Discrete Applied Mathematics 160 (2012) 2329–2338

Contents lists available at SciVerse ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Analysis of a multi-category classifier

Martin Anthony [a,*], Joel Ratsaby [b]

[a] Department of Mathematics, The London School of Economics and Political Science, Houghton Street, London WC2A2AE, UK
[b] Electrical and Electronics Engineering Department, Ariel University Center of Samaria, Ariel 40700, Israel

ARTICLE INFO

Article history:
Received 15 February 2012
Received in revised form 5 July 2012
Accepted 8 July 2012
Available online 26 July 2012

Keywords:
Multi-category classification
Box clustering
Generalization error

ABSTRACT

The use of boxes for pattern classification has been widespread and is a fairly natural way in which to partition data into different classes or categories. In this paper we consider multi-category classifiers which are based on unions of boxes. The classification method studied may be described as follows: find boxes such that all points in the region enclosed by each box are assumed to belong to the same category, and then classify remaining points by considering their distances to these boxes, assigning to a point the category of the nearest box. This extends the simple method of classifying by unions of boxes by incorporating a natural way (based on proximity) of classifying points outside the boxes. We analyze the generalization accuracy of such classifiers and we obtain generalization error bounds that depend on a measure of how definitive is the classification of training points.

1. Box-based multi-category classifiers

Classification in which each category or class is a union of boxes is a long-studied and natural method for pattern classification. It is central, for instance, to the methods used for logical analysis of data (see, for example [9,10,15,5]) and has been more widely studied as a geometrical classifier (see [11], for instance). More recently, unions of boxes have been used in combination with a nearest-neighbor (or proximity) paradigm for binary classification [6] and multi-category classification [13], enabling meaningful classification for points of the domain that lie outside any of the boxes.

In this paper, we analyze multi-category classifiers of the type described by Felici et al. [13]. In that paper, they describe a set of classifiers based on boxes and nearest-neighbors, where the metric used for the nearest-neighbor measure is the Manhattan (or taxicab) metric. (We give explicit details shortly.) They use an agglomerative box-clustering method to produce a set of candidate classifiers of this type. They then select from these one that is, in a sense they define, optimal. First they focus on the classifiers which are, with respect to the two dimensions of the error on the sample, E, and complexity (number of boxes), B, Pareto-optimal. Among these they then select a classifier that minimizes an objective function of the form $(E - E_0)^2 + (B - B_0)^2$ (effecting a tradeoff between the error and complexity) and, if there is more than one such classifier, they choose that which minimizes E. They provide some experimental evidence that this approach works. Here, we obtain generalization error bounds for the box-based classifiers of the type considered in [13], within a version of the standard PAC model of probabilistic learning. The bounds we obtain depend on the error and complexity and they improve (that is, they decrease) the more 'definite' is the classification of the sample points.

Suppose points of $[0, 1]^n$ are to be classified into C classes, which we will assume are labeled $1, 2, \ldots, C$. We let $[C]$ denote the set $\{1, 2, \ldots, C\}$.

A *box* (or, more exactly, an axis-parallel box) in $\mathbb{R}^n$ is a set of the form

$$\mathbf{I}(u, v) = \{x \in \mathbb{R}^n : u_i \le x_i \le v_i,\ 1 \le i \le n\},$$

* Corresponding author. Tel.: +44 2079557623; fax: +44 2079556877.
E-mail addresses: m.anthony@lse.ac.uk (M. Anthony), ratsaby@ariel.ac.il (J. Ratsaby).

