mathcal{D}}{\operatorname{argmax}}\, E\left[\frac{I(A=\mathcal{D}(X))}{A\pi+(1-A)/2}R\right].$$

Note that $\mathcal{D}^*$ does not change if R is replaced by $R + c$ for any constant c. Thus, without loss of generality, we assume that R is nonnegative in the following.

2.2 Outcome Weighted Learning (OWL) for Estimating Optimal ITR

Assume that we observe i.i.d data (X_i, A_i, R_i), $i = 1, \ldots, n$ from the two-arm randomized trial described above. Previous approaches to estimating optimal ITR first estimate $E(R|X, A)$, using the observed data via parametric or semiparametric models, and then estimate the optimal decision rule by comparing the predicted value $E(R|X, A = 1)$ versus $E(R|X, A = -1)$ (Robins 2004; Moodie et al. 2009; Qian & Murphy 2011). As discussed before, these approaches indirectly estimate the optimal ITR, and are likely to produce a suboptimal ITR if the model for R given (X, A) is overfitted. As an alternative, we propose a nonparametric approach which directly maximizes the value function based on an outcome weighted learning method.

To illustrate our approach, we first notice that searching for the optimal ITR, $\mathcal{D}^*$, which maximizes $\mathcal{V}(\mathcal{D})$, is equivalent to finding $\mathcal{D}^*$ that minimizes

$$E[R|A=1] + E[R|A=-1] - \mathcal{V}(\mathcal{D}) = E\left[\frac{I(A \neq \mathcal{D}(X))}{A\pi+(1-A)/2}R\right].$$

The latter can be viewed as a weighted classification error, for which we want to classify A using X but we also weigh each misclassification event by $R/(A\pi + (1 - A)/2)$. Hence, using the observed data, we approximate the weighted classification error by

$$n^{-1}\sum_{i=1}^{n}\frac{R_i}{A_i\pi+(1-A_i)/2}I(A_i \neq \mathcal{D}(X_i))$$

and seek to minimize this expression to estimate $\mathcal{D}^*$. Since $\mathcal{D}(x)$ can always be represented as sign($f(x)$), for some decision function f, minimizing the above expression for $\mathcal{D}^*$ is equivalent to minimizing

$$n^{-1}\sum_{i=1}^{n}\frac{R_i}{A_i\pi+(1-A_i)/2}I(A_i \neq \text{sign}(f(X_i))) \quad (2.1)$$

to obtain the optimal f^*, and then setting $\mathcal{D}^*(x) = \text{sign}(f^*(x))$.

The above minimization also has the following interpretation. That is, we intend to find a decision rule which assigns treatments to each subject only based on their prognostic information. For subjects observed to have a large reward, this rule is apt to recommend the same treatment assignments that the subject has actually received; however, for subjects with small rewards, the rule is more likely to give the opposite treatment assignment to what they received. In other words, if we stratify subjects into different strata based on the rewards, we will expect that the optimal ITR misclassifies less subjects in the high reward stratum as compared to the low reward stratum.

In the machine learning literature, (2.1) can be viewed as a weighted summation of 0–1 loss. It is well known that minimizing (2.1) is difficult due to the discontinuity and non-convexity of 0–1 loss. To alleviate this difficulty, one common approach is to find a convex surrogate loss for the 0–1 loss in (2.1) and develop a tractable estimation procedure (Zhang 2004; Lugosi & Vayatis 2004; Steinwart 2005). Among many choices of surrogate loss, one of the most popular is the hinge loss used in the context of the support vector machine (Cortes & Vapnik 1995), which we will adopt in this paper. Furthermore, we penalize the complexity of the decision function in order to avoid overfitting. In other words, instead of minimizing (2.1), we aim to minimize

$$n^{-1}\sum_{i=1}^{n}\frac{R_i}{A_i\pi+(1-A_i)/2}(1-A_i(f(X_i)))^{+}+\lambda_n\|f\|^2, \quad (2.2)$$

where $x^+ = \max(x, 0)$ and $\|f\|$ is some norm for f. In this way, we cast the problem of estimating the optimal ITR into a weighted classification problem using support vector machine techniques.

2.3 Linear Decision Rule for Optimal ITR

Suppose that the decision function $f(x)$ minimizing (2.2) is a linear function of x, that is, $f(x) = \langle\beta, x\rangle + \beta_0$, where $\langle\cdot,\cdot\rangle$ denotes the inner product in Euclidean space. Then the corresponding ITR will assign a subject with prognostic value X into treatment 1 if $\langle\beta, X\rangle + \beta_0 > 0$ and −1 otherwise.

In (2.2), we define $\|f\|$ as the Euclidean norm of β. Following the usual SVM, we introduce a slack variable ξ_i for subject i to allow a small portion of wrong classification. Denote $C > 0$ as the classifier margin. Then minimizing (2.2) can be rewritten as

$$\max_{\beta,\beta_0,\left\|\beta\right\|=1} C \text{ subject to } A_i(\langle\beta, X_i\rangle+\beta_0) \geq C(1-\xi_i), \xi_i \geq 0, \sum \frac{R_i}{\pi_i}\xi_i<s,$$

where $\pi_i = \pi I(A_i = 1)+(1 - \pi)I(A_i = -1)$ and s is a constant depending on λ_n. This is equivalent to

$$\min \frac{1}{2}\left\|\beta\right\|^2 \quad \text{subject to} \quad A_i(\langle\beta, X_i\rangle+\beta_0) \geq (1-\xi_i), \xi_i \geq 0, \sum \frac{R_i}{\pi_i}\xi_i<s,$$

that is

$$\min \frac{1}{2}\left\|\beta\right\|^2+\kappa\sum_{i=1}^{n}\frac{R_i}{\pi_i}\xi_i \quad \text{subject to} \quad A_i(\langle\beta, X_i\rangle+\beta_0) \geq (1-\xi_i), \xi_i \geq 0,$$

where $\kappa > 0$ is a tuning parameter and R_i/π_i is the weight for the i^{th} point. We observe that the main difference compared to standard SVM is that we weigh each slack variable ξ_i with R_i/π_i.

After introducing Lagrange multipliers, the Lagrange function becomes:

$$\frac{1}{2}\left\|\beta\right\|^2+\kappa\sum_{i=1}^{n}\frac{R_i}{\pi_i}\xi_i-\sum_{i=1}^{n}\alpha_i\{A_i(X_i^T\beta+\beta_0)-(1-\xi_i)\}-\sum_{i=1}^{n}\mu_i\xi_i,$$

with $\alpha_i \geq 0$, $\mu_i \geq 0$. Taking derivatives with respect to (β, β_0) and ξ_i, we have $\beta=\sum_{i=1}^{n}\alpha_i A_i X_i$, $0=\sum_{i=1}^{n}\alpha_i A_i$ and $\alpha_i = \kappa R_i/\pi_i - \mu_i$. Plugging these equations into the Lagrange function, we obtain the dual problem

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

subject to $0 \leq \alpha_i \leq \kappa R_i/\pi_i$, $i = 1, \ldots, n$, and $\sum_{i=1}^{n}\alpha_i A_i=0$. Quadratic programming algorithms from many widely available software packages can be used to solve this dual problem. Finally, we obtain that

$$\widehat{\beta}=\sum_{\widehat{\alpha}_i>0}\widehat{\alpha}_i A_i X_i,$$

and $\hat{\beta}_0$ can be solved using the margin points ($0 < \hat{\alpha}_i$, $\hat{\xi}_i = 0$) subject to the Karush-Kuhn-Tucker conditions (Page 421, Hastie, Tibshirani & Friedman 2009). The decision rule is given by $\text{sign}\{\langle\hat{\beta}, X\rangle + \hat{\beta}_0\}$. Similar to the traditional SVM, the estimated decision rule is determined by the support vectors with $\hat{\alpha} > 0$.

2.4 Nonlinear Decision Rule for Optimal ITR

The previous section targets a linear boundary of prognostic variables. This may not be practically useful since the dimension of the prognostic variables can be quite high and complicated relationships may be involved between the desired treatments and these variables. However, we can easily generalize the previous approach to obtain a nonlinear decision rule for obtaining the optimal ITR.

We let $k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, called a kernel function, be continuous, symmetric and positive semidefinite. Given a real-valued kernel function k, we can associate with it a *reproducing kernel Hilbert space* (RKHS) $\mathcal{H}_k$, which is the completion of the linear span of all functions $\{k(\cdot, x), x \in \mathcal{X}\}$. The norm in $\mathcal{H}_k$, denoted by $\|\cdot\|_k$, is induced by the following inner product,

$$\langle f, g \rangle_k = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j k(x_i, x_j),$$

for $f(\cdot) \sum_{i=1}^{n} \alpha_i k(\cdot, x_i)$ and $g(\cdot) \sum_{j=1}^{m} \beta_j k(\cdot, x_j)$.

We note that our decision function $f(x)$ is from $\mathcal{H}_k$ equipped with norm $\|\cdot\|_k$. Thus since any function in $\mathcal{H}_k$ takes the form $\sum_{i=1}^{m} \alpha_i k(\cdot, x_i)$, it can be shown that the optimal decision function is given by

$$\sum_{i=1}^{n} \widehat{\alpha}_i A_i k(X, X_i) + \widehat{\beta}_0,$$

where $(\hat{\alpha}_1, \ldots, \hat{\alpha}_n)$ solves

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j A_i A_j k(X_i, X_j)$$

subject to $0 \le \alpha_i \le \kappa R_i / \pi_i$, $i = 1, \ldots, n$, and $\sum_{i=1}^{n} \alpha_i A_i = 0$. We note that if we choose $k(x, y) = \langle x, y \rangle$, then the obtained rule reduces to the previous linear rule.

3. THEORETICAL RESULTS

In this section, we establish consistency of the optimal ITR estimated using OWL. We further obtain a risk bound for the estimated ITR and show how the bound can be improved for certain specific, realistic situations.

3.1 Notation

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

$$\mathscr{R}(f)=E\left[\frac{R}{A\pi+(1-A)/2}I(A\neq \operatorname{sign}(f(X)))\right]$$

and the minimal risk (called Bayes risk in the learning literature) as $\mathcal{R}^* = \inf_f\{\mathcal{R}(f)|f: \mathcal{X} \to \mathbb{R}\}$. Thus, for the optimal ITR $\mathcal{D}^*(x) = \operatorname{sign}(f^*(x))$ (called the Bayes classifier in the learning literature), $\mathcal{R}^* = \mathcal{R}(f^*)$. In terms of the value function, we note that $\mathcal{V}(\mathcal{D}^*) - \mathcal{V}(\mathcal{D}) = \mathcal{R}(f) - \mathcal{R}(f^*)$.

In the OWL approach, we substitute 0–1 loss $I(A \neq \operatorname{sign}(f(X)))$ by a surrogate loss, $\varphi(Af(X))$, where $\varphi(t) = (1 - t)^+$. Thus we define the φ-risk

$$\mathscr{R}_\varphi(f)=E\left[\frac{R}{A\pi+(1-A)/2}\varphi(Af(X))\right],$$

and, similarly, the minimal φ–risk as $\mathscr{R}_\varphi^*=\inf_f\{\mathscr{R}_\varphi(f)|f:\mathscr{X}\to\mathbb{R}\}$.

