

# Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification

R. Sathya

Professor, Dept. of MCA,  
Jyoti Nivas College (Autonomous), Professor and Head,  
Dept. of Mathematics,  
Bangalore, India.

Annamma Abraham

Professor and Head, Dept. of Mathematics  
B.M.S. Institute of Technology,  
Bangalore, India.

**Abstract:** This paper presents a comparative account of unsupervised and supervised learning models and their pattern classification evaluations as applied to the higher education scenario. Classification plays a vital role in machine based learning algorithms and in the present study, we found that, though the error back-propagation learning algorithm as provided by supervised learning model is very efficient for a number of non-linear real-time problems, KSOM of unsupervised learning model, offers efficient solution and classification in the present study.

**Keywords –** Classification; Clustering; Learning; MLP; SOM; Supervised learning; Unsupervised learning;

## I. INTRODUCTION

Introduction of cognitive reasoning into a conventional computer can solve problems by example mapping like pattern recognition, classification and forecasting. Artificial Neural Networks (ANN) provides these types of models. These are essentially mathematical models describing a function; but, they are associated with a particular learning algorithm or a rule to emulate human actions. ANN is characterized by three types of parameters; (a) based on its interconnection property (as feed forward network and recurrent network); (b) on its application function (as Classification model, Association model, Optimization model and Self-organizing model) and (c) based on the learning rule (supervised/ unsupervised /reinforcement etc.) [1].

All these ANN models are unique in nature and each offers advantages of its own. The profound theoretical and practical implications of ANN have diverse applications. Among these, much of the research effort on ANN has focused on pattern classification. ANN performs classification tasks obviously and efficiently because of its structural design and learning methods. There is no unique algorithm to design and train ANN models because, learning algorithm differs from each other in their learning ability and degree of inference. Hence, in this paper, we try to evaluate the supervised and unsupervised learning rules and their classification efficiency using specific example [3].

The overall organization of the paper is as follows. After the introduction, we present the various learning algorithms used in ANN for pattern classification problems and more specifically the learning strategies of supervised and unsupervised algorithms in section II.

Section III introduces classification and its requirements in applications and discusses the familiarity distinction between supervised and unsupervised learning on the pattern-class information. Also, we lay foundation for the construction of classification network for education problem of our interest. Experimental setup and its outcome of the current study are presented in Section IV. In Section V we discuss the end results of these two algorithms of the study from different perspective. Section VI concludes with some final thoughts on supervised and unsupervised learning algorithm for educational classification problem.

## II. ANN LEARNING PARADIGMS

Learning can refer to either acquiring or enhancing knowledge. As Herbert Simon says, Machine Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more efficiently and more effectively the next time.

ANN learning paradigms can be classified as supervised, unsupervised and reinforcement learning. Supervised learning model assumes the availability of a teacher or supervisor who classifies the training examples into classes and utilizes the information on the class membership of each training instance, whereas, Unsupervised learning model identify the pattern class information heuristically and Reinforcement learning learns through trial and error interactions with its environment (reward/penalty assignment).

Though these models address learning in different ways, learning depends on the space of interconnection neurons. That is, supervised learning learns by adjusting its inter connection weight combinations with the help of error signals where as unsupervised learning uses information associated with a group of neurons and reinforcement learning uses reinforcement function to modify local weight parameters.

Thus, learning occurs in an ANN by adjusting the free parameters of the network that are adapted where the ANN is embedded.

This parameter adjustment plays key role in differentiating the learning algorithm as supervised or unsupervised models or other models. Also, these learning algorithms are facilitated by various learning rules as shown in the Fig 1 [2].

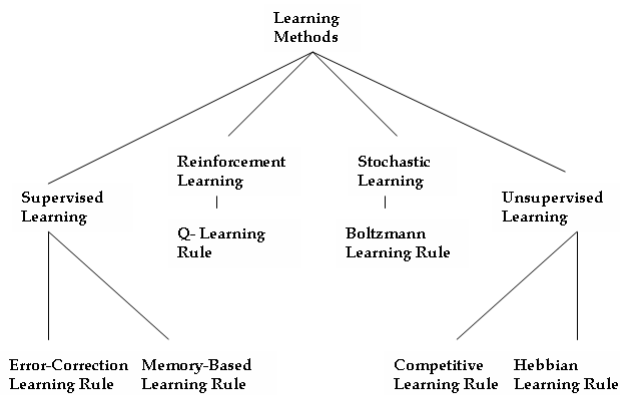


Fig. 1. Learning Rules Of ANN

### A. Supervised Learning

Supervised learning is based on training a data sample from data source with correct classification already assigned. Such techniques are utilized in feedforward or MultiLayer Perceptron (MLP) models. These MLP has three distinctive characteristics:

1. One or more layers of hidden neurons that are not part of the input or output layers of the network that enable the network to learn and solve any complex problems
2. The nonlinearity reflected in the neuronal activity is differentiable and,
3. The interconnection model of the network exhibits a high degree of connectivity

These characteristics along with learning through training solve difficult and diverse problems. Learning through training in a supervised ANN model also called as error back-propagation algorithm. The error correction-learning algorithm trains the network based on the input-output samples and finds error signal, which is the difference of the output calculated and the desired output and adjusts the synaptic weights of the neurons that is proportional to the product of the error signal and the input instance of the synaptic weight. Based on this principle, error back propagation learning occurs in two passes:

**Forward Pass:** Here, input vector is presented to the network. This input signal propagates forward, neuron by neuron through the network and emerges at the output end of the network as output signal:  $y(n) = \phi(v(n))$  where  $v(n)$  is the induced local field of a neuron defined by  $v(n) = \sum w(n)y(n)$ . The output that is calculated at the output layer  $o(n)$  is compared with the desired response  $d(n)$  and finds the error  $e(n)$  for that neuron. The synaptic weights of the network during this pass are remains same.

**Backward Pass:** The error signal that is originated at the output neuron of that layer is propagated backward through network. This calculates the local gradient for each neuron in each layer and allows the synaptic weights of the network to undergo changes in accordance with the delta rule as:

$$\Delta w(n) = \eta * \delta(n) * y(n).$$

This recursive computation is continued, with forward pass followed by the backward pass for each input pattern till the network is converged [4-7].

Supervised learning paradigm of an ANN is efficient and finds solutions to several linear and non-linear problems such as classification, plant control, forecasting, prediction, robotics etc [8-9]

### B. Unsupervised Learning

Self-Organizing neural networks learn using unsupervised learning algorithm to identify hidden patterns in unlabelled input data. This unsupervised refers to the ability to learn and organize information without providing an error signal to evaluate the potential solution. The lack of direction for the learning algorithm in unsupervised learning can sometime be advantageous, since it lets the algorithm to look back for patterns that have not been previously considered [10]. The main characteristics of Self-Organizing Maps (SOM) are:

1. It transforms an incoming signal pattern of arbitrary dimension into one or 2 dimensional map and perform this transformation adaptively
2. The network represents feedforward structure with a single computational layer consisting of neurons arranged in rows and columns.
3. At each stage of representation, each input signal is kept in its proper context and,
4. Neurons dealing with closely related pieces of information are close together and they communicate through synaptic connections.

The computational layer is also called as competitive layer since the neurons in the layer compete with each other to become active. Hence, this learning algorithm is called competitive algorithm. Unsupervised algorithm in SOM works in three phases:

**Competition phase:** for each input pattern  $x$ , presented to the network, inner product with synaptic weight  $w$  is calculated and the neurons in the competitive layer finds a discriminant function that induce competition among the neurons and the synaptic weight vector that is close to the input vector in the Euclidean distance is announced as winner in the competition. That neuron is called best matching neuron, i.e.  $x = \arg \min \|x - w\|$ .

**Cooperative phase:** the winning neuron determines the center of a topological neighborhood  $h$  of cooperating neurons. This is performed by the lateral interaction  $d$  among the cooperative neurons. This topological neighborhood reduces its size over a time period.

**Adaptive phase:** enables the winning neuron and its neighborhood neurons to increase their individual values of the discriminant function in relation to the input pattern through suitable synaptic weight adjustments,  $\Delta w = \eta h_{(x)}(x - w)$ .

Upon repeated presentation of the training patterns, the synaptic weight vectors tend to follow the distribution of the input patterns due to the neighborhood updating and thus ANN learns without supervisor [2].

Self-Organizing Model naturally represents the neuro-biological behavior, and hence is used in many real world applications such as clustering, speech recognition, texture segmentation, vector coding etc [11-13].

