

## AN ALGORITHM FOR DETERMINING IDENTITY OF NEAREST-NEIGHBOR AND POTENTIAL FUNCTION DECISION RULES\*

SARGUR N. SRIHARI

Department of Computer Science, SUNY at Buffalo, Amherst, NY 14226, U.S.A.

THOMAS SNABB

Department of Mathematics and Statistics,  
University of Michigan, Dearborn, MI 48128, U.S.A.

and

LEE J. WHITE

Department of Computer and Information Science,  
The Ohio State University, Columbus, OH 43210, U.S.A.

(Received 10 January 1980)

**Abstract** - The nearest-neighbor and potential function decision rules are nonparametric techniques that partition the feature space based on a set of labelled sample points. Determining whether the partitions of the two rules are identical for a given set of points is an interesting problem in computational geometry. Here, a relationship between the two methods in terms of subclasses and composite classes is developed. Considering an exponential potential function, necessary and sufficient conditions for identity of their decision surfaces are obtained. Based on conditions of *symmetry*, *weighting*, and the *Voronoi region* of a point, an algorithm for establishing identity in  $\mathbb{R}^d$  is introduced.

Computational geometry    Nearest-neighbor rule    Potential function classifier    Decision rule  
equivalence    Pattern recognition algorithms    Identity conditions

### 1. INTRODUCTION

The nearest-neighbor (n-n) decision rule<sup>(1,2)</sup> and the potential function (pf) classifier<sup>(3,4)</sup> are two of the more important methods of nonparametric pattern classification. Both methods require a set of correctly labelled sample points from each category. For each test point  $x \in \mathbb{R}^d$  the n-n rule determines the nearest sample point and assigns  $x$  to the corresponding category. The pf rule computes the "potential" induced at  $x$  by each prototype according to a function and assigns  $x$  to the category whose cumulative potential at  $x$  is maximum. With appropriate choice of potential function parameters the pf rule is equivalent to a "two-step" rule<sup>(5)</sup>, which is the Bayes decision rule based on density estimates of  $x$  obtained as linear combinations of kernel functions centered at the prototypes.

The large sample performance relationship between the two methods is well known, i.e., the n-n error rate is bounded above by twice the minimal error rate and two step rules are optimal. This asymptotic bound on

the n-n error rate can be tightened by using a generalization, viz.,  $k$ -nearest-neighbor rule, in which classification is based on a majority of the  $k$  nearest neighbors of  $x$ . There exists little published analytical work, however, of the small sample performance relationship and the relative computational complexities of these methods for arbitrary dimensions. Here we demonstrate that the single n-n and pf decision rules are special cases of parametric methods based on subclasses and composite classes respectively and that the time complexities of direct implementation of the two rules are similar. We then attempt to obtain precise conditions under which their two-class feature space partitions are identical. These conditions, which formalize a heuristic comparison of the decision surfaces of the two methods previously given in<sup>(6)</sup>, allow the determination from a finite sample set whether the performance of the two rules will be identical for arbitrary parent distributions of the samples.

### 2. THE TWO DECISION RULES

#### *Composite class and subclass discrimination*

Many pattern classification problems can be regarded as a case of discriminating between composite

\*This work was supported in part by grants NIH5-507-RR07066-13 and HL 18968.

© 1979 IEEE. Reprinted with permission from IEEE Computer Society Conference on Pattern Recognition and Image Processing, 6-8 August 1979, Chicago, Illinois.

classes.<sup>(7)</sup> An abstraction of such a problem is to assign each  $x \in \mathbb{R}^d$  into either  $C_1$  or  $C_2$  where  $C_i$  includes  $n$  subclasses  $C_{ij}$ ,  $\Pi_i$  denotes the prior probability of  $C_i$ ,  $\Pi_{ij}$  the prior probability of  $C_{ij}$  when  $C_i$  is true and  $p_{ij}(x)$  the subclass-conditional probability density of  $x$ .

Let  $\delta_1$  and  $\delta_2$  be two statistical decision rules that map  $\{x\}$  to  $\{C_1, C_2\}$  as follows:  $\delta_1$  assigns  $x$  to the class that includes the subclass with the maximum a posteriori probability, and  $\delta_2$  assigns  $x$  to the class with maximum a posteriori probability. Thus if  $g_{ij}(x) = sp_{ij}(x) \Pi_{ij} \Pi_i$  where  $s$  is a scale factor independent of  $x$ , then

$$\delta_1(x) = C_k \text{ if for some } l \text{ and } k$$

$$g_{kl}(x) = \max_{i,j} \{g_{ij}(x)\};$$

and

$$\delta_2(x) = C_k \text{ if for some } k,$$

$$g_k(x) = \max_i \left\{ \sum_j g_{ij}(x) \right\}.$$

Let  $P_e[\delta_1]$  and  $P_e[\delta_2]$  be the respective error probabilities of  $\delta_1$  and  $\delta_2$ .

*Assertion.*  $P_e[\delta_1] \geq P_e[\delta_2]$ .

