

ART1 Network: Application in Wine Classification

N. C. Kavuri and Madhusree Kundu

Abstract—Unsupervised neural network (NN) based on Adaptive Resonance Theory (ART1) was successfully implemented as an alternative to statistical classifier in order to discriminate among the 178 samples of wine possessing 13 numbers of feature variables. A pattern recognition tool, principal component analysis (PCA) was applied to reduce the dimensionality of the feature variables by 5; out of which the first 2 numbers of principal components captured over 55.4 % of the variance of the dataset of wine. Supervised non-hierarchical K-means clustering was used to designate the classes available among the wine samples, hence discrimination. Supervised hierarchical clustering technique was also applied for discrimination with a mention of their classification level in the produced dendograms. After the discrimination made by hierarchical as well as non- hierarchical clustering, the ART1 classifier was designed.

Index Terms—ART1, Dendograms, K-means clustering, Wine, PCA, SQC.

I. INTRODUCTION

Classification is one of the conscious or unconscious activities prevailing on the earth. It is the prominent research area involving neural networks, though; there exist quite a large number of conventional classifiers based on statistical probabilities. The determination of quality of food stuffs, water and beverages and maintenance of their corresponding standards is an urgent criterion needed to be ensured. Statistical Quality Control (SQC) has been designed to sample and analyze a large population on an infrequent basis. The classification aspect, an integral part of SQC assigns the attribute or quality of the product to a predefined class based on the parametric values or features that influence the product quality. Those parametric values can be determined by various classical analytical techniques; such as various chromatography and spectrometry. These components can be numerous or unknown, besides, it is impractical and very hard to correlate and compare the results of instrumental analysis to biological sensing [1]. Now-a-days, the use of sensor arrays for producing features followed by the multivariate data analysis (MVDA) to discriminate among various samples leads to a successful design of a classifier. The use of various decision rules qualifies the classifier to be used for authentication purpose. Principal component analysis (PCA) and different clustering techniques in food analysis grew rapidly in the last decade. It has been used in

wine analysis by researchers like Buratti et al., 2007; Parra et al., 2006; Buratti et al., 2004; Riul et al., 2004; Di Natale et al., 2004; Legin et al., 2003; Di Natale et al., 2000 [2]-[8].

Discrimination and classification of the feature variables produced from multisensory array owes a profound debt to the multivariate statistics these days. In these procedures, an underlying probability model must be assumed in order to calculate the posterior probability upon which the classification decision is made. One major limitation of the statistical methods is that they work well only when the underlying assumptions are satisfied. Discrimination is concerned with separating distinct sets of objects (or observations) on a one-time basis in order to investigate observed differences when casual relationships are not well understood. The operational objective of classification is to allocate new objects (observations) to predefined groups based on a few well defined rules evolved out of discrimination analysis of allied group of observations.

Neural networks, either supervised or unsupervised have emerged out as an important and competing tool to the conventional statistical classifiers. The advantage of neural networks lies in the following theoretical aspects. First, neural networks are data driven self-adaptive methods in that they can adjust themselves to the data without any explicit specification of functional or distributional form for the underlying model. Second, they are universal functional approximators; neural networks can approximate any function with arbitrary accuracy [9]-[11]. Since any classification procedure seeks a functional relationship between the group membership and the attributes of the object, accurate identification of this underlying function is doubtlessly important. Very complex and non-linear process identification and function approximation which is rather difficult by first principle models, are conveniently done by ANN. Finally, neural networks are able to estimate the posterior probabilities, which provide the basis for establishing classification rule and performing statistical analysis [12].

