Holographic Neural Networks as Nonlinear Discriminants for Chemical Applications

Frank R. Burden

Chemistry Department, Monash University, Clayton, Victoria 3168, Australia

Received June 17, 1997[⊗]

A holographic neural network has been investigated for use as a discriminant. Six sets of artificial data and two data sets of infrared spectra, reduced using principal component analysis, of prepared cervical smears were analyzed by four regular discriminant methods as well as by the holographic neural network method. In all cases, it was found that the holographic neural network method gave comparable, and in some cases superior, results to the other discriminant methods. The holographic neural network method is simple to apply and has the advantage that it can be easily refined when new data become available without disturbing the original mapping. It is suggested that the holographic neural network method should be seriously considered when discrimination methods need to be applied.

INTRODUCTION

The use of chemical data for discriminating between classes of biological samples,[1] which may relate to human health, must be approached with great care. False classifications have the potential to result in the withholding of medical treatment from patients with severe illness. Other less sensitive areas where chemical data can be used to discriminate can have major financial implications such as the classification of diseased food stocks or area of origin.[2]

Automated chemical analysis often makes use of spectral techniques by which means a large amount of data can be gathered for each sample. Any excess of data can be reduced by use of principal component analysis, which removes noise and reduces the dimensionality of a problem. For instance, infrared spectra may be recorded at up to 2000 frequencies but after principal component analysis (PCA) reduce down to less than ten[1] components, and often to two or three, indicating that there are very few independent variables carrying information.

If the samples have been classified, then the scores from the PCA can be subjected to various forms of parametrized or nonparametrized discriminant analysis. Unparametrized methods, such as *k*-nearest-neighbors, *k*NN, can often perform well, though for small data sets there can be an unacceptable chance of misclassification, whereas for large data sets the computational time for classifying each new sample may become too long.

If a model of the classification is sought whereby data from different locations, or even different years at the same location, can be compared, then a parametrized method is needed. In this case, the various classes are represented by a set of coefficients, which is smaller than the set of classified samples.

Most parametrized models depend upon Bayes' theorem,[3] though often with assumptions made about the variance within the classes and of the data as a whole. There are very many discriminant methods, the most popular of which, judging from their inclusion in commercial software, embrace linear and quadratic discriminant analysis, LDA, QDA, and soft independent modeling by class analogy, SIMCA. LDA *and* SIMCA are linear models, while QDA allows for quadratic interactions. However, in some cases the data may be modeled best by several independent variables with nonlinear terms and cross-terms that may go beyond the quadratic. It is common knowledge that nonlinear multivariate analysis is difficult to manage and is often unstable with respect to the addition or removal of terms in a power series expansion.

Holographic neural networks,[4,5] HNN, while unfamiliar in the chemical literature, provide a potential, and certainly an alternative, solution to the problem. HNNs have been developed for use in pattern matching in such areas as handwriting and facial recognition and are readily adapted for the problems outlined above. As stated in ref 4, "the use of the term 'holographic' is suggested by the similarity to a class of mathematics found within electromagnetic field theory, and in the apparent enfolding of information within optical holograms". They rely on the overlaying of a large number of terms in this holographic sense, with various weighting, from which the pattern is reproduced. The training of an HNN is accomplished with very few iterations, often two or three, and the final model is expressed as a set of term coefficients together with associated weights. It should be noted that, "the holographic neural theory is fundamentally different from the standard connectionist models in artificial neural system theory".[4] It is unfortunate that there is not a useful body of literature that can be referred to for the HNN technique, though ref 4 does provide a more comprehensive outline than can be supplied here. Nevertheless, the discussion below does contain all that is necessary to implement the scheme.

The use of the term "neural network" here should not be confused with its more familiar use, as in "back-propagation neural network" and the like. The mapping achieved here is by a totally different algorithm.

In this paper the basic theory of HNNs is described and then applied to several sets of artificial data as well as to

[⊗] Abstract published in *Advance ACS Abstracts,* December 1, 1997.

