INTRODUCTION.

The accomplishments of the work performed with whole or partial support of the Grant AFOSR 72-2371 are outlined in Section I. The papers published with grant support are given in Section II along with those that are in some prepublication stage. Section III lists the technical personnel supported by the grant during each year period.
SECTION I. ACCOMPLISHMENTS.

The main grant research effort has been in the area of nonparametric discrimination and the closely related problem of density estimation. In the discrimination problem a statistician makes an observation $X$, a random vector with values in $\mathbb{R}^d$, and wishes to estimate its state $\theta$, a random variable taking values in $\{1, \ldots, M\}$. All that he knows about the distribution of $(X, \theta)$ is that which can be inferred from a sample $(X_1, \theta_1), \ldots, (X_n, \theta_n)$ of size $n$ drawn from that distribution. The sample, commonly called data, is assumed to be independent of $(X, \theta)$. Using $X$ and the data the statistician makes a randomized decision $\hat{\theta}$ for $\theta$ where his rule is any procedure which determines the probability distribution $\delta = (\delta_1, \ldots, \delta_M)$ for $\hat{\theta}$ given $X$ and the data. In particular,

$$\delta: \mathbb{R}^d \times (\mathbb{R}^d \times \{1, \ldots, M\})^n \rightarrow [0,1]^M,$$

$$\sum_{j=1}^M \delta_j = 1,$$

and

$$P(\hat{\theta} = j | X, (X_1, \theta_1), \ldots, (X_n, \theta_n)) = \delta_j, \quad 1 \leq j \leq M.$$ 

For his data and rule, the probability of error is

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

a random variable whose value is the limiting frequency of errors made when a large number of independent observations have their states estimates with $\delta$ and the given data.

The questions which have been considered in the past usually are of the asymptotic type. For example, what does $L_n$ converge to as $n$ tends
to infinity and if, say, \( L_n \xrightarrow{n} L \) in probability, how does \( L \) compare to the Bayes probability of error \( L^* \). Other asymptotic studies have been concerned with how one estimates \( L \), or \( L^* \) if it is different from \( L \), from the data. Investigators have also been recently concerned with estimating \( L_n \) from the data, sometimes called error estimation. If \( \hat{L}_n \) is a particular estimate of \( L_n \) the statistician would like to know how small

\[
P[|\hat{L}_n - L_n| \geq \epsilon]
\]

is for a given \( \epsilon > 0 \). The difficulty is that the distribution of \((X, \theta)\) is unknown and the best that one can hope for is a distribution-free upper-bound for (1) which depends only on \( n, \epsilon \) and \( d \) and which tends to zero with \( n \) for each fixed \( \epsilon \) and \( d \).

With grant support, a summary of all recent work on nonparametric discrimination, including density estimation, was completed [12]. In [20, 41] a class of rules, which are a natural generalization of k-nearest neighbor rules and which are called voting rules, are discussed and conditions are given for which \( L_n \xrightarrow{n} L^* \) in probability and with probability one. In [19, 43], the asymptotic properties of k-nearest neighbor rules are investigated for the nonparametric estimation problem (\( \theta \) now takes values in \( \mathbb{R}^P \) rather than in \{1, \ldots, M\}).

A consideration always present in nonparametric discrimination is how to implement rules derived from large amounts of data. For example, if one uses the k-nearest neighbor rule with the data, a large \( n \) presents difficulties in that both storage requirements and computation times increase with \( n \). In order to keep the implementation requirements within reason and
still retain the appeal of k-nearest neighbor rules, various procedures for condensing or editing the data before the k-nearest neighbor rule is applied have been suggested. In [5], a simple argument for the asymptotic property of one of the most appealing of these schemes is given while in [23], a basic flaw in the original argument used for obtaining the asymptotic properties of the edited nearest neighbor rule is pointed out.

In error estimation, the first distribution-free bound for (1) was given for k-local rules in [29]† where \( \hat{L}_n \) is the deleted estimate of \( L_n \). A rule is k-local if the decision \( \hat{\theta} \) depends only on \( X \) and the pairs \( (X_i, \theta_i) \) for which \( X_i \) is one of the k-closest to \( X \) from \( X_1, \ldots, X_n \). The bound for (1) given in [29] is of the form \( A/nC^2 \) where \( A \) is an explicitly given small constant depending only on \( M \) and \( k \). Thus deleted estimates appear to be very good estimates of \( L_n \) when k-local rules are used. This has also been confirmed in the extensive simulation [39] which shows that the bound of [29] is pessimistic. The deleted estimate has also been shown to be an asymptotically consistent estimate of the Bayes risk \( L^* \) for a wide variety of rules [2]. Using the resubstitution estimate of \( L_n \), an exponential bound for (1) has been found for linear rules and condensed nearest neighbor rules [17, 22, 26, 38]. Bounds of the type \( A/\sqrt{n} \) have also been obtained for (1) for the popular class of two-step rules [42].

