REPOI AD-A273 259

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE

August 1993

3. REPORT TYPE AND DATES COVERED

Professional Paper

4. TITLE AND SUBTITLE

JOHNSON DISTRIBUTIONS FOR FITTING WEIGHTED SUMS OF SIGMOIDED NEURON OUTPUTS

5. FUNDING NUMBERS

PR: ST56/ST06
PE: 0603741D
WU: DN302231

6. AUTHOR(S)

W. C. Torrez and J. T. Durham

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Command, Control and Ocean Surveillance Center (NCCOSC)
RDT&E Division
San Diego, CA 92152–5001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Space and Naval Warfare Systems Command
Washington, DC 20363–5100

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

In this paper, it is shown that a continuum of distributions best characterizes the hidden layer outputs of a multilayer perceptron when trained as a 0-1 classifier and tested with a range of signal-to-noise ratio (SNR) input distributions. A four parameter system of transformed normal distributions, known as the Johnson system of distributions, is utilized to illustrate the shape of output distributions as a function of input SNR levels.

Published in *Proceedings of the World Congress on Neural Networks*, July 1993, Vol. IV, pp. 715–720.

14. SUBJECT TERMS

Neural Networks
Active Signal Processing

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAME AS REPORT

Best
Available
Copy

UNCLASSIFIED

21a. NAME OF RESPONSIBLE INDIVIDUAL	21b. TELEPHONE (include Area Code)	21c. OFFICE SYMBOL
W. C. Torrez	(619) 553-2020	Code 733

DTIC QUALITY INSPECTED 5

Accesion For	
NTIS CRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	20

93-29028

93 11 26 066

NSN 7540-01-280-5500

Standard form 298 (BACK)

UNCLASSIFIED

Johnson Distributions for Fitting Weighted Sums of Sigmoided Neuron Outputs

W. C. Torrez and J. T. Durham
Signal and Information Processing Division
NCCOSC RDT &E Division (NRaD)
Mailing Address: Commanding Officer (Code 733)
NCCOSC RDT&E DIV
53560 Hull Street
San Diego, CA 92152-5001

January 11, 1993

Abstract

In this paper, it is shown that a continuum of distributions best characterizes the hidden layer outputs of a multilayer perceptron when trained as a 0-1 classifier and tested with a range of signal-to-noise ratio (SNR) input distributions. A four parameter system of transformed normal distributions, known as the Johnson system of distributions, is utilized to illustrate the shape of output distributions as a function of input SNR levels.

1 Introduction

In this paper, a feedforward multilayer perceptron trained as 0-1 classifier with backpropagation of error is considered. It will be shown that the Johnson system of continuous distributions [1] can be used to characterize the continuum of signal to noise ratio (SNR) in terms of the third and fourth order moments of the distribution of pre-squashed neuron outputs (i.e., weighted sums of neuron inputs).

The Johnson system of distributions is generated by transformations of the form

$$Z = \gamma + \eta k(x; \lambda, \epsilon),$$

where Z is a standard normal variate. The parameters ϵ and λ are location and scale parameters, respectively, while η and γ are shape parameters. Johnson [2] suggested the following three functions k to cover a wide range of possible shapes:
k_1 defines the S_U distribution, where

$$k_1(x; \lambda, \epsilon) = sinh^{-1}\left(\frac{x-\epsilon}{\lambda}\right);$$

k_2 defines the S_B distribution, where

$$k_2(x;\lambda,\epsilon) = ln\left(\frac{x-\epsilon}{\lambda+\epsilon-x}\right);$$

and k_3 defines the S_L distribution, where

$$k_3(x;\lambda,\epsilon) = ln\left(\frac{x-\epsilon}{\lambda}\right).$$

The S_B family is bounded on $(\epsilon, \epsilon+\lambda)$ and the S_U family is unbounded on $(\epsilon, \epsilon+\lambda)$, where $\epsilon > x$ and $\lambda > 0$. The S_L distributions divide the skewness-by-kurtosis plane into two regions such that the S_B distributions lie in one of the regions and the S_U distributions lie in the other.

