

The Capacity of Multilevel Threshold Functions

SVERRIR 'OLAFSSON AND YASER S. ABU-MOSTAFA

Abstract—The investigation of information capacity and computation power of threshold elements is motivated by the recent trend of neural-network approaches to pattern recognition. In this correspondence, we estimate lower and upper bounds for the capacity of multilevel threshold elements, using two essentially different enumeration techniques. The results correct a previously published estimate and indicate that adding threshold levels enhances the capacity more than adding variables.

Index Terms—Classifier, information capacity, multilevel threshold element, neural networks, parallel hyperplanes, pattern recognition.

I. INTRODUCTION

Threshold functions have somewhat regained interest in recent years. After considerable activity in the early 1960's, the subject all but vanished in the following decade. Current interest is spurred, among other things, by the unexpected difficulty of realizing efficient pattern recognition schemes on conventional digital computers. As an alternative, threshold functions offer a reasonable discrimination capability with simplicity and low cost. When combined in networks, many together create powerful computation networks and storage devices with interesting properties (e.g., [1], [7]). Single threshold devices, the most prevalent type, are found in various forms in nature (most notably the human brain) and can be easily realized physically; that is where the bulk of research has centered [1]–[4], [10]. Multithreshold devices, however, have drawn less enthusiasm. Among their qualities, though, is that given enough thresholds, a single multithreshold element can realize any given function operating on a finite domain.

There has recently been an intensive interest in threshold logic as the main component of neural network models (e.g., [7]). These models provide a direction for pattern recognition systems with distinct natural advantages. The capacity of these models, as well as their computing power, are directly related to the number of threshold functions [8].

The ability of multilevel threshold devices to simulate a larger number of functions compared to single-threshold devices is vital for the capacity and capabilities of neural network models based on threshold logic. It is therefore of practical as well as theoretical interest to estimate the number of functions that can be modeled as multilevel threshold functions for a given number of inputs and threshold levels.

In a recent paper [5], the number of dichotomies separable by a multilevel threshold element was derived, unfortunately, incorrectly. Here we will show the claimed result to be a valid lower bound. In addition, we will derive an upper bound using an essentially different argument. We will demonstrate that the exact number of multilevel threshold functions depends strongly on the relative topology of the input set; simple criteria like *general position* do not fully specify the problem (as opposed to the single-level case).

Manuscript received June 13, 1986; revised March 16, 1987. Recommended for acceptance by J. Kittler. This work was supported by the President's Fund, Caltech.

S. Olafsson was with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125. He is now with Rockwell International, Semiconductor Products Division, P.O. Box C, Newport Beach, CA 92658.

Y. S. Abu-Mostafa is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125.

IEEE Log Number 8715810.

II. DEFINITIONS AND PROBLEM DESCRIPTION

Say we have a set S of input points x_1, x_2, \dots, x_N in \mathbf{R}^n . We define a k -level threshold function by a weight vector $w \in \mathbf{R}^n$ and a threshold vector $t = (t_1, t_2, \dots, t_k)$ as:

$$f_{w,t}(x) \begin{cases} +1, & \text{if } t_i \leq w \cdot x < t_{i+1}, & \text{for } i \text{ even} \\ -1, & \text{if } t_i \leq w \cdot x < t_{i+1}, & \text{for } i \text{ odd} \end{cases} \quad (1)$$

where we take $t_0 = -\infty, t_{k+1} = +\infty$, and $w \cdot x$ denotes the inner product in \mathbf{R}^n . Equivalently, we can define a k -level threshold element (k-TE) as an operation partitioning S into two sets, S^+ and S^- , where in terms of $f_{w,t}$,

$$\begin{aligned} S^+ &= \{x \in S \mid f_{w,t}(x) = +1\} \\ S^- &= \{x \in S \mid f_{w,t}(x) = -1\} \end{aligned} \quad (2)$$

We say a dichotomy $\{S^+, S^-\}$ of S is *separable* by a k-TE if there exists a k -level threshold function $f_{w,t}$ defining S^+ and S^- .