A number of researchers have illustrated the connection of neural networks to traditional statistical methods. For example, Gallinari et. al. (1991) have presented analytical results that established a link between discrimination analysis and multilayer perceptrons (MLP) used for classification problems [13]. Cheng and Titterington (1994) made a detailed analysis and comparison of various neural network models with traditional statistical methods [14]. They have shown strong associations of the feed forward neural networks with discrimination analysis and regression, and unsupervised networks such as self-organizing neural networks with cluster analysis. Zhang (2000) made a comprehensive review on neural network (NN) as classifier, the issue of posterior probability distribution, the link between NN and conventional statistical classifiers, learning

Manuscript received May 12, 2011; revised May 28, 2011.
Naga C. Kavuri, M.Tech student., e-mail: biochaitanya@gmail.com
Madhusree Kundu, Associate Professor, Department of Chemical Engineering, National Institute of Technology, Rourkela, India.
Corresponding Author.,
E-mail: mkundu@nitrrkl.ac.in; madhushreek@yahoo.com.
Fax: 0661-2462999

and generalization tradeoff in classification, the feature variable selection, as well as effect of misclassification cost [15].

In the present classification problem taken up, the supervised statistical K -means clustering technique, as well as an unsupervised neural network based classifier ART1 were used. The present work utilized PCA for hierarchical as well a non-hierarchical clustering techniques to discriminate among the 178 numbers of available wine samples on the basis of certain selected features. Once the clustering was done, the data (13 attributes of each sample and their corresponding cluster numbers) were divided randomly in to different training & testing sets required for the ART1 networks. The *Adaptive Resonance Theory based Neural Networks* (ART1) were trained with randomly chosen fraction of wine data sample and simulated efficiently with the randomly selected fraction of samples which were not present during the training of the network. The random selection of data was done by the method described by Box and Muller (1958) and Devroye (1986) [16],[17]. ART1 accepts the inputs and targets only in the form of binary data (i.e. 0 and 1). So processing of the data according to the demand of network architecture plays an important role in their successful implementation.

II. PCA AND K-MEANS CLUSTERING

A wine dataset of 178 numbers of samples containing 13 numbers of features, like proline, magnesium content. Ash alkalinity, pH etc was considered. PCA is a multivariate statistical technique that can extract the essential features from a data set by reducing its dimensionality without compromising any valuable information of it. Principal components (PCs) are a new set of coordinates orthogonal to each other. The first PC is the direction of largest variation in the data set. The projection of original data on the PCs produces the score data or transformed data as a linear combination of those fewer mutually orthogonal dimensions. PCA technique was applied on the auto-scaled data matrix to determine the principal eigenvectors, associated Eigen values and scores or the transformed data set.

Clustering technique is more primitive in that; no a-priori assumptions are made regarding the group structures. Grouping of the data can be made on the basis of similarities or distances (dissimilarities). Hierarchical clustering techniques are processed either by a series of successive mergers or a series of successive divisions. Agglomerative hierarchical methods start with individual objects ensuring as much number of clusters as objects initially. Besides hierarchical clustering, non- hierarchical method, K -means clustering was also applied in this work. The number of clusters can be pre-specified or can be determined iteratively as a part of the clustering procedure. The K -means clustering proceeds in three steps, which are as follows,

1. Partition of the items in to K initial clusters.
2. Assigning an item to the cluster whose centroid is nearest (distance is usually Euclidian). Recalculation of the centroid for the cluster receiving the new item and for the cluster losing that item.

Repeating the step-2 until no more reassignment takes

place or stable cluster tags are available for all the items.

The K -means clustering has a specific advantage of not requiring the distance matrix as required in hierarchical clustering, hence ensures a faster computation than the latter.