Recall that the estimated optimal ITR is given by $\operatorname{sign}(\hat{f}_n(X))$, where

$$\widehat{f_n}=\underset{f\in\mathscr{H}_k}{\operatorname{argmin}}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i}\{1-A_if(X_i)\}^+ +\lambda_n\|f\|_k^2\right\}. \quad (3.1)$$

3.2 Fisher Consistency

We establish Fisher consistency of the decision function based on surrogate loss $\varphi(t)$. Specifically, the following result holds:

Proposition 3.1—For any measurable function f, *if* $\tilde{f}$ minimizes $R_\varphi(f)$, then $\mathcal{D}^*(x) = \operatorname{sign}(\tilde{f}(x))$. Proof. First, we note

$$\mathscr{D}^*(x)=\operatorname{sign}\{E[R|X=x,A=1]-E[R|X=x,A=-1]\}.$$

Next, for each $x \in \mathcal{X}$,

$$E\left(R\frac{\varphi(Af(X))}{A\pi+(1-A)/2}\middle|X=x\right)=E(R|A=1,X=x)(1-f(x))+E(R|A=-1,X=x)(1+f(x))$$
$$=((E(R|A=-1,X=x)-E(R|A=1,X=x))f(x)+E(R|A=-1,X=x)+E(R|A=1,X=x)).$$

Therefore, $\tilde{f}(x)$, which minimizes $\mathcal{R}_\phi(f)$, should be positive if $E(R|A = 1, X = x) > E(R|A = -1, X = x)$ and negative if $E(R|A = 1, X = x) < E(R|A = -1, X = x)$. That is, $\tilde{f}(x)$ has the same sign as $\mathcal{D}^*(x)$. The result holds.

The proposition is analogous to results for SVM, for example, Lin (2002). This theorem justifies the validity of using $\varphi(t)$ as the surrogate loss in OWL.

3.3 Excess Risk for $\mathcal{R}(f)$ and $\mathcal{R}_\varphi(f)$

The following result shows that for any decision function f, the excess risk of f under 0–1 loss is no larger than the excess risk of f under the hinge loss. Thus, the loss of the value function due to the ITR associated with f can be bounded by the excess risk under the hinge loss. The proof of the theorem can be found in the Appendix.

Theorem 3.2—For any measurable f: $\mathcal{X} \to \mathbb{R}$ and any probability distribution for (X, A, R),

$$\mathscr{R}(f) - \mathscr{R}^* \leq \mathscr{R}_\varphi(f) - \mathscr{R}_\varphi^*. \quad (3.2)$$

The proof follows the general arguments of Bartlett, Jordan & McAuliffe (2006), in which they bound the risk associated with 0–1 loss in terms of the risk from surrogate loss, utilizing a convexified variational transform of the surrogate loss. In our proof, we extend this concept to our setting by establishing the validity of a weighted version of such a transformation.

3.4 Consistency and Risk Bounds

The purpose of this section is to establish the consistency of $\hat{f}_n$, and, moreover, to derive the convergence rate of $\mathcal{R}(\hat{f}_n) - \mathcal{R}^*$.

First, the following theorem shows that the risk due to $\hat{f}_n$ does converge to $\mathcal{R}^*$, and, equivalently, the value of $\hat{f}_n$ converges to the optimal value function. Results on consistency of the SVM have been shown in current literature, for example, Zhang (2004). Here we apply the empirical process techniques to show that the proposed OWL estimator is consistent. The proof of the theorem is deferred to the Appendix.

Theorem 3.3—Assume that we choose a sequence $\lambda_n > 0$ such that $\lambda_n \to 0$ and $\lambda_n n \to \infty$. Then for all distributions P, we have that in probability,

$$\lim_{n\to\infty} \left\{ \mathscr{R}_\varphi(\widehat{f_n}) - \inf_{f \in \mathscr{H}_k} \mathscr{R}_\varphi(f) \right\} = 0,$$

where $\bar{\mathcal{H}}_k$ denotes the closure of $\mathcal{H}_k$. Thus, if f^* belongs to the closure of $\limsup_{n\to\infty} \mathcal{H}_k$, where $\mathcal{H}_k$ can potentially depend on n, we have $\lim_{n\to\infty} \mathscr{R}_\varphi(\widehat{f_n}) = \mathscr{R}_\varphi^*$ in probability. It then follows that $\lim_{n\to\infty} \mathcal{R}(\hat{f}_n) = \mathcal{R}^*$ in probability.

One special situation where f^* belongs to the limit space of $\mathcal{H}_k$ is when we choose $\mathcal{H}_k$ to be an RKHS with Gaussian kernel and let the kernel bandwidth decrease to zero as $n \to \infty$. This will be shown in Theorem 3.4 below.

We now wish to derive the convergence rate of $\mathcal{R}(\hat{f}_n) - \mathcal{R}^*$ under certain regularity conditions on the distribution P. Specifically, we need the following "geometric noise" assumption for P (Steinwart & Scovel 2007): Let

$$\eta(x) = \frac{E[R|X=x, A=1] - E[R|X=x, A=-1]}{E[R|X=x, A=1] + E[R|X=x, A=-1]} + 1/2, \quad (3.3)$$

then $2\eta(x) - 1$ is the decision boundary for the optimal ITR. We further define $\mathcal{X}^+ = \{x \in \mathcal{X} : 2\eta(x) - 1 > 0\}$, and $\mathcal{X}^- = \{x \in \mathcal{X} : 2\eta(x) - 1 < 0\}$. A distance function to the boundary between $\mathcal{X}^+$ and $\mathcal{X}^-$ is $\Delta(x) = \tilde{d}(x, \mathcal{X}^+)$ if $x \in \mathcal{X}^-$, $\Delta(x) = \tilde{d}(x, \mathcal{X}^-)$ if $x \in \mathcal{X}^+$ and $\Delta(x) = 0$ otherwise, where $\tilde{d}(x, \mathcal{O})$ denotes the distance of x to a set $\mathcal{O}$ with respect to the Euclidean norm. Then the distribution P is said to have geometric noise exponent $0 < q < \infty$, if there exists a constant $C > 0$ such that

$$E\left[\exp\left(-\frac{\Delta(X)^2}{t}\right)|2\eta(X)-1|\right] \le Ct^{qd/2}, t>0. \quad (3.4)$$

In some sense, this geometric noise exponent describes the behavior of the distribution in a neighborhood of the decision boundary. It is affected by how fast the density of the distance $\Delta(X)$ decays along the boundary. For example, assume the boundary is linear, in which case $\Delta(x) = |2\eta(x)-1|$. If for the density of $\Delta(X)$, defined as $f(u)$, we have $f(u) \sim u^p$ when u is close to 0, then we can show $q = (p + 2)/d$. Larger p corresponds to a faster decaying rate of the density, resulting in a larger q accordingly. Another example is distinctly separable data, i.e., when $|2\eta(x) - 1| > \delta > 0$, for some constant δ, and η is continuous, q can be arbitrarily large.

In addition to this specific assumption for P, we also restrict the choice of RKHS to the space associated with Gaussian Radial Basis Function (RBF) kernels, i.e.,

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

where $\sigma_n > 0$ is a parameter varying with n. The tuning parameter σ_n is related to approximation properties of Gaussian RBF kernels. When σ_n goes large, only observations in the small neighborhood contribute to the prediction, in which case we obtain a non-linear decision boundary or even non-parametric decision rule. If σ_n does not diverge, then points further away can contribute to the prediction, resulting in a nearly linear boundary. One advantage of using the Gaussian kernel is that we can determine the complexity of $\mathcal{H}_k$ in terms of capacity bounds with respect to the empirical L^2-norm, defined as

$$\|f-g\|_{L_2(P_n)}=\left(\frac{1}{n}\sum_{i=1}^{n}|f(X_i)-g(X_i)|^2\right)^{1/2}.$$

For any $\varepsilon > 0$, the covering number of functional class $\mathcal{F}$ with respect to $L_2(P_n)$, $N(\mathcal{F}, \varepsilon, L_2(P_n))$, is the smallest number of $L_2(P_n)$ ε-balls needed to cover $\mathcal{F}$, where an $L_2(P_n)$ ε-ball around a function $g \in \mathcal{F}$ is the set $\{f \in \mathcal{F} : \|f - g\|_{L_2(P_n)} < \varepsilon\}$.

Specifically, according to Theorem 2.1 in Steinwart & Scovel (2007), we have that for any $\varepsilon > 0$,

$$\sup_{P_n}\log N(B_{\mathcal{H}_k}, \varepsilon, L_2(P_n)) \le c_{\nu,\delta,m}\sigma_n^{(1-\nu/2)(1+\delta)d}\varepsilon^{-\nu}, \quad (3.5)$$

where $B_{\mathcal{H}_k}$ is the closed unit ball of $\mathcal{H}_k$, and ν and δ are any numbers satisfying $0 < \nu \le 2$ and $\delta > 0$.

Under the above conditions, we obtain the following theorem:

Theorem 3.4—Let P be a distribution of (X, A, R) satisfying condition (3.4) with noise exponent $q > 0$. Then for any $\delta > 0$, $0 < \nu < 2$, there exists a constant C (depending on ν, δ, d and π) such that for all $\tau \geq 1$ and $\sigma_n = \lambda_n^{-1/(q+1)d}$,

$$Pr^*(\mathscr{R}(\widehat{f_n}) \leq \mathscr{R}^* + \varepsilon) \geq 1 - e^{-\tau},$$

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

$$\varepsilon = C\left[\left(\frac{1}{\lambda_n}\right)^{\frac{2}{2+\nu}+\frac{(2-\nu)(1+\delta)}{(2+\nu)(1+q)}}\left(\frac{1}{n}\right)^{\frac{2}{2+\nu}} + \left(\frac{1}{\lambda_n}\right)^{\frac{q}{q+1}}\frac{\tau}{n} + \lambda_n^{\frac{q}{q+1}}\right].$$

The first two terms bound the stochastic error, which arises from the variability inherent in a finite sample size and which depends on the complexity of $\mathscr{H}_k$ in terms of covering numbers, while the third term controls the approximation error due to using $\mathscr{H}_k$, which depends on both σ_n and the noise behavior in the underlying distribution. We expect better approximation properties when the RKHS is more complex, but, conversely, we also expect larger stochastic variability. Using the above expression, an optimal choice of λ_n that balances bias and variance is given by

$$\lambda_n = n^{-\frac{2(1+q)}{(4+\nu)q+2+(2-\nu)(1+\delta)}},$$

so the optimal rate for the risk is

$$\mathscr{R}(\widehat{f_n}) - \mathscr{R}^* = O_p\left(n^{-\frac{2q}{(4+\nu)q+2+(2-\nu)(1+\delta)}}\right).$$

In particular, when data are well separated, q can be sufficiently large and we can let (δ, ν) be sufficiently small. Then the convergence rate almost achieves the rate $n^{-1/2}$. However, if the marginal distribution of $\mathscr{X}$ has continuous density along the boundary, it can be calculated that $q = 2/d$. In this case, the convergence rate is approximately $n^{-2/(d+2)}$. Clearly, the speed of convergence is slower with larger dimension of the prognostic variable space.

To prove Theorem 3.4, we note that according to Theorem 3.2, it suffices to prove the result for the excess φ risk. We also use the fact that

$$\mathscr{R}_\varphi(\widehat{f_n}) - \mathscr{R}_\varphi^* = \mathscr{R}_\varphi(\widehat{f_n}) - \inf_{\mathscr{H}_k}\mathscr{R}_\varphi(f) + \inf_{\mathscr{H}_k}\mathscr{R}_\varphi(f) - \mathscr{R}_\varphi^*.$$

We will then bound the first difference on the right-hand side using the empirical counterpart plus the stochastic variability due to the finite sample approximation. The latter can be controlled using large deviation results from empirical processes and some preliminary bound for $\|\hat{f}_n\|_k$. The second difference on the right-hand side will be bounded

by using the approximation property of the RHKS and the geometric noise assumption of the underlying distribution P. The proof is modified based on Vert & Vert (2006) and Steinwart & Scovel (2007), where the weights in the loss function are taken into consideration. The details are provided in the Appendix.