*Proof.* Any choice of  $\{g_{ij}\}$  and  $x$  such that

$$g_{kl}(x) = \max_{i,j} \{g_{ij}(x)\}$$

and

$$\sum_j g_{kj} < \sum_j g_{mj},$$

$m \neq k$  is sufficient to demonstrate that  $\delta_1 \neq \delta_2$ . Since  $\delta_2$  is the Bayes decision rule the assertion follows. Whether strict inequality of the assertion holds is dependent on the specific probabilities and distributions, however it is of interest to note that strict inequality can hold even for Gaussian  $p_{ij}(x)$  with identical scalar covariance matrices.<sup>(8)</sup>

*Decision surfaces*

Decision rules  $\delta_1$  and  $\delta_2$  yield two different methods of nonparametric classification if we let  $g_{ij}$  be a function of sample  $a_{ij} \in \mathbb{R}^d$  as

$$g_{ij}(x) = K([x, a_{ij}]),$$

where  $[x, y] = \|x - y\|^2$  is the squared Euclidean norm. If  $K$  has the exponential form

$$K(\lambda) = \xi \eta^{-\mu \lambda}, \tag{1}$$

where  $\xi, \eta$  and  $\mu$  are positive real, then  $\delta_1$  becomes the n-n rule of associating  $x$  with the class label of its nearest sample. The n-n decision surface (n-n surface) has the piecewise linear form

$$\min_j \{[x, a_{1j}]\} = \min_j \{[x, a_{2j}]\}. \tag{2}$$

By definition, the pf rule computes the potential at  $x$  as

$$\sum_j K([x, a_{ij}]),$$

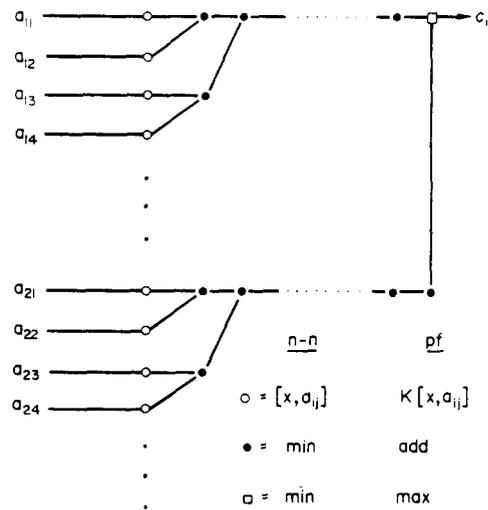


Fig. 1. Tree implementation.

where  $K$  is a potential function that varies inversely with its argument, and associates  $x$  with the class with larger potential; thus  $\delta_2$  reduces to the pf rule. The pf decision surface (pf surface) has the continuously differentiable form

$$\sum_j K([x, a_{1j}]) = \sum_j K([x, a_{2j}]). \tag{3}$$

As  $h \rightarrow \infty$  the pf surface defined by equation (3) and the exponential potential function approaches the n-n surface,<sup>(9)</sup> thus the case of interest is when  $h$  is finite.

*Implementation*

In terms of algebraic complexity, a direct implementation of the n-n rule requires the computation of  $2n$  distances and performing  $2n - 1$  comparisons whereas the pf rule involves computation of  $2n$  potential functions,  $2n - 2$  additions and a single comparison. Due to min, max and add being associative operators both rules can be implemented on a parallel processor using the tree structure shown in Fig. 1. The resulting time complexity is  $1 + \lceil \log_2 n \rceil$  when implemented with degree of parallelism  $2n$ .

Direct implementation of both decision rules requires the storage of  $2n$  samples. Storage efficiency of the n-n rule is achieved by storing hyperplane segments of the n-n surface instead of the samples; though best known algorithms for determining such segments are  $O(n \log n)$  in  $\mathbb{R}^2$  and  $O(n^5)$  in  $\mathbb{R}^3$ .<sup>(10)</sup> A storage efficient indirect implementation of the pf rule is based on storing coefficients in a polynomial expansion of the  $g_i(x)$ .<sup>(9)</sup>

**3. IDENTITY CONDITIONS**

We develop here conditions on sample points in  $\mathbb{R}^d$  for decision surface identity of the n-n and pf rules; the results obtained pertain to the general form of  $K$  as

well as to its exponential form (1) as indicated. The following notation is employed:  $A = \{a_i\} \in C_1$ ,  $B = \{b_i\} \in C_2$ ,  $U = \{u_i\} = A \cup B$  and  $P(a_i, b_j)$  denotes the hyperplane that orthogonally bisects the line segment joining  $a_i$  and  $b_j$ , or

$$P(a_i, b_j): (a_i - b_j)^T [x - (a_i + b_j)/2] = 0.$$

Without loss of generality it is assumed that there are  $n$  distinct points in each class.

*Symmetry condition*

When  $n=1$ , or  $A = \{a_1\}$ ,  $B = \{b_1\}$ , the two surfaces are always identical to  $P(a_1, b_1)$ . The following two lemmas consider the case  $n=2$  for collinear and non-collinear points; their proofs are given in<sup>(6)</sup>.

*Lemma 1.* If  $n=2$  and all points are collinear, identity holds if points are linearly separable and  $[a_1, a_2] = [b_1, b_2]$ .

A corollary of lemma 1 is that if  $n=2$  with linearly separable collinear points such that for each  $a_i$  there exists a unique  $b_i$  equidistant from the n-n hyperplane  $D$ , then the pf surface is also  $D$ . This corollary can be generalized to non-collinear points and  $n>2$ . Next consider  $n=2$  and points in general position, i.e., not all on a single line.

