Pattern Analysis & Applications (1999)2:53–58

Introducing Locality and Softness in Subspace Classification*

Thiagarajan Balachander and Ravi Kothari

Artificial Neural Systems Laboratory, Department of Electrical & Computer Engineering & Computer Science, University of Cincinnati, Cincinnati, OH, USA

Abstract: Subspace classifiers classify a pattern based on its distance from different vector subspaces. Earlier models of subspace classification were based on the assumption that individual classes lie in unique subspaces. In later extensions, locality was introduced into subspace classification allowing for a class to be associated with more than one sub manifold. The local subspace classifier is thus a piecewise linear classifier, and is more powerful when compared to the linear classification performed by global subspace methods. We present extensions to the basic subspace method of classification based on introducing locality and softness in the classification process. Locality is introduced by (subspace) clustering the patterns into clusters, and softness is introduced by allowing a pattern to be associated with more than one cluster. Our motivation for introducing both locality and softness is based on the premise that by introducing locality, it is possible to reduce the bias though at the cost of a possible increase in variance. By introducing softness (or aggregation), the variance can be reduced. Consequently, by introducing both locality and softness, we avoid the possibility of high variance that locality typically introduces. We derive appropriate algorithms to construct a local and soft model of subspace classifiers and present results obtained with the proposed algorithm.

Keywords: Bias; Local classifier; Subspace classification; Subspace clustering; Variance

1. INTRODUCTION

Subspace methods of pattern recognition are based on the hypothesis that individual classes lie in different vector subspaces. Once the projection subspace for each class is determined, a pattern is assigned to a class based on its distance from the different vector subspaces.

One of the first subspace classifiers proposed was CLAFIC (CLAss Featuring Information Compression) [1]. In CLAFIC, the subspace corresponding to each class is determined by a principal component analysis of patterns belonging to that class. However, the drawback of CLAFIC is that the subspaces corresponding to each class are determined without taking into account the features of the other class(es). Consequently, improvements over CLAFIC have been proposed and can be categorised into: (i) those that directly determine the subspaces given the training data set [2]; or (ii) those which adaptively compute the subspaces [2,3]. All these subspace methods of pattern recognition are however global in nature. For each class there exists only one projection subspace over the entirety of the input space.

Local subspace methods of pattern recognition are based on a more general data model, and allow patterns of the same class to be associated with more than one sub-manifold. The sub-manifold that a particular pattern is associated with depends upon its location in the input space. The local subspace classifier is thus a piecewise linear classifier (or at most can be a piecewise quadratic classifier when the dimensionality of the subspaces for the different classes in a cluster are not the same), and is more powerful when compared to the linear classification performed by global subspace methods.

In Laaksonen [4], locality is introduced by combining subspace projection methods with nearest neighbour classification techniques. For each input vector to be classified, subspaces are formed for each class by identifying patterns from that class which are close to the input pattern. As in any nearest neighbour technique, the entire training data has to be stored and prototype vectors for calculating the local subspaces have to be identified every time a new vector is to be classified. This might be a disadvantage

Received: 4 November 1998
Received in revised form: 7 December 1998
Accepted: 7 December 1998
** Presented at ICAPR '98*

when training data set size is large. Local subspace pattern classification has also been presented by Prakash and Murty [5], where a dynamic clustering [6] method is used to partition the data and a projection subspace for clusters in each class is obtained. However, local subspaces rather than local linear manifolds are used to model the data, which constrains the generality of the data model (although it is mentioned that local linear manifold is a better way to model the data). In Kohonen et al [7], the Adaptive Subspace SOM (ASSOM) is introduced, in which each cluster learns a specific low dimensional representation of different temporal events from a stream of consecutive input patterns (like speech inputs). The units (clusters) get tuned to different features in the inputs, but ASSOM also imposes a spatial ordering in the clusters through self organisation.

Though a motivation for introducing local subspace classifiers can be that it is a piecewise linear classifier (as opposed to a linear classifier) and hence more powerful, here we motivate locality (and softness) based on a different consideration. We note that the goal of constructing a classifier is so that reliable generalisation performance (i.e. low prediction error) can be obtained. The prediction error is the error made by the classifier on patterns that were not part of the training set. In general, the prediction error (for squared error and 0–1 cost functions) can be decomposed into (squared) bias and variance components (see elsewhere for an overview [8–11]). This framework, often called the bias/variance framework, allows one to obtain average case performance in the small sample case, and is based on obtaining a classifier trained with different training data sets (each with N patterns), and computing the expected performance.

