Learning Optimal Personalized Treatment Rules Using Robust Regression Informed K-NN

Ruidi Chen
Division of Systems Engineering,
Boston University, Boston, MA
rchen15@bu.edu.

Ioannis Ch. Paschalidis
Department of Electrical and Computer Engineering,
Division of Systems Engineering,
and Department of Biomedical Engineering,
Boston University, Boston, MA
yannisp@bu.edu

Abstract

We develop a prediction-based prescriptive model for learning optimal personalized treatments for patients based on their Electronic Health Records (EHRs). Our approach consists of: (i) predicting future outcomes under each possible therapy using a robustified nonlinear model, and (ii) adopting a randomized prescriptive policy determined by the predicted outcomes. We show theoretical results that guarantee the out-of-sample predictive power of the model, and prove the optimality of the randomized strategy in terms of the expected true future outcome. We apply the proposed methodology to develop optimal therapies for patients with type 2 diabetes or hypertension using EHRs from a major safety-net hospital in New England, and show that our algorithm leads to a larger reduction of the HbA_{1c}, for diabetics, or systolic blood pressure, for patients with hypertension, compared to the alternatives. We demonstrate that our approach outperforms the standard of care under the robustified nonlinear predictive model.

1 Introduction

Learning the optimal treatment policies from large amounts of clinical data presents challenges due to (i) the significant "noise" caused by recording errors, missing values, and large variability across patients, (ii) the lack of counterfactual information, and (iii) the underlying nonlinearity whose parametric form is not known a priori. Attempts on handling these issues have been focusing on denoising [2, 16] and robust nonlinear learning [24, 15]. A popular research direction adopts a dynamic point of view and models the optimal treatment problem through reinforcement learning [4, 18, 13, 14] and A-learning [12, 15]. There are also works adopting a Bayesian approach [10, 21], deep learning [2, 11], and miscellaneous statistical inference/machine learning methods [19, 17, 23, 22, 7] for constructing effective prescriptive rules.

We are interested in learning optimal treatment decisions from Electronic Health Records (EHRs) using computationally efficient and interpretable predictive algorithms that address the aforementioned challenges. We start with a regularized *Least Absolute Deviation (LAD)* [20] model that is obtained as a result of a *Distributionally Robust Optimization (DRO)* [5] problem and thus hedges against significant noise and the presence of outliers. We then predict future outcomes under each treatment by identifying neighbors through an LAD-weighted metric and fitting a *K-Nearest Neighbors (K-NN)* regression model [1] that captures the local nonlinearity in a non-parametric way. A randomized policy is then developed to prescribe each therapy with a probability inversely proportional to its exponentiated predicted outcome.

Our method is similar to [3] where the K-NN regression with an *Ordinary Least Squares (OLS)*-weighted metric was used to learn the optimal treatment for type 2 diabetic patients. The key dif-

Machine Learning for Health (ML4H) Workshop at NeurIPS 2018.

ferences lie in that: (i) we adopt a robustified regression procedure that is immunized against high noise and is thus more stable and reliable; (ii) we propose a randomized prescriptive policy that adds robustness to the methodology whereas [3] deterministically prescribed the treatment that was predicted to result in the lowest HbA_{1c}; (iii) we show rigorous theoretical results that guarantee the optimality of the randomized strategy, and (iv) the prescriptive rule in [3] was activated when the improvement of the recommended treatment over the standard of care exceeded a certain threshold whereas our method looks into the improvement over the previous regimen. This distinction makes our algorithm applicable in the scenario where the standard of care is unknown or ambiguous. Further, we derive a closed-form expression for the threshold level, which greatly improves the computational efficiency compared to [3] where a threshold was selected by cross-validation.

2 Methods

Given a training set of EHRs, we group the patients based on the treatment they receive at the time of the visit. Within each treatment group $m \in [M]$, where $[M] \triangleq \{1, \ldots, M\}$, denote by $\mathbf{x}_m \in \mathbb{R}^p$ a patient-specific feature vector that includes the patient characteristics (e.g., gender, age, race, and *Body Mass Index (BMI)*) and diagnosis history. Our goal is to predict the future outcome y_m under treatment m. Assume the following nonlinear regression model:

$$y_m = \mathbf{x}_m' \boldsymbol{\beta}_m^* + h_m(\mathbf{x}_m) + \epsilon_m,$$

where prime denotes transpose, $\boldsymbol{\beta}_m^*$ is the coefficient vector that captures the linear trend, $h_m(\mathbf{x}_m)$ describes the nonlinear fluctuation in y_m, and ϵ_m is the noise term with zero mean and standard deviation η_m that expresses the intrinsic randomness of y_m and is assumed to be independent of $\mathbf{x}_m$.