Some authors use the terms *threshold function* and *threshold element* strictly when the input set consists of the 2^n vectors with binary coordinates. We will use these terms for arbitrary input sets, denoting explicitly if the inputs are binary.

One may visualize a k-TE as separating the points in S according to which region in \mathbf{R}^n they belong (Fig. 1), the regions being defined by the k parallel hyperplanes $x \cdot w - t_i = 0, i = 1, 2, \dots, k$. Traversing through the hyperplanes in the direction of w , we alternately assign points falling between two adjacent hyperplanes to one set, then the other after going through a hyperplane.

One might also consider the more general case of partitioning S into $k + 1$ subsets rather than just two. However, the results are not significantly different in the case of interest, namely when $N \gg n$ and k .

We are interested in finding $M(N, n, k)$, the number of dichotomies of N points in \mathbf{R}^n separable by a k-TE.

Let us first consider an example which shows that the result in [5] is incorrect. The equation given was

$$M(N, n, k) = 2 \sum_{i=0}^{n+k-1} \binom{N-1}{i}. \quad (3)$$

Let $N = 5, n = 2, k = 2$, and let the 5 points be in *general position* (i.e., no $n + 1$ points lie in the same hyperplane), as in Fig. 2. In this case, the hyperplanes are two simple lines in the two-dimensional plane. It is clear that the dichotomies $\{\emptyset, S\}$ and $\{S, \emptyset\}$ are separable, and we can also separate any one point from the others by drawing the two lines on each side of the point such that no other point is in between [Fig. 2(a)]. To separate two points, we draw the lines just above and just below a line segment joining the points [Fig. 2(b), (c)]. If another point is caught in between, we move the lines closer until it is outside. If that is not possible, three of the points lie on a line and so S is not in general position. Thus we can separate any one and two points. But then, by relabeling, we can separate any three and any four. Hence all dichotomies are separable, and so $M(5, 2, 2) = 2^5 = 32$. But from (3) we get $M(5, 2, 2) = 2 \sum_{i=0}^3 \binom{4}{i} = 2 \cdot (1 + 4 + 6 + 4) = 30 < 32$. Hence, (3) is incorrect.

Furthermore, in [5] it is claimed that if the N points are in general position, this number is unique for given N, n, k . That is untrue, and can be demonstrated by the simple example of Fig. 3, where $N = 6, n = 2, k = 2$. If we let the points be vertices of a hexagon, not all the dichotomies are separable, namely $\{ade \mid bcf\}$ in Fig. 3(a). However, in Fig. 3(b) with one point inside, all 2^6 dichotomies are separable. In both cases, the points are in general position. We therefore need a stronger topological condition in order to find the maximum number of separable dichotomies, a condition we will not attempt to find here. Fig. 4 shows the distribution of the number of k-TE's with $k = 2, n = 2$, where the input points were chosen at random with uniform density from the unit square.

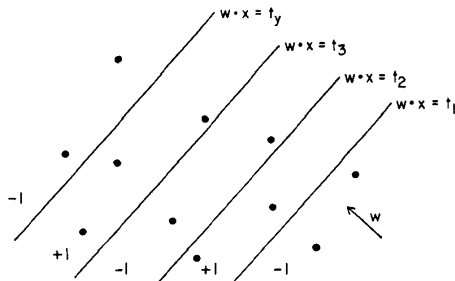


Fig. 1. Geometric interpretation of a k-TE.

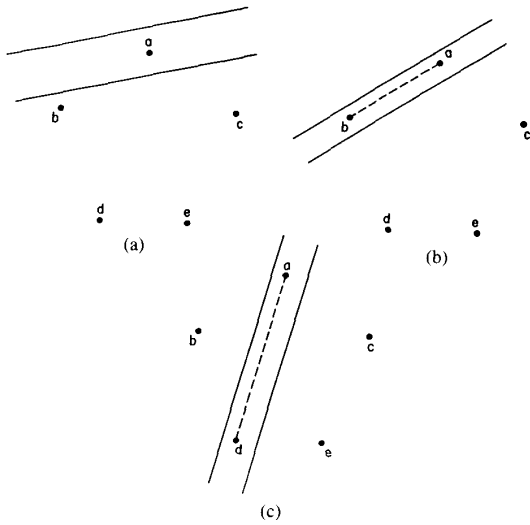


Fig. 2. All dichotomies of five points in two dimensions are separable by a bilevel TE. (a) $\{a | bcde\}$, (b) $\{ab | cde\}$, (c) $\{ad | bce\}$.