† Ironically, this first and, in some ways, best paper that I've been involved in for this area has had to undergo several revisions with long editorial delays between them. The third revision is now being looked at.
In nonparametric density estimation one seeks to estimate the density \( f \) of the sample \( X_1, \ldots, X_n \). Two popular estimates have been investigated under this grant. The kernel estimate \( f_n \) is given by

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

where \( K \), the kernel, is a probability density on \( \mathbb{R}^d \) and \( \{h_n\} \) is a sequence of positive numbers tending to zero with \( n \). In [33] conditions are given which insure that

\[
f_n(x) \rightarrow f(x) \quad \text{w.p.1}
\]

\[
\sup_x |f_n(x) - f(x)| \rightarrow 0 \quad \text{w.p.1}.
\]

In both cases, these conditions are weaker than any in the literature. Additionally, in [34] conditions are given which insure

\[
\int_{\mathbb{R}^d} |f_n(x) - f(x)| dx \rightarrow 0 \quad \text{w.p.1}.
\]

Similar types of results are developed in [36]. The sequence \( \{h_n\} \) in (2) is chosen without regard to \( X_1, \ldots, X_n \), something which one would like to be able to do. In [11] conditions are given which allow \( h_n = h_n(X_1, \ldots, X_n) \) while still yielding (3) and (4). These results are extended in [40].

The nearest neighbor estimate of \( f \) is given by

\[
g_n(x) = \frac{k_n/n}{V_n}
\]

where \( \{k_n\} \) is a sequence of positive integers with \( k_n \leq n \) and
and $V_n$ is the volume of the smallest sphere, centered at $x$, which contains $k_n$ of the points $X_1, \ldots, X_n$. In [3] conditions are given for

$$g_n(x) \xrightarrow{p} f(x) \text{ w.p.1}$$

while in [22] conditions are given for

$$\sup_x |g_n(x) - f(x)| \to 0 \text{ w.p.1}.$$ 

Other results related to density estimation may be found in [4,31,32,35] while [27] contains a density estimation result which applies directly to the clustering problem.

Random search deals with the problem of locating the minimum of an unknown function $g$ defined on $\mathbb{R}^d$. The function $g$ is not assumed to possess any of the usual analytical properties, such as convexity, that are invoked when one is searching for extrema. For this purpose one generates a sequence $Z_1, Z_2, \ldots$ of random vectors with values in $\mathbb{R}^d$ and with $Z_n$ representing the estimate of the location of a global minimum after $n$ steps. The goal is to produce a sequence for which $g(Z_n)$ converges with probability one, to the essential infimum of $g(x)$. All random search methods generate a trial vector $Z_{n+1}^*$ from $Z_1, \ldots, Z_n$ and then let

$$Z_{n+1} = \begin{cases} 
Z_n & \text{if } g(Z_n) < g(Z_{n+1}^*) \\
Z_{n+1}^* & \text{if } g(Z_{n+1}^*) \leq g(Z_n)
\end{cases}.$$ 

A further complication can arise when, for any $x$, one cannot observe $g(x)$, but only a sample with distribution function $G_x$ and mean $g(x)$. 
If $F_n$ denotes the distribution function of $X_{n+1}^*$ given $X_1, \ldots, X_n$ and $\hat{g}(x)$ denotes the sample mean of a sample of a size $\lambda_n$ with the distribution function $G_x$ then Devroye [13], for

$$X_{n+1} = \begin{cases} X_{n+1}^* & \text{if } \hat{g}(X_{n+1}^*) < \hat{g}(X_n) - \epsilon_n \\ X_n & \text{otherwise} \end{cases}$$

gave conditions on $F_n$, $\{\epsilon_n\}$, $\{\lambda_n\}$ which insure that

$$g(X_n) \overset{D}{\to} \text{ess inf } g(x) \text{ w.p.1.}$$

Further refinements may be found in [14,18,21,30,37].

The grant has also been used to partially support several applied pattern recognition projects. Under the direction of J.K. Aggarwal, a flying spot scanner system with color capabilities has been fabricated and used to analyze aerial color infrared photographs [24,28], primarily for the detection of diseases in citrus trees. Additionally, a computer analysis of planar curvilinear moving images was undertaken in [25].

Finally, the grant has supplied partial support for system theory work in [6,7], for the analysis and design of digital filters in [1,8,10, 15,16], and the analysis of computer storage systems in [9].
SECTION II. PAPERS PUBLISHED WITH AFOSR GRANT 72-2371 SUPPORT.

If a paper was published in the proceedings of a conference and later in a refereed journal, only the listing for the journal publication is given below.