We follow the robust predictive procedure proposed in [5] to estimate the linear coefficient $\boldsymbol{\beta}_m^*$. It reduces to solving the following regularized LAD problem:

$$\inf_{\boldsymbol{\beta}_m} \frac{1}{N_m} \sum_{i=1}^{N_m} |y_{mi} - \mathbf{x}_{mi}' \boldsymbol{\beta}_m| + r_m \|(-\boldsymbol{\beta}_m, 1)\|_2, \tag{1}$$

where $(\mathbf{x}_{mi}, y_{mi}), i \in [N_m]$, are the feature-outcome vectors of patients who receive treatment m, and r_m is a regularization penalty. Solving (1) gives us a robust estimator of $\boldsymbol{\beta}_m^*$, which we denote by $\hat{\boldsymbol{\beta}}_m$. It measures the relative significance of the features in terms of their relevance to y_m, and will be used to identify the nearest neighbors in the nonlinear non-parametric K-NN regression model. Specifically, we consider the following weighted distance metric,

$$\|\mathbf{x} - \mathbf{x}_{mi}\|_{\hat{\mathbf{W}}_m}^2 = (\mathbf{x} - \mathbf{x}_{mi})' \hat{\mathbf{W}}_m (\mathbf{x} - \mathbf{x}_{mi}), \tag{2}$$

where $\hat{\mathbf{W}}_m$ is a diagonal matrix with diagonal elements $(\hat{\beta}_{m1})^2, \ldots, (\hat{\beta}_{mp})^2$, where $\hat{\beta}_{mi}$ is the i-th element of $\hat{\boldsymbol{\beta}}_m$. Given a patient with features $\mathbf{x}$, we find her K_m nearest neighbors using (2) within each treatment group m. The predicted future outcome for $\mathbf{x}$ if treatment m is prescribed, denoted by $\hat{y}_m(\mathbf{x})$, is computed as the average outcome among the K_m nearest neighbors, i.e.,

$$\hat{y}_m(\mathbf{x}) = \frac{1}{K_m} \sum_{i=1}^{K_m} y_{m(i)}, \tag{3}$$

where $y_{m(i)}$ is the outcome of the i-th nearest neighbor to $\mathbf{x}$. (3) can be viewed as a local smooth estimator in the neighborhood of $\mathbf{x}$. The selected neighbors are similar to $\mathbf{x}$ in the features that are predictive of the outcome.

Next consider a randomized policy that prescribes treatment m with probability $e^{-\xi \hat{y}_m(\mathbf{x})} / \sum_{j=1}^M e^{-\xi \hat{y}_j(\mathbf{x})}$, with ξ some pre-specified positive scalar. Theorem 2.1 characterizes the expected *true* outcome under this policy.

Theorem 2.1. *Given any fixed predictor $\mathbf{x} \in \mathbb{R}^p$, denote its predicted and true future outcome under treatment m by $\hat{y}_m(\mathbf{x})$ and $y_m(\mathbf{x})$, respectively. Assume that we adopt a randomized strategy that prescribes treatment m with probability $e^{-\xi \hat{y}_m(\mathbf{x})} / \sum_{j=1}^M e^{-\xi \hat{y}_j(\mathbf{x})}$, for some $\xi \geq 0$. Assume $\hat{y}_m(\mathbf{x})$*

and $y_m(\mathbf{x})$ are non-negative, $\forall m \in [M]$. The expected true outcome under this policy satisfies:

$$\sum_{m=1}^{M} \frac{e^{-\xi \hat{y}_m(\mathbf{x})}}{\sum_j e^{-\xi \hat{y}_j(\mathbf{x})}} y_m(\mathbf{x}) \leq y_k(\mathbf{x}) + \left(\hat{y}_k(\mathbf{x}) - \frac{1}{M}\sum_{m=1}^{M} \hat{y}_m(\mathbf{x})\right) + \xi\left(\frac{1}{M}\sum_{m=1}^{M} \hat{y}_m^2(\mathbf{x}) + \sum_{m=1}^{M} \frac{e^{-\xi \hat{y}_m(\mathbf{x})}}{\sum_j e^{-\xi \hat{y}_j(\mathbf{x})}} y_m^2(\mathbf{x})\right) + \frac{\log M}{\xi}, \tag{4}$$

for any $k \in [M]$.

Theorem 2.1 says that the expected *true* outcome of the randomized policy is no worse than the true outcome of any treatment k plus two components, one accounting for the gap between the *predicted* outcome under k and the average predicted outcome, and the other depending on the parameter ξ. Thinking about choosing $k = \arg\min_m y_m(\mathbf{x})$, if $\hat{y}_k(\mathbf{x})$ is below the average predicted outcome (which should be true if we have an accurate prediction), it follows from (4) that the randomized policy leads to a nearly optimal future outcome by an appropriate choice of ξ.