III. ART1 NETWORK

The stability-plasticity dilemma remained unresolved for many conventional artificial neural networks. The ability of a net to learn new patterns equally well at any stage of learning without washing away the previously learnt patterns is called its plasticity. A stable net does not return any pattern to a previous cluster. Some nets achieve stability gradually adjusting their learning rates provided the same training set is presented many times before them. Those conventional nets cannot learn a pattern while presented first time before them. A real network is constantly exposed to changing patterns; it may never see the same training vector twice. Under such a circumstance the back propagation networks can learn nothing with continuously modifying their weights without a respite of getting a stationary setting. ART1 nets are designed to be both plastic and stable. The ART1 network is a vector classifier. It accepts an input vector and classifies it as one of the categories depending upon which of the stored pattern it resembles within a specified tolerance otherwise a new category is created by storing that pattern as an input vector. No stored pattern is modified if it does not match the current input vector within a specified tolerance; hence the stability- plasticity dilemma is solved. ART1 is designed for classifying binary vectors. The classification process through ART involves three steps; recognition, comparison and the search phase. During learning one input vector is presented to the network. The degree of similarity is controlled by vigilance parameter ρ (0-1).

A. Basic Architecture

The ART1 network consists of three major components accompanying groups of neurons.

- Input processing field-F1 layer
- Cluster units –F2 layer
- Reset mechanism

1) Input processing layer:

It is divided into two layers.

- Input portions – F1(a): Represents the given input vector
- Interface portion –F1(b): Exchanges the input portion signal with the F2 layer

2) Cluster units –F2 layer

This is a competitive layer. The cluster unit with largest net input is selected to learn the input pattern. The activation of all other F2 units is set to zero. F1(b) is connected to F2 layer through bottom-up weights b_{ij} and F2 layer is connected to F1(b) layer by top down weight t_{ji}

3) Reset Mechanism

Depending on the similarity between the top down weight and the input vector, the cluster unit is allowed to learn a pattern or not. This is done at the reset unit, based on the signals it receives from the input and interface portion of the F1 later. If the cluster unit is not allowed to learn, it becomes inhibited and a new cluster unit is selected for learning. It

dictates the three possible states for F2 layer neurons; they are namely active, inactive and inhibited. The difference between the inactive and inhibited is that for both the cases activation state of F2 unit is zero. In its inactive state, the F2 neurons are available in next competition during the presentation of current input vector which is not possible when the F2 layer is inhibited.

B. Algorithm

The binary input vector is presented to F1 (a) layer and is then passed on to F1 (b) layer. The F1 (b) layer sends signal to F2 layer over weighted interconnection path (Bottom-up weights). Each F2 unit calculates the net input. The node with the largest input is the winner and its activation state is 1. All the other nodes in F2 layer are considered to have activation state of 0 but not inhibited and the reset is true. The winning node of F2 layer alone is eligible to learn the input pattern. The signal is sent from F2 layer to F1 (b) through weighted interconnections which are top down weights. The activation vector X of the F1 (b) layer are considered to be active if they receive non-zero weights both from F1 (a) and F2 layer.

The norm of the vector $\|X\|$ renders the number of components in which the top-down weight vector for the winning unit (t_{ji}) and the input vector S are both 1. Depending upon the ratio of norm of x to norm of S ($\|x\|/\|s\|$), either the weights of the winning cluster units are adjusted or the reset mechanism is rescheduled. The whole process is repeated until either a match is found or all neurons in the F2 layer are inhibited. The algorithm flow chart for ART1 is as follows,

Step1. Initialization of parameters and weights

$$L > 1 \& 0 < \rho \leq 1$$

$$0 < b_{ij}(0) < \frac{L}{L-1+n}$$

where n is the number of components in the input vector

$$t_{ji}(0) = 1$$

For each training input

Step2. Activation states of all F2 neurons are set to zero and all F1 (a) neurons are assigned to the input vector S.

Step3. Computation of norm of S ($\|S\|$) = $\sum s_i$

Step4. Sending signals from F1 (a) to F1 (b) layer, $x_i = s_i$

For each F2 node that is not already inhibited

Step5. Calculation of net input of that particular F2 node provided the 'reset' is true.

$$y_j = \sum_i (b_{ij} x_i)$$

Step6. Finding highest y_j among all y_j 's.

