Measuring the error of linear separators on linearly inseparable data

Boris Aronov * Delia Garijo † Yurai Núñez-Rodríguez‡ David Rappaport §
Carlos Seara ¶ Jorge Urrutia‖

Abstract

Given a set of red points R and blue points B we seek to determine a linear separator of the sets, although the sets may be linearly inseparable. Thus we obtain a partition into two parts, where each part may contain some misclassified points. We determine the error of the partition based on the amount of work needed to move the misclassified points. We consider several different measures of work and provide algorithms to find linear separators that minimize the error under these different measures.

1 Introduction

Let $R = \{r_1, \ldots, r_p\}$ be a set of p red points and $B = \{b_1, \ldots, b_q\}$ a set of q blue points in $\mathbb{R}^d$. Let $n = p+q$ and assume that the points are *disjoint* and in general position, that is, no $d+1$ of the points lie in the same hyperplane in $\mathbb{R}^d$. We say that R and B are *linearly separable* if there is a hyperplane, a *linear separator*, that partitions $\mathbb{R}^d$ such that each part contains only red or only blue points. If there is no linear separator for R and B, then we say that the sets are *linearly inseparable*. Given sets R and B we attempt to find a linear separator, and if no such separator exists we would like to produce an *approximate separator* with the “best” approximation. The notion of “best” is left intentionally informal as the precise properties that should be optimized are application dependent. In this paper we will examine several different criteria for choosing an approximate separator.

Let $P = \{p_1, \ldots, p_n\} = R \cup B$. Let H denote a hyperplane which misclassifies P and partitions P into two non-empty subsets. It will be useful to call one of the subsets the *left* subset and say that red points in it are *well classified* and blue points in it are *misclassified*. This language is hardly a loss of generality as we can always just label the majority color in some side red, and name that side the *left side*. The other subset is refereed to as *right* and plays a complementary role. Given P and H we may be left with a set, Ξ, the subset of P that is misclassified by H, that is, H is an approximate separator. We use $s(H)$ to represent the quality of H as an approximate separator, a cost of H. Our goal is to find a hyperplane that minimizes the cost under one of the following assumptions.

1. $s(H)$ is the maximum Euclidean distance from H to a point in Ξ: MinMax criterion.
2. $s(H)$ is the sum of the Euclidean distances from H to every point in Ξ: MinSum criterion.
3. $s(H)$ is the sum of squares of the Euclidean distances from H to every point in Ξ: MinSum^2 criterion.
4. $s(H)$ is simply the cardinality of Ξ: MinMis criterion.

*Dept. of Computer and Information Science, Polytechnic University, USA, aronov@cis.poly.edu.
†Depto. de Matemática Aplicada I, Universidad de Sevilla, dgarijo@us.es.
‡School of Computing, Queen’s University, Kingston, Canada, yurai@cs.queensu.ca.
§School of Computing, Queen’s University, Kingston, Canada, daver@cs.queensu.ca.
¶Dept. de Matemàtica Aplicada II, Universitat Politècnica de Catalunya, Spain, carlos.seara@upc.edu.
‖Instituto de Matemáticas, Universidad Nacional Autónoma de México, urrutia@matem.unam.mx.

2 MinMax

In this section we concentrate in what may be considered as the simplest version of the problem as our solutions are straightforward and succinct. We will also be setting a pattern that we follow in all subsequent versions. We first consider the problem in $\mathbb{R}$. In each case the one-dimensional algorithm for finding an optimal approximate separator offers intuitive insight to solutions in higher dimensions. This phenomenon is easily explained by the following observation.

Observation 2.1. Suppose we have obtained an optimal approximate separator H for P in $\mathbb{R}^d$. Let P^- (H^-) denote a projection orthogonal to H of P (H) to $\mathbb{R}^{d-1}$. Maintaining the same notation, repeatedly project into a lower dimension until we get to $\mathbb{R}$. Then H^- is an optimal approximate separator of P^- and furthermore $s(H) = s(H^-)$.

