The International Journal Of Engineering And Science (Ijes)
||Volume||2 ||Issue|| 3 ||Pages|| 33-38 ||2013||
Issn: 2319 – 1813 Isbn: 2319 – 1805

Artificial Neural Network Approach to Clustering

[1,]Farhat Roohi
Department of Electronics and IT, University of Kashmir, J&K, India

--Abstract--
In the field of computing, learning and artificial intelligence the developments have seen a transition from developing powerful computing powers, programmed closed systems devoid of learning capability to attempts to model the human brain as far as its learning and functioning is concerned. Attempts at this have to make models which have inbuilt self learning mechanism capable to learn from data continuously and modify the rules and output accordingly. Towards this end many attempts have been made amongst which neural networks based on their robustness and emulating capability of human brain deserve special attention. These networks which are combinations and interconnections of large number of artificial neurons mimic the learning capability and mechanism of human learning. They form the base of artificial intelligence and are central to various clustering programmes. Clustering, which is at the base of most of the data analysis, decision making, designing, forecasting, problems finds wide applications of neural networks. They have greatly facilitated the clustering function in both supervised and unsupervised kind of learning. They are of great help when the number of clusters is not known before hand. Against this backdrop this paper attempts to gain an insight into ANN and its basic components with main focus on its application in clustering.

Keywords - *Artificial Neural Network (ANN), cluster analysis, competitive learning, self-organizing map.*

Date Of Submission: 27 February , 2013 ⟺ Date Of Publication:25 March 2013

I. INTRODUCTION

Faced with the inability of the conventional computer systems to learning or be intelligence, the scholars and experts have been continuously trying to emulate the intelligence and learning capability of humans in computers. The usual computer system operates through sequential linear processing technologies; applying formulas, decision rules, and algorithms instructed by users to produce outputs from the inputs. They are good at numerical computation. This human like learning and intelligence capability is built into the computer system by artificial neural networks (ANNs). They have the capability of improving their own rules as they function and thereby produce better output as they proceed.

An ANN is basically a computer program designed to learn like the human brain. While the neural network (NN) formed in this system pertains to ANN, it forms the foundation of Artificial Intelligence. It mimics the human brain on two grounds. First, it obtains knowledge by means of the network through a learning process. Secondly, for storing the acquired knowledge the interneuron connection strengths are utilized. The greatest advantage of NN is that it makes data to speak for itself and does not require a priori assumption about the problem and like human brain attempts to learn continuously through experiences to make better decisions over a period of time. It attempts to filter out patterns present in the existing data and then extrapolates it to the future. Thus as data increases the quality of output also improves, as the system and rules learns from the data. In fact the basic research in this field has been inspired by the research to understand the learning mechanism of human brain.

Like human brain which has diversified applications, ANN also displays interesting learning capability and thus finds applications and research interests in varied fields like science, space, medicine, business, finance economics, econometrics, etc. Given the diversified applications of ANN research activities and interests in this field particularly for the last two decades have surged. Against this backdrop the current paper attempts to gain an insight into NN and its basic components. Since ANN works by filtering out patterns so the current paper mainly focuses on its application in data clustering.

II. ARTIFICIAL NEURAL NETWORK

ANN, an information processing model, is inspired by the way biological nervous system, such as the brain, processes the information. It is a mathematical model that seeks to simulate the structure and functionalities of biological neural networks. The unique structure of its information processing system is its key feature. ANN is comprised of a large number of vastly interconnected processing elements, called neurons, which work in unity to solve some specific problems. They learn by examples, like humans do, and are configured through a learning process for specific applications, like pattern recognition or data classification. Learning in the ANNs happen in the same way as it takes place in the biological systems, which involves adjustments to the interneuron synaptic connections.The basic building block of every artificial neural network is an artificial neuron. It is a simple mathematical model/ function, which has three simple sets of rules: multiplication, summation and activation. At the entry level each input to the artificial neuron is multiplied with individual weight to get a weighted input value. These weighted input values and the bias are summed up through a summation function that is in the middle section of artificial neuron. At the final step or the exit of the artificial neuron the weighted sum of previously weighted inputs and bias passes through the activation function, also known as transfer function (Fig. 1.) [1].