In consideration of the health care costs and treatment transients, it is not desired to switch patients' treatments too frequently. We thus set a threshold level for the expected improvement on the outcome, below which the randomized strategy will be frozen and the current therapy will be continued. Specifically,

$$m_{\mathrm{f}}(\mathbf{x}) = \begin{cases} m, \text{ w.p. } \frac{e^{-\xi \hat{y}_m(\mathbf{x})}}{\sum_{j=1}^{M} e^{-\xi \hat{y}_j(\mathbf{x})}}, & \text{if } \sum_k \frac{e^{-\xi \hat{y}_k(\mathbf{x})}}{\sum_j e^{-\xi \hat{y}_j(\mathbf{x})}} \hat{y}_k(\mathbf{x}) \leq x_{\mathrm{co}} - T(\mathbf{x}), \\ m_{\mathrm{c}}(\mathbf{x}), & \text{otherwise,} \end{cases}$$

where $m_{\mathrm{f}}(\mathbf{x})$ and $m_{\mathrm{c}}(\mathbf{x})$ are the future and current treatments of patient $\mathbf{x}$, respectively; x_{co} represents the current observed outcome, which is assumed to be one of the components of $\mathbf{x}$, and $T(\mathbf{x})$ is some threshold level which is determined in Theorem 2.2.

Theorem 2.2. *Assume that the distribution of the predicted outcome $\hat{y}_m(\mathbf{x})$ conditional on $\mathbf{x}$, is sub-Gaussian, and its ψ_2-norm is equal to $\sqrt{2}C_m(\mathbf{x})$, for any $m \in [M]$ and any $\mathbf{x}$. Given a small $0 < \bar{\epsilon} < 1$, if*

$$\mathbb{P}\Big(\sum_k \frac{e^{-\xi \hat{y}_k(\mathbf{x})}}{\sum_j e^{-\xi \hat{y}_j(\mathbf{x})}} \hat{y}_k(\mathbf{x}) > x_{co} - T(\mathbf{x})\Big) \leq \bar{\epsilon},$$

then,

$$T(\mathbf{x}) = \max\Big(0,\ \min_m\Big(x_{co} - \mu_{\hat{y}_m}(\mathbf{x}) - \sqrt{-2C_m^2(\mathbf{x})\log(\bar{\epsilon}/M)}\Big)\Big),$$

where $\mu_{\hat{y}_m}(\mathbf{x}) = \mathbb{E}[\hat{y}_m(\mathbf{x})|\mathbf{x}]$.

Notice that as $\xi \to \infty$, the randomized policy will assign probability 1 to the treatment with the lowest predicted outcome, which is equivalent to the deterministic policy used in [3]. In this case, slight modification to the threshold level $T(\mathbf{x})$ is given as follows:

$$T(\mathbf{x}) = \max\Big(0,\ \min_m\Big(x_{\mathrm{co}} - \mu_{\hat{y}_m}(\mathbf{x}) - \sqrt{-2C_m^2(\mathbf{x})\log\bar{\epsilon}}\Big)\Big).$$

3 Results

We apply our method to develop optimal prescriptions for patients with type-2 diabetes and/or hypertension. The data used for the study come from the Boston Medical Center – the largest safety-net hospital in New England – and consist of EHRs containing the patients' medical history in the period 1999–2014. We consider the following sets of features for building the predictive model: (i) demographic information, including gender, age and race; (ii) measurements, including systolic blood pressure and diastolic blood pressure, Body Mass Index (BMI) and heart rate; (iii) lab tests, including blood chemistry tests and hematology tests, and (iv) diagnosis history. For the diabetic patients, we consider both oral, e.g., metformin, pioglitazone, and sitagliptin, etc., and injectable prescriptions, e.g., insulin. A complete list of the oral prescriptions can be found in the Appendix. For the hypertension patients, six types of prescriptions are considered: ACE inhibitor, Angiotensin Receptor Blockers (ARB), calcium channel blockers, thiazide and thiazide-like diuretics, α-blockers and β-blockers.

The diabetes dataset contains 12,016 patient visits with 63 normalized features, and the hypertension dataset contains 26,128 patient visits with the same set of features. We randomly split both datasets into a training (80%) and a test set (20%), where the training set is used to tune the parameters (e.g., the ℓ_2 regularizer r_m, the number of neighbors K_m, and the exponent ξ in the randomized policy) and train the models, and the test set is used to evaluate the performance. We regress the tuned K_m against $\sqrt{N_m}$ and use the derived linear equation to determine the number of neighbors to be used.

We will compare our method with several alternatives that replace our predictive model, which we refer to as RLAD+K-NN, with a different learning procedure such as LASSO, CART, and OLS+KNN [3]. Both deterministic and randomized prescriptive policies are considered using predictions from these models. To evaluate the performance, we need to determine the effects of the counterfactual treatments. By assessing the predictive power of several models on the non-grouped datasets in terms of their R^2 and out-of-sample estimation errors (see Appendix), we choose the RLAD+K-NN model that excels in all metrics to impute the outcome for an unobservable treatment m, where the number of neighbors should be selected to fit the size of the test set.

The average improvement in outcomes (the predicted *future* outcome under the recommended therapy minus the *current* observed outcome) on the test set is shown in Table 1, where the numbers outside the parentheses are the mean values over 5 repetitions, and the numbers in the parentheses are the corresponding standard deviations. We note that HbA_{1c} is measured in percentage while systolic blood pressure in mmHg. We also list the results from the *standard of care*, and the *current prescription* which prescribes $m_{\mathrm{f}}(\mathbf{x}) = m_{\mathrm{c}}(\mathbf{x})$ with probability one, i.e., always continuing the current drug regimen.