Thus every optimal solution in higher dimensions has an equivalent one-dimensional solution, but the number of possible candidate solutions that we can evaluate to determine an optimal one usually increases with the dimension, i.e., bigger dimension implies more possible candidates.

2.1 $d = 1$

Observation 2.2. An optimal MinMax approximate separator is the mean of the leftmost blue point and the rightmost red point and is found in $\Theta(n)$ time.

The observation is justified by the fact that the cost $s(H)$ is realized by the misclassified point that is furthest away from H. Placing H at the mean of the extremes is obviously the cost minimizing approximate separator.

2.2 $d = 2$

As in the one-dimensional version of this problem the optimal approximate separator is the mean of extreme points. In $\mathbb{R}^2$ we must consider combinations of an extreme point of R paired with an extreme and antipodal extreme point of B. We then pick the pair that are closest together. When closest antipodal extreme points come from the same set they determine the width of the set. In our case the pairs come from two different sets, but it is a routine matter to show that in $\mathbb{R}^2$ there are only $O(n)$ antipodal pairs that need to be examined. These points can be found using the well known algorithmic technique that is described as the method of *rotating calipers*. See for example Tousaint [12].

Thus the crux of our method examines antipodal pairs from R and B. We begin by computing the convex hulls $\mathrm{CH}(B)$ and $\mathrm{CH}(R)$ represented as the lists of the extreme points of B and R respectively stored in order around the boundary of the corresponding convex hulls. There are many well known convex hull algorithms for a set of n points and the most efficient use $O(n \log h)$ time where h represents the number of extreme points of the set. See for example Chan's algorithm [2]. Once we have the hulls we pick fixed antipodal directions for $\mathrm{CH}(B)$ and $\mathrm{CH}(R)$, and then systematically march around each of the boundaries. As we never backtrack along either of the hulls, we make the full revolution in $O(n)$ steps. At each step we maintain the distance between the antipodal points, and when we complete a full revolution we have the overall minimum antipodal distance. Thus we have the following observation.

Observation 2.3. We can obtain a MinMax optimal approximate separator in $O(n \log n)$ time, using $O(n)$ space. If we use h to denote the sum of the number of the blue and red extreme points then we can say that the complexity is in $O(n \log h)$ time.

When the number of extreme points of R and B are in $\Omega(n)$ then we can demonstrate that $\Omega(n \log n)$ is a lower bound for find a best MinMax optimal approximate separator. We use a reduction from the Max–Gap problem for points on the first quadrant of the unit circle [11] which is known to have an $\Omega(n \log n)$ lower bound in the algebraic computation tree model.

2.3 $d \geq 3$

The minimum width problem for a set of n points P in $\mathbb{R}^3$ can be solved in $O(m+n \log n)$ time and $O(n)$ space, where m is the number of antipodal edge-edge pairs of $CH(P)$ (see Houle and Toussaint [9]). Houle and Toussaint perform an exhaustive analysis and show that there are polyhedra that have $\Theta(n^2)$ antipodal edge-edge pairs. Thus it follows that with separate polyhedra with a sum total of n edges, we may have $\Theta(n^2)$ edge pairs as well. Thus the search space for determining an antipodal pair with minimal distance is in $\Theta(n^2)$, and by using the techniques developed by Houle and Toussaint we can obtain an optimal approximate MinMax separator in $O(m+n \log n)$ time. With respect to the problem of computing the width of a set of points, a series of papers derived improved sub-quadratic algorithms, the best of which requires $O(n^{3/2+\delta})$ expected time (see [3]). For $d \geq 4$, the $O(n^{\lceil d/2 \rceil})$ time algorithm can be achieved by realizing the solution space as a convex polytope in $d+1$ variables/dimensions, and applying an optimal half-space intersection algorithm.

