



Training Algorithms for Supervised Machine Learning: Comparative Study

Dr. Rafiqul Zaman Khan¹, Md. Haider Khalaf Jabbar²

¹Associate Professor,

Department of Computer Science,

Aligarh Muslim University, Aligarh-202002 (U.P), INDIA

rzk32@yahoo.co.in, rzkhan.cs@mail.amu.ac.in

²Research Scholar,

University of Misan, Misan, Iraq and

Department of Computer Science

Aligarh Muslim University, Aligarh-202002 (U.P), INDIA

haider.allamy@gmail.com, he25_allamy@yahoo.com

ABSTRACT

Supervised machine learning is an important task for learning artificial neural networks; therefore a demand for selected supervised learning algorithms such as back propagation algorithm, decision tree learning algorithm and perceptron algorithm has been arise in order to perform the learning stage of the artificial neural networks. In this paper; a comparative study has been presented for the aforementioned algorithms to evaluate their performance within a range of specific parameters such as speed of learning, overfitting avoidance, and their accuracy. Besides these parameters we have included their benefits and limitations to unveil their hidden features and provide more details regarding their performance. We have found the decision tree algorithm is the best as compared with other algorithms that can solve the complex problems with a remarkable speed.

Keywords

artificial neural networks, supervised learning, back propagation, Perceptron, Decision Tree learning algorithm.

Academic Discipline And Sub-Disciplines

Computer science, Expert System and Artificial Intelligence

SUBJECT CLASSIFICATION

Artificial Intelligent, Machine Learning

TYPE (METHOD/APPROACH)

Comparative Study

Council for Innovative Research

Peer Review Research Publishing System

Journal: [International Journal of Management & Information Technology](#)

Vol.4, No.3

editor@cirworld.com

www.cirworld.com, member.cirworld.com

2. LEARNING

The ability to learn new knowledge is a basically trait of intelligence. For learning network, we need to know two things: first we must have a model of environment in which neural network will be operated that means we must know the learning paradigm and second, we must know how the weight of neural network will update and by which rule the weight will be updated [14]. There are three ways to learn neural network (learning paradigm); supervised learning, unsupervised learning and hybrid learning. Each paradigm is working for different tasks based on one of the rules: Error Correction, Boltzmann, Hebbian and Competitive learning [1] [14]. The details are explained and showing in the Table 1.

Table 1. Rule of learning and Learning Algorithms for machine learning

Paradigm	Rule of learning	Learning Algorithms
Supervised	Error-correction	Perceptron learning algorithm, back-propagation & Adaline, madeleine.
	Boltzmann	Boltzmann learning algorithm
	Hebbian	Linear discriminant analysis
	Competitive	Learning vector quantization.
Unsupervised	Error-correction	Sammon's projection
	Hebbian	Principal component analysis
	Competitive	Associative memory learning
		Vector quantization & Kohonen's SOM
Hybrid	Error-correction & competitive	RBF learning algorithm

2.1 Supervised Learning

Supervised learning is a machine learning technique that uses to train the artificial neural networks by the value of input and output by findings the relationship between the input (same time call independent variable) and the output (same time call dependent variable), that means making map between input and output for the artificial system that want to learn. Supervised can predict the value of output for system given new input [1] [12] [10] [16]. Supervised learning is a method for training artificial neural networks when the input and output is known, we use supervised learning when we have a set of training data, this training data is a collection of few input data that connected with few output data, this training data it will use by learning algorithms like back propagation and perceptron, which we have discussed about them in section 3 and 4 [3] [6] [7] [12]. The process of training will be done when the neural network will use the input data to produce its own output data and then compares this with the desired output data or the target output data. If there is no difference between an actual output data and the target output data, that means the learning for this data is completed and no need to continue the training for this data, otherwise, the connection weights will be changed by the algorithm of training to reduce the difference between the input and target output data (see fig 1).