The bias term measures the deviation of the averaged classifier output (i.e. the average output obtained from classifiers trained on different data sets, each with N patterns) from the true output. The variance term, on the other hand, is an average of the squared difference between a given classifier output and the averaged classifier output. Thus, the variance term reflects the sensitivity of the classifier to variations in the training data set. For example, when a classifier pays no attention to the training data, one gets zero variance but high bias. On the other hand, when more attention is paid to the training data (by increasing the complexity of the classifier), one gets lower bias but higher variance. In most situations, there is a trade-off between bias and variance, and a careful balance between them is required to ensure good prediction.

Our motivation for introducing locality and softness in the subspace model of classification is as follows. By introducing locality, it is possible to reduce the bias though at the cost of a possible increase in variance. By introducing softness (or aggregation), the variance can be reduced. Previous localisations of CLAFIC which do not introduce softness are thus susceptible to the high variance that locality typically introduces.

We have organised the rest of the paper as follows. We first review CLAFIC and then propose a new subspace classification method, which we call as the Soft Regional Subspace Classification (SRSC) based on introducing locality and softness in CLAFIC. Some aspects of this are reported in our earlier work [12,13]. We present comparative results of SRSC performance both in terms of classification accuracy, and in terms of the bias/variance decomposition of the prediction error. Finally, we present our conclusions.

2. BACKGROUND – CLAFIC

We introduce some notation before proceeding further. Input patterns are denoted by x where $x \in \Re^n$. There are a total of k classes denoted by $\omega^{(1)}, \ldots, \omega^{(k)}$, and each class is represented by a $n^{(i)}$ dimensional subspace $\mathcal{L}^{(i)}$.

In CLAFIC, the goal is to maximise the average projection of the vectors of a given class $\omega^{(i)}$ on their own subspace $\mathcal{L}^{(i)}$. A suitable cost function can then be defined as

$$J\uparrow = \sum_{i=1}^{k} E[x^T P^{(i)} x \mid x \in \omega^{(i)}] \tag{1}$$

where, the notation $J\uparrow$ $J(\downarrow)$ is used to denote maximisation (or minimisation) of J. J is maximised in Eq. (1) by finding a set of orthonormal basis vectors $\{u_i^{(1)}, \ldots, u_i^{(n^{(i)})}\}$ and computing a unique projection matrix for the subspace $\mathcal{L}^{(i)}$ as $P^{(i)} = \sum_{j=1}^{n^{(i)}} u_i^{(j)} u_i^{(j)T}$. A pattern x is said to belong to class $\omega^{(i)}$ if $x^T P^{(i)} x > x^T P^{(j)} x$, $\forall$ j $\neq$ i. The basis vectors for the ith class subspace of CLAFIC can be shown to be (see, for example, Oja [2]) the eigenvectors corresponding to the $n^{(i)}$ largest eigenvalues of the class correlation matrix, given by

$$Q^{(i)} = E[xx^T \mid x \in \omega^{(i)}] \tag{2}$$

However, if the value of $n^{(i)}$ is larger than half the dimensionality of the input space (n/2), the problem can be restated as a minimisation problem whereby we seek to minimise the projection of the patterns on their orthogonal subspaces $\tilde{\mathcal{L}}^{(i)}$. Thus,

$$J\downarrow = \sum_{i=1}^{k} E[x^T \tilde{P}^{(i)} x \mid x \in \omega^{(i)}] \tag{3}$$

where $\tilde{P}^{(i)} = \sum_{j=1}^{\tilde{n}^{(i)}} \tilde{u}_i^{(j)} \tilde{u}_i^{(j)T}$, and the tildes in the notation correspond to the orthogonal subspace $\tilde{\mathcal{L}}^{(i)}$, where $\tilde{\mathcal{L}}^{(i)} \perp \mathcal{L}^{(i)}$. A pattern x can then be said to belong to class $\omega^{(i)}$ if $x^T P^{(i)} x < x^T P^{(j)} x$, $\forall\ j \neq i$. The basis vectors for the ith class orthogonal subspace of CLAFIC are then the eigenvectors corresponding to the $\tilde{n}^{(i)}$ smallest eigenvalues of the class correlation matrix given by Eq. (2).