*Lemma 2.* If  $n=2$  and points are in general position, identity holds iff

- (i)  $[a_1, a_2] = [b_1, b_2]$ ,
- (ii)  $[a_2, b_1] = [a_1, b_2]$ ,
- (iii)  $a_1 + b_1 \neq a_2 + b_2$

and either

(iv)  $w^l(a_2 - r) > 0 > w^l(b_2 - r)$ ,

or

(v)  $[a_1, b_1] = [a_2, b_2]$ ,

where

$$w = a_1 - b_1, \quad r = \frac{1}{2}(a_1 + b_1),$$

and

$$[a_1, b_1] = \min_j \{[a_1, b_j]\}$$

The different cases of identity arising from lemma 2 are shown for  $\mathbb{R}^2$  in Fig. 2. In order to generalize these results to the case  $n>2$ , consider introducing one point to each class in the cases of lemma 2 such that the surfaces are unchanged. If point  $a^1$  is introduced into one of the  $C_1$  regions (half-space or quarter-space) defined by the separating hyperplanes, then by introducing points  $b^1, b^2$  at the mirrored positions of  $a^1$  with respect to  $D_1$  and  $D_2$ , and a point  $a^2$  at the mirrored position of  $b^2$  in  $D_2$ , surface identity is maintained. This symmetry of points with respect to hyperplanes provides a sufficient condition for identity, as expressed in Theorem 1.

*Theorem 1.* Let the n-n surface consist of a set of hyperplanes  $\{D_j\}$ . If for each  $D_j$ , given  $a \in A$  there exists

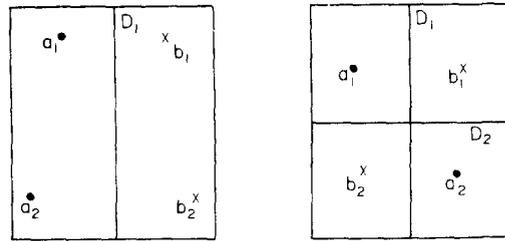


Fig. 2. Identity with  $\mathbb{R}^2$  with  $n=2$ .

a unique  $b \in B$  satisfying  $D_j = P(a, b)$  then the n-n and pf surfaces are identical.

*Proof.* For each  $x \in D_j$  the given condition implies that for every  $a$  there is a unique  $b$  such that  $K([x, a]) = K([x, b])$ , thus  $x$  belongs to the pf surface. These are the only points in the pf surface due to continuity and single-valuedness of the pf surface between the nearest unlike pair of points defining  $D_j$ . Q.E.D.

Thus when the n-n surface consists of a set of hyperplanes  $D_1, \dots, D_p$ , if  $A$  and  $B$  can be partitioned into disjoint subsets  $A_1, \dots, A_c$  and  $B_1, \dots, B_c$  respectively, where each  $A_i$  and  $B_i$  has  $p$  points each and for each  $D_j$  given  $a \in A_i$  there exists a unique  $b \in B_j$  such that  $D_j = P(a, b)$ , then identity holds. Thus for the symmetry condition to hold it is necessary that  $n=pc$ .

Figure 3 demonstrates some types of identity surface possible in  $\mathbb{R}^2$  for the following configurations:

(a)  $n=3, p=1, A_1 = \{a_1\}, A_2 = \{a_2\}, A_3 = \{a_3\}$ ,

$$B_1 = \{b_1\}, B_2 = \{b_2\}, B_3 = \{b_3\},$$

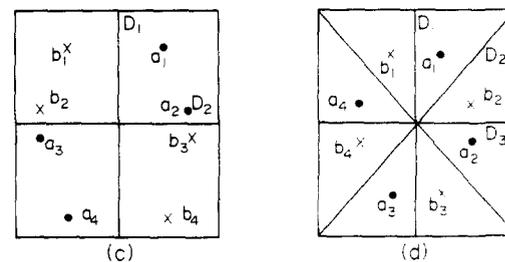
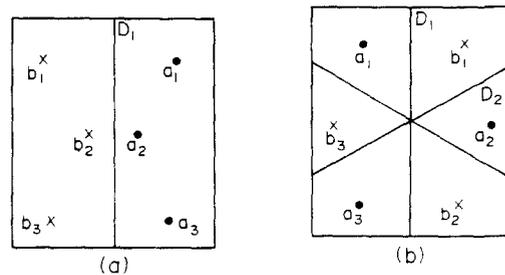


Fig. 3. Identity with symmetry.