Table 1: The reduction in HbA_{1c}/systolic blood pressure for various models.

	Diabetes		Hypertension	
	Deterministic	Randomized	Deterministic	Randomized
LASSO	-0.51 (0.16)	-0.51 (0.16)	-4.71 (1.09)	-4.72 (1.10)
CART	-0.45 (0.13)	-0.42 (0.14)	-4.84 (0.62)	-4.87 (0.66)
OLS+K-NN	-0.53 (0.13)	-0.53 (0.13)	-4.33 (0.46)	-4.33 (0.47)
RLAD+K-NN	-0.56 (0.06)	-0.55 (0.08)	-6.98 (0.86)	-7.22 (0.82)
Current prescription	-0.22 (0.04)		-2.52 (0.19)	
Standard of care	-0.22 (0.03)		-2.37 (0.11)	

Several observations are in order: (i) all models outperform the current prescription and the standard of care; (ii) the RLAD+K-NN model leads to the largest reduction in outcomes with a relatively stable performance; and (iii) the randomized policy achieves a similar performance (slightly better on the hypertension dataset) to the deterministic one. We expect the randomized strategy to win when the effects of several treatments do not differ much, in which case the deterministic algorithm might produce misleading results. The randomized policy could potentially improve the out-of-sample (generalization) performance, as it gives the flexibility of exploring options that are suboptimal on the training set, but might be optimal on the test set. The advantages of the RLAD+K-NN model are more prominent on the hypertension dataset due to the fact that we considered a finer classification of the prescriptions for patients with hypertension, while for diabetic patients, we only distinguish between the oral and injectable prescriptions.

4 Conclusions

We developed a prediction-based randomized prescriptive algorithm that determines the probability of prescribing each treatment based on the predictions from an RLAD+K-NN model which takes into account both the high noise level and the nonlinearity of the data. A deterministic variant was obtained as a special case of the randomized strategy. Theoretical guarantees on the optimality of the prescriptive algorithm and a closed-form expression for the threshold level were provided. The proposed approach was applied to two datasets obtained from the Boston Medical Center: one with diabetic patients and the other for patients with hypertension, providing numerical evidence for the superiority of our algorithm in terms of the improvement in outcomes.

Acknowledgments

We thank Henghui Zhu who preprocessed the electronic health records to produce the datasets we used in this paper to validate the methods. We also thank Dr. Theofanie Mela and Dr. Rebecca Mishuris for useful discussions.

References

[1] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

[2] Onur Atan, J Jordan, and Mihaela van der Schaar. Deep-treat: Learning optimal personalized treatments from observational data using neural networks. AAAI, 2018.

[3] Dimitris Bertsimas, Nathan Kallus, Alexander M Weinstein, and Ying Daisy Zhuo. Personalized diabetes management using electronic medical records. *Diabetes care*, page dc160826, 2016.

[4] Bibhas Chakraborty, Susan Murphy, and Victor Strecher. Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical methods in medical research*, 19(3):317–343, 2010.

[5] Ruidi Chen and Ioannis Ch Paschalidis. A robust learning approach for regression models based on distributionally robust optimization. *Journal of Machine Learning Research*, 19(13), 2018.

[6] Elad Hazan. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

[7] Yicheng He, Junfeng Liu, Lijun Cheng, and Xia Ning. Drug selection via joint push and learning to rank. *arXiv preprint arXiv:1801.07691*, 2018.

[8] Peter J Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, pages 73–101, 1964.

[9] Peter J Huber. Robust regression: asymptotics, conjectures and monte carlo. *The Annals of Statistics*, pages 799–821, 1973.

[10] Juhee Lee, Peter F Thall, Yuan Ji, and Peter Müller. Bayesian dose-finding in two treatment cycles based on the joint utility of efficacy and toxicity. *Journal of the American Statistical Association*, 110(510):711–722, 2015.

[11] Shuhan Liang, Wenbin Lu, and Rui Song. Deep advantage learning for optimal dynamic treatment regime. *Statistical Theory and Related Fields*, pages 1–9, 2018.

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

[13] Niranjani Prasad, Li-Fang Cheng, Corey Chivers, Michael Draugelis, and Barbara E Engelhardt. A reinforcement learning approach to weaning of mechanical ventilation in intensive care units. *arXiv preprint arXiv:1704.06300*, 2017.

[14] Aniruddh Raghu, Matthieu Komorowski, Leo Anthony Celi, Peter Szolovits, and Marzyeh Ghassemi. Continuous state-space models for optimal sepsis treatment - a deep reinforcement learning approach. In *MLHC*, 2017.

[15] Chengchun Shi, Ailin Fan, Rui Song, Wenbin Lu, et al. High-dimensional a-learning for optimal dynamic treatment regimes. *The Annals of Statistics*, 46(3):925–957, 2018.

