Abstract

K-nearest neighbor (KNN) is one of the accepted classification tools. Classification is one of the foremost machine-learning tools used in the field of medical data mining. However, one of the most complicated tasks in developing a KNN is determining the optimal number of nearest neighbors, which is usually obtained by repeated experiments for different values of K, till the minimum error rate is achieved. This paper describes the novel approach of finding optimal number of nearest neighbors for KNN classifier by combining Akaike’s information criterion (AIC) and the golden-section search technique. The optimal model developed was used for categorization of a variety of medical data garnered from UC Irvine Machine Learning Repository.

Reference

Combining Akaike’s Information Criterion (AIC) and the Golden-Section Search Technique to find Optimal Numbers of K-Nearest Neighbors

Second Intl Symp Inf Theory, 267-81.

Index Terms

Computer Science
Data Mining

Key words

Medical Data mining
K-nearest neighbor (KNN)
Akaike’s information criterion (AIC)
Golden-selection Ratio