- $D_1 = P(a_1, b_1) = P(a_2, b_2) = P(a_3, b_3)$ .
- (b)  $n=3, p=3, A_1 = \{a_1, a_2, a_3\}, B_1 = \{b_1, b_2, b_3\}$ ,  
 $D_1 = P(a_1, b_1) = P(a_2, b_2) = P(a_3, b_3)$ ,  
 $D_2 = P(a_1, b_2) = P(a_2, b_1) = P(a_3, b_3)$ ,  
 $D_3 = P(a_1, b_3) = P(a_2, b_2) = P(a_3, b_1)$ .
- (c)  $n=4, p=2, A_1 = \{a_1, a_4\}, A_2 = \{a_2, a_3\}$ ,  
 $B_1 = \{b_1, b_4\}, B_2 = \{b_2, b_3\}$   
 $D_1 = P(a_1, b_1) = P(a_2, b_2) = P(a_3, b_3) = P(a_4, b_4)$ ,  
 $D_2 = P(a_1, b_4) = P(a_2, b_3) = P(a_3, b_2) = P(a_4, b_1)$ .
- (d)  $n=4, p=4, A_1 = \{a_1, a_2, a_3, a_4\}$ ,  
 $B_1 = \{b_1, b_2, b_3, b_4\}$ ,  
 $D_1 = P(a_1, b_1) = P(a_2, b_4) = P(a_3, b_3) = P(a_4, b_2)$ ,  
 $D_2 = P(a_1, b_2) = P(a_2, b_1) = P(a_3, b_4) = P(a_4, b_3)$ ,  
 $D_3 = P(a_1, b_3) = P(a_2, b_2) = P(a_3, b_1) = P(a_4, b_4)$ ,  
 $D_4 = P(a_1, b_4) = P(a_2, b_3) = P(a_3, b_2) = P(a_4, b_1)$ .

**Weighting condition**

When the potential function is of the exponential form we can exhibit asymmetry of points when identity holds. We need the following necessary and sufficient conditions for identity when the n-n surface is a single hyperplane.

**Theorem 2.** If the n-n surface is a single hyperplane  $D$  then identity with the exponential pf surface holds if for every  $u \in U$

$$\sum_{y \in u_A} K([y, \gamma]) = \sum_{z \in u_B} K([z, \gamma]) \quad (4)$$

where  $u_A = \{x \mid x \in A \text{ and } x \text{ lies on } L(u)\}$ ,  $u_B = \{x \mid x \in B \text{ and } x \text{ lies on } L(u)\}$ ,  $\gamma = \text{point of intersection of } L(u) \text{ and } D$ , and  $L(u)$  is the normal from  $u$  to  $D$ .

For the proof of Theorem 2 refer to the Appendix. It can be further shown that when the n-n surface consists of more than one hyperplane, then for identity with the pf surface each hyperplane has to be of infinite extent and satisfy weighting condition (4). It is clear that when the symmetry condition holds the weighting condition is satisfied with respect to each  $D_j$ . Figure 4 demonstrates two types of identity possible with asymmetry for  $n = 3$  and  $4$  in  $\mathbb{R}^2$  and the following configurations.

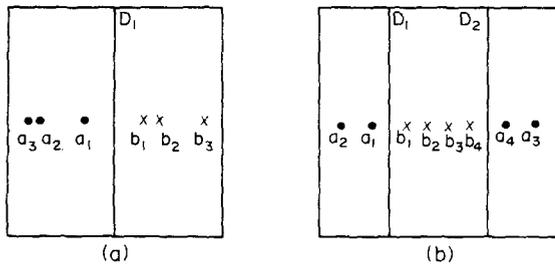


Fig. 4. Identity with non-symmetry.

(a)  $D = \{D_1\}$  with points on a single line in the configuration  $a_3 a_2 a_1 b_1 b_2 b_3$  such that  $D_1 = P(a_1, b_1)$ ,  $D_1 \neq P(a_i, b_j)$  for all  $i, j > 1$  and (4) holds.

(b)  $D = \{D_1, D_2\}$  with points on a single line in the configuration  $a_2 a_1 b_1 b_2 b_3 b_4 a_4 a_3$  such that  $D_1 = P(a_1, b_1)$ ,  $D_2 = P(a_4, b_4)$  and (4) holds for both  $D_1$  and  $D_2$ . The existence of configuration 4(b) can be shown as follows. Let  $[a_2, a_1] = [a_4, a_3] = \beta$ ,  $[a_1, b_1] = [a_4, b_4] = 2\alpha$  and  $[b_1, b_2] = [b_2, b_3] = [b_3, b_4] = \alpha$ . Thus  $d_1 = \frac{1}{2}(a_1 + b_1)$  and  $d_2 = \frac{1}{2}(a_4 + b_4)$  are n-n points. For identity we need

$$K[a_2, d_1] + K[a_3, d_1] + K[a_4, d_1] = K[b_2, d_1] + K[b_3, d_1] + K[b_4, d_1],$$

and

$$K[a_3, d_2] + K[a_2, d_2] + K[a_1, d_2] = K[b_3, d_2] + K[b_2, d_2] + K[b_1, d_2].$$

The two equations are identical since  $[a_2, d_1] = [a_3, d_2]$ ,  $[a_3, d_1] = [a_2, d_2]$ ,  $[a_4, d_1] = [a_1, d_2]$  and  $[b_2, d_1] = [b_3, d_2]$ ,  $[b_3, d_1] = [b_2, d_2]$ ,  $[b_4, d_1] = [b_1, d_2]$ . The equation has a unique solution; for example, if  $\alpha = 1$ , and  $K[x, y] = e^{-|x-y|^2}$ , then  $B \approx 0.099832$  satisfies the equation.