[16] Chengchun Shi, Rui Song, and Wenbin Lu. Robust learning for optimal treatment decision with np-dimensionality. *Electronic journal of statistics*, 10:2894, 2016.

[17] Chengchun Shi, Rui Song, Wenbin Lu, and Bo Fu. Maximin projection learning for optimal treatment decision with heterogeneous individualized treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2018.

[18] Susan M Shortreed, Eric Laber, Daniel J Lizotte, T Scott Stroup, Joelle Pineau, and Susan A Murphy. Informing sequential clinical decision-making through reinforcement learning: an empirical study. *Machine learning*, 84(1-2):109–136, 2011.

[19] Lan Wang, Yu Zhou, Rui Song, and Ben Sherwood. Quantile-optimal treatment regimes. *Journal of the American Statistical Association*, 113(523):1243–1254, 2018.

[20] Li Wang, Michael D Gordon, and Ji Zhu. Regularized least absolute deviations regression and an efficient algorithm for parameter tuning. In *Data Mining, 2006. ICDM'06. Sixth International Conference on*, pages 690–700. IEEE, 2006.

[21] Yingfei Wang and Warren Powell. An optimal learning method for developing personalized treatment regimes. *arXiv preprint arXiv:1607.01462*, 2016.

[22] Yuanjia Wang, Haoda Fu, and Donglin Zeng. Learning optimal personalized treatment rules in consideration of benefit and risk: with an application to treating type 2 diabetes patients with insulin therapies. *Journal of the American Statistical Association*, 113(521):1–13, 2018.

[23] Yuanjia Wang, Peng Wu, Ying Liu, Chunhua Weng, and Donglin Zeng. Learning optimal individualized treatment rules from electronic health record data. In *Healthcare Informatics (ICHI), 2016 IEEE International Conference on*, pages 65–71. IEEE, 2016.

[24] Ying-Qi Zhao, Donglin Zeng, Eric Benjamin Laber, Rui Song, Ming Yuan, and Michael Rene Kosorok. Doubly robust learning for estimating individualized treatment with censored data. *Biometrika*, 102(1):151–168, 2014.

Appendix

Proof of Theorem 2.1

Proof. The proof borrows ideas from Theorem 1.5 in [6]. Define $W_m \triangleq e^{-\xi \hat{y}_m(\mathbf{x})} / \sum_{j=1}^M e^{-\xi \hat{y}_j(\mathbf{x})}$, and $\phi \triangleq \sum_{m=1}^M e^{-\xi \hat{y}_m(\mathbf{x})} e^{-\xi y_m(\mathbf{x})}$. Then,

$$\begin{aligned}
\phi &= \Big(\sum_{j=1}^M e^{-\xi \hat{y}_j(\mathbf{x})}\Big) \sum_{m=1}^M W_m e^{-\xi y_m(\mathbf{x})} \\
&\le \Big(\sum_{j=1}^M e^{-\xi \hat{y}_j(\mathbf{x})}\Big) \sum_{m=1}^M W_m \big(1 - \xi y_m(\mathbf{x}) + \xi^2 y_m^2(\mathbf{x})\big) \\
&= \Big(\sum_{j=1}^M e^{-\xi \hat{y}_j(\mathbf{x})}\Big) \Big(1 - \xi \sum_{m=1}^M W_m y_m(\mathbf{x}) + \xi^2 \sum_{m=1}^M W_m y_m^2(\mathbf{x})\Big) \\
&\le \Big(\sum_{j=1}^M e^{-\xi \hat{y}_j(\mathbf{x})}\Big) e^{-\xi \sum_{m=1}^M W_m y_m(\mathbf{x}) + \xi^2 \sum_{m=1}^M W_m y_m^2(\mathbf{x})},
\end{aligned}$$

where the first inequality uses the fact that for $x \ge 0$, $e^{-x} \le 1 - x + x^2$, and the last inequality is due to the fact that $1 + x \le e^x$. Next let us examine the sum of exponentials:

$$\begin{aligned}
\sum_{j=1}^M e^{-\xi \hat{y}_j(\mathbf{x})} &\le \sum_{j=1}^M \Big(1 - \xi \hat{y}_j(\mathbf{x}) + \xi^2 \hat{y}_j^2(\mathbf{x})\Big) \\
&= M\Big(1 - \xi \frac{1}{M} \sum_{j=1}^M \hat{y}_j(\mathbf{x}) + \xi^2 \frac{1}{M} \sum_{j=1}^M \hat{y}_j^2(\mathbf{x})\Big) \\
&\le M e^{-\xi \frac{1}{M} \sum_{j=1}^M \hat{y}_j(\mathbf{x}) + \xi^2 \frac{1}{M} \sum_{j=1}^M \hat{y}_j^2(\mathbf{x})}.
\end{aligned}$$