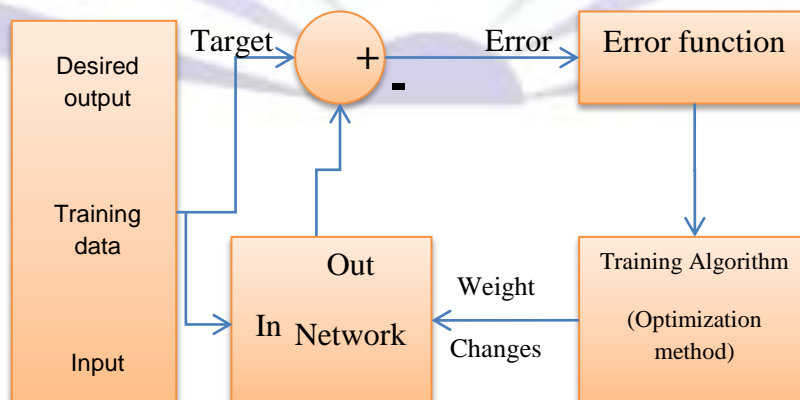


Fig 1: The process of supervised learning

3. BACK-PROPAGATION

The perceptron that uses in today's artificial neural networks was conceived by Rosenblatt in 1950 and he proposed it in 1962 [3] [15]. The back-propagation or (back error-propagation) is the most commonly adopted Multi Layers Perceptron (MLPs) training algorithm, it was first presented by werbos in 1974 and then it was independently re-invented in 1986 by Rumelhart et al. Neural networks are a mathematical model inspired by biological neural networks that shows the set of connected input/output units in which each connection has a weight associated with it, this algorithm is one of the neural used supervised learning method. The learning of neural network will be done by adjusting the weight of neural network [13] [15] [8]. Back propagation uses the target values to calculate the mean square error of the artificial neural networks and level of individual in a population [3] [4] [6].

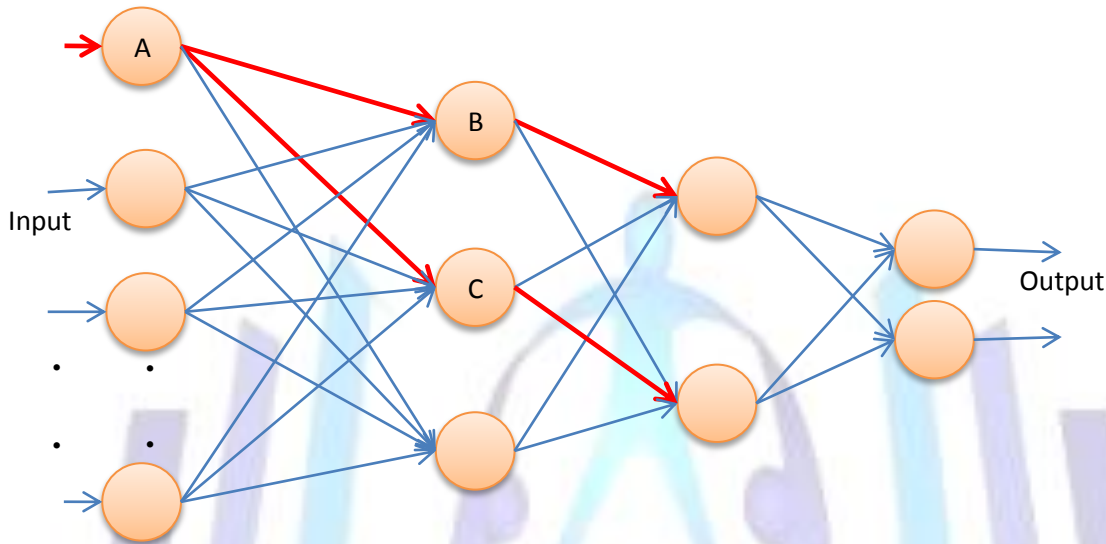


Fig 2: Architecture of back propagation (MLP)

The back propagation algorithm is the most used algorithms for learning in neural network especially for Multi Layers Perceptron (MLPs) [8].

The algorithm is summarized as follow [18] [19]:

Step 1: Initialize all weights for the neural network and bias.

Step 2: Select the input data to the neural network and apply to input layer of network.

Step 3: Activation of the units of the output layer through the network.

Step 4: Calculate the error for the neuron. We mean by errors are the deference between the desired output and neural output.

$$Error_B = Output_B (1 - Output_B)(target_B - Output_B) \tag{1}$$

Calculate what is the output should have been and a scaling factor, how much lower or higher. The output must be adjusted to match the desired output.