**Necessary and sufficient conditions**

We have shown that the weighting condition is necessary for identity. Thus a possible procedure for determining decision surface identity is to generate the n-n surface and to test whether each hyperplane of the n-n surface satisfies the weighting condition. Since generating the n-n surface is a difficult task, we can take a simpler approach of generating a set of candidate hyperplanes and checking if weighting holds for each member of this set. However when the weighting condition is satisfied with respect to a hyperplane  $D$ , it is insufficient to conclude that  $D$  is part of the n-n surface. In order to see this consider the case shown in Fig. 5 where  $n=4, D = P(a_1, b_1) = P(a_2, b_2)$  and the remaining points are chosen to be collinear such that equation (4) holds and the point of intersection  $\gamma$  satisfies

$$\min_i \{[\gamma, a_i]\} = [\gamma, a_3] < [\gamma, b_3].$$

It follows that  $\gamma$  belongs to the pf surface but not the n-n surface.

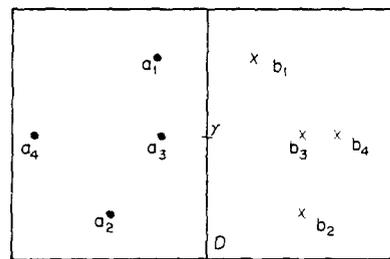


Fig. 5. Insufficiency of weighting.

We next develop necessary and sufficient conditions for a hyperplane to belong to the n-n surface. These conditions can be used to provide necessary and sufficient conditions for surface identity. First we characterize the n-n surface in terms of the Voronoi region<sup>(11)</sup>. For each  $u \in \{u_k\} = \{a_i\} \cup \{b_j\}$  define the Voronoi region,

$$v(u) = \{x | [x, u] \leq \min_k \{[x, u_k]\}\}$$

and let  $v^\circ(u)$  denote the interior of the region.

Note that  $v(u)$  consists of all points in  $\mathbb{R}^d$  closest to  $u$ . If we let  $h(u_k, u_j)$  be the closed half-space of  $P(u_k, u_j)$  that contains  $u_k$  then

$$v(u_k) = \bigcap_{j \neq k} h(u_k, u_j).$$

In particular the boundary of  $v(u_k)$  consists of segments of hyperplanes determined by the  $h(u_k, u_j)$ s. Thus  $v(a_i) \cap v(b_j)$  either consists of a segment of  $P(a_i, b_j)$  or it is empty.

*Lemma 3.* The n-n surface consists of

$$\bigcup_{i,j} \{v(a_i) \cap v(b_j)\}.$$

*Proof.* Let  $D$  be the n-n surface. If  $x \in D$  then there exist  $a_k$  and  $b_l$  such that

$$[x, a_k] = \min_i \{[x, a_i]\} = \min_j \{[x, b_j]\} = [x, b_l].$$

It follows that  $x \in v(a_k) \cap v(b_l)$ . On the other hand if  $x \in v(a_k) \cap v(b_l)$  for some  $a_k$  and  $b_l$  then

$$\min_i \{[x, a_i]\} = [x, a_k] = [x, b_l] = \min_j \{[x, b_j]\},$$

which implies that  $x \in D$ .

We now state a theorem which specifies conditions for a hyperplane to be part of the n-n surface.

*Theorem 3.* A hyperplane  $D$  is part of the n-n surface if and only if

- (i) there exists an unlike pair  $a$  and  $b$  such that  $D = P(a, b)$ , and
- (ii) for arbitrary  $a_i$  either there exists a  $b_j$  such that  $D = P(a_i, b_j)$  or  $v^\circ(a_i) \cap D = \phi$ .

*Proof.* First assume that  $D$  is part of the n-n surface. Condition (i) is then obvious. To show (ii) let  $a_i$  be arbitrary. If there exists  $b_j$  such that  $v(a_i) \cap v(b_j) \neq \phi$  then  $v(a_i) \cap v(b_j)$  is a segment of  $P(a_i, b_j)$ . Since by lemma 3  $v(a_i) \cap v(b_j)$  is also a segment of the n-n surface, it follows that  $P(a_i, b_j) = D$  or  $v^\circ(a_i) \cap D = \phi$ . If on the other hand  $v(a_i) \cap v(b_j) = \phi$  for all  $j$  then by lemma 3  $v(a_i) \cap D = \phi$ .

Next assume (i) and (ii) are satisfied. We show for all  $x \in D$  that

$$\min_i \{[x, a_i]\} = \min_j \{[x, b_j]\}.$$

The proof is by contradiction. Suppose for some  $x \in D$  there exists  $a_k$  such that

$$[x, a_k] = \min_i \{[x, a_i]\} < \min_j \{[x, b_j]\}.$$

This implies  $x \in v^\circ(a_k)$  so  $v^\circ(a_k) \cap D \neq \phi$ . However, the strict inequality implies  $D \neq P(a_k, b_j)$  for all  $j$ , so by (ii)

$v^\circ(a_k) \cap D = \phi$ , a contradiction. *Q.E.D.*

If we examine Fig. 5 in light of Theorem 3 we see that for  $a_3$  there is no  $b_j$  for which  $D = P(a_3, b_j)$  but  $v^\circ(a_3) \cap D \neq \phi$  so  $D$  cannot be the n-n surface. If  $a_3, a_4, b_3, b_4$  are translated so their Voronoi regions do not intersect  $D$ , yet the weighting at  $\lambda$  is satisfied, then  $D$  is the identity surface. Finally, for inclusion in the n-n surface, a hyperplane  $D$  has to satisfy the condition that if  $a$  and  $b$  are the nearest unlike pair on any line normal to  $D$  then either  $v(a) \cap D = v(b) \cap D = \phi$  or  $D = P(a, b)$ .