3 MinSum

In this section we study the problem of minimizing the sum of distances of those points that have to be moved for R and B to become linearly separable, that is, using the MinSum criterium. Most of the proofs in dimension $d = 1$ can be extended for $d \geq 2$ as a consequence of Observation 2.1.

3.1 $d = 1$

Let R and B be inseparable sets of red and blue points on a line. We want to find the optimal separator a according to the MinSum criterium. Assume that the red (blue) region is on the left (right) of a, and that a is strictly between two consecutive points. The following lemma characterizes the solution of the MinSum problem.

Lemma 3.1. *An optimal separator point a lies between two consecutive points of $R \cup B$ such that the number of misclassified blue points is equal to the number of misclassified red points.*

Observation 3.2. If there exists an optimal separator a between consecutive points, say s and t, then any point $p \in (s, t)$ is an optimal separator. In fact, we can take the closed interval $[s, t]$ treating s as belonging to the left region and t belonging to the right region.

Lemma 3.3. *Let $|R| = p$, $|B| = q$. The* MinSum *problem in $\mathbb{R}$ has infinitely many optimal solutions. An optimal separator is any point a of the closed interval defined by the two consecutive points of $R \cup B$ in the positions p and $p+1$ counting from left to right.*

Theorem 3.4. *The 1-dimensional* MinSum *problem can be solved in $\Theta(n)$ time.*

3.2 $d = 2$

Let ℓ be an the optimal separator line for R and B according to the MinSum criterion. Let ℓ^+ (ℓ^-) be the open half-plane above (below) ℓ. The following two lemmas are straightforward consequence from Observation 2.1 and Lemmas 3.1 and 3.3.

Lemma 3.5. *The optimal direction to move the misclassified points is the orthogonal direction to ℓ.*

Lemma 3.6. *The number of misclassified blue points on ℓ^+ is equal to the number of misclassified red points on ℓ^-. The 2-dimensional* MinSum *problem has an infinite number of optimal separators.*

Let $\ell := ax + y + b = 0$ and $p_i = (x_i, y_i) \in \ell^+$ ($p_j = (x_j, y_j) \in \ell^-$) a point of $P = R \cup B$ which satisfies $ax_i + y_i + b > 0$ ($ax_j + y_j + b < 0$). Assign a value δ_i to each point p_i, red or blue, such that $\delta_i = 0$ for each red point in ℓ^- (well classified) and $\delta_i = 1$ for each blue point in ℓ^- (misclassified). Analogously, assign $\delta_j = 0$ to each blue point in ℓ^+ (well classified) and $\delta_j = 1$ to each red point in ℓ^+

(misclassified). A red or blue point $p_i = (x_i, y_i)$ on ℓ satisfies $ax_i + y_i + b = 0$ and it is always consider well classified, i.e., $\delta_i = 0$. If the number of misclassified points on ℓ^- and ℓ^+ are equal, then the sum of distances of the points of $P = R \cup B$ to ℓ is the function:

$$s(\ell) = \frac{1}{\sqrt{a^2+1}} \left(a(\sum(\delta_i x_i - \delta_j x_j)) + \sum(\delta_i y_i - \delta_j y_j) \right).$$

Let us write $A = \sum(\delta_i x_i - \delta_j x_j)$ and $B = \sum(\delta_i y_i - \delta_j y_j)$, which are constants for a fixed bipartition. Then, the function $s(\ell)$ only depends on the parameter a, the slope of ℓ, so we write $s(\ell) = s(a)$.

$$s(\ell) = s(a) = \frac{Aa + B}{\sqrt{a^2+1}}, \qquad s'(a) = \frac{-Ba + A}{(a^2+1)^{3/2}}.$$

The value $a = A/B$ is a minimum of the function $s(a)$, so line $\ell_0 := A/Bx + y + b = 0$ is the optimal separator for the bipartition defined by ℓ. This fixes the slope of ℓ_0 but we still need to determine a value (or range of values) of b. The family of parallel lines $A/Bx + y + b = 0$, $b \in [b_1, b_2]$, is defined by the supporting lines with slope A/B of the left and right part of the bipartition of P produced by ℓ.