S0095-2338(97)00286-2 CCC: $15.00
Published on Web 01/19/1998

some previously published data of the infrared spectra of treated cervical smears with a cancer related classification.[1]

THEORY

The basis of the holographic method[4] relies on transforming the data to vectors in the complex plane. The vectors have both direction and magnitude, which correspond respectively to the magnitude and weighting of the original data. In this paper, the original[4] HNN terminology has been retained to avoid confusion, though it may prove unfamiliar to many chemists. Let the independent variables representing the data be related to the dependent variables by

$$r_l = f(s_1,s_2,s_3,...,s_{N_S}) \tag{1}$$

where there are N_S independent variables, or stimuli, s_k, and N_R dependent variables, or responses, r_l. Each variable is now transformed to a vector in the complex plane by either

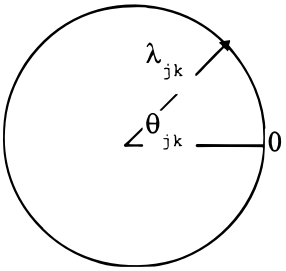
$$s_k \Rightarrow \lambda_k e^{i\theta_k} \tag{2}$$

or

$$r_l \Rightarrow \gamma_l e^{i\phi_l} \tag{3}$$

with λ_k and γ_l being the associated weightings, which are usually set to unity.

For the jth particular stimulus and response set (data point) s_{jk} and r_{jl}, where $k = 1...N_S$, $l = 1...N_R$, $j = 1...M$, become θ_{jk} and φ_{jl} with associated weightings λ_{jk} and γ_{jl}. Diagrammatically for stimulus s_{jk}



The particular jth stimulus to response relationship, s_{jk} to r_{jl}, is encoded as $\lambda_{jk}\gamma_{jl}e^{i(\varphi_{jl}-\theta_{jk})}$. The total data set of M relationships, S_k to R_l, is encoded as

$$\sum_{j=1}^{M} \lambda_{jk}\gamma_{jl}e^{i(\varphi_{jl}-\theta_{jk})} \tag{4}$$

which is a vector sum in the complex plane and produces a resultant vector of length

$$\chi_{kl} = [[\sum_{j=1}^{M} \lambda_{jk}\gamma_{jl}\cos(\varphi_{jl} - \theta_{jk})]^2 + [\sum_{j=1}^{M} \lambda_{jk}\gamma_{jl}\sin(\varphi_{jl} - \theta_{jk})]^2]^{1/2} \tag{5}$$

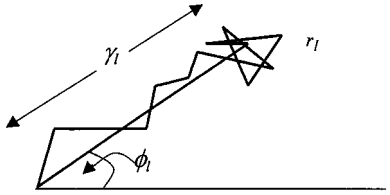


Figure 1. Buildup of a response vector (diagrammatic illustration of eq 7) from the individual contributions.

and angle

$$\eta_{kl} = \tan^{-1}\left[\frac{[\sum_{j=1}^{M} \lambda_{jk}\gamma_{jl}\sin(\varphi_{jl} - \theta_{jk})]}{[\sum_{j=1}^{M} \lambda_{jk}\gamma_{jl}\cos(\varphi_{jl} - \theta_{jk})]}\right] \tag{6}$$

taking care to ascribe the correct quadrant to the tangent.

In eq 5, χ_{kl} gives the average relationship between stimulus variables, s_k, and response variables, r_l. This can be represented in matrix terms as $\mathbf{X}+ = \mathbf{S}^T\cdot\mathbf{R}$, where the $+$ sign represents the sum expressed in eq 4, $\sum_{j=1}^{M}\lambda_{jk}\gamma_{jl}e^{i(\varphi_{jl}-\theta_{jk})}$. The matrix $\mathbf{X}$ represents the memory and may be thought of in terms of a cortex cell.

Decoding, that is the evaluation of a new response, not necessarily included in the training set, r_l*, when presented with a new stimuli, s_k*, is given by

$$r_l^* = \frac{1}{M}\sum_{k=1}^{M} \chi_{kl}\lambda_k^* e^{(\eta_{kl}+\theta_k^*)} = \gamma_l^* e^{i\varphi_l^*} \tag{7}$$

where θ_k* is produced from s_k* using eq 2 and λ_k* and the weighting of s_k* will usually, but not necessarily, be set to unity. The actual evaluation the magnitude to r_l* in eq 7 is made in a similar manner to eqs 5 and 6, the weighting giving the degree of confidence in the estimate of r_l*.

As explained above, the holographic nature of the encoding process is achieved through the addition of weighted vectors in the complex plane. The encoding matrix $\mathbf{X}$ is of dimensions $N_S \times N_R$ and may be stored in two parts, the magnitude and weighting of each element, or alternatively as matrices of sines and of cosines.

The resonse r* produced from a new stimulus s* is built up in a holographic manner from $\mathbf{S}^*\cdot\mathbf{X}$, where each term contributes its weight and magnitude to the building up of the response vector. The buildup of the response vector r_l* is shown diagrammatically in Figure 1, where the main components add to the length, γ_l, whereas the less important components tend to cancel one another out.