### III. CLASSIFICATION

Classification is one of the most frequently encountered decision making tasks of human activity. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object. There are many industrial problems identified as classification problems. For examples, Stock market prediction, Weather forecasting, Bankruptcy prediction, Medical diagnosis, Speech recognition, Character recognitions to name a few [14-18]. These classification problems can be solved both mathematically and in a non-linear fashion. The difficulty of solving such problem mathematically lies in the accuracy and distribution of data properties and model capabilities [19].

The recent research activities in ANN prove, ANN as best classification model due to the non-linear, adaptive and functional approximation principles. A Neural Network classifies a given object according to the output activation. In a MLP, when a set of input patterns are presented to the network, the nodes in the hidden layers of the network extract the features of the pattern presented. For example, in a 2 hidden layers ANN model, the hidden nodes in the first hidden layer forms boundaries between the pattern classes and the hidden nodes in the second layer forms a decision region of the hyper planes that was formed in the previous layer. Now, the nodes in the output layer logically combines the decision region made by the nodes in the hidden layer and classifies them into class 1 or class 2 according to the number of classes described in the training with fewest errors on average. Similarly, in SOM, classification happens by extracting features by transforming of m-dimensional observation input pattern into q-dimensional feature output space and thus grouping of objects according to the similarity of the input pattern.

The purpose of this study is to present the conceptual framework of well known Supervised and Unsupervised learning algorithms in pattern classification scenario and to discuss the efficiency of these models in an education industry as a sample study. Since any classification system seeks a functional relationship between the group association and attribute of the object, grouping of students in a course for their enhancement can be viewed as a classification problem [20-22]. As higher education has gained increasing importance due to competitive environment, both the students as well as the education institutions are at crossroads to evaluate the performance and ranking respectively. While trying to retain its high ranking in the education industry, each institution is trying to identify potential students and their skill sets and group them in order to improve their performance and hence improve their own ranking. Therefore, we take this classification problem and study how the two learning algorithms are addressing this problem.

In any ANN model that is used for classification problem, the principle is learning from observation. As the objective of

the paper is to observe the pattern classification properties of those two algorithms, we developed Supervised ANN and Unsupervised ANN for the problem mentioned above. A Data set consists of 10 important attributes that are observed as qualification to pursue Master of Computer Applications (MCA), by a university/institution is taken. These attributes explains, the students' academic scores, priori mathematics knowledge, score of eligibility test conducted by the university. Three classes of groups are discovered by the input observation [3]. Following sections presents the structural design of ANN models, their training process and observed results of those learning ANN model.

### IV. EXPERIMENTAL OBSERVATION

#### A. Supervised ANN

A 11-4-3 fully connected MLP was designed with error back-propagation learning algorithm. The ANN was trained with 300 data set taken from the domain and 50 were used to test and verify the performance of the system. A pattern is randomly selected and presented to the input layer along with bias and the desired output at the output layer. Initially each synaptic weight vectors to the neurons are assigned randomly between the range [-1,1] and modified during backward pass according to the local error, and at each epoch the values are normalized.

Hyperbolic tangent function is used as a non-linear activation function. Different learning rate were tested and finally assigned between [0.05 - 0.1] and sequential mode of back propagation learning is implemented. The convergence of the learning algorithm is tested with average squared error per epoch that lies in the range of [0.01 - 0.1]. The input patterns are classified into the three output patterns available in the output layer. Table I shows the different trial and error process that was carried out to model the ANN architecture.

TABLE I: SUPERVISED LEARNING OBSERVATION

No. of hidden neurons	No. of Epochs	Mean-squared error	Correctness on training	Correctness on Validation
3	5000 - 10000	0.31 - 0.33	79%	79%
<b>4</b>	<b>5000 - 10000</b>	<b>0.28</b>	<b>80% - 85%</b>	<b>89%</b>
5	5000 - 10000	0.30 - 0.39	80% - 87%	84%

#### B. Unsupervised ANN

Kohonen's Self Organizing Model (KSOM), which is an unsupervised ANN, designed with 10 input neurons and 3 output neurons. Data set used in supervised model is used to train the network. The synaptic weights are initialized with  $1/\sqrt{}$  (number of input attributes) to have a unit length initially and modified according to the adaptability.

