Identity Conditions for Nearest-Neighbor and Potential-Function Classifiers*

SARGUR N. SRIHARI

Department of Computer Science, SUNY at Buffalo, Amherst, New York 14226

LEE J. WHITE

Department of Computer and Information Science, The Ohio State University, Columbus, Ohio 43210

and

THOMAS SNABB

Department of Mathematics and Statistics, University of Michigan, Dearborn, Michigan 48128

Communicated by John M. Richardson

ABSTRACT

The nearest-neighbor rule and the potential-function classifier are nonparametric discrimination methods that require the storage of a set of sample patterns. Here, a relationship between the two methods in terms of subclasses and superclasses is developed. Considering an exponential potential function, necessary and sufficient conditions for identity of their decision surfaces are obtained. Based on these conditions, an algorithm for establishing identity is introduced.

1. INTRODUCTION

The nearest-neighbor decision rule [1, 2] and the potential-function classifier [3, 4] are two nonparametric classification methods. There exists little published analytical work concerning conditions under which the performances of the two classifiers are identical, with the exception of a heuristic comparison of their decision surfaces given in [5]. Here we demonstrate a

*This research was supported in part by AFOSR 74-2611.

 0020-0255/79/07021-11$01.75

relationship between the two methods, and obtain conditions under which their two-class decision surfaces are identical. These conditions provide the basis for an algorithm that determines from the design samples whether the two classifiers will result in identical decision surfaces.

2. PRELIMINARIES

Consider the discrimination problem with classes C_1 and C_2, where class C_i has n subclasses C_{ij}. Let π_i denote the prior probability of C_i, π_{ij} the prior probability of subclass C_{ij} when C_i is true, and $p_{ij}(\mathbf{x})$ the subclass-conditional probability density function of the d-component pattern $\mathbf{x}$.

Let D_1 and D_2 be two parametric decision rules designed for the above problem as follows. D_1 assigns $\mathbf{x}$ to the class associated with the *subclass* with the maximum *a posteriori* probability, and D_2 assigns $\mathbf{x}$ to the *class* with the maximum *a posteriori* probability. That is, D_1 chooses class C_i corresponding to $\max_{i,j}\{g_{ij}(\mathbf{x})\}$, and D_2 chooses class C_i corresponding to $\max_i\{\Sigma_j g_{ij}(\mathbf{x})\}$, where $g_{ij}(\mathbf{x})=p_{ij}(\mathbf{x})\cdot\pi_{ij}\cdot\pi_i$. In general, D_1 and D_2 have different decision boundaries, as shown in Fig. 1 for particular univariate Gaussian $p_{ij}(\mathbf{x})$ with mean $\mathbf{a}_{ij}$, and a common variance. The locations of $\mathbf{a}_{ij}$ are such that D_1 assigns the interval (d_2,d_3) to C_1 [Fig. 1(a)] and D_2 assigns it to C_2 [Fig. 1(b)].

D_1 and D_2 yield two different methods of nonparametric classification if we let g_{ij} be a function of sample $\mathbf{a}_{ij}$, as

$$g_{ij}(\mathbf{x})=K[\mathbf{x},\mathbf{a}_{ij}]=\exp(-h\cdot[\mathbf{x},\mathbf{a}_{ij}]), \tag{1}$$

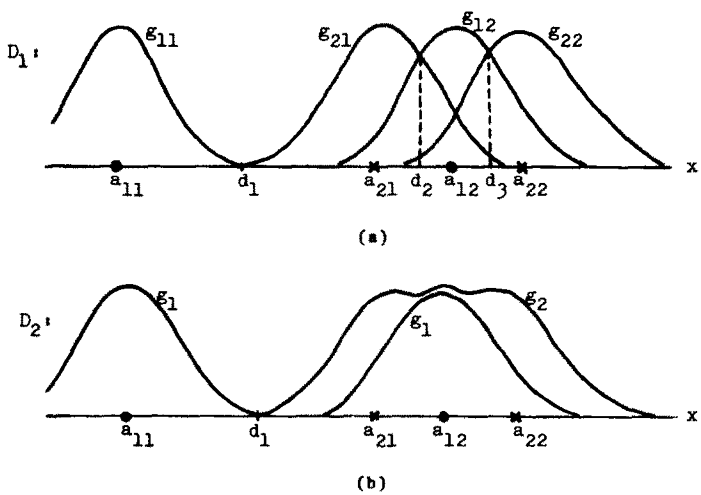

Fig. 1. Example of nonidentity of decision rules D_1 and D_2.