Enhanced Encoding. The above scheme needs to be supplemented in three ways. Firstly, the values of θ_{jk} and φ_{jl} of eqs 2 and 3 must be scaled to the range $0-2\pi$. This is can be done using a sigmoid, or similar, scaling function,

$$\theta_{jk} = \frac{2\pi}{1 + e^{(s_{jk}-\bar{s}_k)/\sigma_k}} \tag{8}$$

where $\bar{s}_k$ is the mean value of the s_{jk}'s and σ_k is the standard deviation, though the scaling may take other forms.

Table 1. Statistics[a] for Various Numbers of Independent Variables, N_s, and Orders of Statistics, O_s

N_s	max $O_s = 3$	max $O_s = 4$	max $O_s = 5$	max $O_s = 6$
2	18	28	40	54
3	38	68	110	166

[a] See eq 13 and related discussion.

Secondly, the encoding is also greatly enhanced by the use of higher order stimuli. Extra terms such as $...s_i^l s_j^m s_k^n...$ are included and encoded as $...\lambda_i^l \lambda_j^m \lambda_k^n e^{i(...l\theta_i + m\theta_j + n\theta_k)}$ and are known as *statistics*. Incorporation of these statistics also ensures a greater symmetry in the scheme of vectors about zero, which generally gives results that are more reliable.

Thirdly, symmetry can also be enhanced by the use of supplementary complex conjugate terms so that $e^{-i\theta}$ is included as well as $e^{i\theta}$ and so on. Reference 4 incorporates a table showing the number of statistics arising from various numbers of independent and dependent variables. Table 1 gives the statistics used in this work.

Iterations. Since each term χ_{kl} in eq 5 is built up from the M stimuli or data points, its contribution to the r^* response from the stimulus s^* has been averaged via the vector summation. The x_{lk}'s, of the memory matrix $\mathbf{X}$, can be refined to ensure an overall improved mapping of the network training set stimuli to the training set responses.

The original encoding expressed in eqs 5 and 6 represent the zeroth iteration

$$\mathbf{X}^0 += (\mathbf{S}^0)^{\mathrm{T}} \cdot \mathbf{R} \tag{9}$$

$\mathbf{X}^0$ is now used to produce an estimate of the training set responses, $\mathbf{R}^1$, where

$$\mathbf{R}^1 = \frac{1}{M}\mathbf{S}^0 \cdot \mathbf{X}^0, \quad \mathbf{X}^1 += (\mathbf{S}^0)^{\mathrm{T}} \cdot \mathbf{R}^1, \quad \text{and again}$$

$$\mathbf{R}^2 = \frac{1}{M}\mathbf{S}^0 \cdot \mathbf{X}^1 \quad \text{and so on} \tag{10}$$

In the author's experience, the convergence, to an acceptable tolerance, is rapid though it may be that some ill-conditioned problems do not converge as rapidly. In the present work, two iterations were found to be sufficient in all cases.

Implementation. Although the scheme might appear to be somewhat complex at first sight, it is remarkably easy to encode and interpret, with the bulk of the coding done in the sine and cosine representations of the complex algebra as expressed in eqs 5 and 6. The work illustrated here was carried out using a small *MATLAB*[6] program, written by the author for investigative purposes. A commercial program, HNeT,[7] is also available, though it is not set up for discriminant work.

RESULTS

The HNN method has been applied to eight sets of data; the first six were generated using random numbers, while the last two were derived by principal component analysis, PCA, from 120 samples of the spectra of processed cervical smears, related to the screening for cancer.[1,9] All results were produced using a "leave-one-out" cross-validation.

The randomly generated samples are grouped as follows and summarized in Table 1. They have been generated

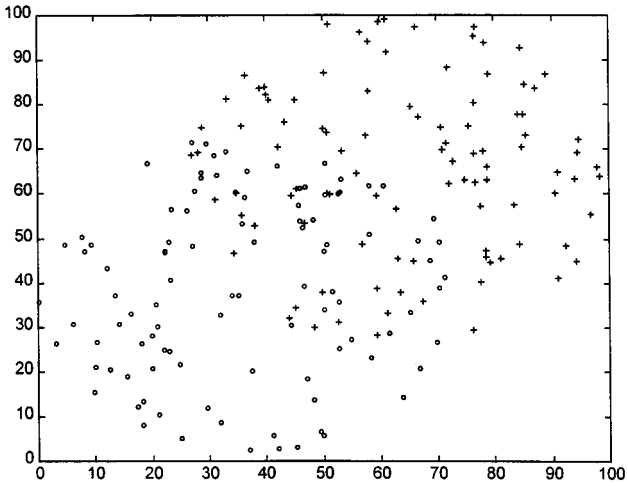


Figure 2. Data set 1 (Table 2), overlapping sets.