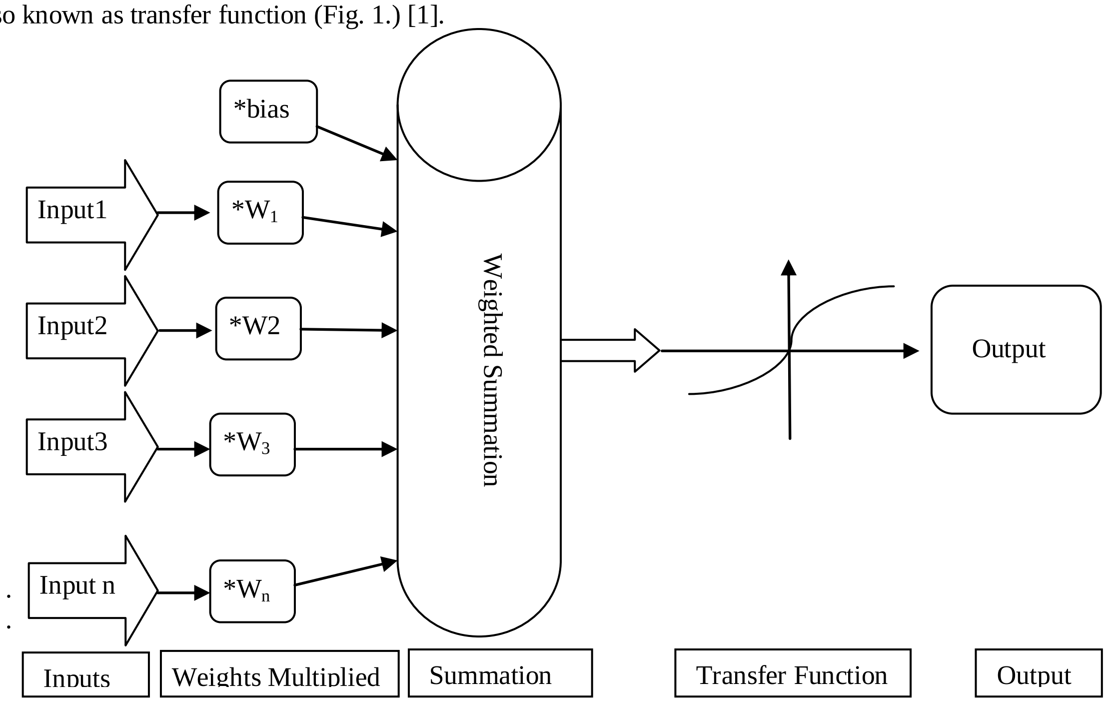

Figure: 1. A Simple Model of working Neuron

This simple looking artificial neuron shows its real power when connected with other neurons in a network. The information processing in it can be simply understood by viewing neurons taking information by way of inputs which are multiplied by individual weights at the entry and fed into the body of the artificial neuron where these weighted inputs are summed along with bias are processed and passed through transfer function to output. The artificial neuron can be mathematically modeled as [1]:

$$y(k) = F(\sum_{i=0}^{m} wi(k).xi(k) + b)$$

Where:

$xi(k)$ is input value in discrete time k and i ranges from 0 to m,

$wi(k)$ is the weight value in discrete time k and i ranges from 0 to m,

b is the bias,

F is a the transfer function,

y(k) is output value in discrete time k.