Step7. Re-computation of x of F1 (b) layer. $x_i = s_i t_{ji}$

Step8. Computation of the norm of vector $x = \|X\| = \sum x_i$

Step9. Test for reset, if $\left(\frac{\|X\|}{\|S\|}\right) < \rho$, the j^{th} node is inhibited.

Continue from Step 5. If $\left(\frac{\|X\|}{\|S\|}\right) > \rho$

Step10. Updation of weights for node j, $b_{ij(new)} = \frac{Lx_i}{L-1+\|x\|}$

& $t_{ji(new)} = x_i$

Test for stopping criterion:

- No change in top-down or bottom up weights.
- No reset
- Maximum number of epochs exceeded

Normalization of data matrix:

All the elements in the scaled matrix are lying between 0-1. The linear scaling function for zero to one transforms a variable x_k into x_k^* in the following way:

$$x_{k,j}^* = \frac{x_{k,j} - \min(x_k)_{for\ all\ j's}}{\max(x_k)_{for\ all\ j's} - \min(x_k)_{for\ all\ j's}}$$

where, k and j are column and row of the data matrix respectively.

C. Conversion of scaled data matrix into binary matrix:

The elements of the above scaled matrix which are below 0.5 are given an attribute of '0' & elements including 0.5 and above are considered as '1' in the binary data matrix.

The 8%, 20%, 30%, 44%, 56% and 92% of 178 binary data samples were randomly chosen as training data sets as well as target data sets ($n \times 14$ matrix) for the ART1 network. The first 13 columns of the data set formed the n number of input feature vectors and 14th column serves as n numbers of targets or class tags. Three different classes of training and testing pools were created out of 178 samples or feature vectors to design three different classifiers as ART1-1, ART1-2, & ART1-3. In a particular data pool; either training or testing, the presence of any one of the three classes of feature vectors are targeted as '1' and any other class of feature vectors apart from that category in that pool are targeted as '0'. For two different vigilance parameters and training data pools, the ART1 networks were trained and simulated with corresponding testing pools to determine their performances.

IV. RESULTS AND DISCUSSION

The application of PCA on the auto-scaled matrix containing 178 samples of wine resulted in 5 numbers of PCS. Table 1, which is a list of Eigen values corresponding to the column vectors of the loading matrix, reveals that the features (variables) like proline, magnesium, ash-alkalinity, color intensity, and maleic acid had captured almost 80% of the variance of the data thus became the dominant PCs. The first two PCs captured 55.4% of variance in the data, hence, it was decided to transform the original data matrix (178x13) to a (178x2) score matrix for clustering analysis. The scores along pc1 & pc2 were chosen to represent the transformed data.

TABLE 1 PRINCIPAL EIGEN VALUES AND PERCENTAGE VARIANCE CAPTURED BY PRINCIPAL COMPONENTS.

Component	PC1	PC2	PC3	PC4	PC5
Eigen values	4.706	2.497	1.446	0.919	0.853
% Variance	36.19	19.20	11.12	7.06	6.50
Cumulative % Variance	36.19	55.40	66.53	73.59	80.16

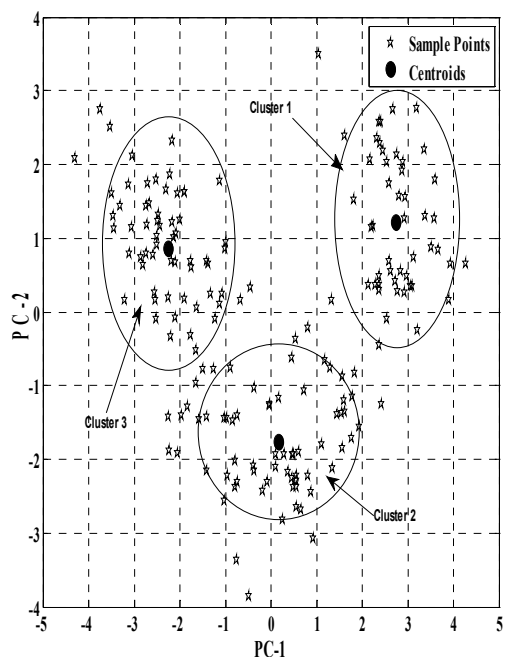


Fig. 1 Discrimination and clustering of scores along PC1-PC2

TABLE 2 STATISTICS FOR K-MEANS CLUSTERING (SCORE ALONG PC1-PC2).