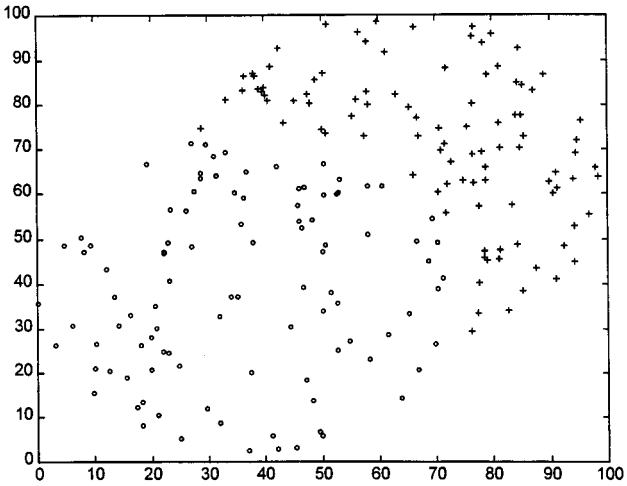


Figure 3. Data set 2 (Table 1), non-overlapping sets.

within a 100 × 100 grid in the two-dimensional case and a 100 × 100 × 100 grid in the three-dimensional case.

The eight data sets were analyzed by five methods: (i) a k-neighbors method, kNN (it was found that $k = 7$ gave the best discrimination in most of the cases presented here, the optimized values of k are given in the tables of results); (ii) a linear discriminant analysis, LDA, the results of which can be understood with the aid of a ruler on the two-dimensional plots in Figures 2−4; (iii) a quadratic discriminant analysis, QDA; (iv) a soft independent modeling by class analogy, SIMCA; and (v) a holographic neural network, HNN.

The holographic method, HNN, was applied with two orders of statistics, 4 and 5, and warrants a full description. As will be seen later, the HNN method does better in some situations.

Each of the HNN calculations needs to be supplied with five parameters:

(a) As given in eq 8, the stimuli and responses of the HNN were scaled using a sigmoidal function before eqs 2 and 3 were implemented. Adjustable parameters C_S and C_r have been incorporated to improve the distribution of the data about $\theta = 0$ and $\phi = 0$. For example in the case of C_s,

$$\theta = 2\pi(1 - e^{[-c_s(s-\mu)]/\sigma}) \tag{11}$$

Optimal values of C_s and C_r were found to be either 1.1 or 1.2, and the values used are given in the tables. The final

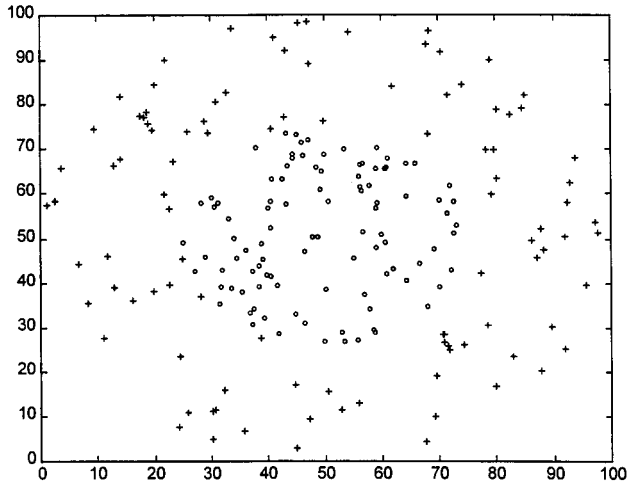


Figure 4. Data set 3 (Table 2), encompassing sets.

results were obtained after applying the inverse transformation to the responses.

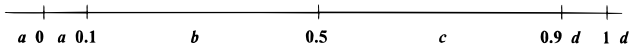
(b) as with most iterative schemes, a damping factor is necessary, and in this case it was set to 0.5, though the convergence, as described above under Iterations, is not very sensitive to it. In all cases, the number of iterations never exceeded 2.

(c) Finally, the maximum order of statistics needs to be set. Since the actual number of statistics is given by

$$\sum_{O_s=1}^{\max O_s} \frac{(N_s + O_s - 1)!}{(N_s - 1)!O_s!} \tag{12}$$

where N_S is the number of independent variables and O_S is the order of the statistics. In the present case $N_S = 2$ or 3 and O_S was found to be optimal at either 4 or 5. The number of statistics is doubled when, as in this case, the complex conjugates are included. The use of complex conjugates adds some symmetry to the representation of the stimuli about the line $\theta = 0$ and leads to a more accurate mapping.[4] The number of statistics, including complex conjugates, for each case, is shown in Table 1. It is evident that the number of statistics grows rapidly, both with the number of independent variables and with the order of the statistics. *It should be noted that* although the number of statistics, or weights, used in the *holographic process* will normally be much larger than in the case of the other discriminant models, there is no cause for concern other than, as in the case of back-propagation neural networks,[8] when the number of statistics approaches the number of individual training stimuli. O_S values in excess of 5 were considered here, and shown by experience, to give rise to more statistics than are warranted by the size of the data sets, and O_S values less than 4 were found to be insufficient.