On the other hand, for any $k \in [M]$,

$$e^{-\xi \hat{y}_k(\mathbf{x}) - \xi y_k(\mathbf{x})} \le \phi \le M e^{-\xi \frac{1}{M} \sum_{j=1}^M \hat{y}_j(\mathbf{x}) + \xi^2 \frac{1}{M} \sum_{j=1}^M \hat{y}_j^2(\mathbf{x}) - \xi \sum_{m=1}^M W_m y_m(\mathbf{x}) + \xi^2 \sum_{m=1}^M W_m y_m^2(\mathbf{x})}. \tag{5}$$

Taking the logarithm on both sides of (5) and dividing by ξ, we obtain

$$\frac{1}{M}\sum_{m=1}^{M}\hat{y}_m(\mathbf{x}) + \sum_{m=1}^{M}\frac{e^{-\xi\hat{y}_m(\mathbf{x})}}{\sum_j e^{-\xi\hat{y}_j(\mathbf{x})}}y_m(\mathbf{x}) \leq \hat{y}_k(\mathbf{x}) + y_k(\mathbf{x}) + \xi\Big(\frac{1}{M}\sum_{m=1}^{M}\hat{y}_m^2(\mathbf{x}) + \sum_{m=1}^{M}\frac{e^{-\xi\hat{y}_m(\mathbf{x})}}{\sum_j e^{-\xi\hat{y}_j(\mathbf{x})}}y_m^2(\mathbf{x})\Big) + \frac{\log M}{\xi}.$$

□

Proof of Theorem 2.2

Proof. We first review the definition of sub-Gaussian random variables.

Definition 1 (Sub-Gaussian random variable)**.** *A random variable $y \in \mathbb{R}$ with mean $\mu_y \triangleq \mathbb{E}(y)$ is sub-Gaussian if there exists some positive constant C such that the tail of y satisfies:*

$$\mathbb{P}(|y-\mu_y| \geq t) \leq 2\exp(-t^2/(2C^2)), \forall t \geq 0. \tag{6}$$

The smallest constant $\sqrt{2}C$ satisfying (6) is called the sub-Gaussian norm, or the ψ_2-norm of y, denoted as $\|\|y\|\|_{\psi_2}$.

By the sub-Gaussian assumption we have:

$$\begin{aligned}\mathbb{P}\Big(\sum_k \frac{e^{-\xi\hat{y}_k(\mathbf{x})}}{\sum_j e^{-\xi\hat{y}_j(\mathbf{x})}}\hat{y}_k(\mathbf{x}) > x_{\text{co}} - T(\mathbf{x})\Big) &\leq \mathbb{P}\Big(\max_k \hat{y}_k(\mathbf{x}) > x_{\text{co}} - T(\mathbf{x})\Big)\\ &= \mathbb{P}\Big(\bigcup_k \big\{\hat{y}_k(\mathbf{x}) > x_{\text{co}} - T(\mathbf{x})\big\}\Big)\\ &\leq \sum_k \mathbb{P}\big(\hat{y}_k(\mathbf{x}) > x_{\text{co}} - T(\mathbf{x})\big)\\ &\leq \sum_k \exp\Big(-\frac{\big(x_{\text{co}} - T(\mathbf{x}) - \mu_{\hat{y}_k}(\mathbf{x})\big)^2}{2C_k^2(\mathbf{x})}\Big),\end{aligned} \tag{7}$$

where $\mu_{\hat{y}_k}(\mathbf{x}) = \mathbb{E}[\hat{y}_k(\mathbf{x})|\mathbf{x}]$. Note that the probability in (7) is taken with respect to the measure of the training samples. We essentially want to find the largest threshold $T(\mathbf{x})$ such that the probability of the expected improvement being less than $T(\mathbf{x})$ is small. Given a small $0 < \bar{\epsilon} < 1$, if

$$\mathbb{P}\Big(\sum_k \frac{e^{-\xi\hat{y}_k(\mathbf{x})}}{\sum_j e^{-\xi\hat{y}_j(\mathbf{x})}}\hat{y}_k(\mathbf{x}) > x_{\text{co}} - T(\mathbf{x})\Big) \leq \bar{\epsilon},$$

then from (7), we get:

$$\sum_k \exp\Big(-\frac{\big(x_{\text{co}} - T(\mathbf{x}) - \mu_{\hat{y}_k}(\mathbf{x})\big)^2}{2C_k^2(\mathbf{x})}\Big) \leq \bar{\epsilon}. \tag{8}$$

A relaxation of (8) implies that:

$$\exp\Big(-\frac{\big(x_{\text{co}} - T(\mathbf{x}) - \mu_{\hat{y}_m}(\mathbf{x})\big)^2}{2C_m^2(\mathbf{x})}\Big) \leq \frac{\bar{\epsilon}}{M},\ \forall m,$$

which yields that,

$$T(\mathbf{x}) \leq x_{\text{co}} - \mu_{\hat{y}_m}(\mathbf{x}) - \sqrt{-2C_m^2(\mathbf{x})\log(\bar{\epsilon}/M)},\ \forall m.$$

Given that $T(\mathbf{x})$ is non-negative, we take:

$$T(\mathbf{x}) = \max\Big(0,\ \min_m\Big(x_{\text{co}} - \mu_{\hat{y}_m}(\mathbf{x}) - \sqrt{-2C_m^2(\mathbf{x})\log(\bar{\epsilon}/M)}\Big)\Big).$$

The parameters $\mu_{\hat{y}_m}(\mathbf{x})$ and $C_m(\mathbf{x})$ for $m = 1, \ldots, M$ can be estimated through Algorithm 1.

