Proceedings of Machine Learning Research vol 291:1–2, 2025 38th Annual Conference on Learning Theory

Fast and Multiphase Rates for Nearest Neighbor Classifiers

Pengkun Yang YANGPENGKUN@TSINGHUA.EDU.CN
Department of Statistics and Data Science, Tsinghua University
Jingzhao Zhang JINGZHAOZ@MAIL.TSINGHUA.EDU.CN
IIIS, Tsinghua University and Shanghai Qi Zhi Institute

Editors: Nika Haghtalab and Ankur Moitra

Abstract

We study the scaling of classification error rates with respect to the size of the training dataset. In contrast to classical results where rates are minimax optimal for a problem class, this work starts with the empirical observation that, even for a fixed data distribution, the error scaling can have *diverse* rates across different ranges of sample size. To understand when and why the error rate is non-uniform, we provide a fine-grained framework for analyzing nearest neighbor classifiers. With the proposed analysis, we achieve the following key results.

1. **Instance learnability and general error rates.** By characterizing the distribution of the instance learnability determined by the data distribution, we provide an error rate that can have diverse convergence rates across different ranges of sample size n.
2. **Optimal error rates for the logistic problem with a norm design.** With the above tools, we first analyze the standard logistic regression model, demonstrating that k-NN classifiers, although agnostic to the underlying linear structure, can achieve the parametric rate
$$\mathbb{E}[R(\hat{f}^{\mathsf{NN}}_{n,k})] - R^* = \mathcal{O}(d/n),$$
which matches the minimax lower bound up to constant factors Hsu and Mazumdar (2024); Kuchelmeister and van de Geer (2024).
3. **Provable two-phase rates for a modified logistic problem.** In the *second* application, we show that a slight variation of the logistic regression model can produce multiphase rates. Specifically, we rotate the first two coordinates of the normal logistic regression model, such that the optimal decision boundary is still smooth and linear. Then we show in both the experiments and theorems, that the error rate initially follows the parametric rate before eventually slowing to a nonparametric rate. In an additional theorem, we prove that this final phase is unavoidably slower than any polynomial rate.

We note that benign conditions for k-NN has been long studied Györfi et al. (2002); Tsybakov (2004); Massart and Nédélec (2006); Kpotufe (2011); Samworth (2012); Chaudhuri and Dasgupta (2014); Gottlieb et al. (2016); Gadat et al. (2016); Xue and Kpotufe (2018); Ashlagi et al. (2021); Györfi and Weiss (2021); Hanneke et al. (2023). Under the additional smoothness β and noise condition α, the convergence rate can be improved to $\mathcal{O}\left(n^{-\frac{\beta(1+\alpha)}{2\beta+d}}\right)$ or even $\mathcal{O}\left(\exp\left(-nc^{\frac{\beta(1+\alpha)}{2\beta+d}}\right)\right)$ for $c \in (0,1)$. However, this does not close the gap between parametric and nonparametric rates. For this logistic regression model, we observe that $\alpha = 1$ and $\beta = 1$. This results in an sample complexity *exponentially* large in d for achieving any nontrivial error $\epsilon < 0.5$.

In contrast, as shown in our experiments and theorems, nearest neighbor classifiers can achieve near optimal error with *polynomial* sample size. We approach the benign condition in an orthogonal direction and highlight the complexity of instance-wise data distribution in determining the test error. In this way, we show that the sample complexity of nearest neighbor classifier can depend polynomially on d for standard logistic regression models. [1]

Keywords: Generalization error, scaling laws, nearest neighbor classifiers

1. Extended abstract. Full version appears as arXiv 2308.08247 v2

References

Yair Ashlagi, Lee-Ad Gottlieb, and Aryeh Kontorovich. Functions with average smoothness: structure, algorithms, and learning. In *Conference on Learning Theory*, pages 186–236. PMLR, 2021.

Kamalika Chaudhuri and Sanjoy Dasgupta. Rates of convergence for nearest neighbor classification. *Advances in Neural Information Processing Systems*, 27, 2014.

Sébastien Gadat, Thierry Klein, and Clément Marteau. Classification in general finite dimensional spaces with the k-nearest neighbor rule. *The Annals of Statistics*, 44(3):982–1009, 2016.

Lee-Ad Gottlieb, Aryeh Kontorovich, and Pinhas Nisnevitch. Nearly optimal classification for semimetrics. In *Artificial Intelligence and Statistics*, pages 379–388. PMLR, 2016.

László Györfi and Roi Weiss. Universal consistency and rates of convergence of multiclass prototype algorithms in metric spaces. *J. Mach. Learn. Res.*, 22:151–1, 2021.

László Györfi, Michael Kohler, Adam Krzyzak, Harro Walk, et al. *A distribution-free theory of nonparametric regression*, volume 1. Springer, 2002.

Steve Hanneke, Aryeh Kontorovich, and Guy Kornowski. Near-optimal learning with average h\" older smoothness. *arXiv preprint arXiv:2302.06005*, 2023.

Daniel Hsu and Arya Mazumdar. On the sample complexity of parameter estimation in logistic regression with normal design. In *Conference on learning theory*, 2024.

Samory Kpotufe. k-NN regression adapts to local intrinsic dimension. *Advances in neural information processing systems*, 24, 2011.

Felix Kuchelmeister and Sara van de Geer. Finite sample rates for logistic regression with small noise or few samples. *Sankhya A*, pages 1–70, 2024.

Pascal Massart and Élodie Nédélec. Risk bounds for statistical learning. *The Annals of Statistics*, 34 (5):2326 – 2366, 2006. doi: 10.1214/009053606000000786. URL https://doi.org/10.1214/009053606000000786.

Richard J Samworth. Optimal weighted nearest neighbour classifiers. *The Annals of Statistics*, 40 (5):2733–2763, 2012.

Alexander B Tsybakov. Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics*, 32(1):135–166, 2004.

Lirong Xue and Samory Kpotufe. Achieving the time of 1-NN, but the accuracy of k-NN. In *International Conference on Artificial Intelligence and Statistics*, pages 1628–1636. PMLR, 2018.