2 Fitting Johnson Distributions by the Method of Quantile Matching

Figure 1 presents the skewness-by-kurtosis plane for a set of distributions of weighted sums of squashed hidden layer outputs. As noted on the plot, for each distribution, the relative signal level (RSL) of the signals varies from 0 dB to -6 dB in decrements of 2 dB; the skewness and kurtosis of the noise only distribution is also noted on the plot. This plot clearly indicates the relationship of the skewness and kurtosis of these distributions to the SNR. As the RSL (i.e., SNR) decreases, the skewness ranges from negative to positive values while at the same time, the kurtosis first decreases when RSL is about -4 dB and then increases as the RSL continues to decrease.

The parameters $\eta, \gamma, \lambda, \epsilon$ are estimated by a refinement of the method of quantile matching as described by Slifker and Shapiro [3]. Let $x_1, ..., x_n$ be the given sample.
(1) For a given unit normal quantile z, calculate $p_z = P(Z < z), p_{3z} = P(Z < 3z)$, where Z is a standard normal variate. Set $p_{-z} = 1 - p_z, p_{-3z} = 1 - p_{3z}$.
(2) From the data, calculate the sample quantiles $Q(p_\zeta), \zeta = \pm z, \pm 3z$, as follows: Calculate $i = Np_\zeta + \frac{1}{2}$; then $Q(p_\zeta) = x_{(i)}$, where $x_{(i)}$ is the i^{th} ordered observation in the sample. Since i in general will not be an integer, it will be necessary to interpolate.
(3) Calculate $p = Q(p_z) - Q(p_{-z}), m = Q(p_{3z}) - Q(p_z), n = Q(p_{-z}) - Q(p_{-3z})$. We then check $c = \frac{mn}{p^2}$:

$If\ c > 1$, then use the S_U parameters.

$If\ c < 1$, then use the S_B parameters.

$If\ c = 1$, then use the S_L parameters.

(4) The formulas for the estimates of the parameters η, γ, λ, and ϵ are given in Reference [3] and will not be repeated here.

For an example distribution, Figure 2(a) shows the values of c as a function of the range of z. Here z ranges from 0.01 to 0.80, in increments of Δz=0.01. Figure 2(b) shows the corresponding Kolmogorov-Smirnov (KS) distance measuring the goodness of fit of the calculated Johnson distribution to the sample data. The authors have observed that these results depend on the resolution of the step size in z. Under the assumption that the quantile matching method places us within the neighborbood of a better fit, a stochastic optimization procedure was applied. This procedure is described in the next section.

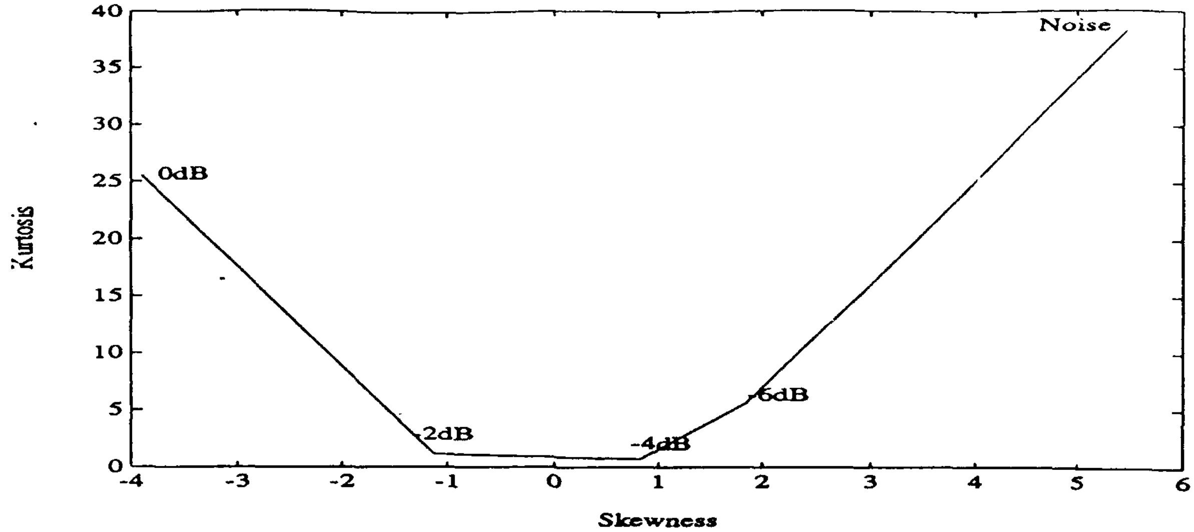


Figure 1: Skewness by Kurtosis Values of Noise Only Distribution and Signal Distributions Indexed by RSL.