Algorithm 1: 2d-MinSum-optimal-separator

1. Compute the p and $(p+1)$-levels in $\mathcal{A}(P)$ and the sequence $(a_1, \ldots, a_{n_p-1})$. Let e_i and e'_i be the lines which intersection point is a_i.
2. Let $\ell_1 := \mathcal{D}(a_1)$. In $O(n)$ time do the following: (1) compute in the primal the bipartition $\{P_{1,1}, P_{1,2}\}$ of P produced by ℓ_1: $P_{1,1}$ is the subset of points of P above ℓ_1 including the point $\mathcal{D}(e_1)$, and $P_{1,2}$ is the subset of points of P below ℓ_1 including the point $\mathcal{D}(e'_1)$; (2) compute the function $s(a)$, the value A/B which minimizes $s(a)$, check that $A/B \leq a_1$; and (3) do $m := A/B$, $s(m) := s(A/B)$, and $ind := 1$ the index of the cell C_1.
3. **For** $i := 2$ to n_p (sweeping the p-level from left to right and stopping at a_i) **do**:
   (a) Update the changes of the function $s(a)$: only the two points $\mathcal{D}(e_{i-1})$ and $\mathcal{D}(e'_{i-1})$ interchange their half-planes in the bipartition $\{P_{i,1}, P_{i,2}\}$ given by $\ell_i := \mathcal{D}(a_i)$ with respect to the bipartition $\{P_{i-1,1}, P_{i-1,2}\}$ working with the neighbor cells C_{i-1} and C_i. Thus, only two values of δ_j in the function $s(a)$ has to be updated.
   (b) Compute the new value A/B which minimizes the function $s(a)$, and check that $a_{i-1} \leq A/B \leq a_i$ (if $i = n_p$ check that $a_{n_p} \leq A/B$). Compute the value $s(A/B)$. If $s(A/B) < s(m)$, then do $m := A/B$, $s(m) := s(A/B)$, and $ind := i$.
4. Let m, $s(m)$, and ind be the current values. Compute the intersection points of the vertical line $x = m$ with the boundary of the cell C_{ind}. Let (m, b_1) and (m, b_2) be these intersection points. Any line of the family of parallel lines $\ell = mx + y + b = 0$, $b \in [b_1, b_2]$ is an optimal separator for P and MinSum $= s(m)$.

Theorem 3.7. *The* 2*-dimensional* MinSum *problem can be solved deterministically in time* $O(n \log n + n^{4/3} \log^{1+\epsilon} n)$ *for an arbitrarily small constant* $\epsilon > 0$ *or in* $O(n \log n + n^{4/3})$ *expected time.*

3.3 $d = 3$

Let π a plane which is an optimal separator for $P = R \cup B$ according to the MinSum criterion. By the same reasoning, an optimal separator plane π with normal vector $\vec{v}$ can be translated along the direction given by $\vec{v}$ until it bumps into a point in P, and all these planes are optimal.

Let $\pi := ax + by + z + c = 0$ a plane with normal vector $(a, b, 1)$ and $p_i = (x_i, y_i, z_i) \in \pi^+$ ($p_j = (x_j, y_j, z_j) \in \pi^-$) a point of $P = R \cup B$ which satisfies $ax_i + by_i + z_i + c > 0$ ($ax_i + by_i + z_i + c < 0$). Let δ_i and δ_j be defined as in the subsection above depending on whether the point is well or misclassified. We recall that a point $p_i = (x_i, y_i, z_i)$ on π satisfies $ax_i + by_i + z_i + c = 0$ and it is always consider well classified, so $\delta_i = 0$. If the number of misclassified points in both half-spaces of π are equal, the sum of distances between misclassified points of P and the plane π is the function:

$$s(\pi) = \frac{1}{\sqrt{a^2+b^2+1}}\left(a(\sum(\delta_i x_i - \delta_j x_j)) + b(\sum(\delta_i y_i - \delta_j y_j)) + \sum(\delta_i z_i - \delta_j z_j)\right).$$