where $[\mathbf{x},\mathbf{y}]=\|\mathbf{x}-\mathbf{y}\|^2$, and h is a positive scalar. D_1 reduces to the *nearest-neighbor* (n−n) rule of associating an unknown $\mathbf{x}$ with the class label of its nearest sample, having decision surface

$$\min_j \{[\mathbf{x},\mathbf{a}_{1j}]\} = \min_j \{[\mathbf{x},\mathbf{a}_{2j}]\}, \tag{2}$$

which is piecewise linear due to the discontinuous choice function min. The *potential function* (pf) classifier computes the potential at $\mathbf{x}$ as $\Sigma_j K[\mathbf{x},\mathbf{a}_{ij}]$, where K is a potential function that varies inversely with its argument, and associates $\mathbf{x}$ with the class with larger potential [6]. Thus D_2 reduces to the pf decision rule having a decision surface with the continuously differentiable form

$$\sum_j K[\mathbf{x},\mathbf{a}_{1j}] = \sum_j K[\mathbf{x},\mathbf{a}_{2j}]. \tag{3}$$

As $h\to\infty$ the pf surface defined by (3) and (1) approaches the n−n surface [7]; thus we restrict ourselves to the case of finite h.

3. IDENTITY CONDITIONS

We develop here conditions on sample patterns $\{\mathbf{a}_i\}\in C_1$, $\{\mathbf{b}_i\}\in C_2$ for identity of decision surfaces of the n−n and pf classifiers, where the results obtained pertain to the general form of K as well as to its exponential form, as indicated.

In the case $n=1$, or C_1 is characterized by $\mathbf{a}_1$ and C_2 by $\mathbf{b}_1$ in R^d, the surfaces of the two classifiers are identical to the hyperplane that orthogonally bisects the line segment joining $\mathbf{a}_1$ and $\mathbf{b}_1$, or

$$(\mathbf{a}_1-\mathbf{b}_1)^t\left(\mathbf{x}-\frac{\mathbf{a}_1+\mathbf{b}_1}{2}\right)=0,$$

thereby making identity independent of the samples. In the case $n=2$ we consider the two cases where all patterns are and are not on a single line.

Lemma 1. *If* $\{\mathbf{a}_i\}\in C_1$, $\{b_i\}\in C_2$, $i=1,2$, *are points on a line, the n−n and pf boundaries are identical iff the points are linearly separable and* $[\mathbf{a}_1,\mathbf{a}_2]=[\mathbf{b}_1,\mathbf{b}_2]$.

Proof. First assume identical boundaries. The points can be in one of three distinct configurations: $\mathbf{a}_2\mathbf{a}_1\mathbf{b}_1\mathbf{b}_2$, $\mathbf{a}_1\mathbf{b}_1\mathbf{b}_2\mathbf{a}_2$, and $\mathbf{a}_1\mathbf{b}_1\mathbf{a}_2\mathbf{b}_2$, the other permutations being identical to one of these except for labeling. With either $\mathbf{a}_1\mathbf{b}_1\mathbf{b}_2\mathbf{a}_2$ or $\mathbf{a}_1\mathbf{b}_1\mathbf{a}_2\mathbf{b}_2$, one of the n-n points is $\mathbf{d}_1=\frac{1}{2}(\mathbf{a}_1+\mathbf{b}_1)$, which is not a pf point, since $[\mathbf{d}_1,\mathbf{a}_1]$, $[\mathbf{d}_1,\mathbf{b}_1]$, and $[\mathbf{d}_1,\mathbf{a}_2]\neq[\mathbf{d}_1,\mathbf{b}_2]$. Thus the only possibility is $\mathbf{a}_2\mathbf{a}_1\mathbf{b}_1\mathbf{b}_2$, which is linearly separable. With this configuration, let $\mathbf{d}_1$ and $\mathbf{d}_2$ be the respective n–n and pf points. If boundaries are identical, $\mathbf{d}_1=\mathbf{d}_2=\frac{1}{2}(\mathbf{a}_1+\mathbf{b}_1)$. Since $K[\mathbf{d}_1,\mathbf{a}_1]+K[\mathbf{d}_1,\mathbf{a}_2]=K[\mathbf{d}_1,\mathbf{b}_1]+K[\mathbf{d}_1,\mathbf{b}_2]$, it follows that $[\mathbf{a}_2,\mathbf{a}_1]=[\mathbf{b}_2,\mathbf{b}_1]$.