## 5. IDENTITY DETERMINATION

A procedure for determining whether a given set of points yield identical n-n and pf surfaces can now be formulated. Essentially, the identity surface has to consist of a finite set of hyperplanes of infinite extent, each of which has to be a n-n surface hyperplane as specified by Theorem 3 and also satisfy the weighting condition along each line perpendicular to it as specified by Theorem 2.

To begin, a set of candidate hyperplanes  $D_1, D_2, \dots, D_p$  are obtained as follows.  $D_1 = P(a, b)$  where

$$[a, b] = \min_{i,j} \{[a_i, b_j]\}.$$

The approach of using an exhaustive search to determine the nearest unlike pair  $(a, b)$  requires  $n^2/4$  distances to be computed; however recently-developed algorithms have time complexity  $O(n \log^{d-1} n)$ <sup>(12)</sup> and  $O(n \log n)$ <sup>(13)</sup>. Let the feature space partition due to  $D_1$  be  $H_1 = \{H_1^A, H_1^B\}$ , where  $H_1^A = h(a, b)$  and  $H_1^B = h(b, a)$ . If there are points of more than one class in any block of partition  $H_1$ , determine  $D_2$  as the perpendicular bisector of the nearest such pair, giving the partition refinement  $H_2 = H_1 \cap \{H_2^A, H_2^B\}$ . This process is continued until  $D_p$  with partition  $H_{p-1} \cap \{H_p^A, H_p^B\}$  separates the classes. It is clear that segments of each of the resulting  $D_j$  have to be part of the n-n surface. Furthermore, when identity holds the  $D_j$  are the only members of the identity surface.

Algorithm 1 tests whether a candidate hyperplane  $D$  determined by the above method satisfies symmetry, weighting and Voronoi conditions. First a check is made to determine if the sufficient condition of symmetry is satisfied. If so the algorithm terminates, otherwise point sets  $B_k$  that lie on parallel planes  $R_k$  orthogonal to  $D$  are determined. Subsets of  $B_k$  that lie on lines perpendicular to  $D$  are tested to determine whether the weighting condition holds and if points nearest to  $D$  have their Voronoi regions intersect  $D$ .

*Algorithm 1.* Hyperplane Identity

- 1(a) (obtain matched pairs). Determine the partition  $(U_1, U_2)$  of  $U$  as
  - $U_1 = \{x | P(x, y) = D, x \in C_i, y \in C_j, i \neq j\}$  and  $U_2 = U - U_1$ ;
- 1(b) (symmetry?). If  $U_2 = \phi$ ,  $D$  is an identity hyper-

plane; else determine hyperplane  $R_1$  that is orthogonal to  $D$  and passes through a pair  $x, y \in U$  such that  $P(x, y) = D$ ;

- 1(c) [initialize for loop]. Set  $k = 1$  and  $B_1 = B'_1 = \{x | x \in U \text{ and } x \text{ lies on } R_1\}$ ;
- 1(d) [obtain points on line]. For  $u \in B_k$  determine the set  $S(u) = \{x | x \in B_k \text{ and } x \text{ lies on } L(u), \text{ the normal from } u \text{ to } D\}$ ;
- 1(e) [necessary conditions]. If the following two constraints do not hold,  $D$  is not an identity hyperplane,

weighting:

$$\sum_{y \in A \cap S(u)} K([y, \gamma]) = \sum_{z \in B \cap S(u)} K([z, \gamma])$$

Voronoi: either  $v^{\circ}(a) \cap D = v^{\circ}(b) \cap D = \phi$  or  $[a, \gamma] = [b, \gamma]$  where  $\gamma = L(u) \cap D$ ,  $[a, \gamma] = \min \{[y, \gamma] | y \in A \cap S(u)\}$  and  $[b, \gamma] = \min \{[z, \gamma] | z \in B \cap S(u)\}$ ;

- 1(f) [next line]. Set  $B_k = B_k - S(u)$ . If  $B_k \neq \phi$  go to 1(d); else increment  $k$  by one, and determine

$$B'_k = B_k = x, x \in \{U - B'_{k-1}, \dots, -B'_1\},$$

and

$$[x, R_{k-1}] = \min_{x \in U} \{[x, R_{k-1}]\}$$

where  $[x, R_k] = \min \{[x, y] | y \in R_k\}$ ;

- 1(g) [next plane?]. If  $B_k = \phi$ ,  $D$  is an identity hyperplane; else determine hyperplane  $R_k$  containing points in  $B_k$  and parallel to  $R_{k-1}$ , and go to 1(d).

If any candidate hyperplane is not an identity hyperplane then identity does not hold. On the other hand when each candidate hyperplane is an identity hyperplane, the two surfaces are identical.

## 6. RESULTS AND DISCUSSION

When samples satisfy certain location constraints the n-n and pf decision surfaces are identical. Symmetry of samples with respect to a set of hyperplanes is sufficient to determine identity. A weighting condition is necessary for a n-n decision hyperplane to be a pf decision hyperplane. The Voronoi regions of points provides a necessary and sufficient condition for a hyperplane to belong to the n-n decision surface. These conditions form the basis for an effective procedure to determine identity.