Results of the network depends on the presentation pattern of the input vector for small amount of training data hence, the training patterns are presented sequentially to the NN.

Euclidean distance measure was calculated at each iteration to find the winning neuron. The learning rate parameter initially set to 0.1, decreased over time, but not decreased below 0.01. At convergence phase it was maintained to 0.01 [11]. As the competitive layer is one dimensional vector of 3 neurons, the neighborhood parameter has not much influence on the activation. The convergence of the network is calculated when there were no considerable changes in the adaptation. The following table illustrates the results:

TABLE II. UNSUPERVISED LEARNING OBSERVATION

Learning rate parameter	No. of Epochs	Correctness on training	Correctness on Validation
.3 - .01	1000 - 2000	85%	86%
<b>.1 - .01</b>	<b>1000 - 3000</b>	<b>85% - 89%</b>	<b>92%</b>

## V. RESULTS AND DISCUSSION

In the classification process, we observed that both learning models grouped students under certain characteristics say, students who possess good academic score and eligibility score in one group, students who come from under privileged quota are grouped in one class and students who are average in the academics are into one class.

The observation on the two results favors unsupervised learning algorithms for classification problems since the correctness percentage is high compared to the supervised algorithm. Though, the differences are not much to start the comparison and having one more hidden layer could have increased the correctness of the supervised algorithm, the time taken to build the network compared to KSOM was more; other issues we faced and managed with back-propagation algorithm are:

1. *Network Size*: Generally, for any linear classification problem hidden layer is not required. But, the input patterns need 3 classifications hence, on trail and error basis we were confined with 1 hidden layer. Similarly, selection of number of neurons in the hidden layer is another problem we faced. As in the Table I, we calculated the performance of the system in terms of number of neurons in the hidden layer we selected 4 hidden neurons as it provides best result.
2. *Local gradient descent*: Gradient descent is used to minimize the output error by gradually adjusting the weights. The change in the weight vector may cause the error to get stuck in a range and cannot reduce further. This problem is called local minima. We overcame this problem by initializing weight vectors randomly and after each iteration, the error of current pattern is used to update the weight vector.

3. *Stopping Criteria*: Normally ANN model stops training once it learns all the patterns successfully. This is identified by calculating the total mean squared error of the learning. Unfortunately, the total error of the classification with 4 hidden neuron is 0.28, which could not be reduced further. When it is tried to reduce minimum the validation error starts increasing. Hence, we stopped the system on the basis of correctness of the validation data that is shown in the table 89%. Adding one more neuron in the hidden layer as in the last row of Table I increase the chance of over fitting on the train data set but less performance on validation.
4. The only problem we faced in training of KSOM is the declaration of learning rate parameter and its reduction. We decreased it exponentially over time period and also we tried to learn the system with different parameter set up and content with 0.1 to train and 0.01 at convergence time as in Table II.

Also, unlike the MLP model of classification, the unsupervised KSOM uses single-pass learning and potentially fast and accurate than multi-pass supervised algorithms. This reason suggests the suitability of KSOM unsupervised algorithm for classification problems.

As classification is one of the most active decision making tasks of human, in our education situation, this classification might help the institution to mentor the students and improve their performance by proper attention and training. Similarly, this helps students to know about their lack of domain and can improve in that skill which will benefit both institution and students.

## VI. CONCLUSION

Designing a classification network of given patterns is a form of learning from observation. Such observation can declare a new class or assign a new class to an existing class. This classification facilitates new theories and knowledge that is embedded in the input patterns. Learning behavior of the neural network model enhances the classification properties. This paper considered the two learning algorithms namely supervised and unsupervised and investigated its properties in the classification of post graduate students according to their performance during the admission period. We found out that though the error back-propagation supervised learning algorithm is very efficient for many non- linear real time problems, in the case of student classification KSOM – the unsupervised model performs efficiently than the supervised learning algorithm.

## REFERENCES

- [1] L. Fu., Neural Networks in Computer Intelligence, Tata McGraw-Hill, 2003.
- [2] S. Haykin, Neural Networks- A Comprehensive Foundation, 2<sup>nd</sup> ed., Pearson Prentice Hall, 2005.
- [3] R. Sathya and A. Abraham, "Application of Kohonen SOM in Prediction", In Proceedings of ICT 2010, Springer-Verlag Berlin Heidelberg, pp. 313–318, 2010.
- [4] R. Rojas, The Backpropagation Algorithm, Chapter 7: Neural Networks, Springer-Verlag, Berlin, pp. 151-184, 1996.