3. SOFT REGIONAL SUBSPACE CLASSIFICATION (SRSC)

We derive SRSC from CLAFIC by introducing locality and softness as described below. Specifically, by (subspace) clustering the patterns into l clusters we introduce locality. Within each cluster, we obtain the sub manifold for each class. Softness is introduced by allowing a pattern to be

associated with more than one cluster. The cost function representing this modification can be written as

$$J\downarrow = \sum_{c=1}^{l} \sum_{i=1}^{k} E[\alpha_c(x)(x-r_c)^T \tilde{P}_c^{(i)}(x-r_c) \mid x \in \omega^{(i)}] \qquad (4)$$

where $\alpha_c(x)$ denotes the 'amount' by which a pattern belongs to a cluster with centre r_c. $\alpha_c(x) \in \{0,1\}$ subject to the constraint that $\Sigma_c \alpha_c(x) = 1$ corresponds to hard clustering which is the case when a given pattern is associated with only one cluster. This corresponds to a precursor to the SRSC algorithm called RSC that we had reported earlier [12]. When $\alpha_c(x) \in [0,1]$ subject to the constraint that $\Sigma_c \alpha_c(x) = 1$, we allow a pattern to be associated with more than one cluster. We enforce both of these constraints in SRSC by using a softmax function given by

$$\alpha_j(x) = \frac{e^{-\gamma d(x,r_j)}}{\sum_{c=1}^{l} e^{-\gamma d(x,r_c)}} \qquad (5)$$

where $d(x,r_c) = (x-r_c)^T \tilde{P}_c^{(i)}(x-r_c)$ for $x \in \omega^{(i)}$, and γ is a constant (which controls the membership degree). A smaller γ implies that there is a greater chance for a specific training pattern vector to be softly assigned equally to more than one cluster. A higher γ introduces a lesser degree of softness in clustering.

Thus, additionally, over CLAFIC, patterns within the same class are assigned to different clusters in a soft fashion leading to a projection matrix $\tilde{P}$ for each class $\omega^{(i)}$ and for each cluster $C^{(c)}$. A pattern is then said to belong to class $\omega^{(i)}$ if

$$(x-r_c)^T \tilde{P}_c^{(i)}(x-r_c) < (x-r_{c'})^T \tilde{P}_{c'}^{(j)}(x-r_{c'}), \quad \forall\, j \neq i,\ c,c' \in \{1,2,\ldots,L\} \qquad (6)$$

The basis vectors of the cth cluster ith class orthogonal subspace are then given by the eigenvectors corresponding to the $\tilde{n}_c^{(i)}$ minimum eigenvalues of the *weighted class cluster covariance matrix* (where $\tilde{n}_c^{(i)}$ is the dimensionality of the subspace for the cth cluster ith class), given by

$$Q_c^{(i)} = E[\alpha_c(x)(x-r_c)(x-r_c)^T \mid x \in \omega^{(i)}] \qquad (7)$$

To calculate $Q_c^{(i)}$ it is first necessary to cluster the training data. One could (soft) cluster the training data and calculate the cluster centres using a standard unsupervised clustering algorithm (like C-means). However, such an unsupervised clustering scheme is likely to produce sub-optimal clusters (sub-optimal in the sense of the minimisation of the classification error). Hence, we propose an iterative scheme, which at each step, tries to directly minimise the classification error in the clustering procedure. This scheme is based on the batch mode Generalized Lloyd's Algorithm (GLA) [14] to iteratively cluster the patterns in the input space. The distortion measure that is necessary for the clustering, or Vector Quantisation, is then the classification error defined in Eq. (4). This modification is distinct from methods that use annealing for performing vector quantisation, where more cooperation is encouraged in the initial stages of the vector quantisation process, and as training proceeds the competition is increased to avoid local minima problem. Final memberships in such approaches are crisp, whereas the final memberships in SRSC can remain soft.