Cluster Identity	No. of Samples	Cluster Centroid
1	49	(2.7362, 1.2108)
2	65	(0.1623, -1.7626)
3	64	(-2.2598, 0.8632)

To represent the score data the bi-plot (Fig.1) is chosen, which clearly shows the presence of three clusters in the data. With the K-means clustering, 3 numbers of clusters were resulted as shown in Table 2. The stable K-means statistics of the score along pc1-pc2 are presented in Table 2, which presented all the 3 cluster centroids, the number of data points pertaining to each cluster with a mention to the individual

sample numbers belonging to that cluster.

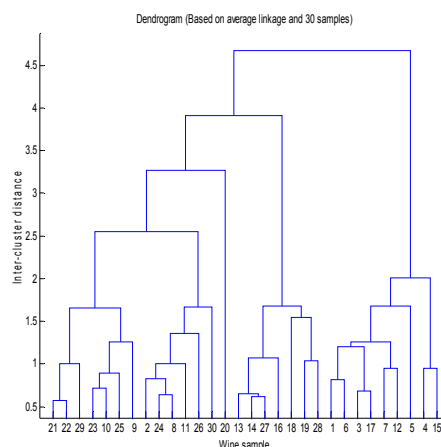


Fig. 2 Dendrogram on score along PC1-PC2

As a part of hierarchical clustering, the distance matrix or a dissimilarity matrix was determined, which was symmetric along the diagonal (all the diagonal elements were zero). A hierarchical cluster tree was then created with that distance matrix to form the dendrogram (Fig. 2) originated from the scores along PC1- PC2. Dendrogram consists of many U-shaped lines connecting objects in a hierarchical tree. The height of each U represents the distance between the two objects being connected. If there are 30 or fewer data points in the original dataset, each leaf in the dendrogram corresponds to one data point. If there are more than 30 data points, the complete tree can look crowded, and dendrogram collapses lower branches as necessary, so that some leaves in the plot correspond to more than one data point. The dendrograms were created using 1-30 numbers of samples for its clarity in representation. Fig. 2 demonstrates that over an inter-cluster distance of 4.5 there is a merger of all the data points (samples) in one group. At an inter-cluster distance of 3.5, there are 3 classes present in the transformed data set along pc1- pc2 coordinates.

A representative (10×10 matrix, 9 feature columns + 1 target column) normalized data matrix is shown in Table 3. Table 4 is a representative (10×12 matrix, 9 feature columns + 3 target columns) binary data matrix. Tables 3 & 4 are significant part of the reported results because the data processing for ART1 network is an important part of the successful classifier design. The performance of ART1 networks developed is presented in Tables 5 and 6, which reflects the ART1 network as an efficient classifier. The efficiency of ART1 network is defined as percentage of correct classification either it is in terms of 0 or 1. Three numbers of distinct classifiers (ART1-1, ART1-2, and ART1-3) were used to discriminate among 3 clusters of wine samples present in the database taken up. The two vigilance parameters (ρ) of 0.4 & 0.7 with 100 iterations were used for the training of the networks. The ART1 networks developed were very robust as reflected by its classification efficiency of 100 % for all combinations of training and testing vectors. A randomly selected 20 % of the data from the data base were used for training each of the ART1-1, ART1-2, and ART1-3 networks (The two vigilance parameters (ρ) of 0.4 & 0.7 with 100 iterations were used for the training of the networks) and simulation for three trained networks were done with the corresponding randomly selected samples containing 20 %

30 %, 40 %, 50 %, 60 % & 70 % data. Table 7 represents the training time & efficiency of the networks for the aforesaid operation.