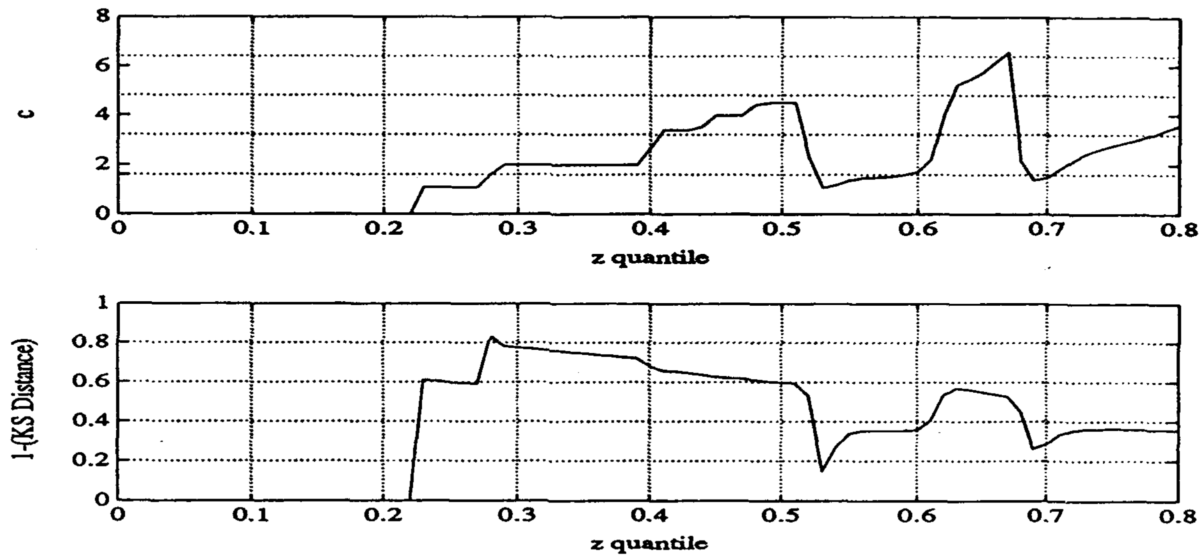


Figure 2: (a) Top: Plot of $c = \frac{mn}{p^2}$ as a Function of the Normal Quantile Value z. (b) Bottom: Plot of 1-(KS Distance) as a Function of the Normal Quantile Value z.

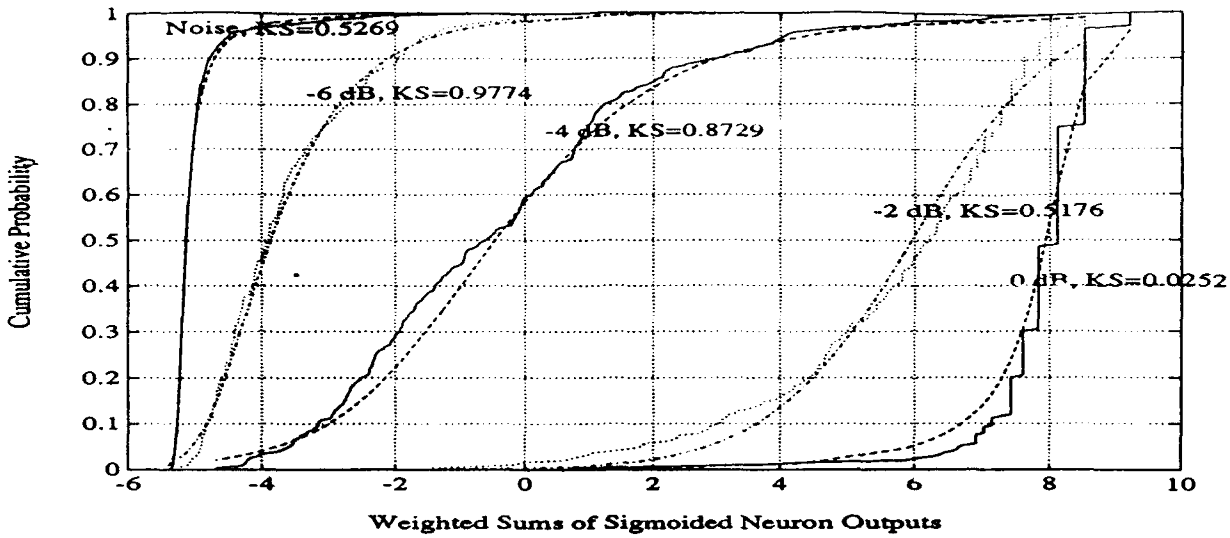


Figure 3: Johnson CDF's Overplotted with the Empirical Distribution Functions.