4. THE COMPLETE SRSC ALGORITHM

Let $X = \{(x^{(1)}, y^{(1)}), (x^{(2)},y^{(2)}), \ldots, (x^{(N)},y^{(N)})\}$ represent the N training patterns and their associated class labels, where $x^{(i)} \in \Re^n$ and $y^{(i)} \in [1,2,\ldots,k]$. Then we would like to come up with the following: (i) the cluster associations, A which is a $N \times l$ matrix and the jth row of the matrix A corresponds to the associations of the jth training vector with each of the l clusters; (ii) the cluster centres r_c; and (iii) the orthogonal space projection matrix $\tilde{P}_c^{(i)}$ corresponding to all classes and clusters such that J is minimised. Adapting the GLA to iteratively refine A, and hence the projection matrices $\tilde{P}_c^{(i)}$ and the cluster centres r_c to minimise J, we can write the complete algorithm as

1. Initialise the l cluster centres r_c to l randomly chosen patterns from the training set and initialise the A matrix by using Eq. (5). However, since the $\tilde{P}_c^{(i)}$ matrices are not initialised, we could just use the Euclidean distance of each training pattern from the cluster centres in Eq. (5) for initialisation of memberships.
2. Use the above memberships and Eq. (7) to calculate $Q_c^{(i)}$ and hence initialise the projection matrices.
3. The update rule for r_c (cluster centre's) can be derived based on modifying Eq. (4) corresponding to each cluster $C^{(c)}$, i.e.

$$J_c\downarrow = \sum_{i=1}^{k} E[\alpha_c(x)(x-r_c)^T \tilde{P}_c^{(i)}(x-r_c) \mid x \in \omega^{(i)}] \qquad (8)$$

which can be minimised using gradient descent to provide an update rule for r_c, i.e.

$$\frac{\partial J_c}{\partial r_c} = \sum_{i=1}^{k} E[2\alpha_c(x)\tilde{P}_c^{(i)}(x-r_c) \mid x \in \omega^{(i)}] \qquad (9)$$

Setting the above to zero we can write,

$$\sum_{i=1}^{k} \tilde{P}_c^{(i)} \left(E[x\alpha_c(x) \mid x \in \omega^{(i)}] - r_c E[\alpha_c(x) \mid x \in \omega^{(i)}]\right) = 0 \qquad (10)$$

and hence we express r_c as

$$r_c = \left(\sum_{i=1}^{k} E[\alpha_c(x) \mid x \in \omega^{(i)}]\tilde{P}_c^{(i)}\right)^{-} \sum_{i=1}^{k} E[x\alpha_c(x) \mid x \in \omega^{(i)}]\tilde{P}_c^{(i)} \qquad (11)$$

where $(\cdot)^-$ is the generalised inverse.

4. Calculate the weighted class cluster covariance matrix $Q_c^{(i)}$, and hence the updated projection matrices $\tilde{P}_c^{(i)}$ (the local subspace dimension is assumed to be the same for all clusters and all classes).

5. Update the cluster associations α_c of individual training patterns $x^{(j)}$ using Eq. (5).
6. Iterate steps 3–5 until the cluster centres stabilise or stop after a maximum number of iterations.

5. SIMULATIONS

We present the efficacy of SRSC using three different data sets which we briefly describe below.

The diabetes data set is an eight dimensional data set based on personal data and medical examinations collected to diagnose diabetes in Pima Indians [15] (this is the 'diabetes1' data set). There are a total of 768 patterns (500 without diabetes and 268 with diabetes).

The lymphoma data set is a 10 dimensional data set, and is based on features extracted from images of cytologic imprints for the purposes of labeling the imprint as benign or malignant [16]. Of the 10 features, four are tone based, another four are texture based, and the remaining two characterise the local and global characteristics of the boundaries of individual cells in the image of the imprint. There are a total of 439 patterns (145 benign and 294 malignant).

This vowel data set is a 10 dimensional data set collected to identify the speaker based on utterances of 11 steady state vowels (classes) [17]. There are a total of 15 speakers and a total of 990 patterns.