Let us write $A = \sum(\delta_i x_i - \delta_j x_j)$, $B = \sum(\delta_i y_i - \delta_j y_j)$, and $C = \sum(\delta_i z_i - \delta_j z_j)$. These values are constants for a fixed bipartition. Then, the function $s(\pi)$ only depends on the parameters a and b, e.i., it only depends on the normal vector $(a, b, 1)$ of π, so we write $s(\pi) = s(a, b)$.

$$s(\pi) = s(a,b) = \frac{Aa + Bb + C}{\sqrt{a^2+b^2+1}}.$$

Algorithm 2: 3d-MinSum-optimal-separator

1. Compute the p and $(p+1)$-levels in $\mathcal{A}(P)$ and the sequence $(a_1, \ldots, a_{n_p})$. Let e_i, e'_i, and e''_i be the three planes which intersection point is a_i.
2. Let $\pi_1 := \mathcal{D}(a_1)$. In $O(n)$ time compute in the primal the bipartition $\{P_{1,1}, P_{1,2}\}$ of P produced by π_1: $P_{1,1}$ is the subset of points of P in π_1^+ including the point $\mathcal{D}(e_1)$, and $P_{1,2}$ is the subset of points of P in π_1^- including the points $\mathcal{D}(e'_2)$ and $\mathcal{D}(e''_3)$ (analogously for $\mathcal{D}(e_2)$ in π_1^+ and $\mathcal{D}(e'_1)$ and $\mathcal{D}(e''_3)$ in π_1^-, and for $\mathcal{D}(e_3)$ in π_1^+, and $\mathcal{D}(e'_1)$ and $\mathcal{D}(e''_2)$ in π_1^-, corresponding to the two neighbor cells sharing a_1). For each case compute the function $s(a, b)$ and the pair which minimizes $s(a, b)$. Let (a_1, b_1) be this pair. Do $(m_1, m_2) := (a_1, b_1)$, $s(m_1, m_2) := s(a_1, b_1)$, and *ind* the index of the current cell.
3. **For** $i := 2$ to n_p **do**:
   (a) Update the changes of the function $s(a, b)$: only the two points interchange their half-spaces in the bipartition $\{P_{i,1}, P_{i,2}\}$ given by $\pi_i := \mathcal{D}(a_i)$ with respect to the bipartition $\{P_{i-1,1}, P_{i-1,2}\}$ working with the neighbor cells. Thus, only two values of δ_j in the function $s(a, b)$ has to be updated.
   (b) Compute the pair (a_i, b_i) which minimizes the function $s(a, b)$, compute the value $s(a_i, b_i)$. If $s(a_i, b_i) < s(m_1, m_2)$, then do $(m_1, m_2) := (a_i, b_i)$, $s(m_1, m_2) := s(a_i, b_i)$, and *ind* the index of the current cell.
4. Let (m_1, m_2), $s(m_1, m_2)$, and *ind* be the current values. To check that $(m_1, m_2, 1)$ is a normal vector between the ones defined by the cell C_{ind} cost the complexity of the cell. To determine the plane π and the infinite solutions we compute the extreme points, (m_1, m_2, c_1) (m_1, m_2, c_2), in the the boundary of the cell C_{ind} with the vertical line passing through the point $(m_1, m_2, 0)$. Any plane of the family of parallel planes $\pi = m_1x + m_2y + z + c = 0$, $c \in [c_1, c_2]$ is an optimal separator for P and MinSum $= s(m_1, m_2)$.