Decision surface identity is sufficient for equal error rates independent of the types of underlying probability distributions. If one restricts consideration to specific types of distributions it is then possible to have equal error rates but non-identical decision surfaces.

In direct implementation, the pf and n-n decision rules are similar in terms of the number of algebraic operations per decision, with the exception of the need to compute the potential function of every sample

distance by the pf rule. Thus when identity is determined the n-n rule may be considered to be somewhat better. We have also quantitatively demonstrated that due to restrictions placed on the points, identity cannot be expected with random samples.

The need for developing theoretical guidelines for classifier comparison has been felt by pattern recognition researchers.<sup>(14)</sup> A solution to the meta-decision problem of analytically determining whether two decision rules are identical or not may be viewed as an aid in this context. They are also a source of challenging problems which can be tackled using the developing tools of computational geometry.

## REFERENCES

1. E. Fix and J. L. Hodges, Jr., Discriminatory analysis: nonparametric discrimination: consistency properties. In *Machine Recognition of Patterns*, A. K. Agrawala, ed., IEEE Press, New York (1977).
2. T. M. Cover and P. E. Hart, Nearest-neighbor pattern classification, *IEEE Trans. Inf. Theory* IT-13, 21-27 (1967).
3. D. A. Bashkurov, E. M. Braverman and I. B. Muchnik, Potential function algorithms for pattern recognition learning machines, *Automn remote Control* 25, 629-631 (1964).
4. M. A. Aizermann, Remarks on two problems connected with pattern recognition. In *Methodologies of Pattern Recognition*, S. Watanabe, ed., pp. 1-10. Academic Press, New York (1969).
5. T. M. Cover and T. J. Wagner, Topics in statistical pattern recognition. In *Communications and Cybernetics*, Vol. 10, K. S. Fu, ed. Springer, Berlin (1976).
6. B. G. Batchelor, A comparison of the decision surfaces of the nearest-neighbor and potential function classifiers, *Inf. Sci.* 5, 171-178 (1973).
7. S. N. Srihari, Pattern classification with composite classes, Proc. IEEECS Conf. on Pattern Recognition and Image Processing, Chicago, IL, pp. 173-180, May (1978).
8. S. N. Srihari, L. J. White and T. Snabb, Identity conditions for nearest-neighbor and potential function classifiers, *Inf. Sci.* 19, 21-31 (1979).
9. D. F. Specht, Generation of polynomial discriminant functions for pattern recognition, *IEEE Trans. Electron. Comput.* EC-16, 308-319 (1967).
10. B. Dasarathy and L. J. White, A characterization of nearest-neighbor rule decision surfaces and a new approach to generate them. *Pattern Recognition* 10, 41-46 (1978).
11. M. I. Shamos and D. Hoey, Closest-point problems, Proc. 16th Ann. IEEE Symp. on Foundations of Computing, pp. 151-162 (1975).
12. J. L. Bentley and M. I. Shamos, Divide-and-conquer in multi-dimensional space, Proc. 8th Symp. on Theory of Computing, pp. 220-230 (1976).
13. J. E. Zolnowsky, Topics in computational geometry, SLAC Report 206, Stanford University, May (1978).
14. L. N. Kanal, Patterns in pattern recognition: 1968-1974, *IEEE Trans. Inf. Theory* IT-20, 697-722 (1974).

## APPENDIX

### Proof of Theorem 2

The necessity of weighting condition (4) is considered first. In general all samples will be located on  $r \leq 2n$  lines perpendicular to  $D$ . Let  $\lambda_k, k = 1, \dots, r$  be samples whose

normals  $L(\lambda_k)$  are distinct and exhaustive, and let  $\gamma_k = L(\lambda_k) \cap D$ . Since we assume  $D$  is the pf surface, we have for all  $x \in D$

$$\sum_{k=1}^r \sum_{y \in \lambda_{kA}} K[x, y] = \sum_{k=1}^r \sum_{z \in \lambda_{kB}} K[x, z], \quad (5)$$

where  $\lambda_{kA}$  and  $\lambda_{kB}$  are  $u_A$  and  $u_B$  with  $u = \lambda_k$ . Defining the constants

$$\alpha_k = \sum_{y \in \lambda_{kA}} K[\gamma_k, y]$$

and

$$\beta_k = \sum_{z \in \lambda_{kB}} K[\gamma_k, z]$$

the necessity of (4) follows by showing  $\alpha_k = \beta_k$ . Orthogonality of  $L(\lambda_k)$  to  $D$  implies that for  $x \in D$ ,  $w \in U \cap L(\lambda_k)$ ,  $[x, w] = [x, \gamma_k] + [\gamma_k, w]$  which implies  $K[x, w] = K[x, \gamma_k] \cdot K[\gamma_k, w]$ . Thus (5) reduces to

$$\sum_{k=1}^r \alpha_k K[x, \gamma_k] = \sum_{k=1}^r \beta_k K[x, \gamma_k].$$

If  $\delta_k = \alpha_k - \beta_k$  then

$$\sum_{k=1}^r \delta_k K[x, \gamma_k] = 0 \quad \forall x \in D.$$

Since the  $\gamma_k$  are distinct functions,  $K[x, \gamma_k]$  are independent, which implies  $\delta_k = 0$ , or  $\alpha_k = \beta_k$ .