5.1. Classification Accuracy

For each data set, we present the accuracy in classification obtained. Determining the number of clusters is a generic open problem in pattern recognition (see Kothari and Pitts [18], however), as is determining the appropriate dimensionality. To get around this limitation, we divided each data set into three parts. We used one part for training, one for validation, and the third for testing (generalisation). We trained many classifiers (with varying number of clusters and varying dimensionality), and used the validation set to choose the best classifier (best in the sense of classification accuracy). This chosen classifier was then used with the testing data set to obtain the generalisation accuracies. To offset the sensitivity due to initialisation, each simulation was repeated five times based on different initialisations, and the average accuracy obtained over the five independent runs is reported.

In each case, we compare the results with CLAFIC and also with a strictly local version of SRSC, i.e. without the soft sharing part. The strictly local version of SRSC is referred to as RSC below. Table 1 shows the accuracies obtained with CLAFIC, RSC and SRSC.

For comparison, we note that the *best* previous performance performance (that we are aware of) obtained with the diabetes data set was 77.3% using a combination of weak classifiers [19], while that for the Vowel data set was 67.0% using Growing Cell Structures [20]. An important point to note is that, since the partitioning of the data may or may not be the same as that used by us, these numbers should be used as qualitative indicators.

5.2. Bias/Variance Decomposition of Prediction Error

We also present the prediction error in terms of bias and variance to demonstrate the reduction in bias due to locality and the reduction in variance due to softness. For these set of results, the dimensionality was fixed. In this case, 10 different training data sets were generated by sampling a subset of the entire data without replacement. The test data was kept fixed: 10 classifiers (with fixed number of clusters and dimensionality) were independently trained, each using one of the training data sets. The performance of these 10 classifiers on the test data was used to obtain the bias and variance components of the prediction error. In this case, we also report the average accuracy (average of the 10 classifier's individual accuracies), which differs somewhat from the accuracy reported in the previous subsection. In the simulations of this subsection, the training data is changed, whereas in the simulations of the previous subsection the training data remained the same and it was the initialisation that changed.

As before, we compare the results with CLAFIC, and also with a strictly local version of SRSC, i.e. without the soft sharing part. The strictly local version of SRSC is referred to as RSC below. Table 2 shows the bias/variance decomposition of the prediction error for CLAFIC, RSC, and SRSC for each of the three data sets.

Consider the specific case of the diabetes data set. Note that the bias of CLAFIC is higher than that obtained by RSC (the latter is based on introduction of locality in CLAFIC). Further, the variance has increased corresponding to the typical situation of increase in variance with locality. This is the susceptibility that localisation based extensions of CLAFIC are subject to. On the contrary, when softness is introduced (SRSC) the variance reduces (as compared to RSC), and the bias increases, though marginally. This highlights the typical compromise that exists between bias and variance.

We also note that the accuracies as reported in Table 2 are more reliable estimates of what might be expected. This is because the process of determining the optimal number of clusters and dimensions (as was done for Table 1) is a computationally expensive procedure. When the results of Table 2 are considered, SRSC outperforms CLAFIC in each of the three data sets.

6. CONCLUSION

We presented a soft and local adaptation of CLAFIC motivated in part by the ability of softness to reduce the variance and locality to reduce the bias. Simulation results with three data sets supports this hypothesis, and we found SRSC to provide better performance than either CLAFIC or RSC. This improvement does come at a slightly increased

Table 1. The generalisation accuracy obtained for the three data sets by SRSC compared to that obtained using RSC and CLAFIC. Here σ denotes the standard deviation of the classification accuracy over the five independent runs. CLAFIC does not have a standard deviation as the projection subspaces are determined solely from the correlation matrices, and consequently, does not have any variance. The number of clusters and dimensions are those that gave the best performance on the validation set.

Data set	Classifier	% Accuracy	σ	# Clusters	# Dimensions
	CLAFIC	60.4	–	1	3
Diabetes	RSC	69.0	4.3	2	1
	SRSC	72.8	3.3	4	1
	CLAFIC	75.6	–	1	2
Lymphoma	RSC	68.0	3.9	9	7
	SRSC	71.6	2.6	9	3
	CLAFIC	35.1	–	1	2
Vowel	RSC	44.0	2.9	5	2
	SRSC	55.7	3.8	7	2

Table 2. The bias and variance components of the prediction error for the three data sets. Here σ denotes the standard deviation of the classification accuracy over the 10 independent runs (each using a differently sampled training data set. The numbers of clusters and dimensions are chosen arbitrarily though close to the number of clusters and dimensions that gave the best performance (see Table 1).