Fig. 3. The number of k-TE's depends on the position of the points in space. (a) The dichotomy $\{ade | bcf\}$ is not separable with a bilevel TE. (b) All dichotomies are separable with bilevel TE.

In a recent correspondence, Krueger [11] points out the above problems with [5] and cites one source of error. In fact, there is a second error in [5], namely Lemma 3. It will be stated in its correct form below.

III. LOWER BOUND

In what follows, when we say hyperplane P is restricted to point x we mean that x lies in P , whatever the orientation (weight vector w) of P . Let us now define the following extension of $M(N, n, k)$ as in [5]. Let $S = \{x_1, x_2, \dots, x_N\}$ and $R = \{y_1, y_2, \dots, y_r\}$

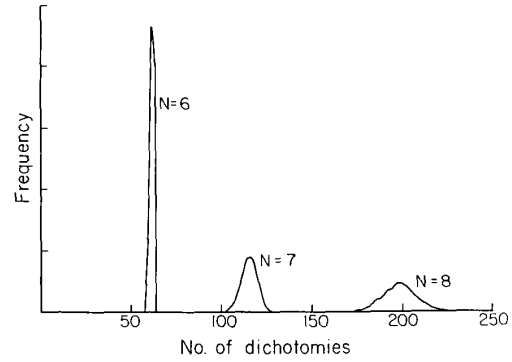


Fig. 4. The number of bilevel threshold functions in R^n .

be some given sets of points in R^n . Furthermore, let P_1, P_2, \dots, P_k be parallel hyperplanes, also in R^n , where r of the k hyperplanes are restricted to points in R (one point for each plane). For this type of "restricted" k-TE, the number of separable dichotomies on S is denoted by $D_r(N, n, k)$. An example of a restricted k-TE is given in Fig. 5.

Claim 1: If the set $S \cup R$ is in general position, then

$$D_r(N, n, k) \geq D_r(N-1, n, k) + D_{r+1}(N-1, n, k). \quad (4)$$

Proof: Consider the set $S' = \{x_1, \dots, x_{N-1}\}$. Pick some point x such that $S' \cup R \cup \{x\}$ is in general position, and let $S = S' \cup \{x\}$. Some of the $D_r(N-1, n, k)$ dichotomies of S' are separable by restricting one of the unrestricted planes, say P , to x , at least $D_{r+1}(N-1, n, k)$ of them. For each of these dichotomies of S' , there correspond two dichotomies of S , since by slightly perturbing P about x we can classify x either into S^+ or S^- without affecting the dichotomy of S' . The number of dichotomies of S' no longer separable when P is restricted to x is $D_r(N-1, n, k) - D_{r+1}(N-1, n, k)$. In these cases we have no freedom in classifying x . Hence the number of dichotomies of S is at least

$$\begin{aligned} D_r(N, n, k) &\geq D_r(N-1, n, k) - D_{r+1}(N-1, n, k) \\ &\quad + 2D_{r+1}(N-1, n, k) \\ &= D_r(N-1, n, k) + D_{r+1}(N-1, n, k). \end{aligned}$$

However, we cannot state equality. Consider a dichotomy $\{S'_1, S'_2\}$ of S' , and assume it is separable in two essentially different ways with (w_a, t_a) and (w_b, t_b) , i.e., one cannot be changed to the other moving the hyperplanes continuously without changing the function along the way. An example is given in Fig. 6, where (a) and (b) show two essentially different ways of separating the set $S'_1 = \{bc\}$. Further assume neither can be realized with any hyperplane restricted to x , and that (w_a, t_a) associates x with S'_1 [Fig. 6(a)], and (w_b, t_b) associates x with S'_2 [Fig. 6(b)]. But in this case the dichotomies $\{S'_1 \cup \{x\}, S'_2\}$ and $\{S'_1, S'_2 \cup \{x\}\}$ of S are both separable, yet are not counted in the previous argument. Hence the inequality. ■

