

AD-A075 062

TEXAS UNIV AT AUSTIN DEPT OF ELECTRICAL ENGINEERING  
THE STUDY OF NONPARAMETRIC PROCEDURES FOR DISCRIMINATION AND CL--ETC(U)  
SEP 79 T J WAGNER

F/G 12/1

AFOSR-77-3385

UNCLASSIFIED

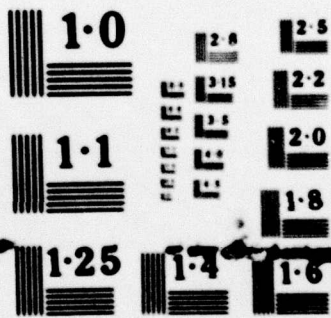
AFOSR-TR-79-1064

NL

| OF |  
AD  
A075062



END  
DATE  
FILMED  
11-79  
DDC



NATIONAL BUREAU OF STANDARDS  
MICROCOPY RESOLUTION TEST CHART

AD A 075 062

UUC FILE COPY

<b>19</b> REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER <b>18</b> AFOSR TR-79-1064	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle)	<b>28</b>	5. TYPE OF REPORT & PERIODICITY <b>9</b> Final rept. 1 Jul 79
6. THE STUDY OF NONPARAMETRIC PROCEDURES FOR DISCRIMINATION & CLUSTERING PROBLEMS	7. AUTHOR(S) T. J. Wagner	8. PERFORMING ORG. REPORT NUMBER 130 Jul 79
9. PERFORMING ORGANIZATION NAME AND ADDRESS The University of Texas at Austin Department of Electrical Engineering Austin, Texas 78712	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F <b>16</b> 2344 A5 <b>17</b> A5	11. CONTRACT OR GRANT NUMBER(S) <b>15</b> AFOSR-77-3385
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332	12. REPORT DATE 9/24/79	13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) <b>LEVEL II</b>	15. SECURITY CLASS. (of this report) UNCLASSIFIED	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of abstract entered in Block 20, if different from Report) <b>11</b> 24 Sep 79		<b>DDC</b> <b>RECEIVED</b> OCT 15 1979 <b>RECEIVED</b> <b>E</b>
18. SUPPLEMENTARY NOTES <b>12</b> 10		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The final report lists the papers and accomplishments supported by AFOSR Grant 77-3385 as well as the technical personnel who received support.		

401997

Jm

**AFOSR-TR- 79 - 1064**

**The Study of Nonparametric Procedures for  
Discrimination & Clustering Problems**

**Final Report for AFOSR Grant 77-3385**

**7/1/77 to 7/30/79**

**T. J. Wagner**

**INTRODUCTION**

The accomplishments of the work performed with the support of AFOSR Grant 77-3385 are outlined in Section I. The papers published with grant support are listed in Section II along with those that are in some prepublication stage. Section III lists the technical personnel supported by the grant. 7

**AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)**

**NOTICE OF TRANSMITTAL TO DDC**

This technical report has been reviewed and is approved for public release IAW AFR 190-12 (7b). Distribution is unlimited.

**A. D. BLOSE**

**Technical Information Officer**

**Approved for public release;  
distribution unlimited.**

## I. Accomplishments

Accession #	NTIS Grant	DDC TAB	Unannounced	Justification
By	Distribution	Availability	Avail and/or	Dist special
				<b>A</b>

In nonparametric density estimation, one is concerned with estimating the density  $f$  from a sample  $X_1, \dots, X_n$  drawn from that density. The kernel estimate of  $f$  is given by

$$f_n(x) = \sum_{i=1}^n K(x-X_i/h_n)/nh_n^d$$

where  $K$ , the kernel, is a bounded probability density,  $\{h_n\}$  is a sequence of positive numbers tending to 0 with  $n$ , and the parameter  $d$  is the dimension of the  $X_i$ . In [1] conditions on  $K$  and  $\{h_n\}$  are given which insure that

$$\int_{\mathbb{R}^d} |f_n(x) - f(x)| dx \xrightarrow{n} 0 \text{ w.p.1.} \quad (1)$$