Data Set	Classifier	% Accuracy	σ	# Clusters	# Dim	$(\text{Bias})^2$	Variance
	CLAFIC	64.8	0.7	1	1	0.3331	0.0190
Diabetes	RSC	67.7	3.6	3	1	0.1859	0.1376
	SRSC	72.0	2.4	3	1	0.2005	0.0792
	CLAFIC	66.1	1.6	1	3	0.3111	0.0275
Lymphoma	RSC	66.5	4.0	9	3	0.1836	0.1517
	SRSC	67.5	4.6	9	3	0.2570	0.0684
	CLAFIC	40.6	1.3	1	2	0.0085	0.0013
Vowel	RSC	47.2	4.1	5	2	0.0049	0.0039
	SRSC	48.9	3.6	5	2	0.0047	0.0038

computational cost. However, once the class cluster subspaces are formed, then time involved for testing a new point is rapid. Since an iterative algorithm was employed for the minimisation of the cost function, a convergence criterion could be specified. In all of our simulations we found it convenient to run the iteration for a fixed number of times (30). The changes in the cluster centres usually stabilised by then. The standard deviations in the accuracies obtained was probably due to the fact that the GLA that was applied for the minimisation procedure did not converge to the global minima. Global optimisation strategies may be employed to improve performance.

SRSC is distinct from soft competitive Vector Quantisation algorithms in which memberships start out to be soft and then are annealed to become crisp. In SRSC the memberships determine how much each of the training patterns influence the projection subspaces and even the final memberships could remain soft (see Fig. 1), as opposed to crisp final memberships in soft competitive Vector Quantisation strategies.

Finally, we also note that extensions to CLAFIC can also be made on the basis of determining the projection subspaces not by simply maximising the expected projection of patterns of a given class, but by maximising the ratios in projection distances of patterns belonging to different classes using Oriented Principal Component Analysis (OPCA). Though it can be shown that the maximum dimensionality of the optimal subspace with OPCA is 1 (adding more dimensions does not increase the performance), we have found that OPCA gives competitive results even with a dimensionality of 1.

Acknowledgement

This research is funded in part by a grant from The Whitaker Foundation.

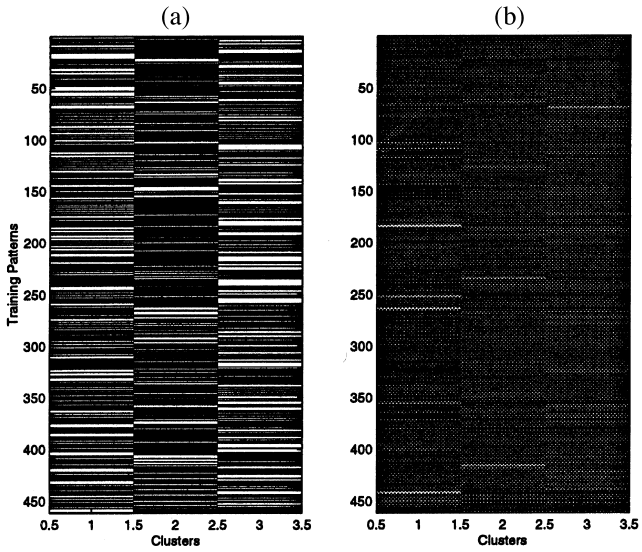

Fig. 1. Pattern memberships in the different clusters for the diabetes data set. (a) Cluster membership in RSC is crisp as shown, (b) cluster membership in SRSC can remain soft as shown here. Lighter shade indicates higher membership.

