A New Variant of the Optimum-Path Forest Classifier

João P. Papa and Alexandre X. Falcão

Institute of Computing, State University of Campinas,
Av. Albert Einstein 1216, Campinas, São Paulo, Brazil
{papa.joaopaulo,alexandre.falcao}@gmail.com

Abstract. We have shown a supervised approach for pattern classification, which interprets the training samples as nodes of a complete arc-weighted graph and computes an optimum-path forest rooted at some of the closest samples between distinct classes. A new sample is classified by the label of the root which offers to it the optimum path. We propose a variant, in which the training samples are the nodes of a graph, whose the arcs are the k-nearest neighbors in the feature space. The graph is weighted on the nodes by their probability density values (pdf) and the optimum-path forest is rooted at the maxima of the pdf. The best value of k is computed by the maximum accuracy of classification in the training set. A test sample is assigned to the class of the maximum, which offers to it the optimum path. Preliminary results have shown that the proposed approach can outperform the previous one and the SVM classifier in some datasets.

1 Introduction

Pattern recognition techniques aim to find decision rules, which can separate samples from distinct classes. The methods can be divided into three categories, *unsupervised, supervised, and semi-supervised*, according to the knowledge about the labels (classes) of the samples in a given training set. Unsupervised approaches have no prior knowledge about the labels, while supervised techniques have fully information about them. Semi-supervised methods use both labeled and unlabeled samples for training. A dataset is usually divided in two parts, a training set and a test set, being the first used to project the classifier and the second used for validation, by measuring its classification errors (accuracy). This process must be also repeated several times with randomly selected training and test samples to achieve a conclusion about the statistics of its accuracy (robustness).

Several approaches for supervised classification have been proposed under certain assumptions about the distribution of the samples in the feature space. Simple techniques can deal with linearly separable classes (Figure 1a), such as the well known perceptrons. Piecewise linearly separable classes (Figure 1b) require more robust techniques, such as Artificial Neural Networks using Multilayer Perceptrons [1]. If the classes have some known shape, one can use Gaussian

G. Bebis et al. (Eds.): ISVC 2008, Part I, LNCS 5358, pp. 935–944, 2008.

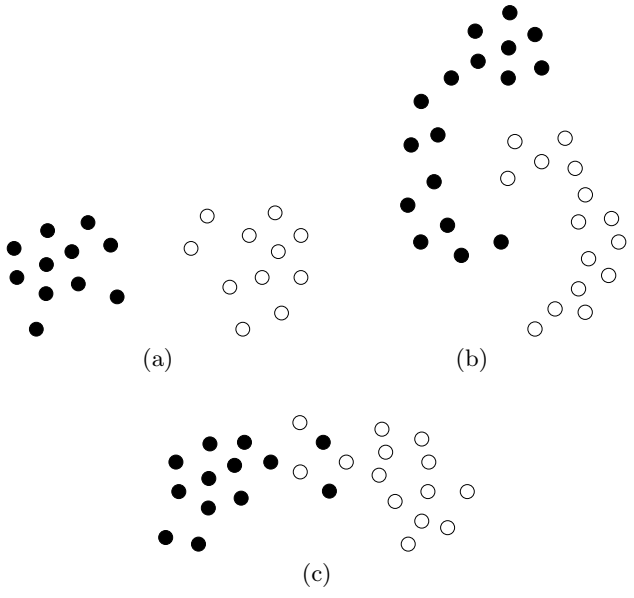

Fig. 1. Examples of some feature spaces: (a) Linearly separable. (b) Piecewise linearly separable. (c) Non separable.

Mixture Models [2], for instance. However, if we have non separable classes (Figure 1c), Support Vector Machines [3] (SVMs) can handle them by non-linearly mapping the samples into a higher-dimension feature space, in which the classes are assumed to be linearly separable.

Unfortunately, practical applications usually involve non separable classes and the assumption of the SVMs about the linear separability in a higher-dimension space does not hold very often. Other techniques try to handle the overlapping problem between classes by taking local decisions, based on the distances between nearby samples (e.g., the k-nearest neighbors [4]). However, as far as we know, the strength of connectivity between samples in the feature space seems to not have caught much attention in supervised classification, except by our previous work [5,6]. This method interprets a training set as a complete graph (i.e., the arcs connect all pairs of nodes), in which the arcs are weighted by the distance between the feature vectors of their corresponding nodes. A path in the graph is a sequence of nodes connecting two terminal samples, each path has a value given by a *path-value function* (e.g., the maximum arc weight along the path), and a path is optimum when its value is minimum. The "strength of connectedness" between two samples is then inversely proportional to the value of an optimum path between them. Prototypes from all classes are identified in the training set, among the closest samples between distinct classes. The prototypes compete with each other, such that each sample is assigned to its most strongly connected prototype, forming an optimum-path forest rooted at the prototypes (an optimal partition of the training set). The classification of a new sample