To prove sufficiency, let $[\mathbf{a}_2,\mathbf{a}_1]=[\mathbf{b}_2,\mathbf{b}_1]=k^2$ for the linearly separable configuration. Thus for point $\mathbf{d}_2$, $K[\mathbf{d}_2,\mathbf{a}_1]+K[\mathbf{d}_2,\mathbf{a}_1-k]=K[\mathbf{d}_2,\mathbf{b}_1]+K[\mathbf{d}_2,\mathbf{b}_1+k]$, a solution to which is $\mathbf{d}_2=\frac{1}{2}(\mathbf{a}_1+\mathbf{b}_1)=\mathbf{d}_1$. If $\mathbf{d}'$ is another solution, then w.o.l.o.g. assume $|\mathbf{d}'-a_1|<|\mathbf{d}'-b_1|$, which implies $K[\mathbf{d}',\mathbf{a}_1]>K[\mathbf{d}',\mathbf{b}_1]$ and $K[\mathbf{d}',\mathbf{a}_2]>K[\mathbf{d}',\mathbf{b}_2]$, which contradicts $\mathbf{d}'$ as a decision point; thus the solution is unique and boundaries are identical.

A corollary of Lemma 1 is that if $n=2$ with linearly separable collinear points such that for each $\mathbf{a}_i$ there exists a unique $\mathbf{b}_i$ equidistant from the n–n hyperplane D, then the pf surface is also D. This corollary can be generalized to $n>2$ and noncollinear points. Lemma 2 considers $n=2$ and points in general position, i.e., not all on a single line. The different cases are illustrated for R^2 in Fig. 2.

LEMMA 2. *If* $\{\mathbf{a}_i\}\in C_1$, $\{\mathbf{b}_i\}\in C_2$, $i=1,2$ *are in general position, the* n-n *and* pf *surfaces are identical iff* (i) $[\mathbf{a}_1,\mathbf{a}_2]=[\mathbf{b}_1,\mathbf{b}_2]$, (ii) $[\mathbf{a}_2,\mathbf{b}_1]=[\mathbf{a}_1,\mathbf{b}_2]$, (iii) $\mathbf{a}_1+\mathbf{b}_1\neq\mathbf{a}_2+\mathbf{b}_2$, *and either* (iv) $\mathbf{w}'(\mathbf{a}_2-\mathbf{r})>0>\mathbf{w}'(\mathbf{b}_2-\mathbf{r})$, *or* (v) $[\mathbf{a}_1,\mathbf{b}_1]=[\mathbf{a}_2,\mathbf{b}_2]$, *where* $\mathbf{w}=\mathbf{a}_1-\mathbf{b}_1$, $\mathbf{r}=\frac{1}{2}(\mathbf{a}_1+\mathbf{b}_1)$, *and* $[\mathbf{a}_1,\mathbf{b}_1]=\min_j[\mathbf{a}_1,\mathbf{b}_j]$.

Proof. The conditions are clearly sufficient, we show only necessity assuming surface identity. Due to the pf surface being continuously differentiable, the identity surface has to consist of hyperplanes of infinite extent, one of which is the perpendicular bisector D_1 of the line segment joining the nearest unlike pair, say $(\mathbf{a}_1,\mathbf{b}_1)$. Since for each point $\mathbf{x}$ on $\mathbf{w}'(\mathbf{x}-\mathbf{r})=0$, $[\mathbf{a}_1,\mathbf{x}]=[\mathbf{b}_1,\mathbf{x}]$, it follows that $[\mathbf{a}_2,\mathbf{x}]=[\mathbf{b}_2,\mathbf{x}]$ for identity. Thus $\mathbf{w}'(\mathbf{x}-\mathbf{r})=0$ is the perpendicular bisector of segment $\mathbf{a}_2\mathbf{b}_2$, from which (i)–(iii) follow.