The term *misclassification* needs some amplification when applied to the HNN results. Whereas the other discriminant methods return a decisive answer for each point to be classified, the HNN can potentially return any answer between 0 and 1, or even outside this range, and a possible interpretation of results that deviate from the extremes, 0 or 1, is the degree of uncertainty with which they are being classified. If the classification domain is represented by four regions *a*, *b*, *c*, and *d* and the tolerance level is set at 0.1, as shown,



then a point that should have been classed as 0 can fall into one of the four regions *a*, *b*, *c*, or *d*. If it falls in region *a*, then it is considered as correctly classified; if it falls in region *b*, then under a stringent classification scheme it can be considered as misclassified as 1, or as a false positive (FP), though under a less stringent scheme it can still be considered as correctly classified. The inverse reasoning applies to points that should have been classified as 1 and regions *c* and *d*, known as a false negative (FN). In this work the training tolerance has been set at two levels, 0.1 to represent a stringent discrimination and 0.5 to represent a similar situation to the other discriminants where each point is definitively ascribed to one class or the other. The actual deviation from the true classification of 0 or 1 can also be thought of as representing the degree of uncertainty in the result and will be discussed further with regard to data sets 7 and 8.

Table 2. Data Sets

data set	dimension	nature	data set	dimension	nature
1	2D	overlapping[a]	5	3D	non-overlapping[e]
2	2D	non-overlapping[b]	6	3D	encompassing[f]
3	2D	encompassing[c]	7	2D	cancer[g]
4	3D	overlapping[d]	8	3D	cancer[h]

[a] Radius = 36.94. Distance between centers = 36.94. Perimeter of each circle touching the center of the other circle. 100 points in each class. Points from each class exist in the overlapping region. [b] As for a. Points of only one class exist in the overlapping region. [c] Circle of radius 25 at the center of a circle of radius 50. Points of class 2 excluded from the inner circle and points of class 1 excluded from the region between the perimeters of the circles. [d] Spheres of radius = 36.94 and distance between centers = 36.94. Perimeter of each sphere touching the center of the other sphere. 200 points in each class. Points from each class exist in the overlapping region. [e] As for *a*. Points of only one class exist in the overlapping region. [f] Sphere of radius 25 at the center of a sphere of radius 50. Points of class 2 excluded from the inner sphere and points of class 1 excluded from the region between the surfaces of the spheres. [g] Cervical smear data from refs 1 and 9. Two principal components used, account for 90% of the variance. [h] Cervical smear data from refs 1 and 9. Three principal components used, account for 94% of the variance.

Analysis of the Data Sets. The results for the eight data sets are presented in Tables 3 and 5−7. The false positives, FP, and negatives, FN, are shown alongside the percentage sensitivity {100 × TP/(TP + FN)} and specificity {100 × FP/(FP + TN)}.

Data Sets 1−3. Two overlapping circles are illustrated in Figures 2−4 and described in Table 2. These are of three types: overlapping circles where each circle, defining the different classes, is filled with 100 randomly generated data points. In data set 1 the overlapping region contains points from both classes. In data set 2 the overlapping region contains points from only one of the classes. In data set 3, one circle encompasses the other and the classes are separated by the boundary of the inner circle.

Table 3 shows that the HNN method can often discriminate as well as the unparametrized *k*NN method and outperform the parametrized methods. This will not always be so since other pairs of the data sets will have their own specific

Table 3. Cross-Validated Misclassification Results, Two-Dimensional Data Sets[a]