3 Stochastic Optimization of the Parameter Estimates

A further refinement for obtaining better density fits is the stochastic optimization of the parameter estimates afforded by randomly perturbing the estimates by a small uniformly random quantity. The results of this parameter optimization are shown in Figure 3. Also noted are the corresponding probability values for these fits. The randomly perturbed parameters were first evaluated with respect to the first and last quantiles of the data. If the fit was within an acceptable range, the entropy of the new distribution was then calculated using the empirical data samples and the new probability density function. If the entropy of the new distribution was greater than that of the best fitting distribution, the KS probability and KS distance were calculated for the new distribution. If the KS probability was within an acceptable range, the new distribution was assigned to be the parameter values. Finally, for each update of the parameters, the best KS fit was separately saved due to the fact that the KS is allowed to degrade as the entropy is maximized. Typically, 1000 different distributions were evaluated with a quadratic or third order cooling schedule. This ad hoc technique was developed for the timely production of improved fits. Such fits were needed to illustrate the change of shape of output distributions. An alternative method which calculates maximum likelihood estimates of the Johnson parameters is a focus of our most recent work [5].

4 Conclusions

As can be seen in Figure 4, the shape of the weighted sums is a function of the SNR of the neural network inputs. For performance measures such as ROC curves and Recognition Differentials, this feature of neural network classifiers illustrates that the overlapped noise and signal distributions are at least four parameter

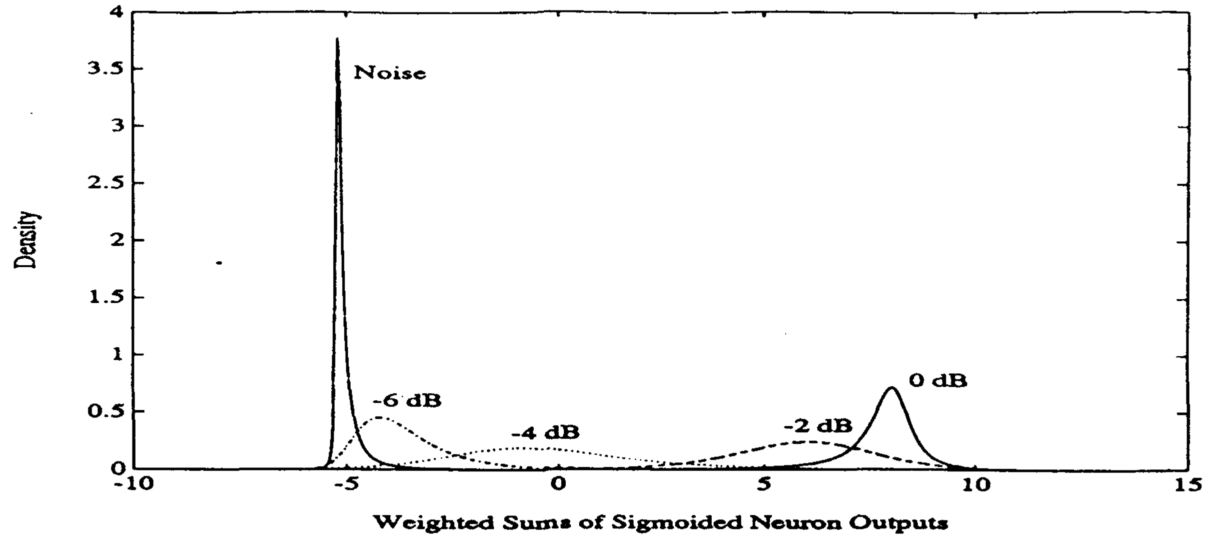


Figure 4: Johnson Densities Corresponding to the Fitted CDF's.

distributions. In previous work (References [1] and [4]), we have shown that for some cases, the two parameter Gaussian or three parameter lognormal distributions can provide reasonable fits. The results of this effort have shown that at least four parameters are generally needed.