Next we need to show that when (4) holds and  $D$  is the n-n surface, point  $d_1 \in D$  implies  $d_1 \in D'$ , the pf surface; and point  $d_2 \notin D$  implies  $d_2 \notin D'$ . Since for  $x \in \lambda_{kA} \cup \lambda_{kB}$ ,  $[x, \gamma] = [x, \gamma] + [\gamma, d_1]$  for any sample  $u$

$$\sum_{y \in u_A} K[y, d_1] = \sum_{z \in u_A} K[z, d_1]. \quad (6)$$

By considering equation (6) for every sample,

$$\sum_{y \in A} K[y, d_1] = \sum_{y \in B} K[z, d_1],$$

or  $d_1 \in D'$ .

Let  $\lambda \in A$  and  $\lambda \neq d_2$ . We will consider two cases, first  $L(d_2) \neq L(\lambda)$ . Let  $t = L(d_2) \cap D$  and  $\gamma = L(\lambda) \cap D$ . For any  $y \in \lambda_A$  an application of the law of cosines yields  $[y, d_2] < [y, t] + [t, d_2]$ . Since  $[y, t] = [y, \gamma] + [\gamma, t]$  we have the inequality  $[y, d_2] < [y, \gamma] + [\gamma, t] + [t, d_2]$ . Next let  $z \in \lambda_B$ . Noting that  $z$  is on the  $C_2$  side of  $D$  we again apply the law obtaining  $[z, d_2] > [z, t] + [t, d_2]$  and finally, as above  $[z, d_2] > [z, \gamma] + [\gamma, t] + [t, d_2]$ . It follows that

$$\sum_{y \in \lambda_A} K[y, d_2] > Q \cdot \sum_{y \in \lambda_A} K[y, \gamma] \quad (7)$$

and

$$\sum_{z \in \lambda_B} K[z, d_2] > Q \cdot \sum_{z \in \lambda_B} K[z, \gamma]$$

where  $Q = K[\gamma, t] + K[t, d_2]$ . Combining inequalities (7) and (4) yields

$$\sum_{y \in \lambda_A} K[y, d_2] > \sum_{z \in \lambda_B} K[z, d_2]. \quad (8)$$

For the second case, assume  $L(d_2) = L(\lambda)$  and  $t = \gamma$ . For any  $y \in \lambda_A$  and  $z \in \lambda_B$  we have  $[y, d_2] < [y, t] + [t, d_2]$  and  $[z, d_2] > [z, t] + [t, d_2]$  which, using the same approach as in the first case, yields equation (8). Since equation (8) is true of every sample  $\lambda$ , it follows that  $d_2$  belongs to the  $C_1$  side of  $D'$ . Similarly  $d_2$  on the  $C_2$  side of  $D$  is on the  $C_2$  side of  $D'$ .

**About the Author** – SARGUR N. SRIHARI was born in Bangalore, India on 7 May 1950. After obtaining a B.E. degree in Electrical Communication Engineering from the Indian Institute of Science, Bangalore, in 1970, he joined The Ohio State University, Columbus, where he graduated with the M.S. and Ph.D. degrees in Computer and Information Science, in 1972 and 1976 respectively. His doctoral thesis was in the design and analysis of implementation algorithms for statistical pattern recognition. During the period 1976–1978 he was an Assistant Professor in the Computer Science Section of the Department of Mathematics at Wayne State University, Detroit, Michigan. Since the autumn of 1978 he has been an Assistant Professor in the Department of Computer Science at the State University of New York at Buffalo, where he teaches and does research in pattern recognition, image processing, artificial intelligence and computed tomography. At present he is working on the development of multidimensional image analysis procedures for computed tomography and on the design of knowledge-based contextual algorithms for text recognition. Dr. Srihari is a member of IEEE, ACM, AAAI and The Pattern Recognition Society.

**About the Author** – THOMAS E. SNABB was born in Highland Park, Michigan, in 1945. He received the B.S., M.A. and Ph.D. degrees from Wayne State University, Detroit, Michigan, in 1968, 1970 and 1976 respectively, all in mathematics.

He has been an assistant professor in the Department of Mathematics and Statistics at the University of Michigan–Dearborn since 1976. His research interests are in the areas of pattern recognition, statistics and algebra. Dr. Snabb is a member of the Institute of Mathematical Statistics.

**About the Author** – L. J. WHITE received the BSEE degree in electrical engineering from the University of Cincinnati in 1962, and the Ph.D. degree in electrical engineering from the University of Michigan in 1967. He is currently Professor and Chairman of the Department of Computer and Information Science at The Ohio State University, and holds a joint appointment in the Department of Electrical Engineering. His current research interests deal with the analysis of algorithms and software analysis and testing. He has published in the areas of pattern recognition, automatic document classification, combinatorial computing and graph theory. He has served as a consultant for the Monsanto Research Laboratory, Rockwell International Corporation, and a number of other industrial firms. Dr. White is a member of the IEEE Computer Society, ACM, SIAM, and Sigma Xi.