			tol = 0.1				tol = 0.5			
method	set	param	FP[b]	FN[b]	sens[b]	spec[b]	FP[b]	FN[b]	sens[b]	spec[b]
kNN	1	$k = 7$[c]	N/A	N/A	N/A	N/A	16	25	77	82
LDA	1		N/A	N/A	N/A	N/A	18	18	82	82
QDA	1		N/A	N/A	N/A	N/A	17	17	83	83
SIMCA	1		N/A	N/A	N/A	N/A	18	19	81	82
HNN[e]	1	$O_s = 4$[d]	16	28	75	84	11	27	77	87
HNN[e]	1	$O_s = 5$[d]	16	31	73	81	13	29	76	84
kNN	2	$k = 3$[c]	N/A	N/A	N/A	N/A	2	1	99	99
LDA	2		N/A	N/A	N/A	N/A	16	2	98	86
QDA	2		N/A	N/A	N/A	N/A	9	3	96	92
SIMCA	2		N/A	N/A	N/A	N/A	0	30	77	100
HNN[e]	2	$O_s = 4$[d]	2	2	98	98	2	2	98	98
HNN[e]	2	$O_s = 5$[d]	2	2	98	98	4	4	96	96
kNN	3	$k = 7$[c]	N/A	N/A	N/A	N/A	7	4	96	93
LDA	3		N/A	N/A	N/A	N/A	51	51	49	49
QDA	3		N/A	N/A	N/A	N/A	4	3	97	96
SIMCA	3		N/A	N/A	N/A	N/A	99	0	100	50
HNN[e]	3	$O_s = 4$[d]	2	21	82	98	2	19	84	98
HNN[e]	3	$O_s = 5$[d]	0	20	83	80	0	19	84	81

[a] Data sets as defined in Table 2 with 100 points in each class. [b] TP, true positive; FP, false positive; TN, true negative; FN, false negative; sens(itivity) = 100 × TP/(TP + FN); spec(ificity) = 100 × TN/(FP + TN). [c] Optimized value of k. [d] O_s = order of statistics. [e] Optimized scaling factors: $C_s = 1.1$; $C_r = 1.1$.

Table 4. Sensitivity of Misclassifications, FP + FN, for Data Set 1, to Data Scaling Factors C_s and C_r in Equation 12, $O_s = 4$

C_r	$C_s = 1.0$	$C_s = 1.1$	$C_s = 1.2$	$C_s = 1.3$
1.0	46	41	43	43
1.1	41	38	38	39
1.2	47	42	43	43

conditions and may suit another particular discriminant. For instance, where the classes are well-separated, then all methods will be equal, whereas if the two classes clearly lay on either side of a straight line, in a two-dimensional case, then the LDA method cannot be improved upon. It can be seen in Table 3 that the QDA method is the best discriminant for set 3, the two-dimensional encompassing sets, whereas the HNN method is best for set 2, the non-overlapping sets. There is also little difference between the results using a tolerance of 0.1 or of 0.5.

Table 4 shows the sensitivity of the total misclassifications, FP + FN, to changes in C_s and C_r, the scaling factors of eq 12. It can be seen that a minimum false alarm rate was achieved in this case with $C_s = 1.1$ and $C_r = 1.1$. These values were also optimal for the other randomly generated data sets 2−6.

Table 5 shows the set of 28 angles and weights, or lengths, for data set 1 using $O_s = 4$ and presented in descending order of weight. These angles and weights are the memory as represented by the **X** matrix in the association $\mathbf{X}+ = \mathbf{S}^{\mathrm{T}}\cdot\mathbf{R}$ of eq 10. The weights are the vector lengths of eq 5, and the angles are expressed in eq 6. Small weights represent weak average associations between the stimuli and responses for the given statistic.

2. Data Sets 4−6. These data sets are the equivalents of data sets 1−3 but in three dimensions with spheres rather than circles. Table 6 shows that the HNN method performs in a manner similar to that discussed above in that HNN performs better than the other methods for set 5, the non-overlapping set, and somewhat better for set 4, the non-overlapping set.

3. Data Sets 7 and 8. These data were taken from a recent paper,[1] together with some more recent additions,[9]

Table 5. Values and Weights[a] for the Statistics of Data Set 1 Using $O_s = 4$

no.	j[b]	k[b]	angle	weight
1	1	0	6.250	8.493
2	−1	0	3.174	8.485
3	0	1	6.234	8.065
4	0	−1	3.190	8.062
5	−2	0	3.083	5.545
6	2	0	0.060	5.517
7	0	−2	2.781	5.152
8	0	2	0.363	5.146
9	1	2	3.744	4.142
10	−1	−2	5.683	4.125
11	44	0	2.598	3.732
12	−4	0	0.543	3.716
13	−3	−1	5.710	3.359
14	3	1	3.711	3.355
15	2	1	3.309	3.024
16	−2	−1	6.121	3.014
17	−2	−2	6.268	2.455
18	2	2	3.151	2.450
19	−1	−3	4.674	1.625
20	1	3	4.757	1.620
21	1	1	1.811	1.618
22	−1	−1	1.347	1.602
23	3	0	5.350	1.100
24	−3	0	4.065	1.097
25	0	−4	6.189	0.996
26	0	4	3.241	0.993
27	0	−3	4.375	0.757
28	0	3	5.020	0.756