In Figure 5, the skewness and kurtosis of the empirical data distributions (dotted line) is plotted with the corresponding fitted Johnsons. Although empirical estimates of skewness and kurtosis can be highly variable, the plot of the empirical estimates illustrates that shape is a function of SNR. The Johnson fits further illustrate this relationship. Note that the -2 dB, -4 dB, -6 dB, and Noise distributions are almost collinear with kurtosis on a log scale. Also note that the -2 dB distribution has the minimum kurtosis and is the most symmetric, i.e. skewness is approximately zero. This observed relationship suggests that in some cases, a change in shape can be easily parametrized as a line in the skewness versus log(kurtosis) plane.

In conclusion, the shape of the distributions of neuron outputs has been shown to be a function of the SNR of the input distributions. For performance measures such as ROC curves and Recognition Differentials (RD) this means that the change of only mean and variance cannot be assumed for a fixed distribution, e.g. Gaussian. Furthermore, a minimal parametrization, i.e. Johnson Distributions, can provide RD estimates which account for such changes in shape due to changes in input signal levels.

5 References

1. Torrez, W. C. and J. T. Durham, 1992: "Performance Measures for Neural Nets Using Johnson Distributions, " To Appear in the *Proceedings*, IEEE International Conf. on Neural Nets, San Francisco, CA, March 1993.

2. Johnson, N. L. , 1949: "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*,

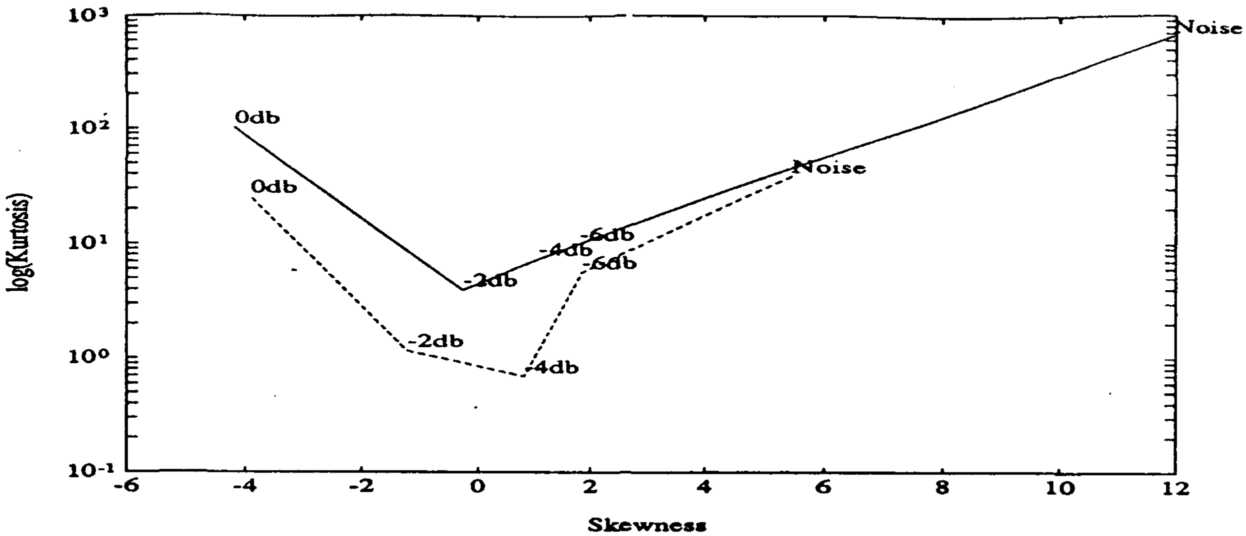


Figure 5: Comparison of Skewness and Kurtosis of Fitted CDF's with the Skewness and Kurtosis of the Data (Dotted Line).