Now we can, by deriving boundary conditions for $N = 1$ and $r = k$, obtain a bound for the desired function, $D_0(N, n, k) = M(N, n, k)$. Clearly, $D_r(1, n, k) = 2$. As for $D_k(N, n, k)$, it is clearly greater than $D_1(N, n, 1)$, which is simply the number of dichotomies separable with a single-level homogeneous threshold function given by [1], namely

$$\begin{aligned} D_k(N, n, k) &\geq D_1(N, n, 1) = C(N, n) \\ &= 2 \sum_{i=0}^{n-1} \binom{N-1}{i}. \end{aligned} \quad (5)$$

A solution to (4) and (5) with inequalities replaced by equalities is given by Takiyama [5] as $C(N, n+k)$, hence we obtain for N points in general position,

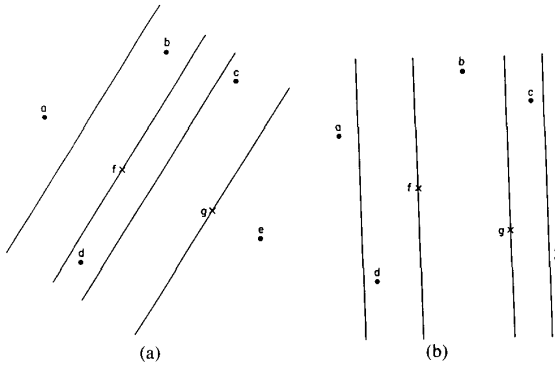


Fig. 5. Two implementations of a restricted k-TE with $k = 4$, $r = 2$, $N = 5$, $n = 2$. (a) $\{ade | bc\}$, (b) $\{abe | cd\}$.

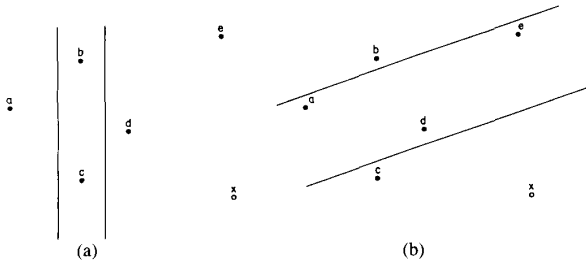


Fig. 6. Two ways of separating $\{b, c\}$, one includes x , one excludes x . (a) $\{bc | adex\}$, (b) $\{bcx | ade\}$.

$$D_0(N, n, k) = M(N, n, k) \geq 2 \sum_{i=0}^{n+k-1} \binom{N-1}{i}. \quad (6a)$$

We would like to obtain a more useful quantity than the above sum for later analysis. It can be easily shown that for n and k fixed,

$$\log \left[2 \sum_{i=0}^{n+k-1} \binom{N-1}{i} \right] = \log (N-1)^{n+k-1} + O(1/N) + c$$

where $O(1/N)$ behaves asymptotically as $1/N$ as $N \rightarrow \infty$ and c depends only on n and k . We will make use of this fact by writing

$$2 \sum_{i=0}^{n+k-1} \binom{N-1}{i} \approx N^{n+k} \quad (6b)$$

for large N and make a similar approximation for the upper bound to facilitate comparison.