0166-218X/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2012.07.010

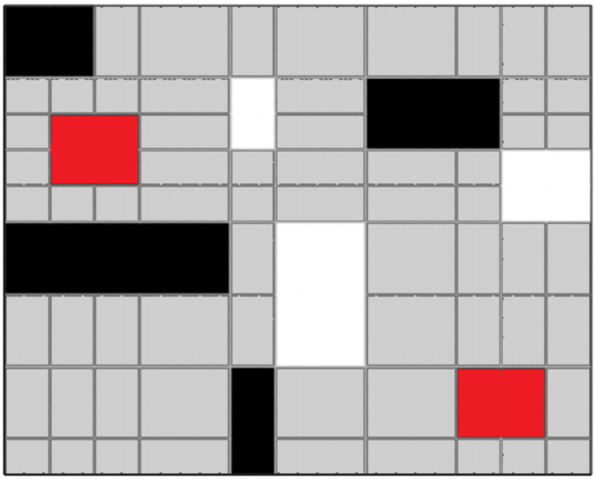

Fig. 1. Boxes of three categories.

where $u, v \in [0, 1]^n$ and $u \leq v$, meaning that $u_i \leq v_i$ for each i. We consider multi-category classifiers which are based on C unions of boxes, as we now describe. For $k = 1, \ldots, C$, suppose that S_k is a union of some number, B_k, of boxes:

$$S_k = \bigcup_{j=1}^{B_k} \mathbf{I}(u(k, j), v(k, j)).$$

Here, the jth box is defined by $u(k, j), v(k, j)$ where $u(k, j), v(k, j) \in [0, 1]^n$ and $u(k, j) \leq v(k, j)$ (so, for each i between 1 and n, $u(k, j)_i \leq v(k, j)_i$). We assume, further, that for $k \neq l$, $S_k \cap S_l = \emptyset$. We think of S_k as being a region of the domain all of whose points we assume to belong to class k. So, as in [13,15], for instance, the boxes in S_k might be constructed by 'agglomerative' box-clustering methods.

To define our classifiers, we will make use of a metric on $[0, 1]^n$. To be specific, as in [13], d will be the d_1 (or 'Manhattan' or 'taxicab') metric: for $x, y \in [0, 1]^n$,

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|.$$

We could equally well (as in [6], where the two-class case is the focus) use the supremum or d_∞ metric, defined by

$$d_\infty(x, y) = \max\{|x_i - y_i| : 1 \leq i \leq n\}$$

and similar results would be obtained. For $x \in [0, 1]^n$ and $S \subseteq [0, 1]^n$, the distance from x to S is

$$d(x, S) = \inf_{y \in S} d(x, y).$$

Let $\mathcal{S} = (S_1, S_2, \ldots, S_C)$ and denote by $h_\mathcal{S}$ the classifier from $[0, 1]^n$ into $[C]$ defined as follows: for $x \in [0, 1]^n$,

$$h_\mathcal{S}(x) = \text{argmin}_{1 \leq k \leq C} d(x, S_k),$$

where if $d(x, S_k)$ is minimized for more than one value of k, one of these is chosen randomly as the value of $h_\mathcal{S}$. So, in other words, the class label assigned to x is k where S_k is the closest to x of the regions $S_1, S_2, \ldots, S_C$. We refer to $B = B_1 + \cdots + B_C$ as the number of boxes in $\mathcal{S}$ and in $h_\mathcal{S}$. We will denote by H_B the set of all such classifiers where the number of boxes is B. The set of all possible classifiers we consider is then $H = \bigcup_{B=1}^{\infty} H_B$.

These classifiers, therefore, are based, as a starting point, on regions assumed to be of particular categories. These regions are each unions of boxes, and the regions do not overlap. (In practice, these boxes and the corresponding regions will likely have been constructed directly from a training sample by finding boxes containing sample points of a particular class, and merging, or agglomerating these; see [13].) See, for example, Fig. 1. The three types of boxes are indicated, and the pale gray region is the region not covered by any of the boxes.

Then, for all other points of the domain, the classification of a point is given by the class of the region to which it is closest (in the d_1 metric). For the initial configuration of boxes indicated in Fig. 1, the final classification of the whole domain is as indicated in Fig. 2. Bounding lines for the boxes have been inserted in these figures to make it easier to see the correspondence between them.

These classifiers seem quite natural, from a geometrical point of view, and unlike 'black-box' classifiers (such as neural networks), can be described and understood: there are box-shaped regions where we assert a known classification, and the classification elsewhere is determined by an arguably fairly sensible nearest-neighbor approach.