V. CONCLUSIONS

The wine data base containing 178 no of samples was reduced to 5 principal dimensions according to PCA analysis. The scores along PC1-PC2 were used for K-means clustering of the present data base, and existence of three clusters was concluded. Once the clustering was over, the data (13 attributes of each sample and their corresponding cluster description) was processed as per the requirement of the ART1 networks and randomly divided in to appropriate training and testing sets for the three numbers of ART1 networks. Because of the presence of 3 clusters in the database; 3 different classes of training and testing pools were created out of 178 samples or feature vectors to form 3 distinct ART1 networks. In a particular data pool; either training or testing, the presence of any one of the three types

of feature vectors are targeted (described) as '1' and any other class of feature vectors present in that pool are targeted as '0'. ART1 networks took the inputs only in the form of binary data (i.e. 0 and 1). For two different vigilance parameters, 100 iterations and various training data pools, the ART1 networks were trained and simulated with corresponding testing data pools. ART1-2, ART1-2, & ART1-3 networks developed were equally efficient and consume very less CPU time; make them lucrative for on-line monitoring & quality control of wine. It can be acclaimed through the present study that the unsupervised NN classifiers are equally efficient as the conventional statistical classifiers; if not better.

ACKNOWLEDGMENT

The authors are grateful to Mr. Palash Kumar Kundu, Electrical Engineering Department, Jadavpur University, Kolkata, West Bengal, for his suggestions and encouragement.

TABLE 3 REPRESENTATIVE NORMALIZED DATA (10×10) MATRIX.

Sample No	Alcohol	Ash	Alcalinity	Magnesium	Phenols	Color	Hue	Dilution	Proline	Cluster #
1	0.641	0.290	0.397	0.449	1.000	0.222	0.561	0.353	0.667	1
2	0.024	0.347	0.000	0.000	0.129	0.111	0.255	0.235	0.019	1
3	0.000	0.815	0.726	0.755	0.161	0.222	0.745	0.471	1.000	1
4	0.725	0.484	0.493	0.571	0.548	1.000	1.000	0.118	0.596	2
5	0.048	1.000	1.000	1.000	0.710	0.222	0.184	1.000	0.365	1
6	0.623	0.331	0.425	0.408	0.516	0.570	0.898	0.706	0.462	1
7	0.737	0.419	0.425	0.347	0.000	0.000	0.010	0.471	0.468	1
8	0.539	0.645	0.644	0.653	0.806	0.074	0.000	0.529	0.000	3
9	1.000	0.234	0.041	0.286	0.032	0.222	0.480	0.412	0.468	1
10	0.419	0.000	0.178	0.490	0.065	0.356	0.653	0.000	0.385	3

TABLE 4 REPRESENTATIVE BINARY DATA (10×12) MATRIX.

Sample No	Alcohol	Ash	Alcalinity	Magnesium	Phenols	Color	Hue	Dilution	Proline	Cluster 1	Cluster 2	Cluster 3
1	1	0	0	0	1	0	1	0	1	1	0	0
2	0	0	0	0	0	0	0	0	0	1	0	0
3	0	1	1	1	0	0	1	0	1	1	0	0
4	1	0	0	1	1	1	1	0	1	0	1	0
5	0	1	1	1	1	0	0	1	0	1	0	0
6	1	0	0	0	1	1	1	1	0	1	0	0
7	1	0	0	0	0	0	0	0	0	1	0	0
8	1	1	1	1	1	0	0	1	0	1	0	1
9	1	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	1	0	0	0	0	1

TABLE 5 PERFORMANCE OF THE ART1 NETWORKS WITH VIGILANCE PARAMETER 0.4 & 100 ITERATIONS.