IV. UPPER BOUND

We will derive the upper bound by considering the number of possible orderings of the N points when projected onto a line in the direction of a given w (see Fig. 7). The weight vectors giving a particular ordering are contained in some region R^n . That region must be bordered by a set of weight vectors W_b such that the inner products of the two or more points in S with a $w \in W_b$ are equal, i.e., for all $w \in W_b$, there are $x_i, x_j \in S$ such that $w \cdot x_i = w \cdot x_j$, or equivalently, $W_b \subseteq \{w | w \cdot (x_i - x_j) = 0, 1 \leq j < i \leq N\}$. Therefore, the boundaries are simply homogeneous hyperplanes orthogonal to the differences of vectors in S . Since there are $\binom{N}{2}$ difference vectors, the number of regions enclosed by these hyperplanes is at most $C(\binom{N}{2}, n)$. For each given weight vector, there are $\binom{N-1}{k}$ different ways of assigning k thresholds (not counting degenerate cases), and so by an apparently large overestimate we get

$$M(N, n, k) \leq \binom{N-1}{k} \cdot 2 \sum_{i=0}^{n-1} \binom{\binom{N}{2} - 1}{i}. \quad (7)$$

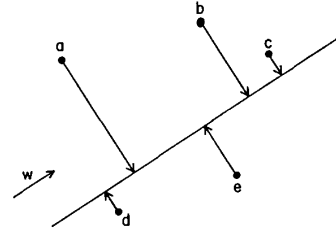


Fig. 7. An ordering d, a, e, b, c .

In the case of 2^n binary inputs, we can reduce the above bound somewhat by noticing that there can only be three different coefficients of the difference vectors, namely $-2, 0, 2$. Not counting $(0, 0, 0)$ and taking pairs of antipodal vectors once, we see there can only be $\frac{1}{2}(3^n - 1)$ difference vectors. Therefore, the number of k-TE's with binary inputs is bounded by

$$cM_b(2^n, n, k) \leq \binom{N-1}{k} \cdot 2 \sum_{i=0}^{n-1} \binom{\frac{1}{2}(3^n - 1) - 1}{i}. \quad (8)$$

If we want to consider the cumulative number of TE's with $\leq k$ thresholds, we simply sum over k and obtain

$$M'(N, n, k) = 2 \sum_{l=0}^k \binom{N-1}{l} \cdot \sum_{i=0}^{n-1} \binom{\binom{N}{2} - 1}{i} \approx N^{2n+k} \quad (9)$$

where the approximation is made in the sense of (6b).

The above can be expressed in the binary case as

$$\begin{aligned} \log_2 M_b(2^n, n, k) \\ \leq (\log_2 3) n^2 + 2k(n - \log_2 k), \\ \text{for } k \leq 2^{n-1} \end{aligned} \quad (10)$$

where we have used the bounds

$$\log_2 \sum_{l=0}^k \binom{2^n - 1}{l} < 2k(n - \log_2 k), \quad \text{for } k \leq 2^{n-1}$$

and

$$\log_2 \sum_{i=0}^{n-1} \binom{\frac{1}{2}(3^n - 1) - 1}{i} < (\log_2 3) n^2.$$

One may consider the structure of the above regions in terms of a graph. Each ordering, i.e., each region of w 's giving the same ordering, is uniquely represented by a vertex. If two vertices, say v_1 and v_2 , represent adjacent regions in R^n separated by the hyperplane $\{w | w \cdot (x_i - x_j) = 0\}$, they have an edge between them which then represents the interchanging of x_i and x_j . Therefore, if the ordering at vertex v_1 is $x_1, \dots, x_i, x_j, \dots, x_N$, the ordering at the adjacent vertex v_2 is $x_1, \dots, x_j, x_i, \dots, x_N$. Thus, to each hyperplane $w \cdot (x_i - x_j) = 0$, there corresponds a set of edges for which x_i and x_j (who must be adjacent on either side) are interchanged. Given this graph for a particular set of points, we have all the information we need to find the exact number of k-TE's.