- [5] M. K. S. Alsmadi, K. B. Omar, and S. A. Noah, "Back Propagation Algorithm: The Best Algorithm Among the Multi-layer Perceptron Algorithm", *International Journal of Computer Science and Network Security*, Vol.9, No.4, 2009, pp. 378 – 383.
- [6] X. Yu, M. O. Efe, and O. Kaynak, "A General Backpropagation Algorithm for Feedforward Neural Networks Learning", *IEEE Trans. On Neural Networks*, Vol. 13, No. 1, 2002, pp. 251 – 254.
- [7] Jin-Song Pei, E. Mai, and K. Piyawat, "Multilayer Feedforward Neural Network Initialization Methodology for Modeling Nonlinear Restoring Forces and Beyond", *4<sup>th</sup> World Conference on Structural Control and Monitoring*, 2006, pp. 1-8.
- [8] Awodele, and O. Jegede, "Neural Networks and Its Application in Engineering", *Proceedings of Informing Science & IT Education Conference (InSITE) 2009*, pp. 83-95.
- [9] Z. Rao, and F. Alvarruiz, "Use of an Artificial Neural Network to Capture the Domain Knowledge of a Conventional Hydraulic Simulation Model", *Journal of Hydroinformatics*, 2007, pg.no 15-24.
- [10] T. Kohonen, O. Simula, "Engineering Applications of the Self-Organizing Map", *Proceeding of the IEEE*, Vol. 84, No. 10, 1996, pp.1354 – 1384
- [11] R. Sathya and A. Abraham, "Unsupervised Control Paradigm for Performance Evaluation", *International Journal of Computer Application*, Vol 44, No. 20, pp. 27-31, 2012.
- [12] R. Pelessoni, and L. Picech, "Self Organizing Maps - Applications and Novel Algorithm Design, An application of Unsupervised Neural Networks in General Insurance: the Determination of Tariff Classes".
- [13] P. Mitra, "Unsupervised Feature Selection Using Feature Similarity", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, pp. 2002.
- [14] Moghadassi, F. Parvizian, and S. Hosseini, "A New Approach Based on Artificial Neural Networks for Prediction of High Pressure Vapor-liquid Equilibrium", *Australian Journal of Basic and Applied Sciences*, Vol. 3, No. 3, pp. 1851-1862, 2009.
- [15] R. Asadi, N. Mustapha, N. Sulaiman, and N. Shiri, "New Supervised Multi Layer Feed Forward Neural Network Model to Accelerate Classification with High Accuracy", *European Journal of Scientific Research*, Vol. 33, No. 1, 2009, pp.163-178.
- [16] Pelliccioni, R. Cotroneo, and F. Pung, "Optimization Of Neural Net Training Using Patterns Selected By Cluster Analysis: A Case-Study Of Ozone Prediction Level", *8<sup>th</sup> Conference on Artificial Intelligence and its Applications to the Environmental Sciences*, 2010.
- [17] M. Kayri, and Ö. Çokluk, "Using Multinomial Logistic Regression Analysis In Artificial Neural Network: An Application", *Ozean Journal of Applied Sciences* Vol. 3, No. 2, 2010.
- [18] U. Khan, T. K. Bandopadhyaya, and S. Sharma, "Classification of Stocks Using Self Organizing Map", *International Journal of Soft Computing Applications*, Issue 4, 2009, pp.19-24.
- [19] S. Ali, and K. A. Smith, "On learning algorithm selection for classification", *Applied Soft Computing* Vol. 6, pp. 119–138, 2006.
- [20] L. Schwartz, K. Stowe, and P. Sendall, "Understanding the Factors that Contribute to Graduate Student Success: A Study of Wingate University's MBA Program".
- [21] B. Naik, and S. Ragothaman, "Using Neural Network to Predict MBA Student Success", *College Student Journal*, Vol. 38, pp. 143-149, 2004.
- [22] V.O. Oladokun, A.T. Adebajo, and O.E. Charles-Owaba, "Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course", *The Pacific Journal of Science and Technology*, Vol. 9, pp. 72–79, 2008.