3.5 Improved Rate with Data Completely Separated

In this section, we show that a faster convergence rate can be obtained if the data are completely separated. We assume

(A1) $\forall x \in \mathcal{X}$, $|\eta(x) - 1/2| \geq \eta_0$, where $\eta(x)$ is defined in (3.3), and η is continuous.

(A2) $\forall x \in \mathcal{X}$, $\min(\eta(x), 1 - \eta(x)) \geq \eta_1$.

Assumption (A1) can be referred as a "low noise" condition equivalent to $|E(R|A = 1, X) - E(R|A = -1, X)| \geq \eta_0$. Thus, a jump of $\eta(x)$ at the level of 1/2 requires a gap between the rewards gained from treatment 1 and −1 on the same patient. This assumption is an adaptation of the noise condition used in classical SVM to obtain fast learning rates and it is essentially equivalent to one of the conditions in Blanchard et al. (2008).

Theorem 3.5—Assume that (*A1*) and (*A2*) are satisfied. For any $\nu \in (0, 1)$ and $q \in (0, \infty)$, let $\lambda_n = O(n^{-1/(\nu+1)})$ and $\sigma_n = \lambda_n^{-1/(q+1)d}$. Then

$$\mathscr{R}(\widehat{f_n}) - \mathscr{R}^* = O_p\left(n^{-\frac{1}{\nu+1}\frac{q}{q+1}}\right).$$

We can let q go to ∞ and ν go to zero, and this theorem shows that the convergence rate for $\mathcal{R}(\hat{f}_n) - \mathcal{R}(f^*)$ is almost n^{-1}, a much faster rate compared to what was given in Theorem 3.4. This result is similar to results for SVM described in Tsybakov (2004), Steinwart & Scovel (2007), and Blanchard et al. (2008).

To prove Theorem 3.5, we can rewrite the minimization problem in (3.1) as:

$$\min_{S \in \mathbb{R}^+} \left\{ \min_{f: \|f\|_k \leq S} \frac{1}{n} \sum_{i=1}^{n} \frac{R_i}{\pi_i} \{1 - A_i f(X_i)\}^+ + \lambda S^2 \right\}.$$

Thus the problem can be viewed in the model selection framework: a collection of models are balls in $\mathcal{H}_k$, and for each model, we solve the penalized empirical ϕ-risk minimization to obtain an estimator $\hat{f}_n$. We can utilize a result for model selection, presented in Theorem 4.3 of Blanchard et al. (2008), to choose the model which yields the minimal penalized empirical ϕ-risk among all the models. We need to verify the conditions required for the theorem based on the weighted hinge loss and the condition on the covering number of functional class $\mathcal{F}$ with respect to $L_2(P_n)$, i.e., condition (3.5). Proof details are provided in the Appendix.

4. SIMULATION STUDY

We have conducted extensive simulations to assess the small-sample performance of the proposed method. In these simulations, we generate 50-dimensional vectors of prognostic variables $X_1, \ldots, X_{50}$, consisting of independent $U[-1, 1]$ variates. The treatment A is generated from $\{-1, 1\}$ independently of X with $P(A = 1) = 1/2$. The response R is normally

distributed with mean $Q_0 = 1 + 2X_1 + X_2 + 0.5X_3 + T_0(X, A)$ and standard deviation 1, where $T_0(X, A)$ reflects the interaction between treatment and prognostic variables and is chosen to vary according to the following four different scenarios:

1. $T_0(X, A) = 0.442(1 - X_1 - X_2)A$.
2. $T_0(X, A)=(X_2-0.25X_1^2-1)A$.
3. $T_0(X, A)=(0.5-X_1^2-X_2^2)(X_1^2+X_2^2-0.3)A$.
4. $T_0(X, A)=(1-X_1^3+\exp(X_3^2+X_5)+0.6X_6-(X_7+X_8)^2)A$.

The decision boundaries in the first three scenarios are determined by X_1 and X_2. Scenario 1 corresponds to a linear decision boundary in truth, where the shape of the boundary in Scenario 2 is a parabola. The third is a ring example, where the patients on the ring are assigned to one treatment, and another if inside or outside the ring. The decision boundary in the fourth example is fairly nonlinear in covariates, depending on covariates other than X_1 and X_2. For each scenario, we estimate the optimal ITR by applying OWL. We use the Gaussian kernel in the weighted SVM algorithm. There are two tuning parameters: λ_n, the parameter for penalty, and σ_n, the inverse bandwidth of the kernel. Since λ_n plays a role in controlling the severity of the penalty on the functions and σ_n determines the complexity of the function class utilized, σ_n should be chosen adaptively from the data simultaneously with λ_n. To illustrate this, Figure 1 shows the contours of the value function for the first scenario with different combinations of (λ_n, σ_n) when $n = 30$. We can see that λ_n interacts with σ_n, with larger λ_n generally coupled with smaller σ_n for equivalent value function levels. In our simulations, we apply a 5-fold cross validation procedure, in which we search over a pre-specified finite set of (λ_n, σ_n) to select the pair maximizing the average of the estimated values from the validation data. In case of tied values for parameter pair choices, we first choose the set of pairs with smallest λ_n and then select the one with largest σ_n.

Additionally, comparison is made among the following four methods:

- the proposed OWL using Gaussian kernel (OWL-Gaussian)
- the proposed OWL using linear kernel (OWL-Linear)
- the l_1 penalized least squares method (l_1-PLS) developed by Qian & Murphy (2011), which approximates $E(R|X, A)$ using the basis function set $(1, X, A, XA)$ and applies the LASSO method for variable selection, and
- ordinary least squares method (OLS), which estimates the conditional mean response using the same basis function set as in 3 but without variable selection.

We consider the OWL with linear kernel (method 2) mainly to assess the impact of different kernels in the weighted SVM algorithm. In this case, there is only one tuning parameter, λ_n, which can be chosen to maximize the value function in a cross-validation procedure. The selection of the tuning parameters in the l_1-PLS approach follows similarly. The last two approaches estimate the optimal ITR using the sign of the difference between the predicted $E(R|X, A = 1)$ and the predicted $E(R|X, A = -1)$. In the comparisons, the performances of the four methods are assessed by two criteria: the first criterion is to evaluate the value function using the estimated optimal ITR when applying to an independent and large validation data; the second criterion is to evaluate the misclassification rates of the estimated optimal ITR from the true optimal ITR using the validation data. Specifically, a validation set with 10000 observations is simulated to assess the performance of the approaches. The estimated value function using any ITR $\mathcal{D}$ is given by $\mathbb{P}_n^*[I(A=\mathcal{D}(X))R/P(A)]/\mathbb{P}_n^*[I(A=\mathcal{D}(X))/P(A)]$ (Murphy et al. 2001), where $\mathbb{P}_n^*$ denotes the

empirical average using the validation data and $P(A)$ is the probability of being assigned treatment A.

For each scenario, we vary sample sizes for training datasets from 30 to 100, 200, 400 and 800, and repeat the simulation 1,000 times. The simulation results are presented in Figures 2 and 3, where we report the mean square errors (MSE) of both value functions and misclassification rates. Simulations show there are no large differences in the performance if we replace the Gaussian kernel with the linear kernel in the OWL. However, there are examples presenting advantages of the Gaussian kernel, which suggests that under certain circumstances, it is useful to have a flexible nonparametric estimation procedure to identify the optimal ITR for the underlying nonparametric structures. As demonstrated in Figure 2 and Figure 3, the OWL with either Gaussian kernel or linear kernel has better performance, especially for small samples, than the other two methods, from the points of view of producing larger value functions, smaller misclassification rates, and lower variability of the value function estimates. Specifically, when the approximation models used in the l_1-PLS and OLS are correct in the first scenario, the competing methods perform well with large sample size; however, the OWL still provides satisfactory results even if we use a Gaussian kernel. When the optimal ITR is nonlinear in X in the other scenarios, the OWL tends to give higher values and smaller misclassification rates. OLS generally fails unless the sample size is large enough since it encounters severe bias for small sample sizes. This is due to the fact that without variable selection for OLS, there is insufficient data to fit an accurate model with all 50 variables included. We also note that l_1-PLS has comparatively larger MSE, resulted from high variance of the method, which may be explained by the conflicting goals of maximizing the value function and minimizing the prediction error (Qian & Murphy 2011). Note that a richer class of basis functions can be used for fitting the regression models. We have tried a polynomial basis and a wavelet basis to see if they could improve the performance. However, as a larger set of basis functions enters the model, we need to take into account higher dimensional interactions which do not necessarily yield better results. Also, we noted that higher variability is introduced with a richer basis for the approximation space (results not shown).

Additional simulations are performed by generating binary outcomes from a logit model. It turns out the OWL procedures outperform a traditional logistic regression procedure (results not shown). Finally, using empirical results, we also verify that the cross validation procedure can indeed identify the optimal pairs (λ_n, σ_n) with the order desired by the theoretical results, i.e., $\sigma_n = \lambda_n^{-(q+1)d}$. The numerical results indicate that $\log_2 \sigma_n$ is linear in $\log_2 \lambda_n$ and the ratio between the slopes is close to the reciprocal ratio between the dimensions of the covariate spaces.

5. DATA ANALYSIS

We apply the proposed method to analyze real data from the Nefazodone-CBASP clinical trial (Keller et al. 2000). The study randomized 681 outpatients with non-psychotic chronic major depressive disorder (MDD), in a 1:1:1 ratio to either Nafazodone, Cognitive Behavioral-Analysis System of Psychotherapy (CBASP) or the combination of Nefazodone and CBASP. The score on the 24-item Hamilton Rating Scale for Depression (HRSD) was the primary outcome, where higher scores indicate more severe depression. After excluding some patients with missing observations, we use a subset with 647 patients for analysis. Among them, 216, 220 and 211 patients were assigned to Nafazodone, CBASP and the combined treatment group respectively. Overall comparisons using t-tests show that the combination treatment had significant advantages over the other treatments with respect to HRSD scores obtained at end of the trial, while there are no significant differences between the nefazodone group and the psychotherapy group.

To estimate the optimal ITR, we perform pairwise comparisons between all combinations of two treatment arms, and, for each two-arm comparison, we apply the OWL approach. We only present the results from the Gaussian kernel, since the analysis shows a similarity with that of the linear kernel. Rewards used in the analyses are reversed HRSD scores and the prognostic variables X consist of 50 pretreatment variables. The results based on OWL are compared to results obtained using the l_1-PLS and OLS methods which use (1, X, A, XA) in their regression models. For comparison between methods, we calculate the value function from a cross-validation type analysis. Specifically, the data is partitioned into 5 roughly equal-sized parts. We perform the analysis on 4 parts of the data, and obtain the estimated optimal ITRs using different methods. We then compute the estimated value functions using the remaining fifth part. The value functions calculated this way should better represent expected value functions for future subjects, as compared to calculating value functions based on the training data. The averages of the cross-validation value functions from the three methods are presented in Table 1.

From the table, we observe that OLS produces smaller value functions (corresponding to larger HRSD in the table) than the other two methods, possibly because of the high dimensional prognostic variable space. OWL performs similarly to l_1-PLS, but gives a 5% larger value function than l_1-PLS when comparing the Combination arm to the Nefazodone arm. In fact, when comparing combination treatment with nefazodone only, OWL recommends the combination treatment to all the patients in the validation data in each round of the cross validation procedure; the OLS assigns the combination treatment to around 70% of the patients in each validation subset; while the l_1-PLS recommends the combination to all the patients in three out of five validation sets, and 7% and 28% to the patients for the other two, indicating a very large variability. If we need to select treatment between combination and psychotherapy alone, the OWL approach recommends the combination treatment for all patients in the validation process. In contrast, the l_1-PLS chooses psychotherapy for 10 out of 86 patients in one round of validation, and recommends the combination for all patients in the other rounds. The percentages of patients who are recommended the combination treatment range from 66% to 85% across the five validation data sets when applying OLS. When the two single treatments are studied, there are only negligible differences in the estimated value functions from the three methods and the selection results also indicate an insignificant difference between them. Thus OWL not only yields ITRs with the best clinical outcomes, but the ITRs also have lowest variability compared to the other methods.