Perhaps it is illustrative to consider the "full permutation graph," namely where all possible orderings are represented by vertices and those connected that have adjacent elements interchanged. We then get something like Fig. 8, where we have for $N = 4$ listed one particular ordering in the leftmost column, those permutations with one interchange of adjacent elements in the next, and so on. In fact, we can define a metric d where $d(p_1; p_2)$ is the minimum number of times one must interchange adjacent elements of p_1 to make it into p_2 , or vice versa. It is easy to see that $d(x_1,$

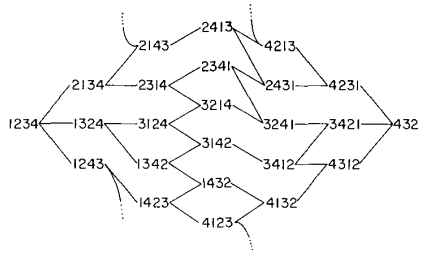


Fig. 8. Permutation graph for four elements.

$x_2, \dots, x_N; x_N, x_{N-1}, \dots, x_2, x_1) = \binom{N}{2}$, thus there are $\binom{N}{2} + 1$ stages in each graph. The graph we obtain in R^n is a subgraph of this full graph, and is determined by the number of dimensions and the position of points in R^n . Given enough dimensions, the full graph is represented, namely if $n \geq N - 1$ (since a region in R^n must be enclosed by at least n hyperplanes, each vertex has n edges, so if $n \geq N - 1$, all the edges, hence all the vertices, must be present). In two dimensions, the subgraph is a single cycle, corresponding to a closed path around the origin.

In considering the number of vertices per stage (column) of the full graph, one finds these numbers, denoted by $\binom{N}{k}$ for the k th column, are related by

$$\binom{N+1}{k} = \sum_{i=k-N}^k \binom{N}{i}$$

and, of course,

$$\sum_{k=0}^N \binom{N}{k} = 2^N$$

Expressing this, row by row, we get something equivalent to a Pascal triangle, where now in the N th row we add $N + 1$ adjacent numbers instead of 2, and get $\binom{N}{2} + 1$ instead of $N + 1$ elements per row, and the sum of row elements is $N!$ instead of 2^N , as can be verified from the following first few rows:

			1															
			1	1														
			1	2	1													
			1	3	5	6	5	3	1									
			1	4	9	15	20	22	20	15	9	4	1					
			1	5	14	29	49	71	90	101	101	90	71	49	29	14	5	1

For a more detailed discussion and proofs, see [9].

V. DISCUSSION

Considering the approximations of (6) and (9) we see that the essential difference between the lower and upper bounds is a factor of N^n . Whether or not all of this factor is really necessary is not quite clear. A more detailed analysis of the structure of the orderings used in Section IV indicates that it is, but no simple method of establishing that has been found. It appears that just by going from one to multiple thresholds there is a significant jump in the number of possibilities. In fact, (6a) holds with strict inequality for $k > 2$, but equality for $k = 1$. Hence adding a second threshold is in fact more effective than adding a dimension, an important fact to note. As another point, while we believed (4) was true with equality as stated in [5], we derived a very generous upper bound to the boundary condition for $r = k$, in order to get an upper bound for $M(N, n, k)$. Essentially, it was of the order $N^{n+k}k^k$, but failed already in the case of $N = 8, n = 2$ and $k = 2$. So expressing M' as a power of $N, M'(N, n, k) \approx N^{\alpha n+k}$, [with \approx interpreted as in (6b)], we are led to conclude that α is indeed greater than 1, and apparently approaches 2 as N, n , and k approach infinity, with n and k growing logarithmically in N . Some numbers for $k = 2, n =$

TABLE 1
COMPARISON OF LOWER AND UPPER BOUNDS FOR $k = 2$

N	n = 2			n = 3	
	Lower Bound	Actual	Upper Bound	Lower Bound	Upper Bound
5	30	32	120	32	552
6	52	58-64	300	62	2120
7	84	102-128	630	114	6330
8	128	170-230	1176	198	15 918

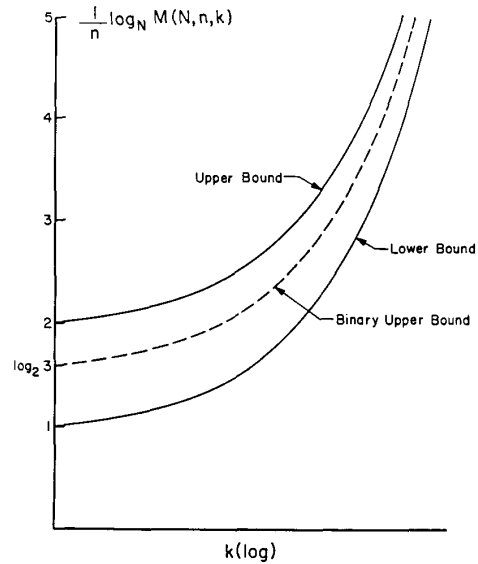


Fig. 9. Plot of upper and lower bounds on $M(N, n, k)$.

