

STATISTICAL ANALYSIS AND DATA MINING

Research Article

Bayesian adaptive nearest neighbor

Ruixin Guo , Sounak Chakraborty 

First published: 24 February 2010

<https://doi.org/10.1002/sam.10067>

Cited by: 1



About



Access

als



ire

Abstract

The k nearest neighbor classification (k - NN) is a very simple and popular method for classification. However, it suffers from a major drawback, it assumes constant local class posterior probability. It is also highly dependent on and sensitive to the choice of the number of neighbors k . In addition, it severely lacks the desired probabilistic formulation. In this article, we propose a Bayesian adaptive nearest neighbor method (BANN) that can adaptively select the shape of the neighborhood and the number of neighbors k . The shape of the neighborhood is automatically selected according to the concentration of the data around each query point with the help of discriminants. The neighborhood size is not predetermined and is kept free using a prior distribution. Thus, we are able to make the model to select the appropriate neighborhood size. The model is fitted using Markov Chain Monte Carlo (MCMC), so we are not using exactly one neighborhood size but a mixture of k . Our BANN model is highly flexible, determining any local pattern in the data - generating process, and adapting it to give an improved prediction. We have applied our model on four simulated data sets with special structures and five real - life benchmark data sets. Our proposed BANN method demonstrates substantial improvement over k - NN and discriminant adaptive nearest neighbor (DANN) in all nine case studies. It also outperforms the probabilistic nearest neighbor (PNN) in most of the data analyses. Copyright © 2010 Wiley Periodicals, Inc. Statistical Analysis and Data Mining 3: 92 - 105, 2010

Citing Literature



About Wiley Online Library



Help & Support



Opportunities



Connect with Wiley



Copyright © 1999-2018 John Wiley & Sons, Inc. All rights reserved

WILEY