References

1. Watanabe S, Pakvasa N. Subspace methods of pattern recognition. Proceedings of the 1st International Conference on Pattern Recognition 1973;25–32
2. Oja E. Subspace Methods of Pattern Recognition. Research Studies Press, Letchworth, UK, 1983
3. Oja E, Kohonen T. The subspace learning method as a formalism for pattern recognition and neural networks. Proceedings of the IEEE International Conference on Neural Networks, 1988; 277–284
4. Laaksonen J. Subspace Methods in Recognition of Handwritten Digits. PhD thesis, Helsinki University of Technology, 1997
5. Prakash M, Murty MN, Extended subspace methods of pattern recognition. Pattern Recognition Letters 1996;17(11):1131–1139
6. Oja E, Parkkinen J. On subspace clustering. In Proceedings of the 7th International Conference on Pattern Recognition, 1984; 692–695
7. Kohonen T, Kaski S, Lappalainen H. Self-organised formation of various invariant-feature filters in the adaptive-subspace som. Neural Computation 1997;9(6):1321–1344
8. Geman S, Bienstock E, Doursat R. Neural networks and the bias/variance dilemma. Neural Computation 1992;4(1):1–58
9. Wolpert DH. On bias plus variance. Neural Computation 1997;9(6):1211–1243
10. Kohavi R, Wolpert DH. Bias variance decomposition for zero-one loss functions. Machine Learning: Proceedings of the 13th International Conference, L. Saitta, ed., Morgan Kaufmann, 1996
11. Kothari R. Prediction error: The bias/variance decomposition, methods of minimization, and estimation. Technical Report TR 222/10/98/ECECS. Department of Electrical and Computer Engineering and Computer Science, University of Cincinnati, 1998
12. Balachander T, Kothari R. Localized soft subspace pattern classification. In Proceedings of the International Conference on Advances in Pattern Recognition, 1998 (in press)
13. Balachander T, Kothari R. Localized subspace pattern classification. Proceedings of the IEEE World Congress on Computational Intelligence (IJCNN), 1998;1804–1809
14. Gersho A, Gray RM. Vector Quantization and Signal Compression. Kluwer Academic, Boston, 1992
15. Prechelt P. PROBEN1 – a set of benchmarks and benchmarking rules for neural-network training algorithms. Technical Report 21/94, Fakultät für Informatik, Universität Karlsruhe, [ftp://ftp.ira.uka.de/pub/neuron/proben1.tar.gz], 1994
16. Kothari R, Balachander T, Lotlikar R, Cualing H. On the separability of grades of lymphoma using higher order cytological features. Technical Report TR 221/10/98/ECECS. Department of Electrical and Computer Engineering and Computer Science, University of Cincinnati, 1998
17. Dataset and results are electronically available from the AI repository, Carnegie Mellon University, [ftp://ftp.cs.cmu.edu/afs/cs/project/connect/bench/]
18. Kothari R, Pitts D. On finding the number of clusters. Pattern Recognition Letters 1999 (in press)
19. Ji C, Ma S. Combinations of weak classifiers. IEEE Transactions on Neural Networks 1997;8(1):32–42
20. Fritzke B. Growing cell structures – a self organizing networks for unsupervised and supervised learning. Neural Networks 1994;7(9):1441–1460

Thiagarajan Balachander received his BE in 1994 from Anna University, India and his MS in 1996 from University of Cincinnati, both in Electrical Engineering. Currently he is a Graduate Research Assistant pursuing a PhD in the Artificial Neural Systems Laboratory, Department of ECECS at the University of Cincinnati. His current research interests lie in pattern recognition, neural networks, and image processing. He is a member of the IEEE. He received the National Merit Scholarship awarded by the Government of India.

Ravi Kothari received his BE degree (with Distinction) in 1986 from Birla Institute of Technology (India), his MS from Louisiana State University in 1988, and his PhD from West Virginia University in 1991, all in Electrical Engineering. His is currently an Associate Professor in the Department of Electrical and Computer Engineering and Computer Science at the University of Cincinnati and Director of the Artificial Neural Systems Laboratory there. His research interests include artificial neural networks, pattern recognition, and image analysis.

He received the Eta Kappa Nu Outstanding Professor Award in 1995, and the William E. Restemeyer Teaching Excellence Award in 1994 from the Department of Electrical and Computer Engineering and Computer Science at the University of Cincinnati.

Dr. Kothari serves on the Editorial Board of the Journal Pattern Analysis and Applications (Springer-Verlag), and is a member of the IEEE, the International Neural Network Society, and the Eta Kappa Nu and Phi Kappa Phi honor societes.

Correspondence and offprint requests to: Professor Ravi Kothari, Department of Electrical & Computer Engineering & Computer Science, P.O. Box 210030, Cincinnati, OH 45221-0030, USA. Email: ravi.kothari@uc.edu