The equation and the model depict that transfer function is the main feature of the model and it can be any mathematical function depending on the nature of the problem. Generally, transfer function is selected from: Step function, Linear function and Non-linear (Sigmoid) function [1]. Step function is a binary function that means only two types of output values e.g. zero and one are possible in this case. It means that if input value meets specific threshold the output value results in one value and if specific threshold is not meet the result will be different output value i.e. zero. Artificial neurons using this type of transfer function are known as perceptrons. Perceptron is generally found in the last layer of artificial neural networks as it is commonly used for solving classification problems. As against this linear transfer functions are usually used in the input layer of artificial neural networks as it is doing simple linear transformation of the sum of weighted inputs and bias. When dealing with non-linear problems, the most commonly used function is sigmoid function, which has easily calculated derivate, that can be important when calculating weight updates [1].The combination of various artificial neurons interconnected with each other give rise to artificial neural network, which are self adaptive data driven non linear approaches. These interconnections or networks can take various topologies or frameworks which along with signal processing can be broadly classified into three basic categories: feed foreword, feedback and self organizing [2]. In feed-forward networks while the set of input signals are transformed into set of output signals the information flows from inputs to outputs in only one direction and the desired input output transformation is usually determined by external, supervised adjustment of the system parameters [3]. In feedback networks the initial activity state of the network is defined by the input information, while after state transitions the asymptotic final state is identified as the outcome of the computation [4]. In this system the information flows not only in one direction from input to output but also in opposite direction. The third kind of network is characterized by the neighboring cells competing in their activities by means of mutual lateral interactions, and develop adaptively into specific detectors of different signal patterns and learning in this category is known as competitive, unsupervised or self-organized [2]. Further, usually in an ANN neurons are placed in input, hidden and output layer, which help in easier handling and mathematical description. Once the architect of ANN is finalized the stage is to make it develop the capability of learning which can be supervised, un-supervised or reinforcement learning. The task of learning irrespective of the method used is to develop proper response to the environment by adjusting the weight and biases on basis of learning data to optimize the objective function.

III. CLUSTERING

Most of the problems that prop up in all the fields of human operations pertain to organize objects or data in different categories are classes with interclass heterogeneity and intraclass homogeneity on the defining criteria. Such problems arise when an object needs to be assigned into a predefined or undefined group or class based on a number of observed attributes related to that object. Several of the problems encountered in business, science, industry, and medicine can be treated as classification or clustering problems, which may include predictions, medical diagnosis, quality control, credit scoring, handwritten character recognition, and speech recognition. It is, however, important to draw the distinction between clustering i.e. unsupervised classification and discriminant analysis i.e. supervised classification. In supervised classification, while a collection of labeled or pre-classified patterns are provided the problem is to label a newly encountered, yet unlabeled, pattern. Characteristically, the given labeled or training patterns are used to learn the descriptions of classes that are in turn used to label a new pattern. Contrary to this, in clustering a collection of unlabeled patterns are grouped into meaningful clusters. In this case also labels are associated with clusters but these category labels are data driven.

Clustering is particularly useful in exploratory pattern-analysis where little prior information is available about the data and it is under such restrictions that it is particularly suitable for the exploration of interrelationships among the data points to make a preliminary assessment of the data structure. Generally pattern clustering activity includes: [5]: (1) pattern representation (optionally including feature extraction and/or selection), (2) definition of a pattern proximity measure appropriate to the data domain, (3) clustering or grouping, (4) data abstraction (if needed), and (5) assessment of output (if needed). Different approaches to clustering data can be broadly classified into hierarchical and partitional, while hierarchical methods produce a nested series of partitions, the partitional methods produce only one. Under these categories are included Agglomerative *vs.* divisive, Monothetic *vs.* polythetic, Hard *vs.* fuzzy, Deterministic *vs.* stochastic, Incremental *vs.* non-incremental [5]. Apart from this many new algorithms have been suggested recently, while some methods represent a modification of classical methods; the others use advanced methods such as neural networks, e.g. represented by Kohonen self-organizing maps, or genetic algorithms. However, ANNs have been used extensively over the past three decades for both classification and clustering [6,7], and have established that neural networks are a promising alternative to various conventional classification methods.