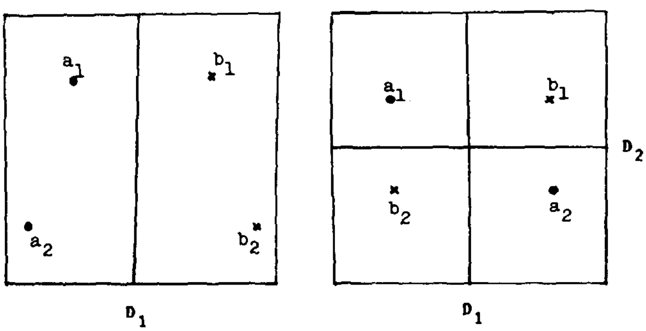

Fig. 2. Identity in R^2 with two references in each class.

If the identity surface consists of only D_1, then points are linearly separable, or (iv) holds. Otherwise the identity surface consists of two hyperplanes, one of which is D_1 and the other is the perpendicular bisector D_2 of segments $\mathbf{a}_1\mathbf{b}_2$ and $\mathbf{a}_2\mathbf{b}_1$ from which (v) follows.

A corollary of Lemma 2 provides the case where points are located on a line in R^d such that the condition of Lemma 1 is satisfied. 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 we introduce a point $\mathbf{a}^1$ into one of the C_1 regions (half space or quarter space) defined by the separating hyperplanes, then by introducing points $\mathbf{b}^1$, $\mathbf{b}^2$ at the mirrored positions of $\mathbf{a}^1$ with respect to D_1 and D_2, and a point $\mathbf{a}^2$ at the mirrored position of $\mathbf{b}^2$ in D_2, we maintain surface identity. Symmetric location of points with respect to hyperplanes provides a sufficient condition for identity, as expressed in Theorem 1.

THEOREM 1. *Let the n-n surface with samples* $\{\mathbf{a}_i\}\cup\{\mathbf{b}_i\}$ *consist of p hyperplanes* $D_j:\mathbf{w}_j^t(\mathbf{x}-\mathbf{r}_j)=0$ $(j=1,\ldots,p)$. *If for each* $\mathbf{a}^k\in\{\mathbf{a}_i\}$, *and* D_j *there exists a unique* $\mathbf{b}^m\in\{\mathbf{b}_i\}$ *satisfying* $\mathbf{w}_j=(\mathbf{a}^k-\mathbf{b}^m)$, *and* $\mathbf{r}_j=\frac{1}{2}(\mathbf{a}^k+\mathbf{b}^m)$, *then the n-n and pf surfaces are identical.*

Proof. For each point $\mathbf{x}$ on D_j, the given condition implies that for every $\mathbf{a}^k$ there is a unique $\mathbf{b}^m$ such that $K[\mathbf{x},\mathbf{a}^k]=K[\mathbf{x},\mathbf{b}^m]$; thus $\mathbf{x}$ is a point of the pf surface. These are the only points on the pf surface, due to continuity and single-valuedness of the pf surface between the pair of nearest unlike points determining D_j.

Theorem 1 implies that if the samples can be partitioned into disjoint subsets $S_1,\ldots,S_c$, where each S_i has $2p$ points $\{\mathbf{a}^1,\ldots,\mathbf{a}^p,\mathbf{b}^1,\ldots,\mathbf{b}^p\}$, and for each D_j and each $\mathbf{a}^k\in S_i$ there exists a unique $\mathbf{b}^m\in S_i$ such that D_j is the perpendicular bisector of the line segment joining $\mathbf{a}^k$ and $\mathbf{b}^m$, then the n–n and pf surfaces are identical.

If we restrict ourselves to the exponential form $K(u)=e^{-u}$, we can exhibit nonsymmetric sample locations for which surfaces are identical. These examples will be based on the following necessary and sufficient conditions for surface identity when the n–n surface is a single hyperplane.

THEOREM 2. *If the n-n surface with samples* $\{\mathbf{a}_{ij}\}\in C_i$, $i=1,2, j=1,\ldots,n$ *is a hyperplane D, then identity with the pf surface with* $K(u)\triangleq e^{-u}$ *holds iff for every* $\boldsymbol{\lambda}\in\{\mathbf{a}_{ij}\}$

$$\sum_{\mathbf{y}\in S_1(\boldsymbol{\lambda})} K[\mathbf{y},\boldsymbol{\gamma}]=\sum_{\mathbf{z}\in S_2(\boldsymbol{\lambda})} K[\mathbf{z},\boldsymbol{\gamma}], \tag{4}$$

where $L(\boldsymbol{\lambda})$ *is a normal from* $\boldsymbol{\lambda}$ *to D,* $\boldsymbol{\gamma}$ *is the intersection point of* $L(\boldsymbol{\lambda})$ *and D, and* $S_i(\boldsymbol{\lambda})=[\mathbf{v}|\mathbf{v}\in\{\mathbf{a}_{ij}\}$, $\mathbf{v}$ *lies on* $L(\boldsymbol{\lambda})]$.

