Abstract

The aim of this paper is to discuss about various feature selection algorithms applied on different datasets to select the relevant features to classify data into binary and multi class in order to improve the accuracy of the classifier. Recent researches in medical diagnose uses the different kind of classification algorithms to diagnose the disease. For predicting the disease, the classification algorithm produces the result as binary class. When there is a multiclass dataset, the classification algorithm reduces the dataset into a binary class for simplification purpose by using any one of the data reduction methods and the algorithm is applied for prediction. When data reduction on original dataset is carried out, the quality of the data may degrade and the accuracy of an algorithm will get affected. To maintain the effectiveness of the data, the multiclass data must be treated with its original form without maximum reduction, and the algorithm can be applied on the dataset for producing maximum accuracy. Dataset with maximum number of attributes like thousands must incorporate the best feature selection algorithm for selecting the relevant features to reduce the space and time complexity. The performance of Classification algorithm is estimated by how accurately it predicts the individual class on particular dataset. The accuracy constrain mainly depends on the selection of appropriate features from the original dataset. The feature selection algorithms play an important role in classification for better performance. The feature selection is one of
the preprocessing techniques in the classification. This research paper deals with different feature selection algorithms and their performance on different dataset.

References

Analysis of Feature Selection Algorithms on Classification: A Survey


- Der-Chiang Li, Chiao-Wen Liu, Susan C. Hu, "A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets," Artificial Intelligence in Medicine, Elsevier, 2011.

Index Terms

Computer Science

Data Mining
### Keywords

Classification  Binary class  Multiclass  feature selection algorithm  Medical high dimensional dataset.