IV. CLUSTER ANALYSIS THROUGH ANN - APPLICATION

Data, being an outcome of a process, possess whole of the information regarding the process but in a hidden form in the complex structures. This data thus needs to be analyzed in order to reveal complex patterns and relationships hidden in it. To achieve this, clustering is used as a valuable tool. It is an unsupervised classification technique that identifies some inherent structures present in a data set based on a similarity or proximity measure. Since all the classification procedures look for an accurate function that underlies the functional relationship between various groups present in the data, artificial neural network proves to be a good option as it can approximate any function with arbitrary accuracy[8,9]. Additionally, neural networks being nonlinear models can be used to model any real world complex process. Their ability to estimate the posterior probabilities provide the basis for establishing classification rule and performing statistical analysis [10]. Above all, the extensive research activities in neural classification have recognized that neural networks are a potential alternative to usual classification methods . Neural networks have proved to be a useful technique for implementing competitive learning based clustering, which have simple architectures. Such networks have an output layer termed as the competition layer. The neurons in the competition layer are fully connected to the input nodes. The lateral connections in this layer are used to perform lateral inhibition. The basic principle underlying competitive learning is the mathematical statistics problem called cluster analysis that is usually based on the minimization of the average of the squared Euclidean distances between the inputs and their closest prototypes such that the input only attracts its winning prototype and has no effect on the non winning prototypes. Patterns are presented at the input and are associated with the output nodes. The weights between the input nodes and the output nodes are iteratively changed until a termination criterion is satisfied. A competitive learning-based neural networks used for clustering include Kohonen's Self-organizing map (SOM), Adaptive resonance theory models and learning vector quantization (LVQ), [11, 12].

V. Self Organizing Feature Maps (SOFM)

Self organizing feature maps (SOFM) also called Kohonen feature maps [11] belong to the category of competitive learning based clustering. The nodes (neurons) of these become specifically tuned to various input patterns. It consists of two layers of neurons, an input layer and a so-called competition layer. When an input pattern is presented to the network, that neuron in the competition layer is determined, the reference vector of which is closest to the input pattern. This neuron is called the winner neuron. The learning is such that only one node (neuron) at the output becomes active corresponding to a particular input pattern The learning in SOFM, however, differs from the other competitive based clustering methods in that not only the weight of the winning neuron is going to change but also those in its neighborhood $N_c(t)$ (with $t_1 < t_2 < t_3 < t_4$) which is defined in terms of some proximity relation. This neighborhood relation is usually represented as a (usually two-dimensional) grid as shown in Fig:2, the vertices of which are the neurons. The grid is most often rectangular or hexagonal. This results in a particular spatial arrangement of the nodes corresponding to a particular domain of input patterns. Thus during the learning process, the weights of all neurons in the competition layer that lie within a certain radius around the winner neuron with respect to this grid are also adapted, although the strength of the adaption depends on their distance from the winner neuron. It thus projects input space on prototypes of a low-dimensional regular grid that can be effectively utilized to visualize and explore properties of the data. The convergence of SOFM is controlled by the learning rate and the neighborhood of the winning node in which learning takes place. The SOM is well suited for clustering data as it preserves the topological and metric relations between the data items. However, for large data bases, SOM is not a good option because the number of nodes in the size of the SOM grid grows exponentially.

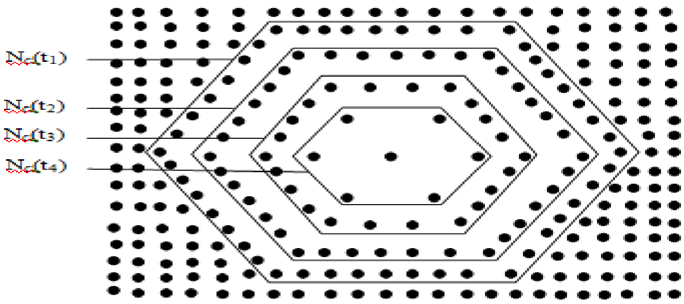

Fig: 2. Topological Neighborhood of Kohonen's Net
Nc(t)

