Hybrid misclassification minimization*

Chunhui Chen and O. L. Mangasarian

Computer Sciences Department, University of Wisconsin, 1210 West Dayton Street, Madison, WI 53706, USA
E.mail: chunhui/olvi@cs.wisc.edu.

Given two finite point sets $\mathcal{A}$ and $\mathcal{B}$ in the $n$-dimensional real space $\mathbb{R}^n$, we consider the NP-complete problem of minimizing the number of misclassified points by a plane attempting to divide $\mathbb{R}^n$ into two halfspaces such that each open halfspace contains points mostly of $\mathcal{A}$ or $\mathcal{B}$. This problem is equivalent to determining a plane $\{x \mid x^Tw = \gamma\}$ that maximizes the number of points $x \in \mathcal{A}$ satisfying $x^Tw > \gamma$, plus the number of points $x \in \mathcal{B}$ satisfying $x^Tw < \gamma$. A simple but fast algorithm is proposed that alternates between (i) minimizing the number of misclassified points by translation of the separating plane, and (ii) a rotation of the plane so that it minimizes a weighted average sum of the distances of the misclassified points to the separating plane. Existence of a global solution to an underlying hybrid minimization problem is established. Computational comparison with a parametric approach to solve the NP-complete problem indicates that our approach is considerably faster and appears to generalize better as determined by tenfold cross-validation.

1. Introduction

A fundamental problem in machine learning is that of discriminating between two given point sets $\mathcal{A}$ and $\mathcal{B}$ in the $n$-dimensional real space $\mathbb{R}^n$. This is typically achieved by constructing a plane

$$x^Tw = \gamma,$$

such that

$$x^Tw > \gamma \quad \text{for } x \in \mathcal{A},$$

$$x^Tw < \gamma \quad \text{for } x \in \mathcal{B}. \quad (2)$$

Here $w$ is the normal to the plane and $(\|\gamma\|/\|w\|)$ is the Euclidean distance from the origin to the plane. In general it is not possible to satisfy (2) except in the special case when the convex hulls of $\mathcal{A}$ and $\mathcal{B}$ do not intersect. Thus, one resorts in the general case to minimizing some error criterion in the satisfaction of (2). The

*This material is based on research supported by Air Force Office of Scientific Research Grant F49620-94-1-0036 and National Science Foundation Grant CCR-9322479.

© J.C. Baltzer AG, Science Publishers
simplest such criterion is to use linear programming in order to construct a plane (1) that minimizes a weighted average of the sum of the distances of the misclassified points to the plane [7,2] as follows:

\[
\min \left\{ \frac{e^T y}{m} + \frac{e^T z}{k} \mid Aw + y \geq e\gamma + e, \ Bw - z \leq e\gamma - e, \ y \geq 0, \ z \geq 0 \right\}. \quad (3)
\]

Here the rows of the matrices \( A \in \mathbb{R}^{m \times n} \) and \( B \in \mathbb{R}^{k \times n} \) represent the \( m \) points in \( A \) and the \( k \) points in \( B \) respectively, while \( e \) is a vector of ones of appropriate dimension. The objective function of (3) represents the sum of the average distances, multiplied by \( \|w\| \), of the misclassified points in \( A \) to the plane \( x^Tw = \gamma + 1 \) and of the misclassified points of \( B \) to the plane \( x^Tw = \gamma - 1 \). If the convex hulls of \( A \) and \( B \) are disjoint, then there are no misclassified points and the linear program (3) yields a zero minimum. However in the general case of intersecting convex hulls, the linear program (3) obtains an approximate separating plane that minimizes an average sum of distances of misclassified points as described above. However, this criterion for discrimination may not minimize the actual number of the misclassified points. The problem of constructing a plane (1) such that the number of misclassified points is minimized, is considerably more difficult and in fact is NP-complete, as shown in proposition 2 of section 2 below. This problem was considered in [8], where a parametric minimization approach was proposed and implemented in [1]. Although the parametric procedure is effective, it is costly computationally, which is to be expected since the underlying problem is NP-complete.

In the present approach we shall propose a fast alternative hybrid criterion that is quite effective in approximately minimizing the number of misclassified points as determined by tenfold cross-validation [15]. The basic idea is to minimize the number of misclassified points by translating the separating plane, and then rotating the plane in order to minimize a weighted average sum of the distances of misclassified points to a separating plane. This hybrid separability criterion leads to an effective finite algorithm for solving the separation problem.

We outline the contents of the paper now. In section 2 we define the misclassification minimization problem (7), and establish the NP-completeness of the equivalent problem (8) in proposition 2. We then define our Hybrid Misclassification Minimization (HMM) problem 3 and establish the existence of a global solution to it in theorem 4, and prescribe a finite hybrid algorithm, HMM algorithm 5, for its approximate solution. Section 3 contains numerical results that indicate that the proposed hybrid algorithm is fast and appears to generalize better than the parametric algorithm misclassification minimization [1].

A word about our notation now. For a vector \( x \) in the \( n \)-dimensional real space \( \mathbb{R}^n \), \( x_+ \) will denote the vector in \( \mathbb{R}^n \) with components \((x_+)_i := \max\{x_i, 0\}\), \( i = 1, \ldots, n \). Similarly \( x_- \) will denote the vector in \( \mathbb{R}^n \) with components \((x_-)_i := (x_i)_+, \ i = 1, \ldots, n \), where \((\cdot)_+\) is the step function defined as one for positive \( x_i \) and zero otherwise. The norm \( \| \cdot \| \) will denote the \( l_2 \) norm, while \( A \in \mathbb{R}^{m \times n} \) will signify a real \( m \times n \) matrix. For such a matrix, \( A^T \) will denote the