2, $N = 6, 7, 8$ are given in Table I, and a plot of the behavior of the bounds is given in Fig. 9.

VI. SUMMARY

We have shown Takiyama's result [5] to be correct as a lower bound for input points in (inhomogeneous) general position, and have given an upper bound for the number of dichotomies separable by a k -level threshold element. Expressing the result asymptotically as $N^{\alpha n+k}$, we find that in the general case $1 \leq \alpha < 2$, and in the binary case $\alpha < \log_2 3$. It can be shown from [1] that for points in *general position*, $\alpha \rightarrow 1$ as $N \rightarrow \infty$ with n fixed when $k = 1$. As k is allowed go grow along with n logarithmically in N , it appears that $\alpha \rightarrow 2$ as $N \rightarrow \infty$.

REFERENCES

- [1] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Trans. Electron. Comput.*, vol. EC-14, pp. 326-334, June 1965.
- [2] M. Minsky and S. Papert, *Perceptrons*. Cambridge, MA: M.I.T. Press, 1969.
- [3] S. Muroga, "Lower bounds on the number of threshold functions and a maximum weight," *IEEE Trans. Electron. Comput.*, vol. EC-14, pp. 136-148, Apr. 1965.
- [4] D. Haring, "Multi-threshold threshold elements," *IEEE Trans. Electron. Comput.*, vol. EC-15, pp. 45-65, June 1965.
- [5] R. Takiyama, "The separating capacity of a multi-threshold threshold element," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, pp. 112-116, Jan. 1985.
- [6] —, "Multiple threshold perceptron," *Pattern Recognition*, vol. 10, pp. 27-30, 1978.

- [7] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci.*, vol. 79, pp. 2554-2558, 1982.
- [8] Y. S. Abu-Mostafa and J. M. St. Jacques, "Information capacity of the Hopfield model," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 461-464, July 1985.
- [9] S. 'Olafsson, "Transpositions and Pascal's triangle," to be published.
- [10] J. Pearl, "Capacity and error estimates for boolean classifiers with limited complexity," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-1, pp. 350-355, 1979.
- [11] F. R. Krueger, "Comments on Takiyama's analysis of the multi-threshold threshold element," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 760-761, Nov. 1986.

Bounds on the Bayes Classification Error Based on Pairwise Risk Functions

F. D. GARBER AND A. DJOUADI

Abstract—Upper and lower bounds on the Bayes risk for multiple, composite-hypothesis classification are obtained. Bounds on the Bayes risk for M simple classes are derived in terms of the risk functions for $(M - 1)$ classes, and so on, until the desired result depends only on the pairwise ($M = 2$) Bayes risks. A method of computing upper and lower bounds on the pairwise Bayes risk for composite classes is developed. Algorithms for computing the upper and lower bounds for the general M -class case and for composite-hypothesis classes are presented. Numerical examples of the application of the bounding techniques to a problem involving the classification of aircraft are discussed. Comparisons of the results for the bounds and other performance measures are given for the most interesting cases.

Index Terms—Asymptotic risk, Bayes risk, bounds on error probability, composite hypothesis classification, nearest neighbor, risk estimation.

I. INTRODUCTION

An important aspect of the design of many pattern recognition systems is the evaluation or estimation of the average performance of the proposed classification algorithm. In situations where the measurements are described parametrically, in terms of the hypothetical observation and a set of random variables with known statistics, the key performance parameter is often chosen to be the average probability of misclassification, which is also the Bayes risk for the common special case of unit costs.