Theorem 3.8. *The* 3*-dimensional* MinSum *problem can be solved in* $O(n^{5/2}\log^6 n)$ *expected time.*

3.4 $d \geq 4$

For dimension $d \geq 4$, the upper bound for the size of the p-level is only slightly better than the $O\left(n^{\lfloor d/2\rfloor}p^{\lceil d/2\rceil}\right)$ [5]. More concretely, an upper bound is $O(n^{d-\alpha_d})$ for a very small $\alpha_d = 1/(4d-3)^d$. As Agarval et al. [1] observed, the bound can be made sensitive to p, namely $O(n^{\lfloor d/2\rfloor}p^{\lceil d/2\rceil-\alpha_d})$. Matoušek et al. [10] give an $O(n^{4-2/45})$ upper bound for $d = 4$.

Theorem 3.9. *The d-dimensional* MinSum *problem can be solved deterministically in time*

$$O\left(n^{\lfloor d/2\rfloor}p^{\lceil d/2\rceil}\left(\frac{\log n}{\log p}\right)^{O(1)}\right) = O(n^d).$$

4 MinSum^2

We now consider the criterion of minimizing the sum of the squared distances of those points that have to be moved until R and B become linear separable, i.e., using the MinSum^2 criterion.

4.1 $d = 1$

Let R and B be two inseparable sets of red and blue points on the x-axis. Assume that an optimal separator point a determines the red region on its left and the blue region on its right. It means that the misclassified red and blue points have to be moved to the left and to the right of a, respectively. Given an optimal separator point a, let $b_1, \ldots, b_k$ be the x-coordinates of the misclassified blue points and $r_1, \ldots, r_m$ the x-coordinates of the misclassified red points. The sum of the squared distances of the misclassified points to a is given by

$$s(a) = (a - b_1)^2 + \cdots + (a - b_k)^2 + (r_1 - a)^2 + \cdots + (r_m - a)^2$$

$$s(a) = (k + m)a^2 - 2a\left(\sum_{i=1}^{k} b_i + \sum_{i=1}^{m} r_i\right) + \left(\sum_{i=1}^{k} b_i^2 + \sum_{i=1}^{m} r_i^2\right).$$

If we put

$$A = \sum_{i=1}^{k} b_i + \sum_{i=1}^{m} r_i, \qquad B = \sum_{i=1}^{k} b_i^2 + \sum_{i=1}^{m} r_i^2,$$

$s(a) = (k+m)a^2 - 2aA + B$, and $s'(a) = (k+m)a - 2A = 0$ which implies that $a = \frac{A}{k+m}$. This value is a minimum since there is at least one misclassified point and so $s''(a) = k + m > 0$. Observe that

$$a = \frac{A}{k+m} = \frac{\sum_{i=1}^{k} b_i + \sum_{i=1}^{m} r_i}{k+m}$$

is the arithmetic median of the misclassified points. Thus, the value a gives us a measure of how much the points are misclassified with respect to their arithmetic median. A two-dimension interpretation is the regression line of the misclassified points.

Let $S = (s_1, \ldots, s_n)$ denote $P = R \cup B$ with $n = p + q$, where the points are ordered from left to right. We use *interval* i to denote the interval between s_i and s_{i+1}. As a shorthand it will be convenient to use Ξ_i to denote the set of misclassified points and N_i the number of misclassified points when the separator is in interval i. The average value of the misclassified points at interval i is

$$a_i = \frac{1}{N_i} \sum_{s \in \Xi_i} s.$$

We say that an average a_i, is *correct* if $s_i < a_i \leq s_{i+1}$. We prove the following theorem.

Theorem 4.1. *There is a unique correct average that defines an optimal solution for the one dimensional* MinSum^2 *problem, and it can be found in linear time.*

