

Influence of Training Sample Preprocessing in Generalization Accuracy of Multilayer Perceptron

Eduardo Gasca and Ricardo Barandela
Lab for Pattern Recognition
Instituto Tecnológico de Toluca, México
e-mail: egasca@it-toluca.edu.mx
rbarandela@hotmail.com

Abstract

The training sample is an important issue in the learning process of the multilayer perceptron neuronal network. For this reason at the present work the behavior of multilayer perceptron (back propagation algorithm) generalization accuracy using different pre-processing methods of training sample was investigated. In the experiments, diverse techniques were used. These were separated in two groups: the first one contains those that select a subset of the original sample, the second clusters techniques whose starting point is a group of codebook prototypes. The tests were carried out with real and artificial data, corresponding to different types of problems. Experimental results show that the combination of both types of procedures gives, in most cases, the best behavior. That is, when it executes an initial filtering with methods of the first group, and later a technique of the second group is applied.