5.1. Adaptive Resonance Theory (ART)

ART networks perform clustering based on the comparison of the input vector with an active code vector. A number of versions of ART networks have been reported in literature. The first ART model, ART1 given by Carpenter and Grossberg in1987 [14] shown in Fig: 3 consists of two layers, the comparison layer and the recognition layer with feedforward (with top up weights tij) as well as feedback connections(with bottom down weights bji) between the nodes. Each output layer node may be visualized as a prototype. The number of neurons in comparison layer are equal to the total number of features and the number of neurons in the recognition layer are equal to the maximum expected number of clusters. As the inputs are presented to the network, the model selects the first input vector as the specimen (code vector) for the first cluster. The first neuron in the recognition layer is made to identify this cluster. The cluster centre is represented by the associated neuron's top down weight vector. When the next input is presented, it is compared with the first cluster specimen. If this input is similar to the specimen, within a specified threshold limit (vigilance parameter), then it is treated as the member of the first cluster. The weight connected to this group is also updated in the light of new input vector. If the new input is not similar to the specimen, it becomes the second cluster in the recognition layer. This process is repeated for all inputs. The network is thus adaptive as it allows for addition of new cluster prototypes to the network. ART learns to cluster the input pattern by making the output neurons compete with each other for the right to react to a particular input pattern. The output neuron which has the weight vector that is most similar to the input vector claims this input pattern by producing an output of '1' and at the same time inhibits other output neurons by forcing them to produce '0's. In ART, only the winning node is permitted to alter its weight vector, which is modified in such a way that is brought near to the representative input pattern in cluster concerned. The output of ART is an indication of membership of the input pattern in a group with similar characteristics. ART networks have the property of classifying large data bases in a stable manner. These networks allow the user to control the degree of similarity between the members of the same cluster by means of a user defined parameter called the vigilance parameter. However, ART nets are order-dependent that is, different partitions are obtained for different orders in which the data is presented to the net. Also, the size and number of clusters generated by an ART net depend on the value chosen for the vigilance threshold, which is used to decide whether a pattern is to be assigned to one of the existing clusters or start a new cluster.

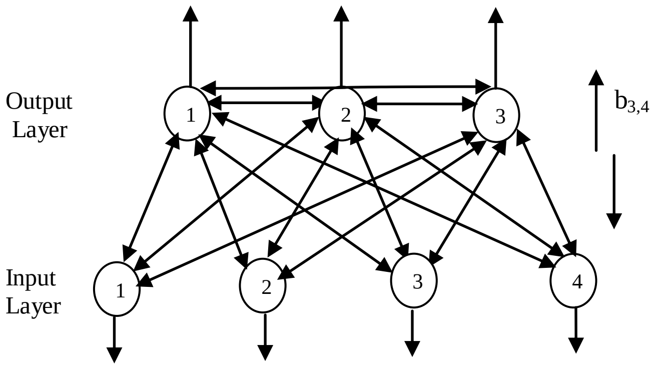

Fig 3: ART1 Network

ART networks are based on Stephen Grossberg's stability plasticity dilemma and are supposed to be stable and plastic [12]. The issue of stability in learning systems comes up when a particular input pattern can fire different output units at different iterations. A system is said to be stable when no pattern in the training data changes its category after a fine number of iterations. The stability is closely associated with the problem of plasticity, which is the ability of the algorithm to adapt to new data.