4.2 $d = 2$

Let $\ell := ax + y + b = 0$ be the optimal separator line according to the MinSum^2 criterion. Let $p_i = (x_i, y_i)$ ($p_j = (x_j, y_j)$) be a point of $R \cup B$ in ℓ^+ (ℓ^-) which satisfies $ax_i + y_i + b > 0$ ($ax_j + y_j + b < 0$). Assign a value δ_i to each point p_i, red or blue, such that $\delta_i = 0$ for each red point in ℓ^+ (well classified) and $\delta_i = 1$ for each blue point in ℓ^+ (misclassified). Analogously, assign $\delta_j = 0$ to each blue point in ℓ^- (well classified) and $\delta_j = 1$ to each red point in ℓ^- (misclassified). A point $p_i = (x_i, y_i)$ on ℓ satisfies $ax_i + y_i + b = 0$ and it is consider well classified, i.e., $\delta_i = 0$. The sum of squared distances between the points of $R \cup B$ and the line ℓ is the function:

$$s(\ell) = \sum_{p_i \in \ell^+} \delta_i \frac{(ax_i + y_i + b)^2}{a^2 + 1} + \sum_{p_j \in \ell^-} \delta_j \frac{(ax_j + y_j + b)^2}{a^2 + 1} =$$

$$= \frac{1}{a^2+1} \left(\sum_{p_i \in R \cup B} \delta_i (a^2 x_i^2 + y_i^2 + b^2 + 2a x_i y_i + 2ab x_i + 2b y_i) \right).$$

Let us write $A = \sum \delta_i x_i^2$, $B = \sum \delta_i y_i^2$, $C = \sum \delta_i$, $D = \sum \delta_i x_i y_i$, $E = \sum \delta_i x_i$, and $F = \sum \delta_i y_i$ which are constants for a given partition. Thus, $s(\ell)$ only depends on a and b, so we write $s(\ell) = s(a,b)$,

$$s(a,b) = \frac{Aa^2 + B + Cb^2 + 2Da + 2Eab + 2Fb}{a^2+1}.$$

PROCEDURE: MinSum^2-OPTIMAL-SEPARATOR

1. Dualize the point set $P = R \cup B$ as red and blue lines obtaining the arrangement $\mathcal{A}(P)$.
2. Visit all the cells in $\mathcal{A}(P)$ and for on each cell do the following calculations:
   (a) Update the changes of the function $s(a,b)$: only a point change in the bipartition $\{P_1, P_2\}$, thus update two values of δ_i.
   (b) Compute the the pair (a_1, b_1) which minimizes $s(a,b)$. Check that the point obtained by dualizing the line $\ell_1 := a_1 x + y + b_1 = 0$ is inside the current cell.
   (c) Compute and update the minimum value of $s(a,b)$ and its corresponding optimal separator $\ell := ax + y + b = 0$.

Theorem 4.2. *The two-dimensional* MinSum^2 *problem can be solved in* $O(n^2)$ *time.*

For $d \geq 3$, similar computations can be extended obtaining an $O(n^d)$ time algorithm.

5 MinMis

Another way of thinking into reach linear separability is to delete the misclassified points, i.e., compute the minimum number of points that have to be deleted in order to get the linear separability of the remaining red and blue points. This problem is equivalent to the problem of computing an optimal separator for B and R which minimize the sum of misclassified points [7].

5.1 $d = 1$

Let R and B be set of red and blue points on a line, respectively. In order to compute the separator a given the minimum number of misclassified points, i.e., the separator a that minimize the sum s of misclassified points in both sides, we sort the points in $O(n \log n)$ time, obtained a linear number of intervals defined by two consecutive points. Any point in an interval gives the same value of s. Thus do a sweep from left to right updating the value s in constant time each time we a point (red or blue) changes from right to left of the current a. Notice that for a given separator a we have r red points and b blue points on its left and $n - r$ red points and $n - b$ blue points on its right. It is clear that we get an overall $O(n \log n)$ time algorithm for computing the optimal separator (either point or interval). We prove that this algorithm is optimal using a reduction to the ϵ-distance problem for points on a line which has an $\Omega(n \log n)$ time lower bound in the algebraic decision tree model [11].