Proof. The necessity of (4) is considered first. In general all samples will be located on $r \leqslant 2n$ lines perpendicular to D. Let λ_k, $k=1,\ldots,r$, be samples whose normals $L(\lambda_k)$ are distinct and exhaustive, and let γ_k be the intersection point of $L(\lambda_k)$ with D. Since we assume D is the pf surface, we have for all $\mathbf{x} \in D$

$$\sum_{k=1}^{r} \sum_{\mathbf{y} \in S_1(\lambda_k)} K[\mathbf{x},\mathbf{y}] = \sum_{k=1}^{r} \sum_{\mathbf{z} \in S_2(\lambda_k)} K[\mathbf{x},\mathbf{z}]. \tag{5}$$

Defining the constants α_k, β_k $(k=1,\ldots,r)$, as

$$\alpha_k \triangleq \sum_{\mathbf{y} \in S_1(\lambda_k)} K[\gamma_k, \mathbf{y}], \beta_k \triangleq \sum_{\mathbf{z} \in S_2(\lambda_k)} K[\gamma_k, \mathbf{z}],$$

the necessity of (4) follows by showing $\alpha_k = \beta_k$. To do so, observe that orthogonality of $L(\lambda_k)$ to D implies that for $\mathbf{x} \in D$, $\mathbf{w} \in S_i(\lambda_k)$, we have $[\mathbf{x},\mathbf{w}] = [\mathbf{x},\gamma_k] + [\gamma_k,\mathbf{w}]$, which implies $K[\mathbf{x},\mathbf{w}] = K[\mathbf{x},\gamma_k] \cdot K[\gamma_k,\mathbf{w}]$. Thus (5) reduces to

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

If $\delta_k \triangleq \alpha_k - \beta_k$, then

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

Since γ_k are distinct, the function $K[\mathbf{x},\gamma_k]$ 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, the pf surface is identical to D. The proof will be to show that an arbitrary point $\boldsymbol{\theta}_1$ on D must also belong to the pf surface, and that a point $\boldsymbol{\theta}_2$ not on D cannot lie on the pf surface.

Since for $\mathbf{x} \in S_i(\lambda)$, $[\mathbf{x},\boldsymbol{\theta}_1] = [\mathbf{x},\gamma] + [\gamma,\boldsymbol{\theta}_1]$, for any sample λ

$$\sum_{\mathbf{y} \in S_1(\lambda)} K[\mathbf{y},\boldsymbol{\theta}_1] = \sum_{\mathbf{z} \in S_2(\lambda)} K[\mathbf{z},\boldsymbol{\theta}_1].$$

By considering equations corresponding to every sample,

$$\sum_{\mathbf{y} \in C_1} K[\mathbf{y},\boldsymbol{\theta}_1] = \sum_{\mathbf{z} \in C_2} K[\mathbf{z},\boldsymbol{\theta}_1];$$

thus $\boldsymbol{\theta}_1$ belongs to the pf surface.

Let $\boldsymbol{\lambda} \neq \boldsymbol{\theta}_2$ be an arbitrary sample on the C_1 side of D. We will consider two cases.

First, let $L(\boldsymbol{\theta}_2) \neq L(\boldsymbol{\lambda})$. Let $\mathbf{t}$ and $\boldsymbol{\gamma}$ be intersection points of $L(\boldsymbol{\theta}_2)$ and $L(\boldsymbol{\lambda})$ with D. For any $\mathbf{y} \in S_1(\boldsymbol{\lambda})$ an application of the law of cosines yields $[\mathbf{y}, \boldsymbol{\theta}_2] < [\mathbf{y}, \mathbf{t}] + [\mathbf{t}, \boldsymbol{\theta}_2]$. Since $[\mathbf{y}, \mathbf{t}] = [\mathbf{y}, \boldsymbol{\gamma}] + [\boldsymbol{\gamma}, \mathbf{t}]$, we have the inequality $[\mathbf{y}, \boldsymbol{\theta}_2] < [\mathbf{y}, \boldsymbol{\gamma}] + [\boldsymbol{\gamma}, \mathbf{t}] + [\mathbf{t}, \boldsymbol{\theta}_2]$. Next let $\mathbf{z} \in S_2(\boldsymbol{\lambda})$. Noting that $\mathbf{z}$ is on the C_2 side of D, we again apply the law, obtaining $[\mathbf{z}, \boldsymbol{\theta}_2] > [\mathbf{z}, \mathbf{t}] + [\mathbf{t}, \boldsymbol{\theta}_2]$ and finally, as above, we have $[\mathbf{z}, \boldsymbol{\theta}_2] > [\mathbf{z}, \boldsymbol{\gamma}] + [\boldsymbol{\gamma}, \mathbf{t}] + [\mathbf{t}, \boldsymbol{\theta}_2]$. It follows that