Training vector %	Testing vector%	ART1-1		ART1-2		ART1-3	
		Training time (s)	Efficiency%	Training time(s)	Efficiency%	Training time(s)	Efficiency%
8	92	0.177	100	0.199	100	0.200	100
20	80	0.275	100	0.371	100	0.305	100
30	70	0.384	100	0.360	100	0.509	100
44	56	1.214	100	1.244	100	1.309	100
56	44	1.44	100	1.419	100	1.59	100
70	30	1.29	100	1.442	100	1.816	100
80	20	1.248	100	1.495	100	1.200	100
92	8	2.646	100	2.690	100	2.015	100

TABLE 6 PERFORMANCE OF THE ART1 NETWORKS WITH VIGILANCE PARAMETER 0.7 & 100 ITERATIONS

Training Vector %	Testing vector%	ART1-1		ART1-2		ART1-3	
		Training time (s)	Efficiency %	Training time (s)	Efficiency %	Training time (s)	Efficiency %
8	92	0.169	100	0.216	100	0.220	100
20	80	0.317	100	0.356	100	0.266	100
30	70	0.456	100	0.376	100	0.577	100
44	56	1.02	100	0.964	100	0.986	100
56	44	1.551	100	1.589	100	1.459	100
70	30	1.558	100	1.591	100	1.808	100
80	20	1.697	100	1.604	100	1.913	100
92	8	2.811	100	2.775	100	2.684	100

TABLE 7 PERFORMANCE OF ART1 NETWORKS TRAINED WITH RANDOMLY SELECTED 20% DATA.

		Test Data set						
		20%	30%	40%	50%	60%	70%	
ART1-1	Computation Time	$\rho=0.4$	0.875	1.0046	0.9709	1.0073	0.872	0.9029
		$\rho=0.7$	0.955	0.9366	1.0522	0.8738	1.0361	0.9216
	Efficiency	$\rho=0.4$	100%	100%	100%	100%	100%	100%
		$\rho=0.7$	100%	100%	100%	100%	100%	100%
ART1-2	Computation Time	$\rho=0.4$	0.892	0.8673	1.0057	0.966	0.9605	0.9626
		$\rho=0.7$	1.0505	1.1874	1.0369	1.0149	1.004	0.9506
	Efficiency	$\rho=0.4$	100%	100%	100%	100%	100%	100%
		$\rho=0.7$	100%	100%	100%	100%	100%	100%
ART1-3	Computation Time	$\rho=0.4$	1.0115	0.9378	1.0767	0.8299	0.9825	0.8963
		$\rho=0.7$	1.0596	0.9239	0.9363	0.9807	1.074	1.0218
	Efficiency	$\rho=0.4$	100%	100%	100%	100%	100%	100%
		$\rho=0.7$	100%	100%	100%	100%	100%	100%