If using the deterministic policy ($\xi \to \infty$),

$$\begin{aligned}\mathbb{P}(\min_m \hat{y}_m(\mathbf{x}) > x_{\text{co}} - T(\mathbf{x})) &= \mathbb{P}\Big(\bigcap_m \big\{\hat{y}_m(\mathbf{x}) > x_{\text{co}} - T(\mathbf{x})\big\}\Big) \\ &\le \mathbb{P}(\hat{y}_m(\mathbf{x}) > x_{\text{co}} - T(\mathbf{x})) \\ &\le \exp\Big(-\frac{\big(x_{\text{co}} - T(\mathbf{x}) - \mu_{\hat{y}_m}(\mathbf{x})\big)^2}{2C_m^2(\mathbf{x})}\Big), \forall m.\end{aligned} \tag{9}$$

Similarly, to make

$$\mathbb{P}\big(\min_m \hat{y}_m(\mathbf{x}) > x_{\text{co}} - T(\mathbf{x})\big) \le \bar{\epsilon},$$

we take

$$T(\mathbf{x}) = \max\bigg(0,\ \min_m\Big(x_{\text{co}} - \mu_{\hat{y}_m}(\mathbf{x}) - \sqrt{-2C_m^2(\mathbf{x})\log\bar{\epsilon}}\Big)\bigg).$$

□

Algorithm 1 Estimating the conditional mean $\mu_{\hat{y}_m}(\mathbf{x})$ and standard deviation $C_m(\mathbf{x})$ of the predicted outcome.

1: **Input:** a feature vector $\mathbf{x}$; a_m: the number of subsamples used to compute $\hat{\boldsymbol{\beta}}_m$, $a_m < N_m$; d_m: the number of repetitions.
2: **for** $i = 1, \ldots, d_m$ **do**
3: Randomly pick a_m samples from prescription group m, and use them to estimate a robust regression coefficient $\hat{\boldsymbol{\beta}}_{m_i}$ through solving (1).
4: The future outcome for $\mathbf{x}$ under prescription m is predicted as $\hat{y}_{m_i}(\mathbf{x}) = \mathbf{x}'\hat{\boldsymbol{\beta}}_{m_i}$.
5: **end for**
6: Estimate the conditional mean of $\hat{y}_m(\mathbf{x})$ as:

$$\mu_{\hat{y}_m}(\mathbf{x}) = \frac{1}{d_m}\sum_{i=1}^{d_m}\hat{y}_{m_i}(\mathbf{x}),$$

and the conditional standard deviation as:

$$C_m(\mathbf{x}) = \sqrt{\frac{1}{d_m - 1}\sum_{i=1}^{d_m}\Big(\hat{y}_{m_i}(\mathbf{x}) - \mu_{\hat{y}_m}(\mathbf{x})\Big)^2}.$$

Predictive Performance of Various Models We use four metrics to evaluate the predictive power of various models on the test set:

- The R-square:

$$\text{R}^2(\mathbf{y}, \hat{\mathbf{y}}) = 1 - \frac{\sum_{i=1}^{N_t}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_t}(y_i - \bar{y})^2},$$

  where $\mathbf{y} = (y_1, \ldots, y_{N_t})$ and $\hat{\mathbf{y}} = (\hat{y}_1, \ldots, \hat{y}_{N_t})$ are the vectors of the true (observed) and predicted outcomes, respectively, with N_t the size of the test set, and $\bar{y} = (1/N_t)\sum_{i=1}^{N_t} y_i$.
- The *Mean Squared Error (MSE)*:

$$\text{MSE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N_t}\sum_{i=1}^{N_t}(y_i - \hat{y}_i)^2.$$

- The *Mean Absolute Error (MeanAE)* that is more robust to large deviations than the MSE in that the absolute value function increases more slowly than the square function over large (absolute) values of the argument.

$$\text{MeanAE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N_t}\sum_{i=1}^{N_t}|y_i - \hat{y}_i|.$$

- The MedianAE which can be viewed as a robust measure of the MeanAE, computing the median of the absolute deviations:

$$\text{MedianAE}(\mathbf{y}, \hat{\mathbf{y}}) = \text{Median}\left(|y_i - \hat{y}_i|, i = 1, \ldots, N_t\right).$$

The out-of-sample performance metrics of the various models on the two datasets are shown in Tables 2 and 3, where the numbers in the parentheses show the improvement of RLAD+K-NN compared against other methods. Huber refers to the robust regression method proposed in [8, 9], and CART refers to the *Classification And Regression Trees*. Huber/OLS/LASSO + K-NN means fitting a K-NN regression model with a Huber/OLS/LASSO-weighted distance metric. We note that in order to produce well-defined and meaningful predictive performance metrics, the datasets used to generate Tables 2 and 3 did not group the patients by their prescriptions. A universal model was fit to all patients using the prescription as one of the predictors. Nevertheless, it would still be considered as a fair comparison as all models were evaluated on the same dataset. The results provide supporting evidence for the validity of our RLAD+K-NN model that outperforms all others in all metrics, and is thus used for predicting the outcomes of counterfactual treatments.