6. DISCUSSION

The proposed OWL procedure appears to be more effective, across a broad range of possible forms of the interaction between prognostic variables and treatment, compared to previous methods. A two-stage procedure is likely to overfit the regression model, and thus cause troubles for value function approximation. The OWL provides a nonparametric approach which sidesteps the inversion of the predicted model required in other methods and benefits from directly maximizing the value function. The convergence rates for the OWL, aiming to identify the best ITR, nearly reach the optimal for the nonparametric SVM with the same type of assumptions on the separations. The rates, however, are not directly comparable to Qian & Murphy's (2011), because we allow for complex multivariate interactions and formulate the problem in a nonparametric framework. The proposed estimator will lead to consistency and fast rate results, but not necessarily the most efficient approach. In some cases when we have knowledge of the specific parametric form, a likelihood based method may be more efficient and aid in the improvement of the estimation. Other possible surrogate loss functions, for example, the negative log-likelihood for logistic regression, can also be useful for finding the desired optimal individualized treatment rules.

Several improvements and extensions are important to consider. An important extension we are currently pursuing is to right-censored clinical outcomes. Another extension involves alleviating potential challenges arising from high dimensional prognostic variables. Recall that the proposed OWL is based on a weighted SVM which minimizes the weighted hinge loss function subject to an l_2 penalty. If the dimension of the covariate space is sufficiently large, not all the variables would be essential for optimal ITR construction. By eliminating the unimportant variables from the rule, we could simplify interpretations and reduce health care costs by only requiring collection of a small number of significant prognostic variables. For standard SVM, the l_1 penalty has been shown to be effective in selecting relevant variables via shrinking small coefficients to zero (Bradley & Mangasarian 1998; Zhu et al. 2003). It outperforms the l_2 penalty when there are many noisy variables and sparse models are preferred. Other forms of penalty have been proposed such as the F_∞ norm (Zou & Yuan 2008) and the adaptive l_q penalty (Liu et al. 2007). In the future, we will examine use of these sparse penalties in the OWL method.

In this paper, we only considered binary options for treatment. When there are more than two treatment classes, although we could do a series of pairwise comparisons as done in Section 5 above, this approach may not be optimal in terms of identifying the best rule considering all treatments simultaneously. It would thus be worthwhile to extend the OWL approach to settings involving three or more treatments. The case of multicategory SVM has been studied recently (Lee, Lin & Wahba 2004; Wang & Shen 2006), and a similar generalization may be possible for finding ITRs involving three or more treatments. Another setting to consider is optimal ITR discovery for continuous treatments such as, for example, a continuous range of dose levels. In this situation, we could potentially utilize ideas underlying support vector regression (Vapnik 1995), where the goal is to find a function that has at most ε deviation from the response. Using a similar rationale as the proposed OWL, we could develop corresponding procedures for continuous treatment spaces through weighing each subject by his/her clinical outcome.

Obtaining inference for individualized treatment regimens is also important and challenging. Due to high heterogeneities among individuals, there may be large variations in the estimated treatment rules across different training sets. Laber & Murphy (2011) construct an adaptive confidence interval for the test error under the non-regular framework. Confidence intervals for value functions help us determine whether essential differences exist among different decision rules. Thus an important future research topic is to derive the limiting distribution of $\mathcal{V}(\hat{\mathcal{D}}_n) - \mathcal{V}(\mathcal{D}^*)$ and to derive corresponding sample size formulas to aid in design of personalized medicine clinical trials.

In some complex diseases, dynamic treatment regimes may be more useful than the single-decision treatment rules studied in this paper. Dynamic treatment regimes are customized sequential decision rules for individual patients which can adapt over time to an evolving illness. Recently, this research area has been of great interest in long term management of chronic disease. See, for example, Murphy et al. (2001), Thall, Sung & Estey (2002), Murphy (2003), Robins (2004), Moodie, Richardson & Stephens (2007), Zhao, Zeng, Socinski & Kosorok (2011). Extension of the proposed OWL approach to the dynamic setting would be of great interest.

Acknowledgments

The first, second and fourth authors were partially funded by NCI Grant P01 CA142538.

References

Bartlett PL, Bousquet O, Mendelson S. Local Rademacher Complexities. The Annals of Statistics. 2005; 33(4):1497–1537.

Bartlett PL, Jordan MI, McAuliffe JD. Convexity, Classification, and Risk Bounds. J of American Statistical Association. 2006; 101(473):138–156.

Blanchard G, Bousquet O, Massart P. Statistical Performance of Support Vector Machines. Annals of Statistics. 2008; 36:489–531.

Bradley, PS.; Mangasarian, OL. Feature Selection via Concave Minimization and Support Vector Machines. Proc. 15th International Conf. on Machine Learning; San Francisco, CA, USA: Morgan Kaufmann Publishers Inc; 1998.

Buzdar AU. Role of Biologic Therapy and Chemotherapy in Hormone Receptor and HER2-Positive Breast Cancer. Annals of Oncology. 2009; 20:993–999. [PubMed: 19150946]

Cai T, Tian L, Uno H, Solomon SD. Calibrating parametric subject-specific risk estimation. Biometrika. 2010; 97(2):389–404. [PubMed: 23049123]

Cortes C, Vapnik V. Support-Vector Networks. Machine Learning. 1995:273–297.

Crits-Christoph P, Siqueland L, Blaine J, Frank A, Luborsky L, Onken LS, Muenz LR, Thase ME, Weiss RD, Gastfriend DR, Woody GE, Barber JP, Butler SF, Daley D, Salloum I, Bishop S, Najavits LM, Lis J, Mercer D, Griffin ML, Moras K, Beck AT. Psychosocial Treatments for Cocaine Dependence. Arch Gen Psychiatry. 1999; 56:493–502. [PubMed: 10359461]

Eagle KA, Lim MJ, Dabbous OH, Pieper KS, Goldberg RJ, de Werf FV, Goodman SG, Granger CB, Steg PG, Joel M, Gore M, Budaj A, Avezum A, Flather MD, Fox KAA. GRACE Investigators. A Validated Prediction Model for All Forms of Acute Coronary Syndrome: Estimating the Risk of 6-Month Postdischarge Death in an International Registry. J Am Med Assoc. 2004; 291:2727–33.

Flume PA, OSullivan BP, Goss CH, Peter J, Mogayzel J, Willey-Courand DB, Bujan J, Finder J, Lester M, Quittell L, Rosenblatt R, Vender RL, Hazle L, Sabadosa K, Marshall B. Cystic Fibrosis Pulmonary Guidelines: Chronic Medications for Maintenance of Lung Health. Am J Respir Crit Care Med. 2007; 176(1):957–969. [PubMed: 17761616]

Grünwald V, Hidalgo M. Developing Inhibitors of the Epidermal Growth Factor Receptor for Cancer Treatment. J Natl Cancer Inst. 2003; 95(12):851–867. [PubMed: 12813169]

Hastie, T.; Tibshirani, R.; Friedman, JH. The Elements of Statistical Learning. 2. New York: Springer-Verlag New York, Inc; 2009.

Insel TR. Translating scientific opportunity into public health impact: a strategic plan for research on mental illness. Archives of General Psychiatry. 2009; 66(2):128–133. [PubMed: 19188534]

Keller MB, Mccullough JP, Klein DN, Arnow B, Dunner DL, Gelenberg AJ, Markowitz JC, Nemeroff CB, Russell JM, Thase ME, Trivedi MH, Zajecka J. A Comparison of Nefazodone, The Cognitive Behavioral-Analysis System of Psychotherapy, and Their Combination for the Treatment of Chronic Depression. The New England Journal of Medicine. 2000; 342(20):1462–70. [PubMed: 10816183]

Laber EB, Murphy SA. Adaptive Confidence Intervals for the Test Error in Classification. To apper in Journal of the American Statistical Association. 2011

Lee Y, Lin Y, Wahba G. Multicategory Support Vector Machines, theory, and application to the classification of microarray data and satellite radiance data. Journal of the American Statistical Association. 2004; 99:67–81.

Lin Y. Support vector machines and the Bayes rule in classification. Data Mining and Knowledge Discovery. 2002; 6:259–275.

Liu Y, Helen Zhang H, Park C, Ahn J. Support vector machines with adaptive L_q penalty. Comput Stat Data Anal. 2007; 51(12):6380–94.

Lugosi G, Vayatis N. On the Bayes-risk consistency of regularized boosting methods. The Annals of Statistics. 2004; 32:30–55.

Marlowe DB, Festinger DS, Dugosh KL, Lee PA, Benasutti KM. Adapting Judicial Supervision to the Risk Level of Drug Offenders: Discharge and 6-month Outcomes from a Prospective Matching Study. Drug and Alcohol Dependence. 2007; 88(Suppl 2 2):S4–S13. [PubMed: 17071020]

Moodie EEM, Platt RW, Kramer MS. Estimating Response-Maximized Decision Rules With Applications to Breastfeeding. Journal of the American Statistical Association. 2009; 104(485): 155–165.

Moodie EEM, Richardson TS, Stephens DA. Demystifying Optimal Dynamic Treatment Regimes. Biometrics. 2007; 63(2):447–455. [PubMed: 17688497]

Murphy SA. Optimal Dynamic Treatment Regimes. Journal of the Royal Statistical Society, Series B. 2003; 65:331–366.

Murphy SA, van der Laan MJ, Robins JM, CPPRG. Marginal Mean Models for Dynamic Regimes. Journal of the American Statistical Association. 2001; 96:1410–23. [PubMed: 20019887]

Piper WE, Boroto DR, Joyce AS, McCallum M, Azim HFA. Pattern of alliance and outcome in short-term individual psychotherapy. Psychotherapy. 1995; 32:639–647.

Qian M, Murphy SA. Performance Guarantees for Individualized Treatment Rules. To appear in the Annals of Statistics. 2011

Robins, JM. Optimal Structural Nested Models for Optimal Sequential Decisions. Proceedings of the Second Seattle Symposium on Biostatistics; Springer; 2004. p. 189-326.

Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large B-cell lymphoma. New England J of Medicine. 2002:1937–47.

Sargent DJ, Conley BA, Allegra C, Collette L. Clinical Trial Designs for Predictive Marker Validation in Cancer Treatment Trials. Journal of Clinical Oncology. 2005; 32:2020–27. [PubMed: 15774793]

Steinwart I. Consistency of Support Vector Machines and Other Regularized Kernel Classifiers. IEEE Transactions on Information Theory. 2005; 51:128–142.

Steinwart I, Scovel C. Fast Rates for Support Vector Machines using Gaussian Kernels. The Annals of Statistics. 2007; 35:575–607.

Thall PF, Sung H-G, Estey EH. Selecting Therapeutic Strategies Based on Efficacy and Death in Multicourse Clinical Trials. Journal of the American Statistical Association. 2002; 97:29–39.

Tsybakov AB. Optimal Aggregation of Classifiers in Statistical Learning. Annals of Statistics. 2004; 32:135–166.

van't Veer LJ, Bernards R. Enabling Personalized Cancer Medicine through Analysis of Gene-Expression Patterns. Nature. 2008; 452:564–570. [PubMed: 18385730]

Vapnik, VN. The nature of statistical learning theory. New York: Springer-Verlag New York, Inc; 1995.