[a] Arranged in order of decreasing weight. [b] Enhanced stimulus s_1^j $s_2^k = \lambda^j\lambda^k e^{i(j\theta_1 + k\theta_2)}$, where s_1 is the first independent variable and s_2 is the second.

where the infrared spectra, recorded at 500 wavelengths, of 120 processed cervical smears were reduced using PCA to a two- or three-dimensional problem. In the analysis presented here it was found that two principal components account for 90% of the variance and that three components account for 94%. Each data point had been ascribed, by a pathologist, to one of two classes relating to the presence or absence of the onset of cancer. (Figure 5). It is of a rather different nature to the random data sets in that the two classes

Table 6. Cross-Validated Misclassification Results, Three-Dimensional Data Sets[a]

			tol = 0.1				tol = 0.5			
method	set	param	FP[b]	FN[b]	sens[b]	spec[b]	FP[b]	FN[b]	sens[b]	spec[b]
kNN	4	$k = 7$[c]	N/A	N/A	N/A	N/A	31	27	86	84
LDA	4		N/A	N/A	N/A	N/A	29	31	85	85
QDA	4		N/A	N/A	N/A	N/A	29	32	84	85
SIMCA	4		N/A	N/A	N/A	N/A	31	32	84	84
HNN[e]	4	$O_s = 4$[d]	34	36	82	83	26	30	85	87
HNN[e]	4	$O_s = 5$[d]	42	44	78	79	27	33	84	86
kNN	5	$k = 3$[c]	N/A	N/A	N/A	N/A	11	15	93	94
LDA	5		N/A	N/A	N/A	N/A	31	47	78	83
QDA	5		N/A	N/A	N/A	N/A	31	42	80	84
SIMCA	5		N/A	N/A	N/A	N/A	49	37	80	77
HNN[e]	5	$O_s = 4$[d]	21	13	93	90	21	11	94	90
HNN[e]	5	$O_s = 5$[d]	23	9	95	89	21	8	96	90
kNN	6	$k = 7$[c]	N/A	N/A	N/A	N/A	25	16	92	88
LDA	6		N/A	N/A	N/A	N/A	100	88	53	53
QDA	6		N/A	N/A	N/A	N/A	24	15	92	89
SIMCA	6		N/A	N/A	N/A	N/A	0	200	50	0
HNN[e]	6	$O_s = 4$[d]	8	45	81	95	8	44	81	95
HNN[e]	6	$O_s = 5$[d]	10	42	82	94	7	41	82	96

[a] Data sets as defined in Table 2 with 100 points in each class. [b] TP, true positive. FP, false, positive; TN, true negative; FN, false negative; sens(itivity) = 100 × TP/(TP + FN); spec(ificity) = 100 × TN/(FP + TN). [c] Optimized value of k. [d] O_s = order of statistics. [e] Optimized scaling factors: $C_s = 1.1$, $C_r = 1.1$.

Table 7. Misclassification Results, Cancer Related Data Sets[a]

			tol = 0.1				tol = 0.5			
method	set	param	FP[b]	FN[b]	sens[b]	spec[b]	FP[b]	FN[b]	sens[b]	spec[b]
kNN	7[f]	$k = 7$[c]	N/A	N/A	N/A	N/A	8	14	69	89
LDA	7		N/A	N/A	N/A	N/A	3	14	72	85
QDA	7		N/A	N/A	N/A	N/A	3	14	72	85
SIMCA	7		N/A	N/A	N/A	N/A	3	22	62	78
HNN[e]	7	$O_s = 4$[d]	8	27	53	87	6	14	70	84
HNN[e]	7	$O_s = 5$[d]	17	28	44	76	5	14	71	84
kNN	8[g]	$k = 3$[c]	N/A	N/A	N/A	N/A	3	14	72	96
LDA	8		N/A	N/A	N/A	N/A	3	13	73	96
QDA	8		N/A	N/A	N/A	N/A	3	13	73	96
SIMCA	8		N/A	N/A	N/A	N/A	6	28	54	73
HNN[e]	8	$O_s = 4$[d]	21	28	39	72	10	7	81	88
HNN[e]	8	$O_s = 5$[d]	18	40	34	69	11	17	62	85

[a] Data sets are defined in Table 2 with 39 points in class 1 and 81 points in class 2. [b] TP, true positive. FP, false positive; TN, true negative; FN, false negative; sens(itivity) = 100 × TP/(TP + FN); spec(ificity) = 100 × TN/(FP + TN). [c] Optimized value of k. [d] O_s = order of statistics. [e] Optimized scaling factors: $C_s = 1.2$; $C_r = 1.2$. [f] Two principal components accounting for 90% of the variance. [g] Three principal components accounting for 94% of the variance.