[9] T.J. Wagner and P.A. Franaszek, "Some distribution-free aspects of
paging algorithm performance," Journal of the Association for Com-
[12] T.M. Cover and T.J. Wagner, "Topics in statistical pattern recogni-
tion," Digital Pattern Recognition, K.S. Fu, Editor, Springer-Verlag,
[14] L.P. Devroye, "Probabilistic search as a strategy selection procedure,
IEEE Transactions on Systems, Man and Cybernetics, SMC-6, 315-321,
1976.
sive digital filters employing floating-point arithmetic," IEEE
digital filters by phase correction," IEEE Transactions on Circuits
in error estimation," IEEE Transactions on Information Theory, IT-22,


Accepted Papers


Papers Submitted


Papers in Preparation

[40] T.J. Wagner and L.P. Devroye, "Generalized kernel estimates."


SECTION III. PERSONNEL SUPPORTED BY AFOSR 72-2371.

6/1/72 - 5/31/73

- T.J. Wagner, Professor 3 summer months (full time)
- J.K. Aggarwal, Professor 3 summer months (full time)
- Baolian Liu, RA* 3 months (50% time)
- D.J. Thompson, RA 8 months (22.5% time)

6/1/73 - 5/31/74

- T.J. Wagner, Professor 3 summer months (full time)
- J.K. Aggarwal, Professor 3 summer months (full time)

6/1/74 - 5/31/75

- T.J. Wagner, Professor 3 summer months (full time)
- J.K. Aggarwal, Professor 3 summer months (full time)
- L.P. Devroye, RA 9 months (50% time)

6/1/75 - 5/31/76

- T.J. Wagner, Professor 2 summer months (full time)
- J.K. Aggarwal, Professor 1½ summer months (2/3 time)
- L.P. Devroye, RA 7 months (50% time)
- O. Teoh, RA 4½ months (25% time)
- M. Day, RA 3½ months (12.5% time)

6/1/76 - 5/31/77

- T.J. Wagner, Professor 2 summer months (full time)
- L.P. Devroye, RA 7 months (50% time)
- M. Ali, PRA** 3 months (full time)
- M. Ali, PRA 2½ months (33.3% time)

* Research Assistant
** Post-doctoral Research Associate
**REPORT DOCUMENTATION PAGE**

<table>
<thead>
<tr>
<th>1. REPORT NUMBER</th>
<th>2. GOVT ACCESSION NO.</th>
<th>3. RECIPIENT'S CATALOG NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFOSR-TR-77-0879</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. TITLE (and Subtitle)</th>
<th>5. TYPE OF REPORT &amp; PERIOD COVERED</th>
</tr>
</thead>
<tbody>
<tr>
<td>THE STUDY OF DISTRIBUTION-FREE PERFORMANCE BOUNDS FOR NONPARAMETRIC DISCRIMINATION ALGORITHMS</td>
<td>Final</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7. AUTHOR(s)</th>
<th>8. CONTRACT OR GRANT NUMBER(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. J. Wagner</td>
<td>AFOSR 72-2371</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. PERFORMING ORGANIZATION NAME AND ADDRESS</th>
<th>10. PROGRAM ELEMENT, PROJECT, TASK AREA &amp; WORK UNIT NUMBERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Texas</td>
<td>61102F 2304/A5</td>
</tr>
<tr>
<td>Department of Electrical Engineering</td>
<td></td>
</tr>
<tr>
<td>Austin, Texas 78712</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>11. CONTROLLING OFFICE NAME AND ADDRESS</th>
<th>12. REPORT DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Force Office of Scientific Research/NM</td>
<td>20 Jun 77</td>
</tr>
<tr>
<td>Bolling AFB, Washington, DC 20332</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>14. MONITORING AGENCY NAME &amp; ADDRESS(If different from Controlling Office)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>15. SECURITY CLASS. (Of this report)</th>
<th>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNCLASSIFIED</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>16. DISTRIBUTION STATEMENT (Of this Report)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approved for public release; distribution unlimited.</td>
</tr>
</tbody>
</table>

| 17. DISTRIBUTION STATEMENT (Of the abstract entered in Block 20, If different from Report) |

<table>
<thead>
<tr>
<th>18. SUPPLEMENTARY NOTES</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>19. KEY WORDS (Continue on reverse side if necessary and identify by block number)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nonparametric discrimination, pattern recognition, clustering, density estimation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>20. ABSTRACT (Continue on reverse side if necessary and identify by block number)</th>
</tr>
</thead>
</table>

The accomplishments of the work performed with the whole or partial support of Grant AFOSR 72-2371, 6/1/72-5/31/77, are outlined along with a listing of the publications and technical personnel supported by the grant.