Vert R, Vert J-P. Consistency and Convergence Rates of One-Class SVMs and Related Algorithms. Journal of Machine Learning Research. 2006; 7:817–854.

Wang L, Shen X. Multi-category Support vector machines, feature selection, and solution path. Statistica Sinica. Statistica Sinica. 2006; 16:617–633.

Zhang T. Statistical behavior and consistency of classification methods based on convex risk minimization. Annals of Statistics. 2004; 32(1):56–85.

Zhao Y, Zeng D, Socinski MA, Kosorok MR. Reinforcement Learning Strategies for Clinical Trials in Nonsmall Cell Lung Cancer. Biometrics. 2011; 67:1422–1433. [PubMed: 21385164]

Zhu J, Rosset S, Hastie T, Tibshirani R. 1-norm Support Vector Machines. Neural Information Processing Systems. 2003:16.

Zou H, Yuan M. The F_∞-norm Support Vector Machine. Statistica Sinica. 2008; 18:379–398.

APPENDIX A. PROOFS

Proof of Theorem 3.2

We consider the case where rewards are discrete. Arguments for the continuous rewards setting follow similarly. Let $\eta_t(x) = p(A = 1|R = r, X = x)$ and $q_t(x) = rp(R = r|X = x)$. We can write

$$\begin{aligned}\mathscr{R}(f)&=E\left[\sum_r rp(R=r|X)E\left(\frac{I(A\neq\operatorname{sign}(f(X)))}{A\pi+(1-A)/2}\middle|R=r,X\right)\right]\\&=E\left[\sum_r q_r(X)\left(\frac{\eta_r(X)}{\pi}I(\operatorname{sign}(f(X))\neq 1)+\frac{1-\eta_r(X)}{1-\pi}I(\operatorname{sign}(f(X))\neq -1)\right)\right]\\&=E[c_0(X)(\eta(X)I(\operatorname{sign}(f(X))\neq 1)+(1-\eta(X))I(\operatorname{sign}(f(X))\neq -1))],\end{aligned}\tag{A.1}$$

where $c_0(x) = \Sigma_r\, q_r(x)[\eta_r(x)/\pi + (1 - \eta_r(x))/(1 - \pi)]$, and $\eta(x)$, defined previously in (3.3), is equal to $\Sigma_r\, q_r(x)\ \eta_r(x)/\pi c_0(x)$. Similarly,

$$\mathscr{R}_\varphi(f)=E[c_0(X)(\eta(X)\varphi(f(X))+(1-\eta(X))\varphi(-f(X)))].$$

We define $C(\eta, \alpha) = \eta\varphi(\alpha) + (1 - \eta)\varphi(-\alpha)$. Then the optimal φ-risk satisfies

$$\mathscr{R}_\varphi^*=E\left[c_0(X)\inf_{\alpha\in\mathbb{R}}C(\eta(X),\alpha)\right]$$

and

$$\mathscr{R}_\varphi-\mathscr{R}_\varphi^*=E\left[c_0(X)\left(C(\eta(X),f(X))-\inf_{\alpha\in\mathbb{R}}C(\eta(X),\alpha)\right)\right].$$

By a result in Bartlett et al. (2006) for a convexified transform of hinge loss, we have

$$2\eta-1=\inf_{\alpha:\alpha(2\eta-1)\leq 0}C(\eta,\alpha)-\inf_{\alpha\in\mathbb{R}}C(\eta,\alpha).\tag{A.2}$$

Thus, according to (A.1) and (A.2), we have

$$\begin{aligned}\mathscr{R}(f)-\mathscr{R}^*&\leq E(I(\operatorname{sign}(f(X))\neq\operatorname{sign}[c_0(X)(\eta(X)-1/2)])|c_0(X)(2\eta(X)-1)|)\\&=E\left[c_0(X)I(\operatorname{sign}(f(X))\neq\operatorname{sign}[c_0(X)(\eta(X)-1/2)])\left(\inf_{\alpha:\alpha(2\eta(X)-1)\leq 0}C(\eta(X),\alpha)-\inf_{\alpha\in\mathbb{R}}C(\eta(X),\alpha)\right)\right]\\&\leq E\left[c_0(X)\left(C(\eta(X),f(X))-\inf_{\alpha\in\mathbb{R}}C(\eta(X),\alpha)\right)\right]\\&=\mathscr{R}_\varphi(f)-\mathscr{R}_\varphi^*.\end{aligned}$$

The last inequality holds because we always have $C(\eta(x), f(x)) \geq \inf_{\alpha\in\mathbb{R}} C(\eta(x), \alpha)$ on the set where $\operatorname{sign}(f(x)) = \operatorname{sign}[c_0(x)(\eta(x) - 1/2)]$ and $C(\eta(x), f(x)) \geq \inf_{\alpha:\alpha(2\eta(x)-1)\leq 0} C(\eta(x), \alpha)$ when $\operatorname{sign}(f(x)) \neq \operatorname{sign}[c_0(x)(\eta(x) - 1/2)]$.

Proof of Theorem 3.3

Define $L_\varphi(f) = R\varphi(Af)/(A\pi + (1 - A)/2)$. By the definition of $\hat{f}_n$, we have for any $f\in\mathcal{H}_k$,

$$\mathbb{P}_n\left(L_\varphi(\widehat{f_n})\right)\leq\mathbb{P}_n\left(L_\varphi(\widehat{f_n})+\lambda_n\|\widehat{f_n}\|^2\right)\leq\mathbb{P}_n\left(L_\varphi(f)+\lambda_n\|f\|^2\right),$$

where $\mathbb{P}_n$ denotes the empirical measure of the observed data. Thus $\limsup_n \mathbb{P}_n(L_\varphi(\hat{f}_n)) \le \mathbb{P}(L_\varphi(f))$. It leads to $\limsup_n \mathbb{P}_n(L_\varphi(\hat{f}_n)) \le \inf_{f\in\mathcal{H}_k} \mathbb{P}(L_\varphi(f))$. Theorem 3.3 holds if we can show $\mathbb{P}_n(L_\varphi(\hat{f}_n)) - \mathbb{P}(L_\varphi(\hat{f}_n)) \to 0$ in probability.

To this end, we first obtain a bound for $\left\|\widehat{f_n}\right\|_k^2$. Since $\mathbb{P}_n(L_\varphi(\widehat{f_n}))+\lambda_n\left\|\widehat{f_n}\right\|^2 \le \mathbb{P}_n(L_\varphi(f))+\lambda_n\|f\|_k^2$ for any $f\in\mathcal{H}_k$, we can select $f=0$ to obtain

$$\left\|\widehat{f_n}\right\|_k^2 \le \frac{1}{\lambda_n}\frac{1}{n}\sum\frac{R_i}{\pi_i}\varphi(0) \le \frac{2}{\lambda_n}\frac{E(R)}{\min\{\pi, 1-\pi\}}.$$

Let $M = 2E(R)/\min\{\pi, 1-\pi\}$ so that the $\mathcal{H}_k$ norm of $\sqrt{\lambda_n}\widehat{f_n}(X)$ is bounded by $\sqrt{M}$. Note that the class $\{\sqrt{\lambda_n}f: \left\|\sqrt{\lambda_n}f\right\|_k \le \sqrt{M}\}$ is contained in a Donsker class. Thus, $\left\{\sqrt{\lambda_n}L_\varphi(f), \left\|\sqrt{\lambda_n}f\right\|_k \le \sqrt{M}\right\}$ is also P-Donsker because $(1 - Af(X))^+$ is Lipschitz continuous with respect to f. Therefore,

$$\sqrt{n}(\mathbb{P}_n-\mathbb{P})L_\varphi(\widehat{f_n}) = \sqrt{\lambda_n^{-1}}\sqrt{n}(\mathbb{P}_n-\mathbb{P})\left[\frac{R}{A\pi+(1-A)/2}\left(\sqrt{\lambda_n}-A\sqrt{\lambda_n}\widehat{f_n}(X)\right)^+\right] = O_p\left(\sqrt{\lambda_n^{-1}}\right).$$

Consequently, from $n\lambda_n \to \infty$, $\mathbb{P}_n(L_\varphi(\hat{f}_n)) - \mathbb{P}(L_\varphi(\hat{f}_n)) \to 0$ in probability.

Proof of Theorem 3.4

First, we have

$$\begin{aligned}\mathcal{R}_\varphi(\widehat{f_n})-\mathcal{R}_\varphi^* &\le \lambda_n\left\|\widehat{f_n}\right\|_k^2+\mathcal{R}_\varphi(\widehat{f_n})-\mathcal{R}_\varphi^* \\ &\le \left[\lambda_n\left\|\widehat{f_n}\right\|_k^2+\mathcal{R}_\varphi(\widehat{f_n})-\inf_{f\in\mathcal{H}_k}(\lambda_n\|f\|_k^2+\mathcal{R}_\varphi(f))\right]+\left[\inf_{f\in\mathcal{H}_k}(\lambda_n\|f\|_k^2+\mathcal{R}_\varphi(f)-\mathcal{R}_\varphi^*)\right]. \quad (A.3)\end{aligned}$$

We will bound each term on the right-hand-side separately in the following arguments.

For the second term on the right-hand-side of (A.3), we use Theorem 2.7 in Steinwart & Scovel (2007) to conclude that

$$\inf_{f\in\mathcal{H}_k}(\lambda_n\|f\|_k^2+\mathcal{R}_\varphi(f)-\mathcal{R}_\varphi^*) = O\left(\lambda_n^{q/(q+1)}\right), \quad (A.4)$$

when we set $\sigma_n = \lambda_n^{-1/(q+1)d}$.

Now we proceed to obtain a bound for the first term on the right-hand-side of (A.3). To do this, we need the useful Theorem 5.6 of Steinwart & Scovel (2007) presented below:

Theorem 5.6, Steinwart & Scovel (2007). *Let $\mathcal{F}$ be a convex set of bounded measurable functions from Z to $\mathbb{R}$ and let $L : \mathcal{F}\times Z \to [0,\infty)$ be a convex and line-continuous loss function. For a probability measure P on Z we define*

$$\mathscr{G} := \{L \circ f - L \circ f_{P,\mathscr{F}} : f \in \mathscr{F}\}.$$

Suppose that there are constants $c \geq 0$, $0 < \alpha < 1$, $\delta \geq 0$ and $B > 0$ with $E_P g^2 \leq c(E_P g)^\alpha + \delta$ and $\|g\|_\infty \leq B$ for all $g \in \mathscr{G}$. Furthermore, assume that $\mathscr{G}$ is separable with respect to $\|\cdot\|_\infty$ and that there are constants $a \geq 1$ and $0 < p < 2$ with

$$\sup_{T \in Z^n} \log N(B^{-1}\mathscr{G}, \varepsilon, L_2(T)) \leq a\varepsilon^{-p}$$

for all $\varepsilon > 0$. Then there exists a constant $c_p > 0$ depending only on p such that for all $n \geq 1$ and all $\tau \geq 1$ we have

$$Pr^*(T \in Z^n : \mathscr{R}_{L,P}(f_{T,\mathscr{F}}) > \mathscr{R}_{L,P}(f_{P,\mathscr{F}}) + c_p \varepsilon(n, a, B, c, \delta, \tau)) \leq e^{-\tau},$$

where

$$\varepsilon(n, a, B, c, \delta, x) := B^{2p/(4-2\alpha+\alpha p)} c^{(2-p)/(4-2\alpha+\alpha p)} \left(\frac{a}{n}\right)^{2/(4-2\alpha+\alpha p)} + B^{p/2}\delta^{(2-p)/4}\left(\frac{a}{n}\right)^{1/2} + B\left(\frac{a}{n}\right)^{2/(2+p)} + \sqrt{\frac{\delta x}{n}} + \left(\frac{c\tau}{n}\right)^{1/(2-\alpha)} + \frac{B\tau}{n}.$$