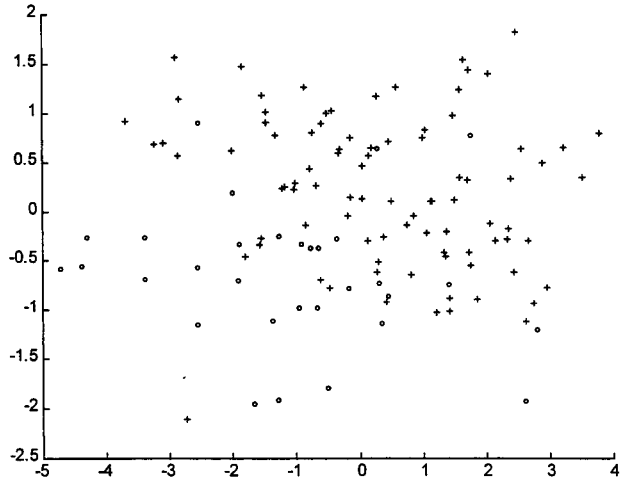


Figure 5. Cancer data (18 +'s and 39 O's), two principal components.

may not fully represent the diversity of the data. There may be, and most likely are, some data points which are from intermediate cases where cancer is not fully developed and the resultant spectra show characteristics of both classes. Table 7 shows that the HNN method still performs at a level equivalent to the other methods when taken at the 0.5 tolerance level. The more stringent tolerance of 0.1 excludes data that lay in the region 0.1−0.9 and may represent dysplastic samples, i.e. samples in which some cells are showing changes that indicate the onset of cancer but are not fully cancerous.

CONCLUSIONS

The holographic neural network, HNN, method is a serious contender as a discriminant method for overlapping data sets of the various types described above. It is a quick method to implement, is simply programmed, and can be easily updated when more training data become available since the new data can be added into the existing model without having to retrain on the whole of the data set, this feature being a great advantage when the data sets become large.

The HNN method gives a *definitive* model in that a set of term values and weights for the statistics used is produced. Unlike the more popular back-propagation artificial neural

networks[7] which usually start off with a random set of weights, and hence may not produce the same model when trained on the same data a second time, the HNNs produce the same results at each separate training on the same data. The term values and weights are also stable with respect to the incorporation of new training data. The only parameters that need to be set are the values of the scaling function terms C_s and C_r, which will normally be small deviations from unity, and the order of statistics, O_s, which must be set, by inspection, according to the size of the data set being investigated.

The drawback of HNNs are largely in their unfamiliarity, a drawback that popular back-propagation artificial neural networks are now overcoming in other areas now that their potential is becoming recognized. The HNNs have the potential in favorable cases, which only more experience will dictate, to perform better than the more conventional parametrized method which include LDA, QDA, and SIMCA.

Work is now underway to explore the use of holographic neural networks in other areas such as quantitative structure−activity relationships, QSAR, as well as discriminant analysis where there are several classes to be discriminated. Comparisons with back-propagation neural networks will be evaluated in future work.

REFERENCES AND NOTES

(1) Wood, B. R.; Quinn, M. A.; Burden, F. R.; McNaughton, D. An Investigation into FTIR Spectroscopy as a Biodiagnostic Tool for Cervical Cancer. *Biospectroscopy* **1996**, *2*.

(2) Aparicio, R.; Ferreiro, L.; Leardi, R.; Forina, M. Building Decision Rules by Chemometric Analysis. *Chemom. Intell. Lab. Syst.* **1991**, *10*, 349−358.

(3) Hand, D. J. *Discrimination and Classification*; Wiley: New York, 1981.

(4) Sutherland, J. G. In *The Holographic Neural Method, in Fuzzy, Holographic and Parallel Intelligence: The Sixth Generation Breakthrough*; Soucek, B., the Iris Group, Eds.; Wiley: New York, 1992.

(5) Sutherland, J. G. Holographic Model of Memory, Learning and Expression. *Int. J. Neur. Syst.* **1990**, *1*, 256−267.

(6) *MATLAB*; The MathWorks Inc.: MA.

(7) *HNeT*; AND America Ltd.: Oakville, Ontario, Canada.

(8) Livingstone, D. *Data Analysis for Chemists*; Oxford Science, OUP: Oxford, U.K., 1995.

(9) Romeo, M.; Wood, B. Private communication, Chemistry Department, Monash University, Clayton, Australia.

CI9702860