entropy

Article

Probabilistic Nearest Neighbors Classification

Bruno Fava [1], Paulo C. Marques F. [2,*] and Hedibert F. Lopes [2]

1 Department of Economics, Northwestern University, Evanston, IL 60208, USA; brunofava@u.northwestern.edu
2 Insper Institute of Education and Research, Rua Quatá 300, São Paulo 04546-042, Brazil; hedibertfl@insper.edu.br
* Correspondence: paulocmf1@insper.edu.br

Abstract: Analysis of the currently established Bayesian nearest neighbors classification model points to a connection between the computation of its normalizing constant and issues of NP-completeness. An alternative predictive model constructed by aggregating the predictive distributions of simpler nonlocal models is proposed, and analytic expressions for the normalizing constants of these nonlocal models are derived, ensuring polynomial time computation without approximations. Experiments with synthetic and real datasets showcase the predictive performance of the proposed predictive model.

Keywords: probabilistic machine learning; nearest neighbors classification; NP-completeness

check for updates

Citation: Fava, B.; F., P.C.F.; Lopes, H.F. Probabilistic Nearest Neighbors Classification. *Entropy* **2024**, *26*, 39. https://doi.org/10.3390/e26010039

Academic Editors: Carlos Alberto De Bragança Pereira, Paulo Canas Rodrigues and Mark Andrew Gannon

Received: 29 November 2023
Revised: 23 December 2023
Accepted: 28 December 2023
Published: 30 December 2023

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

1. Introduction

The now classic nearest neighbors classification algorithm, introduced in a 1951 technical report by Fix and Hodges (reprinted in [1]), marked one of the early successes of machine learning research. The basic idea is that, given some notion of proximity between pairs of observations, the class of a new sample unit is determined by majority voting among its k nearest neighbors in the training sample [2,3]. A natural question is whether it is possible to develop a probabilistic model that captures the essence of the mechanism contained in the classic nearest neighbors algorithm while adding proper uncertainty quantification of predictions made by the model. In a pioneering paper, Holmes and Adams [4] defined a probabilistic nearest neighbors model specifying a set of conditional distributions. A few years later, Cucala et al. [5] pointed out the incompatibility of the conditional distributions specified by Holmes and Adams, which do not define a proper joint model distribution. As an alternative, Cucala et al. developed their own nearest neighbors classification model, defining directly a proper, Boltzmann-like joint distribution. A major difficulty with the Cucala et al. model is the fact that its likelihood function involves a seemingly intractable normalizing constant. Consequently, in order to perform a Bayesian analysis of their model, the authors engaged in a *tour de force* of Monte Carlo techniques, with varied computational complexity and approximation quality.

In this paper, we introduce an alternative probabilistic nearest neighbors predictive model constructed from an aggregation of simpler models whose normalizing constants can be exactly summed in polynomial time. We begin by reviewing the Cucala et al. model in Section 2, showing by an elementary argument that the computational complexity of the exact summation of its normalizing constant is directly tied to the concept of NP-completeness [6]. The necessary concepts from the theory of computational complexity are briefly reviewed. Section 3 introduces a family of nonlocal models, whose joint distributions take into account only the interactions between each sample unit and its r-th nearest neighbor. For each nonlocal model, we derive an analytic expression for its normalizing constant, which can be computed exactly in polynomial time. The nonlocal models are combined in Section 4, yielding a predictive distribution that does not rely on costly Monte