Step 5: Reduce the error by changing the weight let W_{AB}^+ is the new (training) and W_{AB} is initial weight.

$$W_{AB}^+ = W_{AB} + (ERROR_B * output_A) \tag{2}$$

Step 6: Finding the error for the hidden layer. Also change the weights for the hidden layers.

$$Error_A = Output_A (1 - Output_A)(Error_B W_{AB} - Error_C W_{AC}) \tag{3}$$

Step 7: Repeat the steps from step 2 to step 6 to train a network of any number of layers.

4. PERCEPTRON ALGORITHM

Perceptron algorithm has proposed in 1958 by Rosenblatt and its one of the neural networks algorithms that are based on error correction rules. Error correction is using the error signal (d-y) to modify the connection weight to gradually reduce the error. Perceptron consists of a single neuron with adjustable weight, $w_{ij} = 1,2,...,n$, and threshold U, as shown in (Figure 3) given an input vector $x = (x_1, x_2, ..., x_n)$ the net input to neuron is:

$$V = \sum_{i=0}^n W_j X_j - U \tag{4}$$

The output is +1 if $V > 0$ and 0 otherwise [14]. The benefit of perceptron algorithm is its low computational requirement and its ability to guarantee convergence to a solution for linearly separable problem [11].

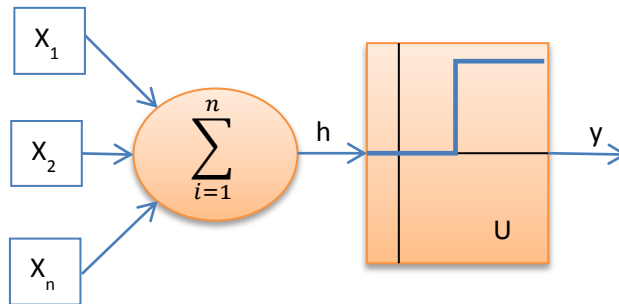


Fig 3: Architecture of Perceptron

Perceptron learning algorithm is summarized as follow [14]:

- 1- Initialize the weight and threshold to small random number.
- 2- Present a pattern vector $(x_1, x_2, \dots, x_n)^t$ and evaluate the output of the neuron.
- 3- Update the weight according to:

$$w_j(t + 1) = w_j(t) + \eta(d - u_0x_j) \tag{5}$$

Where d is the desired output, t is the nitration number, and η ($0.0 < \eta < 1.0$) is the gain (step size).

5. DECISION TREE

The algorithm of decision tree is one of the supervised learning algorithms used like technique machine that build a decision tree from a set of class labeled training sample during the machine learning process. The main purpose of decision tree is the structural information contained in data [9]. Decision tree proposed by Quinlan and there are various Top-down decision tree induces such as ID3(Quinlan 1986), C4.5(Quinlan 1993), CART(Brevnan et al 1984). Decision tree are considered to be one of the most popular approaches for representing classification [1] [15].

The decision tree algorithm is summarized as follow [5] [9]:

If all instance at the current node belong to the same category then create a leaf node if the corresponding class

else

{

Find the feature A that maximized the goodness measure

Make A the decision feature for the current node

for each possible value v of A

{

Add a new branch below node testing for $A = v$

Instances_v := subset of instance with $A = v$

if Instances_v is empty then add a leaf with label the most common value of Target_feature in Instance;

else

{

below the new branch add subtree

DT(Instances_v, Target_feature, Features - {A})

}

}

}



6. COMPARATIVE STUDY

Supervised learning algorithms are used in machine learning for training artificial neural networks. Our comparison subdivided into two sections; first section comparison based on various parameters derived from some theoretical studies showing in table 2 [2] [5] [9] [15] [21] [22] [23]. And second section comparison based on benefits and limitations showing in table 3 [14] [15] [17].