Table 2: Performance of different models for predicting future HbA_{1c} for diabetic patients.

Methods	R^2	MSE	MeanAE	MedianAE
OLS	0.52 (2%)	1.36 (2%)	0.81 (4%)	0.55 (11%)
LASSO	0.52 (2%)	1.37 (2%)	0.80 (3%)	0.54 (9%)
Huber	0.36 (47%)	1.81 (26%)	0.96 (19%)	0.70 (30%)
RLAD	0.50 (4%)	1.40 (4%)	0.78 (1%)	0.50 (1%)
K-NN	0.25 (109%)	2.11 (37%)	1.07 (27%)	0.81 (39%)
OLS+K-NN	0.52 (0%)	1.34 (0%)	0.79 (1%)	0.51 (3%)
LASSO+K-NN	0.52 (1%)	1.36 (1%)	0.79 (1%)	0.50 (1%)
Huber+K-NN	0.51 (3%)	1.38 (3%)	0.81 (3%)	0.53 (7%)
RLAD+K-NN	0.52 (N/A)	1.34 (N/A)	0.78 (N/A)	0.49 (N/A)
CART	0.49 (7%)	1.43 (7%)	0.81 (3%)	0.50 (2%)

Table 3: Performance of different models for predicting future systolic blood pressure for hypertension patients.

Methods	R^2	MSE	MeanAE	MedianAE
OLS	0.31 (14%)	170.80 (6%)	10.09 (7%)	8.15 (9%)
LASSO	0.31 (14%)	170.83 (6%)	10.08 (7%)	8.22 (10%)
Huber	0.22 (62%)	193.54 (17%)	10.70 (12%)	8.61 (14%)
RLAD	0.30 (18%)	173.32 (8%)	10.11 (7%)	8.28 (11%)
K-NN	0.33 (10%)	167.41 (5%)	9.62 (2%)	7.50 (2%)
OLS+K-NN	0.35 (1%)	160.22 (0%)	9.42 (0%)	7.49 (1%)
LASSO+K-NN	0.32 (12%)	169.50 (6%)	9.74 (3%)	7.73 (5%)
Huber+K-NN	0.32 (10%)	167.92 (5%)	9.71 (3%)	7.84 (6%)
RLAD+K-NN	0.36 (N/A)	159.74 (N/A)	9.42 (N/A)	7.38 (N/A)
CART	0.25 (43%)	186.23 (14%)	10.34 (9%)	8.22 (10%)

Oral prescriptions for diabetes. Acarbose, acetohexamide, chlorpropamide, glimepiride, glipizide, glyburide, hydrochloride, metformin, miglitol, nateglinide, pioglitazone, repaglinide, rosiglitazone, sitagliptin, tolazamide, tolbutamide, and troglitazone.

Table 4: Feature importance from the RLAD model for the diabetes dataset.

Features	Regression coefficients
lab test: HbA_{1c}	1.02
lab test: blood glucose	0.27
measurement: systolic blood pressure	0.07
prescription: injectable	0.05
lab test: hematocrit	0.04
lab test: hemoglobin	-0.03
lab test: sodium	0.03
lab test: platelet count	0.03
diagnosis: retinal disorders	0.02
lab test: leukocyte count	0.02
diagnosis: inflammatory and toxic neuropathy	0.02
lab test: mean corpuscular volume	-0.02
lab test: calcium	0.01
lab test: potassium	0.01
diagnosis: disorders of lipoid metabolism	0.01
diagnosis: other disorders of soft tissues	0.01
race: Caucasian	-0.01
diagnosis: obesity	-0.01
age	-0.01
diagnosis: essential hypertension	0.01

Table 5: Feature importance from the RLAD model for the hypertension dataset.

Features	Regression coefficients
measurement: systolic blood pressure	7.62
age	1.87
lab test: sodium	1.29
lab test: hemoglobin	1.26
prescription: calcium channel blockers	0.98
lab test: blood glucose	0.93
lab test: hematocrit	-0.82
sex: female	0.76
lab:mean corpuscular volume	-0.61
diagnosis: asthma	-0.61
prescription: ARB	0.57
diagnosis: cataract	0.57
diagnosis: chronic ischemic heart disease	-0.56
lab test: potassium	0.55
diagnosis: heart failure	-0.53
prescription: diuretics	0.53
diagnosis: cardiac dysrhythmias	-0.51
diagnosis obesity	0.46
race: Caucasian	-0.46
diagnosis: disorders of fluid electrolyte and acid-base balance	0.45