The novelty of this result is that no analytic assumptions are needed for  $f$  to insure (1). In [2], conditions are given which assure that

$$f_n(x) \xrightarrow{n} f(x) \text{ w.p.1} \quad (2)$$

and

$$\sup_x |f_n(x) - f(x)| \xrightarrow{n} 0 \text{ w.p.1.} \quad (3)$$

These conditions were, at the time the paper was written, weaker than any given in the literature to obtain (2) and (3). Also in [2], the practical situation is considered where  $h_n = h_n(X_1, \dots, X_n)$  yet one still obtains (2) and (3). (As noted in the first interim report for this grant, paper [2] is an amalgam of papers [33] and [34] listed in the final report for AFOSR Grant 72-2371.) Additionally, the modes of convergence (1-3) were studied for a recursive version of the kernel estimate in [3].

One should also note here that all of this density estimation work was started on AFOSR Grant 72-2371.

In the nonparametric discrimination problem, the random variable  $\theta$ , known to take values in  $\{1, \dots, M\}$ , is estimated from the random vector  $X$  taking values in  $\mathbb{R}^d$ . All that is known about the distribution of  $(X, \theta)$  is that which can be inferred from a sample  $(X_1, \theta_1), \dots, (X_n, \theta_n)$  of size  $n$  drawn from that distribution. The sample, called data, is assumed to be independent of  $(X, \theta)$ . A discrimination rule is then any procedure which determines an estimate  $\hat{\theta}$  of  $\theta$  from  $X$  and the data. The random variable

$$L_n = P[\hat{\theta} \neq \theta | (X_1, \theta_1), \dots, (X_n, \theta_n)],$$

the probability of error given the data, is important because it measures the performance of the rule with the data that one has.

One of the most studied aspects of discrimination rules is that of asymptotic convergence. For example, for what rules does  $L_n$  converge to  $L^*$ , the Bayes probability of error. The recent work of Stone (Annals of Statistics, vol. 5, pp. 595-645, 1977) showed, among other things, that for the  $k$ -nearest neighbor rule with

$$k = k_n \xrightarrow{n \rightarrow \infty} \infty \tag{4a}$$

and

$$k_n/n \xrightarrow{n \rightarrow \infty} 0 \tag{4b}$$

one has

$$L_n \xrightarrow{n \rightarrow \infty} L^* \text{ in probability} \tag{5}$$

regardless of the distribution of  $(X, \theta)$ . What makes this result interesting is that all previous analytic assumptions about the distribution of  $(X, \theta)$

which were used along with (4) to imply (5) are now removed. Stone's results apply to a wide class of nearest neighbor rules but, for example, do not include the popular potential function rules. In [4] and [5] results which parallel Stone's results for nearest neighbor rules are proven for potential function rules. Devroye [6] has also shown recently that if one replaces (4a) with

$$k_n / \log n \xrightarrow{n \rightarrow \infty} \infty \quad (4a)'$$

then

$$L_n \xrightarrow{n \rightarrow \infty} R^* \text{ with probability one,}$$

again with no assumptions on the distribution of  $(X, \theta)$ . In [6] Devroye has also proven a similar  $\log n$  type of result for potential function rules which extend the results of [5].

Finally, for the single nearest neighbor rule, Devroye [7] has shown that

$$L_n \xrightarrow{n \rightarrow \infty} L \text{ in probability}$$

where  $L$  is an explicitly determined constant depending only on the distribution of  $(X, \theta)$  and

$$L^* \leq L \leq 2L^*(1-L^*).$$

This result again requires no assumptions about the distribution of  $(X, \theta)$  and nicely extends the celebrated result of Cover and Hart.

The nonparametric estimation problem is similar to its discrimination counterpart. Here  $\theta$  takes on real values and, instead of the probability of error, one takes some other criterion appropriate to estimation problems, for example, the mean-squared error. With this criterion it is

natural to (i) estimate  $m(X) = E(\theta|X)$  from the data with  $\hat{m}(X)$  and then (ii) estimate  $\theta$  with  $\hat{\theta} = \hat{m}(X)$ . For this reason, statisticians concentrate on estimating  $m(X)$  from the data and many discrimination results consequently are by-products of regression function estimation. The papers [4-6] deal then with regression function estimation and I have only extracted the details that have impact on the more important discrimination problem. Devroye [8] has also examined some aspects of regression function estimation that don't appear to bear directly on the discrimination problem.