REFERENCES

- [1] A. Legin, A. Rudnitskaya, Y. Vlasov, C. D. Natale, F. Davide, A. D'Amico, "Tasting of beverages using an electronic tongue", *Sens. Actuators. B* 1997, 44: pp. 291-296.
- [2] S. Burrati, D. Ballabio, S. Benedetti, M.S. Cosio, "Prediction of Italian red wine sensorial descriptors from electronic nose, electronic tongue and spectrophotometric measurements by means of Genetic Algorithm regression models", *Food Chemistry* 2007 100: pp. 211-218.
- [3] V.Parra,A.A.Arrieta,J.B.Fernández-Escudero,M.L.Rodríguez-Méndez, J. A. De Saja, "Electronic tongue based on chemically modified electrodes and voltammetry for the detection of adulterations in wines", *Sensors and Actuators. B* 2006, 118: pp. 448-453.
- [4] S. Buratti, S.Benedetti, M. Scampicchio, E.C. Pangerod , "Characterization and classification of Italian Barbera wines by using an electronic nose and an amperometric electronic tongue", *Analytica Chimica Acta*. 2004, 525: pp. 133-139.
- [5] A. Riul, H.C. de Sousa, R. R. Malmegrim, D.S. dos Santos, A,C,P.L.F. Carvalho, F.J. Fonseca, O.N. Oliveira, L.H.C. Mattoso, "Wine classification by taste sensors made from ultra-thin films and using neural networks", *Sensors and Actuators. B* 2004, 98: pp 77-82.
- [6] C. Di Natale, R. Paolesse, M. Burgio, E. Martinelli, G. Pennazza, A. D'Amico, "Application of metalloporphyrins- based gas and liquid sensor arrays to the analysis of red wine", *Analytica Chimica Acta*. 2004, 513: pp 49-56.
- [7] A. Legin, A. Rudnitskaya, L. Lvova, Y. Vlasov, C. Natale, A. D'Amico, "Evaluation of Italian wine by the electronic tongue: recognition, quantitative analysis and correlation with human sensory perception", *Analytica Chimica Acta*. 2003, 484: pp. 33-44.
- [8] C. Natale, R. Paolesse, A. Macagnano, A. Mantini, A. D'Amico, M. Ubigli, A. Legin, L. Lvova, A. Rudnitskaya, Y. Vlasov, "Application of a combined artificial olfaction and taste system to the quantification of relevant compounds in red wine", *Sensors and Actuators. B* 2000, 69: pp. 342-347.
- [9] G. Cybenko, "Approximation by superposition of a sigmoidal function", *Math. Contr. Signals Syst.*, vol. 2, Dec. 1989, pp 303-314
- [10] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks* 1991, 4: pp. 251-257.
- [11] K. Hornik, M. Stinchcombe, H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks* 1989, 2: pp. 359-366.
- [12] M.D., Richard, R. Lippmann, "Neural network classifiers estimate Bayesian a posteriori probabilities" *Neural Comput.* 1989,3: pp. 461-483.
- [13] P. Gallinari, S. Thiria, F. Badran, F. Fogelman-Soulie, "On the relations between discriminant analysis and multilayer perceptrons," *Neural Networks* 1991, 4(3): pp. 349-360.
- [14] B. Cheng, D.M. Titterington,. "Neural networks: A review from a statistical perspective," *Statistical Science* 1994, 9(1): pp. 2-30.
- [15] G P. Zhang, "Neural Networks for Classification: A Survey," *IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews* 2000, 30(4): pp. 451-463.
- [16] G. E. P Box, M.E. Muller, "A note on the generation of random normal deviates," *Annals of Mathematical Statistics* 1958, 29: pp. 610-611.
- [17] L. Devroye, *Non-uniform random variate generation*. Springer, New York, 1986..

Madhusree Kundu was born in Kolkata, 21 January 1965. She did her B.S.C. in Chemistry from university of Calcutta, India. Dr. Kundu got her B.Tech and M.Tech. in Chemical Engineering from the same university in the years 1990 & 1992, respectively. She completed her Doctoral study from Indian Institute of Technology, Kaharagpur, India and was awarded Ph.D. in the year 2005.

She served as a Process Engineer in the Simon carves India Ltd. during 1993-1998. She served as a lecturer and then Assistant Professor in The Birla Institute of Technology & Science, Pilani, India during 2004-2006. Presently, she is the Associate Professor, Department of Chemical Engineering, NIT, Rourkela, Orissa, India. She has published a couple of research articles and book chapters to her credit. Fluid Phase Equilibria, Advanced Process Control, Fault Detection and Diagnosis, & Process Monitoring are the current research interests of her.

Dr. Kundu is the life member of Indian Institute of Chemical Engineers (IICChE).

Naga C. Kavuri, M.Tech Student, Department of Chemical Engineering, NIT, Rourkela, Orissa, India.

E-mail:biochaitanya@gmail.com