In their paper, $f_{P,\mathscr{F}} \in \mathscr{F}$ is a minimizer of $\mathscr{R}_{L,P}(f) = E(L(f, z))$, and $f_{T,\mathscr{F}}$ is similarly defined when T is an empirical measure. To use this theorem, we define $\mathscr{F}$, Z, T, $\mathscr{G}$, $f_{T,\mathscr{F}}$ and $f_{P,\mathscr{F}}$ according to our setting. It suffices to consider the subspace of $\mathscr{H}_k$, denoted by $B_{\mathscr{H}_k}\left(\sqrt{M/\lambda_n}\right)$, as the ball of $\mathscr{H}_k$ of radius $\sqrt{M/\lambda_n}$. Specifically, we let $\mathscr{F}$ be $B_{\mathscr{H}_k}\left(\sqrt{M/\lambda_n}\right)$ and Z be $\mathscr{X}$. The loss function we consider here is $L_\varphi(f) + \lambda_n \|f\|_k^2$ and $\mathscr{G}$ is the function class

$$\mathscr{G}_{\varphi,\lambda_n} = \left\{L_\varphi(f) + \lambda_n \|f\|_k^2 - L_\varphi(f^*_{\varphi,\lambda_n}) - \lambda_n \|f^*_{\varphi,\lambda_n}\|_k^2 : f \in B_{\mathscr{H}_k}\left(\sqrt{M/\lambda_n}\right)\right\},$$

where $f^*_{\varphi,\lambda_n} = \operatorname{argmin}_{f \in B_{\mathscr{H}_k}(\sqrt{M/\lambda_n})} (\lambda_n \|f\|_k^2 + \mathscr{R}_\varphi(f))$. $f_{P,\mathscr{F}}$ and $f_{T,\mathscr{F}}$ correspond to f^*_{φ,λ_n} and $\hat{f}_n$, respectively. Therefore, to apply this theorem, we will show that there are constants $c \geq 0$ and $B > 0$, which can possibly depend on n, such that $E(g^2) \leq cE(g)$ and $\|g\|_\infty \leq B$, $\forall g \in \mathscr{G}_{\phi,\lambda}$. Moreover, there are constants $\tilde{c}$ and $0 < \nu < 2$ with

$$\sup_{P_n} \log N(B^{-1}\mathscr{G}_{\varphi,\lambda_n}, \varepsilon, L_2(P_n)) \leq \tilde{c}\varepsilon^{-\nu},$$

for all $\varepsilon > 0$.

Let C_L denote $\sup\{R/\min(\pi, 1 - \pi)\}$, which is finite provided that R is bounded. Since the weighted hinge loss is Lipschitz continuous with respect to f, with Lipschitz constant C_L, and since $\|f\|_\infty \leq \|f\|_k$ given that $k(x, x) \leq 1$, for any $g \in \mathscr{G}_{\phi,\lambda_n}$, we have

$$\begin{aligned} |g| &\le |L_\varphi(f)-L_\varphi(f^*_{\varphi,\lambda_n})|+\lambda_n\left|\|f\|_k^2-\left\|f^*_{\varphi,\lambda_n}\right\|_k^2\right| \\ &\le C_L|f(x)-f^*_{\varphi,\lambda_n}(x)|+M \\ &\le 2C_L\sqrt{M}\lambda_n^{-1/2}+M. \end{aligned} \quad \text{(A.5)}$$

Therefore, we can set $B=2C_L\sqrt{M}\lambda_n^{-1/2}+M$.

For any $g\in \mathcal{G}_{\phi,\lambda_n}$, we have

$$\begin{aligned} g(f) &\le |L_\varphi(f)-L_\varphi(f^*_{\varphi,\lambda_n})|+\lambda_n\left|\|f\|_k^2-\left\|f^*_{\varphi,\lambda_n}\right\|_k^2\right| \\ &\le C_L|f-f^*_{\varphi,\lambda_n}|+\lambda_n\left\|f-f^*_{\varphi,\lambda_n}\right\|_k\left\|f+f^*_{\varphi,\lambda_n}\right\|_k \\ &=\left(C_L+2\sqrt{M\lambda_n}\right)\left\|f-f^*_{\varphi,\lambda_n}\right\|_k. \end{aligned}$$

Squaring both sides and taking expectations yields

$$E(g^2)\le\left(C_L+2\sqrt{M\lambda_n}\right)^2\left\|f-f^*_{\varphi,\lambda_n}\right\|_k^2. \quad \text{(A.6)}$$

On the other hand, from the convexity of L_φ, we have

$$\begin{aligned} \tfrac{1}{2}(L_\varphi(f)+\lambda_n\|f\|_k^2+L_\varphi(f^*_{\varphi,\lambda_n})+\lambda_n\left\|f^*_{\varphi,\lambda_n}\right\|_k^2) &\ge L_\varphi\left(\tfrac{f+f^*_{\varphi,\lambda_n}}{2}\right)+\lambda_n\tfrac{\left\|f\right\|_k^2+\left\|f^*_{\varphi,\lambda_n}\right\|_k^2}{2} \\ &=L_\varphi\left(\tfrac{f+f^*_{\varphi,\lambda_n}}{2}\right)+\lambda_n\|\tfrac{f+f^*_{\varphi,\lambda_n}}{2}\|_k^2+\lambda_n\|\tfrac{f-f^*_{\varphi,\lambda_n}}{2}\|_k^2 \\ &\ge L_\varphi(f^*_{\varphi,\lambda_n})+\lambda_n\left\|f^*_{\varphi,\lambda_n}\right\|_k^2+\lambda_n\|\tfrac{f-f^*_{\varphi,\lambda_n}}{2}\|_k^2. \end{aligned}$$

Taking expectations on both sides leads to $E(g)\ge\lambda_n\left\|f-f^*_{\varphi,\lambda_n}\right\|_k^2/2$. Combining this with (A. 6), we conclude that $E(g^2)\le cE(g)$, where

$$c=\frac{2}{\lambda_n}\left(C_L+2\sqrt{M\lambda_n}\right)^2. \quad \text{(A.7)}$$

To estimate the bound for $N(B^{-1}\mathcal{G}_{\phi,\lambda_n},\varepsilon,L_2(P_n))$, we first have

$$N(B^{-1}\mathcal{G}_{\varphi,\lambda_n},\varepsilon,L_2(P_n))=N\left(B^{-1}\left\{L_\varphi(f)+\lambda_n\|f\|_k^2:f\in B_{\mathcal{H}_k}\left(\sqrt{M/\lambda_n}\right)\right\},\varepsilon,L_2(P_n)\right).$$

From the sub-additivity of the entropy,

$$\begin{aligned} \log N(B^{-1}\mathcal{G}_{\varphi,\lambda_n},2\varepsilon,L_2(P_n)) &\le \log N\left(B^{-1}\left\{L_\varphi(f):f\in B_{\mathcal{H}_k}\left(\sqrt{M/\lambda_n}\right)\right\},\varepsilon,L_2(P_n)\right) \\ &+\log N\left(\left\{\lambda_n\|f\|_k^2,f\in B_{\mathcal{H}_k}\left(\sqrt{M/\lambda_n}\right)\right\},\varepsilon,L_2(P_n)\right). \end{aligned} \quad \text{(A.8)}$$

Using the Lipschitz-continuity of the weighted hinge loss, we now have that if u, $u' \in B^{-1}\{L_\varphi(f): f \in B_{\mathscr{H}_k}(\sqrt{M/\lambda_n})\}$ with corresponding $f, f' \in B_{\mathscr{H}_k}(\sqrt{M/\lambda_n})$, then $\|u - u'\|_{L_2(P_n)} \le B^{-1}C_L \|f - f'\|_{L_2(P_n)}$, and therefore the first term on the right-hand-side of (A.8) satisfies

$$\begin{aligned}\log N\left(B^{-1}\left\{L_\varphi(f): f \in B_{\mathscr{H}_k}\left(\sqrt{M/\lambda_n}\right)\right\}, \varepsilon, L_2(P_n)\right) &\le \log N\left(B_{\mathscr{H}_k}\left(\sqrt{M/\lambda_n}\right), \tfrac{B\varepsilon}{C_L}, L_2(P_n)\right) \\ &\le \log N\left(B_{\mathscr{H}_k}, \tfrac{B\varepsilon}{C_L\sqrt{M/\lambda_n}}, L_2(P_n)\right) \\ &\le \log N(B_{\mathscr{H}_k}, 2\varepsilon, L_2(P_n)).\end{aligned}$$

The last inequality follows because $B/C_L\sqrt{M/\lambda_n} \ge 2$. It is trivial to see that for the second term on the right hand side of (A.8),

$$\log N\left(\left\{\lambda_n \|f\|_k^2, f \in B\left(\sqrt{M/\lambda_n}\right)\right\}, \varepsilon, L_2(P_n)\right) \le \log\left(\frac{M}{B\varepsilon}\right).$$

Thus,

$$\log N(B^{-1}\mathscr{G}_{\varphi,\lambda_n}, 2\varepsilon, L_2(P_n)) \le \log N(B_{\mathscr{H}_k}, 2\varepsilon, L_2(P_n)) + \log\left(\frac{M}{B\varepsilon}\right).$$

Using (3.5) and a given choice for B, we obtain for all $\sigma_n > 0$, $0 < \nu < 2$, $\delta > 0$, $\varepsilon > 0$,

$$\sup_{P_n} \log N(B^{-1}\mathscr{G}_{\varphi,\lambda_n}, \varepsilon, L_2(P_n)) \le c_2 \sigma_n^{(1-\nu/2)(1+\delta)d} \varepsilon^{-\nu},$$

where c_2 depends on ν, δ and d.

Consequently, from Theorem 5.6 in Steinwart & Scovel (2007), there exists a constant $c_\nu > 0$ depending only on ν such that for all $n \ge 1$ and all $\tau \ge 1$, we have the bound for the first term

$$P^*\left(\lambda_n\left\|\widehat{f_n}\right\|_k^2 + \mathscr{R}_\varphi(\widehat{f_n}) > \inf_{f\in\mathscr{H}_k}(\lambda_n\|f\|_k^2 + \mathscr{R}_\varphi(f)) + c_\nu \varepsilon(n, \tilde{c}, B, c, \tau)\right) \le e^{-\tau},$$

where

$$\varepsilon(n, \tilde{c}, B, c, \tau) = \left(B + B^{\frac{2\nu}{2+\nu}} c^{\frac{2-\nu}{2+\nu}}\right)\left(\frac{\tilde{c}}{n}\right)^{\frac{2}{2+\nu}} + (B+c)\frac{\tau}{n}.$$

With B and c as defined in (A.5) and (A.7), i.e., $\tilde{c} = c_2\sigma_n^{(1-\nu/2)(1+\delta)d}$ and $\sigma_n = \lambda_n^{1/(q+1)d}$, we obtain

$$\varepsilon(n,\tilde{c},B,c,\tau)=C_1\left(\frac{1}{\lambda_n}\right)^{\frac{2}{2+\nu}+\frac{(2-\nu)(1+\delta)}{(2+\nu)(1+q)}}\left(\frac{1}{n}\right)^{\frac{2}{2+\nu}}+C_2\left(\frac{1}{\lambda_n}\right)^{\frac{q}{q+1}}\frac{\tau}{n}, \quad \text{(A.9)}$$

where C_1 and C_2 are constants depending on ν, δ, d, M and π. We complete the proof of Theorem 3.4 by plugging (A.4) and (A.9) into (A.3).