One of the important facets of the discrimination problem is estimating  $L_n$  from the data. Two popular estimates of  $L_n$  are the deleted estimate  $L_n^D$  and the holdout estimate  $L_n^H$ . In [9] exponential bounds for

$$P[|L_n - L_n^H| \geq \epsilon]$$

and

$$P[|L_n - L_n^D| \geq \epsilon]$$

(6)

with the k-nearest neighbor rules are given which are distribution-free (i.e., they depend only on  $n$ ,  $\epsilon$ ,  $k$ ,  $M$  and  $d$ ). While they are exponential in  $n$  they still appear to be rather crude as the simulation study [10] indicates. Distribution-free bounds for (6) have also been found for potential function rules [11] where it is also shown that they are necessarily  $O(1/\sqrt{n})$ . If one considers linear rules, exponential bounds for

$$P[|L_n - L_n^R| \geq \epsilon]$$

have been found which greatly improve earlier ones published [12]. Here  $L_n^R$  is the resubstitution estimate of  $L_n$  from the data.

## II. References

1. "The  $L_1$  Convergence of Kernel Density Estimates," T. J. Wagner and L. P. Devroye, to appear in the September, 1979 issue of the *Annals of Statistics*.
2. "The Strong Uniform Consistency of Kernel Density Estimates," T. J. Wagner and L. P. Devroye, to appear in the *Proceedings of the Fifth International Symposium on Multivariate Analysis*, North Holland Publishing Co., Fall, 1979.
3. "On the Pointwise and Integral Convergence of Recursive Kernel Estimates of Probability Densities," L. P. Devroye, *Utilitas Mathematica*, vol. 15, pp. 113-128, 1979.
4. "On the  $L_1$  Convergence of the Kernel Regression Function Estimates with Applications in Discrimination," T. J. Wagner and L. P. Devroye, to appear in *Zeitschrift für Wahrscheinlichkeits Theorie und Verwandte Gebiete*.
5. "Distribution-Free Consistency Results in Nonparametric Discrimination and Regression Function Estimation," T. J. Wagner and L. P. Devroye, to appear in *Annals of Statistics*, March, 1980.
6. "On the Almost Everywhere Convergence of Nonparametric Regression Function Estimates," L. P. Devroye, submitted to the *Annals of Statistics*.
7. "On the Inequality of Cover and Hart in Nearest Neighbor Discrimination," L. P. Devroye, submitted to the *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
8. "The Uniform Convergence of the Nadaraya-Watson Regression Function Estimate," L. P. Devroye, *Canadian Journal of Statistics*, vol. 6, pp. 179-191, 1978.
9. "Distribution-Free Inequalities for the Deleted and Holdout Error Estimates," T. J. Wagner and L. P. Devroye, *IEEE Transactions on Information Theory*, IT-25, pp. 202-207, 1979.

10. "Risk Estimation for Nonparametric Discrimination and Estimation Rules: A Simulation Study," T. J. Wagner and C. S. Penrod, to appear in the *IEEE Transactions on Information Theory*, November, 1979.
11. "Distribution-Free Performance Bounds with Potential Function Rules," T. J. Wagner and L. P. Devroye, to appear in the *IEEE Transactions on Information Theory*, September, 1979.
12. "Distribution-Free Performance Bounds with the Resubstitution Error Estimate," T. J. Wagner and L. P. Devroye, *IEEE Transactions on Information Theory*, IT-25, pp. 208-210, 1979.

Note: The paper "An Expanding Automaton for Use in Stochastic Optimization" by L. P. Devroye was supported by AFOSR 72-2371. When the six copies with the DD 1473 form were sent in (7/6/78), the grant # was left off, possibly causing this paper to be listed with the above grant.

III. Technical Personnel Supported by AFOSR Grant 77-3385  
during the period 7/1/77 to 7/30/79.

T. J. Wagner, Professor, seven summer months, 7-8/77, 6-8/78, 6-7/79.

Gary L. Wise, Assistant Professor, approximately one summer month during  
the period 6/79 to 7/79.

Luc P. Devroye, Research Associate IV, one summer month, 1978.

Don Halverson, Research Assistant, one-half month.

C. H. Liu, Research Assistant, one and one-half months.

K. Hsu, Research Assistant, one and one-half months.