Theorem 5.1. *The one-dimensional* MinMis *problem can be solved in optimal* $\Theta(n \log n)$ *time.*

5.2 $d = 2$

For the two-dimensional problem we can proceed as for the 2-dimensions MinSum^2 problem counting the number of misclassified points in both sides of the line, updating the minimum of the sum of them. Houle [8] gave an $O(n^2)$ time algorithm for this problem. For a given k, Cole, Sharir and

Yap [6] presented an $O(nk \log n)$ time algorithm to compute the separators misclassifying at most k points. Everett et al. [7] improve upon these results if the number of misclassified points is not too large, getting a $O(nm \log m + n \log n)$ time algorithm, where m is the number of misclassified points. Chan [4] present an algorithm that find a line ℓ that minimizes k, the total number of red points above ℓ and the blue points below ℓ, in $O((n + k^2) \log n)$ expected time and $O(n + k^2)$ space.

Theorem 5.2. *The two-dimensional* MinMis *problem can be solved in* $O(n^2)$ *time.*

For $d \geq 3$ we can proceed as for the 2-dimensions MinSum^2 problem counting the number of misclassified points in both sides of the hyperplane. Thus, the d-dimensional MinMis problem can be solved in $O(n^d)$ time.

Table 1: Summary of results.

Dimension	MinMax	MinSum	MinSum^2	MinMis
$d = 1$	$\Theta(n)$	$\Theta(n)$	$\Theta(n)$	$\Theta(n \log n)$
$d = 2$	$\Theta(n \log n)$	$O(n \log n + n^{4/3} \log^{1+\epsilon} n)$ $O(n \log n + n^{4/3})$ $(*)$	$O(n^2)$	$O(n^2)$
$d = 3$	$O(n^2)$	$O(n^{5/2} \log^6 n)$ $(*)$	$O(n^3)$	$O(n^3)$
$d \geq 4$	$O(n^{\lceil d/2 \rceil})$	$O(n^d)$	$O(n^d)$	$O(n^d)$

$(*)$ expected time

References

[1] P. K. Agarwal, B. Aronov, T. M. Chan, and M. Sharir. On levels in arrangements of lines, segments, planes, and triangles. 2001

[2] T. M. Chan. Optimal output-sensitive convex hull algorithms in two and three dimensions. *Discrete and Computational Geometry*, Vol. 16,1996, pp.361–368.

[3] T. M. Chan. Approximating the diameter, width, smallest enclosing cylinder, and minimum-width annulus. *Sixteenth Annual Symposium on Computational Geometry*, 2000, pp. 300–309.

[4] T. M. Chan. Low-Dimensional Linear Programming with Violations. 2004. A preliminary version of this paper appeared in *Proc. 43rd IEEE Sympos. Found. Comput. Sci.*, 2002.

[5] K. L. Clarkson and P. W. Shor. Applications of random sampling in computational geometry, II. *Discrete Comput. Geom.*, 4, 1989, pp. 387–421.

[6] R. Cole, M. Sharir, and C. K. Yap. On k-hulls and related problems. *SIAM J. Comput.*, 16, 1987, pp. 61–77.

[7] H. Everett, J.-M. Robert, and M. van Kreveld. An optimal algorithm for the ($\leq k$)-levels, with applications to separation and transversal problems. *International Journal of Computational Geometry and Applications*, 6, 1996, pp. 247–261.

[8] M. E. Houle. Algorithms for weak and wide separation of sets. *Discrete Applied Mathematics*, Vol. 45, N. 2, 1993, pp. 139-159.

[9] M. E. Houle and G. T. Toussaint. Computing the width of a set. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 5, pp. 761–765, 1988.

[10] J. Matoušek, M. Sharir, S. Smorodinsky, and U. Wagner. On k-sets in four dimension. Manuscript, 2004.

[11] V. Sacristán. Lower bounds for some geometric problems. *Technical Report MA-IR-98-0034, Universitat Politècnica de Catalunya*, Departament de Matemàtica Aplicada II, 1998.

[12] G. T. Toussaint. Solving geometric problems with the rotating calipers. *Proceedings of IEEE MELECON'83*, Athens, Greece, May 1983, pp. A10.02/1–4.