Proof of Theorem 3.5

We apply Theorem 4.3 in Blanchard et al. (2008) on the scaled loss function $\tilde{L}_\varphi(f) = L_\varphi(f)/C_L$ to obtain the rates in Theorem 3.5. Without loss of generality, we can assume that the Bayes classifier $f^* \in \mathcal{H}_k$, since we can always find $g \in \mathcal{H}_k$ such that $\mathcal{R}_\varphi(g)=\mathcal{R}_\varphi(f^*)=\mathcal{R}^*_\varphi$, provided that $\mathcal{H}_k$ is dense in $C(\mathcal{X})$. Let $\mathcal{S}$ be a countable and dense subset of $\mathbb{R}^+$, and let $B_{\mathcal{H}_k}(S)$ denote the ball of $\mathcal{H}_k$ of radius S. Then $B_{\mathcal{H}_k}(S)$, $S \in \mathcal{S}$ is a countable collection of classes of functions. We can then use Theorem 4.3 in Blanchard et al. (2008) after we verify the following conditions (H1)–(H4):

- **(*H1*)** $\forall S \in \mathcal{S}$, $\forall f \in B_{\mathcal{H}_k}(S)$, $\|\tilde{L}_\varphi(f)\|_\infty \le b_S$, $b_S = 1+ S$;
- **(*H2*)** $\forall f, f' \in \mathcal{H}_k$, $Var(\tilde{L}_\varphi(f) - \tilde{L}_\varphi(f')) \le d^2(f, f')$, $d(f, f') = \|f - f'\|_{L_2(P)}$;
- **(*H3*)** $\forall S \in \mathcal{S}$, $\forall f \in B_{\mathcal{H}_k}(S)$, $d^2(f, f^*) \le C_S E(\tilde{L}_\varphi(f) - \tilde{L}_\varphi(f^*))$, $C_S = 2(S/\eta_0 + 1/\eta_1)$;
- **(*H4*)** Let

$$\xi(x)=\int_0^x \sqrt{\log N(B_{\mathcal{H}_k},\varepsilon,L_2(P_n))}\,d\varepsilon.$$

We have

$$E\left[\sup_{\substack{f\in B_{\mathcal{H}_k}(S)\\ d^2(f,f')\le r}}(P-P_n)(\tilde{L}_\varphi(f)-\tilde{L}_\varphi(f'))\right]\le \inf_{\vartheta>0}\left\{4\vartheta-\frac{12}{\sqrt{n}}\xi(\vartheta)+\frac{12}{\sqrt{n}}\xi\left(\frac{\sqrt{r}}{\sqrt{2}S}\right)\right\}=\psi_S(r).$$

ψ_S, $S \in \mathcal{S}$, is a sequence of sub-root functions, that is, ψ_S is non-negative, nondecreasing, and $\psi_S(r)/\sqrt{r}$ is non-increasing for $r > 0$. Denote x_* as the solution of the equation $\xi(x)= \sqrt{n}x^2$

. If r_S^* denotes the solution of $\psi_S(r) = r/C_S$, then

$$r_S^* \le \inf_{\vartheta>0} C_S\{4\vartheta-12\xi(\vartheta)/\sqrt{n}\}+c^2C_S^2x_*^2.$$

Under these conditions, we define for $n \in \mathbb{N}$ the following quantity:

$$\gamma_n=\inf_{\vartheta>0}\left\{4\vartheta-\frac{12}{\sqrt{n}}\xi(\vartheta)+x_*^2(n)\right\}.$$

Given $\mathcal{H}_k$ is associated with the Gaussian kernel, we can show that $\xi(x) \le \varepsilon^{1-\nu}$ for any $0 < \nu < 2$. Thus, $\gamma_n \le \max(n^{-1/2\nu}, n^{-1/(\nu+1)})$. By the choice of $\lambda_n = O(n^{-1/(\nu+1)})$ for any $\nu \in (0, 1)$, this satisfies

$$\lambda_n \ge c\left(\gamma_n + \eta_1^{-1}\frac{\log(\tau^{-1}\log n) \vee 1}{n}\right).$$

Therefore, according to Theorem 4.3 in Blanchard et al. (2008), the following bound holds with probability at least $1 - \tau$, where $\tau > 0$ is a fixed real number:

$$E(\tilde{L}_\varphi(\widehat{f_n})) - E(\tilde{L}_\varphi(f^*)) \le 2\inf_{f\in\mathcal{H}_k}[E(\tilde{L}_\varphi(f)) - E(\tilde{L}_\varphi(f^*)) + 2\lambda_n\|f\|_k^2] + 4\lambda_n(8 + c\eta_1\eta_0^{-1}).$$

The result does not change after we scale back to the original loss $L_\varphi(f)$. We have shown that $\inf_{f\in\mathcal{H}_k}[\mathcal{R}_\varphi(f) - \mathcal{R}_\varphi(f^*) + 2\lambda_n\|f\|_k^2] = O(\lambda_n^{q/(q+1)})$ in the proof of Theorem 3.4. Thus

$$\mathcal{R}(\widehat{f_n}) - \mathcal{R}^* = O_p(\lambda_n^{q/(q+1)}) = O_p\left(n^{-\frac{1}{\nu+1}\frac{q}{q+1}}\right).$$

The remainder of the proof is to verify conditions (H1)–(H4).

For condition (H1), $\|\tilde{L}_\varphi(f)\|_\infty \le \sup\{R(A\pi + (1 - A)/2)\}(1 + S)/C_L \le 1 + S$, $\|f\|_k \le S$.

For condition (H2), let $d(f, f') = \|f - f'\|_{L_2(P)}$. $L_\varphi(f)$ is a Lipschitz function with respect to f with Lipschitz constant C_L. Then $\tilde{L}_\varphi(f) - \tilde{L}_\varphi(f') \le |f(x) - f'(x)|$. Hence (H2) is easily satisfied.

For condition (H3), the proof is similar to to Lemma 6.4 of Blanchard et al. (2008) with $C_S = 2(S/\eta_1 + 1/\eta_0)$, where η_0 and η_1 are as defined in Assumptions (A1) and (A2) of Section 3.5.

For condition (H4), we introduce the notation for Rademacher averages: let $\varepsilon_1, \ldots, \varepsilon_n$ be n i.i.d Rademacher random variables, independent of (X_i, A_i, R_i), $i = 1, \ldots, n$. For any measurable real-valued function f, the Rademacher average is defined as $\mathcal{L}_n f = n^{-1}\sum_{i=1}^n \varepsilon_i f(X_i)$. Also let $\mathcal{L}_n(\mathcal{F})$ be the empirical Rademacher complexity of function class $\mathcal{F}$, $\mathcal{L}_n\mathcal{F} = \sup_{f\in\mathcal{F}} \mathcal{L}_n f$.

First we have from Lemma 6.7 of Blanchard et al. (2008) that for $f' \in \mathcal{H}_k$,

$$E\left[\sup_{f\in\mathcal{H}_k}(P - P_n)(\tilde{L}_\varphi(f) - \tilde{L}_\varphi(f'))\right] \le 4E[\mathcal{L}_n\{f - f', f \in \mathcal{H}_k\}].$$

Thus for the set $\{f \in \mathcal{H}_k : \|f\|_k \le S, d^2(f, f') \le r\}$ and $f' \in B_{\mathcal{H}_k}(S)$,

$$E\left[\sup_{f\in\mathscr{H}_k}(P-P_n)(\tilde{L}_\varphi(f)-\tilde{L}_\varphi(f'))\right]\le 4E[\mathscr{L}_n\{f-f', f\in\mathscr{H}_k:\|f\|_k\le S, d^2(f,f')\le r\}],$$

the right-hand-side of which is equivalent to $4E\left[\mathscr{L}_n\{f, f\in\mathscr{H}_k:\|f\|_k\le 2S, \|f\|^2_{L_2(P_n)}\le 2r\}\right]$. Now we proceed to show that

$$E\mathscr{L}_n\left\{f\in\mathscr{H}_k:\|f\|_k\le 2S, \|f\|^2_{L_2(P_n)}\le 2r\right\}\le\inf_{\vartheta>0}\left\{4\vartheta+\frac{12}{\sqrt{n}}\int_\vartheta^{\frac{\sqrt{r}}{\sqrt{2}S}}\sqrt{\log N(B_{\mathscr{H}},\varepsilon,L_2(P_n))}d\varepsilon\right\}=\psi_S(r),$$

by slightly modifying the procedure in obtaining Dudley's Entropy Integral for Rademacher complexity of sets of functions. For $j\ge 0$, let $r_j=\sqrt{2}r2^{-j}$ and T_j be a r_j-cover of $B_{\mathscr{H}_k}(2S)$ with respect to the $L_2(P_n)$-norm. For each $f\in B_{\mathscr{H}_k}(2S)$, we can find an $\tilde{f}_j\in T_j$, such that $\|f-\tilde{f}_j\|_{L_2(P_n)}\le r_j$. For any N, we express f as $f=f-\tilde{f}_N+\sum_{j=1}^N(\tilde{f}_j-\tilde{f}_{j-1})$, where $\tilde{f}_0=0$. Note $\tilde{f}_0=0$ is an r_0-approximation of f. Hence,

$$\begin{aligned}\mathscr{L}_n(B_{\mathscr{H}_k}(2S))&=E\left[\sup_{f\in B_{\mathscr{H}_k}(2S)}\frac{1}{n}\sum_{i=1}^n\varepsilon_i\left(f(X_i)-\tilde{f}_N(X_i)+\sum_{j=1}^N(\tilde{f}_j(X_i)-\tilde{f}_{j-1}(X_i))\right)\right]\\&\le E\left[\sup_{f\in B_{\mathscr{H}_k}(2S)}\|\varepsilon\|_{L_2(P_n)}\|f-\tilde{f}_N\|_{L_2(P_n)}\right]+\sum_{j=1}^N E\left[\sup_{f\in B_{\mathscr{H}_k}(2S)}\frac{1}{n}\sum_{i=1}^n\left(\tilde{f}_j(X_i)-\tilde{f}_{j-1}(X_i)\right)\right]\\&\le r_N+\sum_{j=1}^N E\left[\sup_{f\in B_{\mathscr{H}_k}(2S)}\frac{1}{n}\sum_{i=1}^n\left(\tilde{f}_j(X_i)-\tilde{f}_{j-1}(X_i)\right)\right].\end{aligned}$$

Note that

$$\|\tilde{f}_j-\tilde{f}_{j-1}\|^2_{L_2(P_n)}\le\left(\|\tilde{f}_j-f\|_{L_2(P_n)}+\|f-\tilde{f}_{j-1}\|_{L_2(P_n)}\right)^2\le(r_j+r_{j-1})^2=9r_j^2.$$

We therefore have

$$\begin{aligned}\mathscr{L}_n(B_{\mathscr{H}_k}(2S))&\le r_N+\sum_{j=1}^N 3r_j\sqrt{\frac{2\log(|T_j||T_{j-1}|)}{n}}\\&\le r_N+12\sum_{j=1}^N(r_j-r_{j+1})\sqrt{\frac{\log N(B_{\mathscr{H}_k}(2S),r_j,L_2(P_n))}{n}}\\&\le r_N+12\int_{r_{N+1}}^{\sqrt{r}/\sqrt{2}S}\sqrt{\frac{\log N(B_{\mathscr{H}_k},\varepsilon,L_2(P_n))}{n}}d\varepsilon.\end{aligned}$$

For any $\vartheta>0$, we can choose $N=\sup\{j: r_j>2\vartheta\}$. Therefore, $\vartheta<r_{N+1}<2\vartheta$, and $r_N<4\vartheta$. We therefore conclude that

$$\mathscr{L}_n(B_{\mathscr{H}_k}(2S)) \le \inf_{\vartheta>0}\left\{4\vartheta+12\int_{\vartheta}^{\sqrt{r}/\sqrt{2}S}\sqrt{\frac{\log N(B_{\mathscr{H}_k},\varepsilon,L_2(P_n))}{n}}d\varepsilon\right\}$$
$$=\inf_{\vartheta>0}\left\{4\vartheta-\frac{12}{\sqrt{n}}\xi(\vartheta)+\frac{12}{\sqrt{n}}\xi\left(\frac{\sqrt{r}}{\sqrt{2}S}\right)\right\}=\psi_S(r).$$

The function ψ_S is sub-root because log $N(B_{\mathscr{H}_k}, \varepsilon, L_2(P_n))$ is a decreasing function of ε.

