Abstract

Cancer diagnosis is one of the most studied problems in the medical domain. Several researchers have focused in order to improve performance and achieve to obtain satisfactory results. Breast cancer is one of cancer killer in the world. The diagnosis of this cancer is a big problem in cancer diagnosis researches. In artificial intelligence, machine learning is a discipline which allows to the machine to evolve through a process. Machine learning is widely used in bioinformatics and particularly in breast cancer diagnosis. One of the most popular methods is K-nearest neighbors (K-NN) which is a supervised learning method. Using the K-NN in medical diagnosis is very interesting. The quality of the results depends largely on the distance and the value of the parameter "k" which represent the number of the nearest neighbors. In this paper, we study and evaluate the performance of different distances that can be used in the K-NN algorithm. Also, we analyze this distance by using different values of the parameter "k" and by using several rules of classification (the rule used to decide how to classify a sample). Our work will be performed on the WBCD database (Wisconsin Breast Cancer Database) obtained by the university of Wisconsin Hospital.
References

- P. Shi, S. Ray, Q. Zhu, and M. A Kon. Top scoring pairs for feature selection in
machine learning and applications to cancer outcome prediction. BMC Bioinformatics, 12, 2011.

Index Terms

Computer Science
Artificial Intelligence

Keywords
Classification Diagnosis Breast Cancer K-Nearest Neighbors Distance Classification Rule