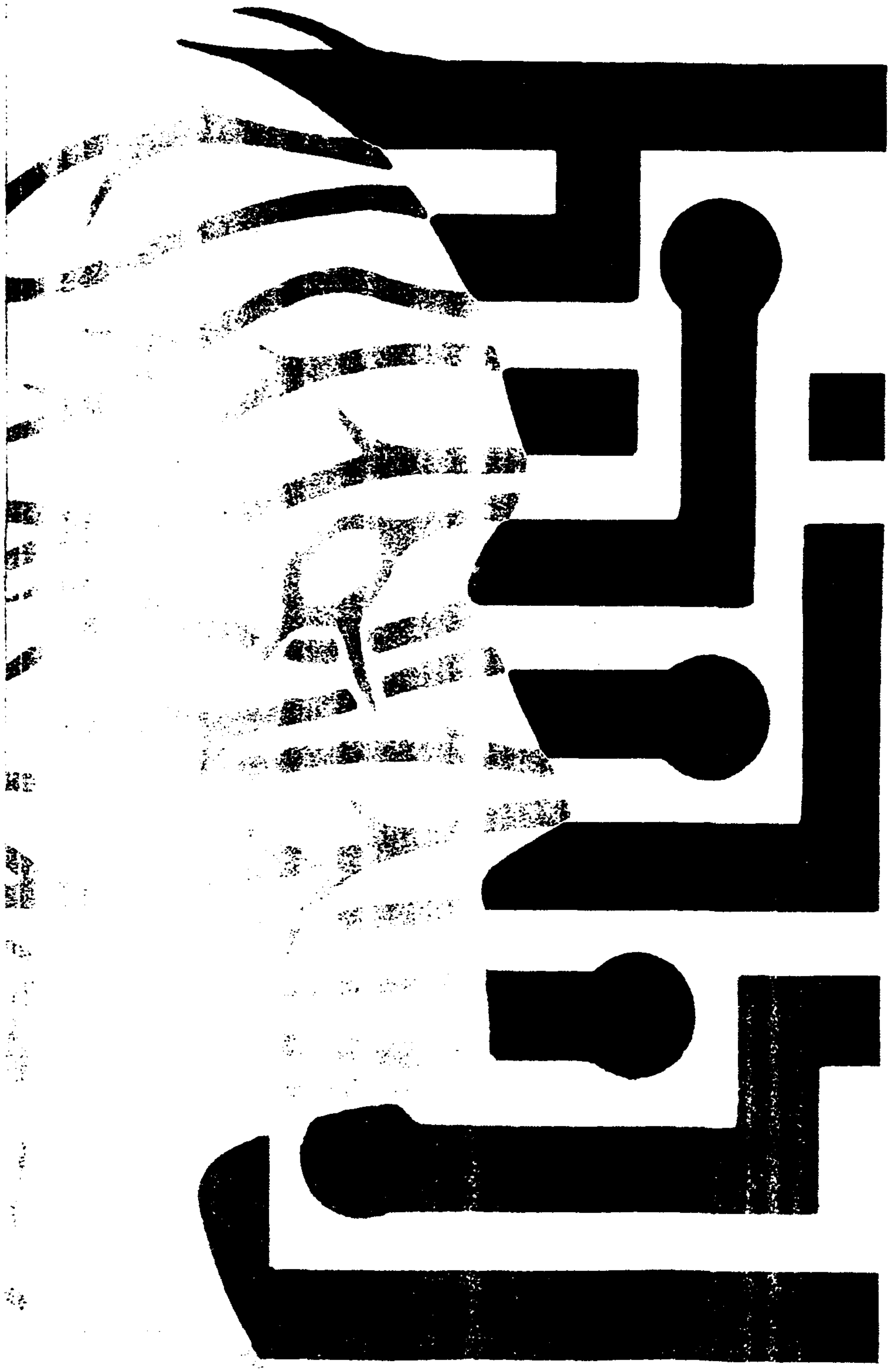
36: 149-176.

3. Slifker, J. F. and S. S. Shapiro, 1980: "The Johnson System: Selection and Parameter Estimation," *Technometrics*, 22: 239-246.

4. Durham, J. T., W. C. Torrez, and E. W. VonColln, 1992: "Performance Analysis of the Air Defense Initiative Neural Network Processor Using Data Sets from the E1 Test," *NRaD Tech. Document.* In Publication.

5. Durham, J. T. and W. C. Torrez, 1993: "A Monte Carlo Method for Fitting Johnson Distributions to Empirical Data," NRad Communication.

Volume IV



World Congress On Neural Networks

1993 International Neural Network Society Annual Meeting

Oregon Convention Center
Portland, Oregon
July 11-15, 1993

LEA