5.2 Learning vector quantization (LVQ)

Learning vector quantization (LVQ) is a method for training competitive layers in a supervised manner (with target outputs) whereby the network learns to classify input vectors into target classes chosen by the user. An LVQ has the same network architecture as that of SOM with the exception that each neuron is specified with some class membership and no assumption is made about the topological structure. An LVQ network consists of a competitive layer and a linear layer. Both the competitive and linear layers have one neuron per class. The competitive layer learns to classify input vectors in much the same way as the competitive layers of Self-Organizing Feature Maps. The linear layer transforms the competitive layer's classes into target classifications defined by the user. The classes learned by the competitive layer are referred to as subclasses and the classes of the linear layer as target classes. Two categories of LVQ models exist, the supervised case which include LVQ1,LVQ2 and LVQ3 [11] and unsupervised like LVQ4 and incremental c means[13].

VI. CONCLUSION

Developing powerful computing powers, and trying to emulate human learning process have been the factors to the present developments in the world. Attempts have been made to place an inbuilt self learning mechanism capable to learn from data continuously and modify the rules and output accordingly in the computer systems. Towards this end many attempts have been made amongst which neural networks, based on their robustness and emulating capability of human brain have found wide applications. These networks which are combinations and interconnections of large number of artificial neurons form the base of artificial intelligence and are central to various clustering programmes. Clustering is at the base of most of the data analysis, decision making, designing, forecasting, problems finds wide applications and are highly facilitated by the artificial neural network. ANNs are now extensively used in both supervised and unsupervised kind of learning. They are of great help when the number of clusters is not known before hand. Considering the significance of this field the present paper has made an attempt to educate about the neural networks, clustering and application of ANN in clustering.

References

[1] Krenker A.; Volk M.; Sedlar U.; Bešter J.; Kos A. (2009). Bidirectional artificial neural networks for mobile-phone fraud detection. ETRI Journal., vol. 31, no. 1, Feb. 2009, pp. 92-94, COBISS.SI-ID 6951764

[2] T Kohonen (1990), The Self Organizing Map, Proceeding of the IEEE, Vol 78, No 9 Sep 1990

[3] D E Rumelhart, G E Hinton and R J Williams, Learning Internal Representations by error propagation, In Parallel Distributed Processing: Explorations in the Microstructure of cognition. Vol. 1, Foundations, D E Rumelhart, J L McClelland and the PDP research group, Eds. Cambridge, Mass, MIT Press, 1986, pp 318-362

[4] J J Hopfield, Neural Networks and Physical Systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. USA, Vol. 79, pp 2554-2558, 1982.

[5] Jain, A. K. and Dubes, R. C. 1988. Algorithms for Clustering Data. Prentice-Hall advanced reference series. Prentice-Hall, Inc., Upper Saddle River, NJ.

[6] Sethi, I. and Jain, A. K., Eds. 1991. Artificial Neural Networks and Pattern Recognition: Old and New Connections. Elsevier Science Inc., New York, NY.

[7] Jain, A. K. and Mao, J. 1994. Neural networks and pattern recognition. In Computational Intelligence: Imitating Life, J. M. Zurada, R. J. Marks, and C. J. Robinson, Eds. 194– 212.

[8] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” Math. Contr. Signals Syst., vol. 2, pp. 303–314, 1989.

[9] K. Hornik, “Approximation capabilities of multilayer feedforward networks,” Neural Networks, vol. 4, pp. 251–257, 1991.

[10] M. D. Richard and R. Lippmann, “Neural network classifiers estimate Bayesian a posteriori probabilities,” Neural Comput., vol. 3, pp. 461–483, 1991.

[11] Kohonen T (1989) Self Organizing and Associative memory 3rd edi Springer Information Science Series, Springer-Verlag NewYork, NY

[12] Carpenter, G. and Grossberg, S. 1990. ART3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. Neural Networks 3, 129–152.

[13] MacQueen JB (1967) Some methods for classification and analysis of multivariate observations. In: Proc 5th Berkeley Symp on Math Statistics and Probability, University of California Press, Berkeley, 281–297

[14] Carpenter GA, Grossberg S (1987) A massively parallel architecture for a selforganizing neural pattern recognition machine. Computer Vision, Graphics, Image Process 37:54–115.