$$\sum_{\mathbf{y} \in S_1(\boldsymbol{\lambda})} K[\mathbf{y}, \boldsymbol{\theta}_2] > Q \cdot \sum_{\mathbf{y} \in S_1(\boldsymbol{\lambda})} K[\mathbf{y}, \boldsymbol{\gamma}],$$

and

$$\sum_{\mathbf{z} \in S_2(\boldsymbol{\lambda})} K[\mathbf{z}, \boldsymbol{\theta}_2] < Q \cdot \sum_{\mathbf{z} \in S_2(\boldsymbol{\lambda})} K[\mathbf{z}, \boldsymbol{\gamma}],$$

where $Q = K[\boldsymbol{\gamma}, \mathbf{t}] + K[\mathbf{t}, \theta_2]$. Combining these inequalities with (4) yields

$$\sum_{\mathbf{y} \in S_1(\boldsymbol{\lambda})} K[\mathbf{y}, \boldsymbol{\theta}_2] > \sum_{\mathbf{z} \in S_2(\boldsymbol{\lambda})} K[\mathbf{z}, \boldsymbol{\theta}_2]. \tag{6}$$

For the second case, assume $L(\boldsymbol{\theta}_2) = L(\boldsymbol{\lambda})$ and $\mathbf{t} = \boldsymbol{\gamma}$. For any $\mathbf{y} \in S_1(\boldsymbol{\lambda})$ and $\mathbf{z} \in S_2(\boldsymbol{\lambda})$, we have $[\mathbf{y}, \boldsymbol{\theta}_2] < [\mathbf{y}, \mathbf{t}] + [\mathbf{t}, \boldsymbol{\theta}_2]$ and $[\mathbf{z}, \boldsymbol{\theta}_2] > [\mathbf{z}, \mathbf{t}] + [t, \boldsymbol{\theta}_2]$ which, using the same approach as in the first case, yields (6).

Since (6) is true of every sample $\boldsymbol{\lambda}$, it follows that $\boldsymbol{\theta}_2$ belongs to the C_1 side of the pf surface. Similarly, $\boldsymbol{\theta}_2$ on the C_2 side of the n–n surface is on the C_2 side of the pf surface, which completes the proof.

It can further be 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 the condition (4). Figure 3 demonstrates the types of identity surfaces possible, for the cases $n=3$ and 4, in R^2 and the following configurations:

(i) D consists of a single hyperplane with pairwise point symmetry;

(ii) D consists of a single hyperplane with points located on a single line in the configuration $a_n \cdots a_1 b_1 \cdots b_n$ such that $d_1 = \frac{1}{2}(\mathbf{a}_1 + \mathbf{b}_1)$, $K[\mathbf{a}_i, \mathbf{d}_1] \neq K[\mathbf{b}_i, \mathbf{d}_1]$, $i > 1$, and $\Sigma_{i>1} K[\mathbf{a}_i, \mathbf{d}_1] = \Sigma_{i>1} K[\mathbf{b}_i, \mathbf{d}_1]$;

(iii) D consists of three hyperplanes with a common intersection such that points are alternately located on the corners of a regular hexagon;

(iv) D consists of a pair of orthogonal hyperplanes such that two points are located symmetrically in each quadrant;