Bayesian decision theory offers a well-motivated methodology for the design of parametric classifiers where the decision problem is posed in probabilistic terms; the performance of the resulting classifier, the Bayes risk, is a standard for evaluating the performance of alternative algorithms [1], [2]. In addition, the sense of optimality implicit in the Bayesian approach is sufficiently flexible to allow applications to a wide variety of practical decision problems [3]. Unfortunately, the exact analytical evaluation of the per-

formance of Bayesian classifiers is often exceedingly difficult, in spite of knowledge concerning the statistical descriptions of the key parameters.

In order to circumvent this difficulty, a number of investigators have been concerned with the approximate evaluation of system performance in terms of error estimates or in terms of bounds on the Bayes risk [2]-[12]. While the search for useful analytical bounds and approximations of system performance is, in itself, a challenging task and the end results may occasionally be less than satisfactory, it is felt that an analytical approach may hold the most potential for identifying the key parameters of a given pattern recognition system and characterizing their influence on system performance in ways that may facilitate the design of more optimal systems.

The present consideration of bounds on the Bayes risk or average classification error probability is primarily motivated by our interest in applications of pattern recognition theory to the problem of designing reliable noncooperative radar target identification systems [13]-[15]. The identification system is to be designed to distinguish measurements of the radar backscatter from an unknown object as belonging to one of a set of M classes, each corresponding to a particular aircraft. The candidate classes are represented by a set of $n \geq 1$ stored reference pattern measurement vectors or prototypes that are obtained prior to operation.

In cases where the aspect angle (azimuth and elevation) of the object are known, the identification system chooses among "simple" classes containing a single prototype corresponding to the known aspect. When the aspect angle of the object is unknown or known only to be within some range of angles, the identification is accomplished by means of comparison to M "composite" classes, each containing n subclasses or prototypes that represent one of the possible aspect angles in the specified range. The nature of this design problem implies we are interested in obtaining bounds and approximations on the performance of classification algorithms for both multiple-hypothesis classification among simple classes and multiple-hypothesis classification among composite classes.

In Section II, upper and lower bounds on the average Bayes risk for M classes are derived as functions of the Bayes risk for $(M - 1)$ classes. A technique is given for a recursive application of this result so that the Bayes risk for M classes may be computed as functions of the $\binom{M}{2}$ pairwise risk functions among the M classes. The resulting upper bound is at least as sharp as bounds computed by means of alternative methods.

The problem of obtaining simple upper and lower bounds on the Bayes risk for composite classes is the subject of Section III. The derivation of the results in this case depends on the established relationships between the Bayes risk and the asymptotic nearest neighbor risk [1]. Finally, the results of the application of the M -class bounds of Section II, in conjunction with the composite-class bounds of Section III, are presented in Section IV. Numerical examples of the most interesting cases showing tradeoffs between the behavior of the bounding techniques are given using data representative of the identification problem discussed above.

II. ERROR BOUNDS FOR M SIMPLE CLASSES

Consider the problem of classifying a vector observation $X \in \mathcal{R}^K$ as belonging to one of M classes, each corresponding to an event ω_i for $i = 1, 2, \dots, M$. Let $p(x/\omega_i)$ denote the conditional probability density function for the random observation $X = x$ defined on the observation space \mathcal{X} , and let $P(\omega_i)$ denote the *a priori* class probabilities for the events ω_i , $i = 1, \dots, M$. The Bayes risk (or average error probability) can then be expressed [4] in integral form as

$$\mathcal{R} = 1 - \sum_{i=1}^M \int_{\Gamma_i} P(\omega_i) p(x/\omega_i) dx \quad (1)$$

Manuscript received September 15, 1986; revised September 14, 1987. Recommended for acceptance by A. K. Jain. This work was supported by the Office of Naval Research under Contract N00014-86-K-0202.

The authors are with the ElectroScience Laboratory, Department of Electrical Engineering, Ohio State University, Columbus, OH 43210.

IEEE Log Number 8718595.