

Summary

Adaptive Weighted Nearest Neighbor (**AWNN**) is an adaptive unsupervised algorithm for density estimation. Motivated by the bias-variance decomposition for the pointwise error of the weighted nearest neighbors for density estimation, we provide an efficient optimization approach to select the weights of nearest neighbors for each instance adaptively from the data. From a theoretical perspective, we establish the convergence rates of **AWNN** in terms of the L_p -norm when the marginal distribution P has unbounded support, which coincides with the minimax lower bound. Moreover, we for the first time manage to explain the benefits of the adaptive method in density estimation. To be specific, we find that to achieve the optimal convergence rate, the condition that **AWNN** requires is weaker than that of the standard weighted k -NN density estimator. In addition, we show that the convergence rate of **AWNN** is no worse than that of the standard weighted k -NN density estimator. In the experiments, we verify the theoretical findings and show the superiority of **AWNN** through both synthetic and real-world data experiments.

Statement of need

A vast literature has focused on density estimation. For example, the most intuitive approach, histogram methods find partitions of input space and estimate density with bins (López-Rubio 2013). Although histogram density estimation enjoys sound theoretical properties, it suffers from low efficiency and boundary discontinuity. Besides, it is sensitive to the choice of bin width. To overcome these issues, the kernel density estimation was proposed in (Rosenblatt 1956; Parzen 1962) with theoretical properties that are well explored. However, when encountering density distributions with varying local properties, the kernel density estimation will have a poor performance. To conquer the weakness, nearest neighbors-based methods for density estimation, proposed in (Loftsgaarden and Quesenberry 1965), were investigated in Biau and Devroye (2015) and successfully applied to many machine learning tasks, like density-based clustering or anomaly detection, see, e.g., (wu2019fast; Gu, Akoglu, and Rinaldo 2019; Zhang et al. 2021).

There are several advantages of k -NN density estimation compared with other methods. First of all, k -NN is a lazy learning method that requires no training stage and has attractive testing stage sample complexity. Moreover, the smoothing of k -NN varies according to the number of observations in a particular region, which can be regarded as a variant of the kernel density estimation with the local choice of the bandwidth (Orava 2011). Furthermore, mild conditions are required about the underlying distribution to provide a convergence guarantee because of k -NN's non-parametric instinct. We refer the reader to (Biau and Devroye 2015) for more discussions. Recently, (Papernot and McDaniel 2018; Göpfert, Wersing, and Hammer 2022) further pointed out that the straightforward logic of k -NN naturally satisfies the requirements of trustworthy AI.

Methods

Before we proceed, we need to introduce some basic notations. For any $x \in \mathbb{R}^d$, we denote $X_{(k)}(x) := X_{(k)}(x; D_n)$ as the k -th nearest neighbor of x in D_n . Then we denote $R_k(x) := R_k(x; D_n)$ as the distance between x and $X_{(k)}(x; D)$, termed as the k -distance of x in D_n . Given n independent identically distributed data $D_n \in \mathbb{R}^d$. For each point $x \in \mathbb{R}^d$, the k -NN density estimator (Loftsgaarden and Quesenberry 1965; Dasgupta and Kpotufe 2014), is defined as follows.

$$f_k(x) = \frac{k/n}{V_d R_k^d(x)},$$

where V_d be the volume of the unit ball in \mathbb{R}^d . An intuitive explanation of this estimator is that from order statistics (David and Nagaraja 2004). We know that $P(B(x, R_k(x)))$ follows Beta Binomial distribution $Beta(k, n-k+1)$. As a result, we have $\mathbb{E}[P(B(x, R_k(x)))] = k/(n+1)$. Therefore, we have the approximation $k/n \approx P(B(x, R_k(x))) \approx f(x)V_d R_k^d(x)$. This yields the standard k -NN estimator (1).

Reference

- Biau, Gérard, Frédéric Chazal, David Cohen-Steiner, Luc Devroye, and Carlos Rodriguez. 2011. “A Weighted k-Nearest Neighbor Density Estimate for Geometric Inference.” *Electronic Journal of Statistics* 5: 204–37.
- Biau, Gérard, and Luc Devroye. 2015. *Lectures on the Nearest Neighbor Method*. Vol. 246. Springer.
- Dasgupta, Sanjoy, and Samory Kpotufe. 2014. “Optimal Rates for k-Nn Density and Mode Estimation.” *Advances in Neural Information Processing Systems* 27.
- David, Herbert A, and Haikady N Nagaraja. 2004. *Order Statistics*. John Wiley & Sons.
- Göpfert, Jan Philip, Heiko Wersing, and Barbara Hammer. 2022. “Interpretable Locally Adaptive Nearest Neighbors.” *Neurocomputing* 470: 344–51.
- Gu, Xiaoyi, Leman Akoglu, and Alessandro Rinaldo. 2019. “Statistical Analysis of Nearest Neighbor Methods for Anomaly Detection.” *Advances in Neural Information Processing Systems* 32.
- Loftsgaarden, Don O, and Charles P Quesenberry. 1965. “A Nonparametric Estimate of a Multivariate Density Function.” *The Annals of Mathematical Statistics* 36 (3): 1049–51.
- López-Rubio, Ezequiel. 2013. “A Histogram Transform for Probabilitydensity Function Estimation.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36 (4): 644–56.
- Orava, Jan. 2011. “K-Nearest Neighbour Kernel Density Estimation, the Choice of Optimal k.” *Tatra Mountains Mathematical Publications* 50 (1): 39–50.
- Papernot, Nicolas, and Patrick McDaniel. 2018. “Deep k-Nearest Neighbors: Towards Confident, Interpretable and Robust Deep Learning.” *arXiv Preprint arXiv:1803.04765*.
- Parzen, Emanuel. 1962. “On Estimation of a Probability Density Function and Mode.” *The Annals of Mathematical Statistics* 33 (3): 1065–76.
- Rosenblatt, Murray. 1956. “Remarks on Some Nonparametric Estimates of a Density Function.” *The Annals of Mathematical Statistics*, 832–37.
- Zhang, Rui, Tao Du, Shouning Qu, and Hongwei Sun. 2021. “Adaptive Density-Based Clustering Algorithm with Shared KNN Conflict Game.” *Information Sciences* 565: 344–69.