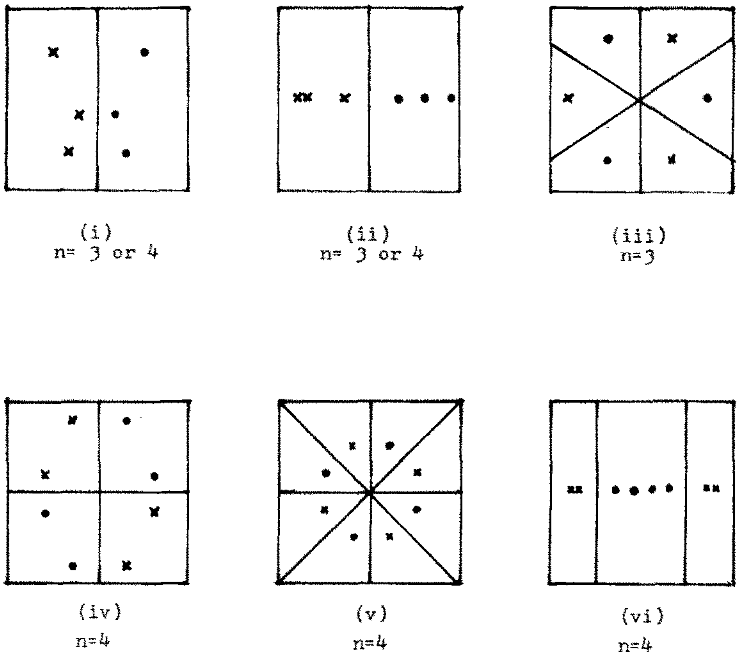

Fig. 3. Identity in R^2 with three or four references in each class.

(v) D consists of four intersecting hyperplanes with a common intersection such that points are located on the alternate corners of a regular octagon;

(vi) D consists of two parallel hyperplanes with points located on a line perpendicular to the two hyperplanes.

The only configurations not satisfying the symmetry sufficient condition of Theorem 1 are cases (ii) and (vi) which satisfy the weighting condition of Theorem 2. The existence of configuration (vi) can be shown as follows. Let points be located on a line in the order $a_2a_1b_1b_2b_3b_4a_4a_3$, with $[a_2,a_1]=[a_4,a_3]=\beta$, $[a_1,b_1]=[b_4,a_4]=2\alpha$, and $[b_1,b_2]=[b_2,b_3]=[b_3,b_4]=\alpha$. The n-n points are thus $d_1=\frac{1}{2}(a_1+b_1)$, $d_2=\frac{1}{2}(b_4+a_4)$. 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

$$[a_2,d_1]=[a_3,d_2], \qquad [a_3,d_1]=[a_2,d_2], \qquad [a_4,d_1]=[a_1,d_2],$$

$$[b_2,d_1]=[b_3,d_2], \qquad [b_3,d_1]=[b_2,d_2], \qquad [b_4,d_1]=[b_1,d_2].$$

The equation has a unique solution; for example, if $\alpha=1$, and $K[x,y]=e^{-|x-y|^2}$, then $\beta\approx 0.099832$ satisfies the equation.

4. DETERMINING IDENTITY

A procedure for determining whether a given set of design samples yield identical n–n and pf surfaces can be formulated. Essentially, the identity surface has to consist of a finite set of hyperplanes of infinite extent, each of which satisfies the condition of equally weighted points along each line perpendicular to it.

A set of candidate hyperplanes $H_0, H_1, \ldots, H_c$ are obtained as follows. H_0 is the perpendicular bisector of the nearest unlike pair, i.e., $H_0:(\mathbf{a}-\mathbf{b})'[\mathbf{x}-\frac{1}{2}(\mathbf{a}+\mathbf{b})]=0$, where $[\mathbf{a},\mathbf{b}]=\min_{i,\,j}[\mathbf{a}_i,\mathbf{b}_j]$. Let the two closed half spaces separated by H_0 be h_{00} and h_{01}. If there are points of more than one class in either half space, determine H_1 as the perpendicular bisector of the closest unlike pair in either half space. Let h_{10} and h_{11} be the half spaces due to H_1. If the regions $h_{0i}\cap h_{1j}$, $i,j=0,1$, are such that one of them contains a pair of unlike points, obtain H_2 as the perpendicular bisector of the nearest unlike pair, and so on.

Algorithm I tests whether a candidate hyperplane determined by the above method satisfies symmetry and weighting conditions. First a check is made to determine if the symmetry condition of Theorem 1 is satisfied. If it is not, point sets B_k that lie on parallel planes P_k orthogonal to H are determined. Subsets of B_k that lie on lines perpendicular to H are tested to determine if the condition (4) holds.

ALGORITHM I (Hyperplane identity).