To show the upperbound of r^*, let $t_S^*=c^2C_S^2x_*^2$. Then $\sqrt{t_S^*}/\sqrt{2}S=cC_Sx_*/\sqrt{2}S$, $C_S/S\ge 1$. Assuming that $c\ge 2$, we have $\sqrt{t_S^*}/\sqrt{2}S \ge x_*$. Since $x^{-1}\xi(x)$ is a decreasing function, it follows that

$$\xi\left(\frac{\sqrt{t_S^*}}{\sqrt{2}S}\right)\le c\frac{C_S}{\sqrt{2}S}\xi(x_*)=\frac{\sqrt{n}}{cSC_S}t_S^*.$$

Therefore, by selecting an appropriate constant c,

$$\psi_S(t_S^*)\le\inf_{\vartheta>0}\left\{4\vartheta-\frac{12}{\sqrt{n}}\xi(\vartheta)\right\}+\frac{12}{cSC_S}t_S^*\le\frac{C_S\inf_{\vartheta>0}\{4\vartheta-12/\sqrt{n}\xi(\vartheta)\}+t_S^*}{C_S}.$$

The desired result follows from the property of sub-root functions, which states that if ψ : $[0, \infty) \to [0, \infty)$ is a sub-root function, then the unique positive solution of $\psi(r) = r$, denoted by r^*, exists, and for all $r > 0$, $r \ge \psi(r)$ if and only if $r^* \le r$ (Bartlett et al. 2005).

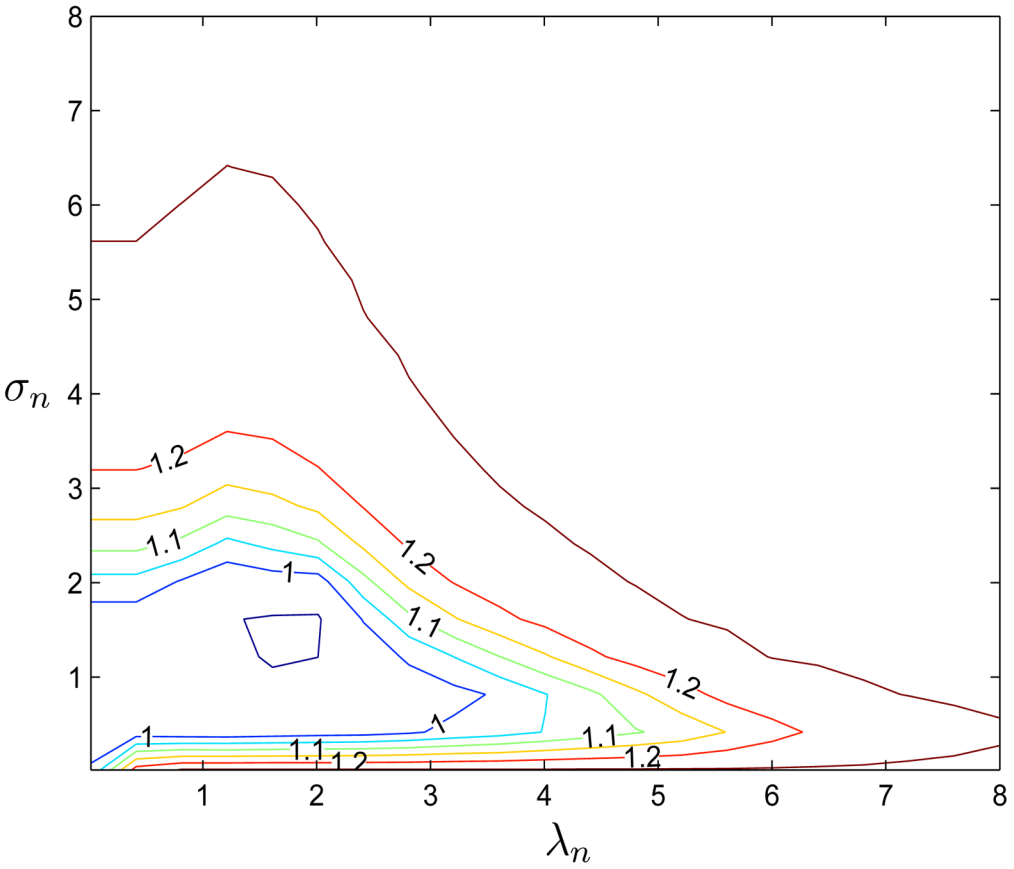

Figure 1.
Contour Plots of Value Function for Example 1 with $\lambda_n \in (0, 10)$ and $\sigma_n \in (0, 10)$

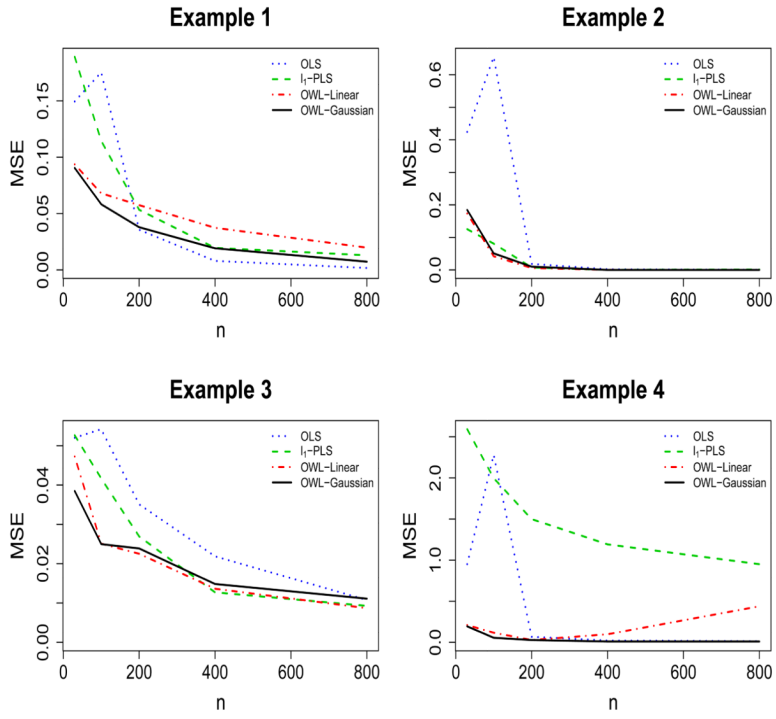

Figure 2.
MSE for Value Functions of Individualized Treatment Rules

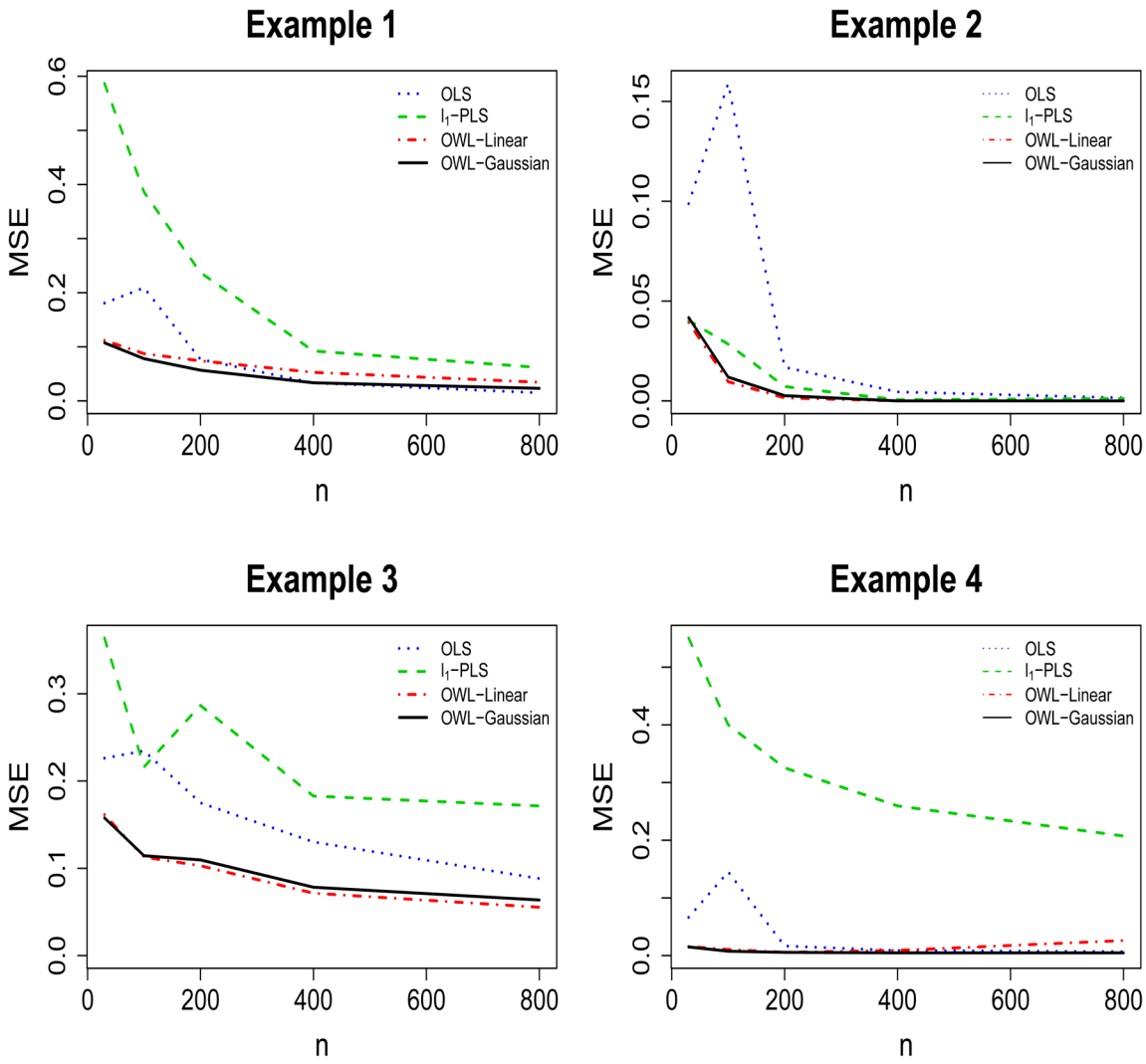

Figure 3.
MSE for Misclassification Rates of Individualized Treatment Rules

Table 1

Mean Depression Scores (the Smaller, the Better) from Cross Validation Procedure with Different Methods

	OLS	l_1-PLS	OWL
Nefazodone vs CBASP	15.87	15.95	15.74
Combination vs Nefazodone	11.75	11.28	10.71
Combination vs CBASP	12.22	10.97	10.86