I1. Let M be a binary relation on $U=\{\mathbf{a}_i\}\cup\{\mathbf{b}_i\}$ that defines matched pairs of points as

$M=[(\mathbf{a}_i,\mathbf{b}_j)|H$ is the perpendicular bisector of the line segment joining $\mathbf{a}_i$ and $\mathbf{b}_j]$.

Determine partition (S_1,S_2) of U as

$$S_1=\{\mathbf{x},\mathbf{y}|\mathbf{x},\mathbf{y}\in U \text{ and } \mathbf{x}M\mathbf{y}\}, \qquad S_2=U-S_1.$$

I2. If $S_2=\varnothing$, H is a pf hyperplane; else, determine hyperplane P_1 that is orthogonal to H, passes through a pair of points $(\mathbf{x},\mathbf{y})$ such that $\mathbf{x}M\mathbf{y}$, and all points in U lie in the same closed half space of P_1.

I3. Initialize for loop: let $k=1$, and $B_1=B_1'=\{\mathbf{x}|\mathbf{x}\in U$, and $\mathbf{x}$ lies on $P_1\}$.

I4. Let $\lambda\in B_k$. Determine the set $S_k(\lambda)$ defined as

$$S_k(\lambda)=\{\mathbf{x}|\mathbf{x}\in B_k, \text{ and } \mathbf{x} \text{ lies on } L(\lambda), \text{ the normal from } \lambda \text{ to hyperplane } H\}.$$

I5. If the following constraint is not satisfied the surfaces are not identical:

$$\sum_{\mathbf{y}\in C_1\wedge S_k(\lambda)} K[\mathbf{y},\gamma]=\sum_{\mathbf{z}\in C_2\wedge S_k(\lambda)} K[\mathbf{z},\gamma]$$

where γ is the point of intersection of $L(\lambda)$ and H.

I6. Let $B_k=B_k-S_k(\lambda)$. If $B_k\neq\varnothing$ go to I4; else increment k by one, and determine

$$B_k'=B_k=\{\mathbf{x}|\mathbf{x}\in U-B_{k-1}'\cdots,-B_1', \text{ and}$$

$$[\mathbf{x},P_{k-1}]=\min_{\mathbf{x}\in U}[\mathbf{x},P_{k-1}]\}.$$

I7. If $B_k\neq\varnothing$, H is a hyperplane of the pf surface; else determine hyperplane P_k containing points in B_k, and parallel to $P_{k-1},\ldots, P_1$.

I8. Go to I4.

In step I6 the notation $[\mathbf{x},P_k]$ corresponds to the distance between point $\mathbf{x}$ and hyperplane P_k. When each n–n hyperplane is also a pf hyperplane, the two surfaces are identical, due to continuity and single-valuedness of the pf surface between a pair of unlike points.

5. CONCLUDING REMARKS

When samples satisfy certain location constraints the decision surface of the pf classifier is identical to that of the n–n decision rule. We have presented symmetry and weighting conditions for identity of the two surfaces. In direct implementation, the pf classifier tends to involve more algebraic operations per decision than the n–n decision rule based on the same references; thus when identity is determined, the latter method is superior. Due to strict restrictions placed on the samples, identity cannot be expected with random samples. In such a case, due to surface identity being sufficient but not necessary for identical performance, generalization of the criterion from surface identity to error-rate identity may be considered.

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, 1977.
2. T. M. Cover and P. E. Hart, Nearest-neighbor pattern classification, *IEEE Trans. Information Theory* IT-13:21–27 (Jan. 1967).
3. D. A. Bashkirov, E. M. Braverman, and I. B. Muchnik, Potential function algorithms for pattern recognition learning machines, *Automat. Remote Control* 25:629–631 (May 1964).
4. M. A. Aizermann, Remarks on two problems connected with pattern recognition, in *Methodologies of Pattern Recognition* (S. Watanabe, Ed.), Academic, New York, 1969, pp. 1–10.
5. B. G. Batchelor, A comparison of the decision surfaces of the nearest-neighbor and potential function classifiers, *Information Sci.* 5:171–178 (1973).
6. W. S. Meisel, *Computer-Oriented Approaches to Pattern Recognition*, Academic, New York, 1972, pp. 98–119.
7. D. F. Specht, Generation of polynomial discriminant functions for pattern recognition, *IEEE Trans. Elec. Comp.* EC-16:308–319 (June 1967).

Received January 1979