Table 2. Comparative study between different supervised algorithms

Algorithms Parameters	Back-propagation	Decision Tree	Perceptron
Proposed by	Rosenblatt/ 1962	Quinlan	Rosenblatt/ 1958
Accuracy in general	Very good	Good	Good
Speed of learning	Average	Very good	Average
Speed of classification	Excellent	Excellent	Excellent
Tolerance to missing values	Average	Very good	Average
Tolerance to irrelevant attributes	Average	Very good	Average
Tolerance to redundant attributes	Good	Good	Good
Tolerance to highly interdependent attributes	Very good	Good	Very good
Dealing with discrete/binary/ continuous attributes	Not discrete	All	Not discrete
Tolerance to noise	Good	Good	Good
Dealing with danger of overfitting	Average	Good	Good
Attempts for incremental learning	Good	Good	Very good
Explanation ability/ transparency of knowledge/classification	Average	Excellent	Average
Support Mlticlassification	Naturally extended	Excellent	Naturally extended

Table 3. Comparative study by benefit and limitation between different supervised algorithms

	Benefit	Limitation
Decision Tree	<ol style="list-style-type: none"> 1. Very simple and fast. 2. Can handle high dimensional data. 3. Representation is easy to understand. 4. Have good accuracy. 5. It Support incremental learning. 	<ol style="list-style-type: none"> 1. It has long training time. 2. Lack of available memory, when dealing with large databases. 3. The resulting regions after partitioning are all hyper rectangles. 4. Most algorithms cannot perform well with problems that require diagonal partitioning. 5. More complex representation.
Back-propagation	<ol style="list-style-type: none"> 1. Able to tolerate noisy data. 2. They can be used when we have the little knowledge of the relationship between attributes and classes. 3. Well suited for continuous valued inputs and outputs. 4. Successful on several real world applications like handwritten character recognition, pathology and laboratory medicine. 	<ol style="list-style-type: none"> 1. Long learning time therefore more suitable for application where this is feasible. 2. Poor interpretability as knowledge is represented. 3. Require number of parameters that are to be determined empirically.
Perceptron	<ol style="list-style-type: none"> 1. Noisy or incomplete inputs will be classified because of their similarity with pure and complete inputs. 2. Relearning after damage can be relatively quick. 	<ol style="list-style-type: none"> 1. Computationally expensive learning process. 2. No guaranteed solution. 3. Scaling problem.



7. CONCLUSION

In this paper we performed a comparative study on various algorithms of supervised learning and we have found the decision tree algorithm is the best as compared with other algorithms that can solve the complex problems with a remarkable speed. As for the back propagation algorithm is the most used algorithm for learning in neural networks especially for multi layers perceptron's (MLPs) and it is very good algorithm for solving simple problems and able to tolerate noisy data. While perceptron algorithm can deal with noisy or incomplete inputs because of their similarity with pure and complete inputs. The benefit of perceptron algorithm is its low computational requirement and its ability to guarantee convergence to a solution for linearly separable problem and Relearning after damage can be relatively quick.

8. REFERENCE

- [1] Krenker, A., Bešter, J., and Kos, A., Introduction to the Artificial Neural Networks, Artificial Neural Networks - Methodological Advances and Biomedical Applications, Prof. Kenji Suzuki (Ed.), ISBN: 978-953-307-243-2, InTech, 2011, DOI: [10.5772/15751](https://doi.org/10.5772/15751).
- [2] Blum, A., Empirical support for Winnow and weighted-majority based algorithms: results on a calendar scheduling domain, Machine Learning, Volume 26, Issue 1, pp. 5-23, 1997 Kluwer Academic Publishers, Boston, Manufactured in The Netherlands, 1997, DOI: [10.1023/A:1007335615132](https://doi.org/10.1023/A:1007335615132).
- [3] Gustav, B., and Linus, K., supervised learning in artificial neural network, A brief introduction, www.idt.mdh.se/kurser/ct3340/archives/ht08/papersRM08/10.pdf.
- [4] İçer, S., Kara, S., and Güven, A., Comparison of multilayer perceptron training algorithm for portal venous Doppler signals in the cirrhosis disease, Expert Systems with Applications, Volume 31, Issue 2, pp. 406-413, ISSN 0957-4174, August 2006.
- [5] Kotsiantis, S. B., Zaharakis, I. D., and Pintelas, P. E., Machine learning: a review of classification and combining techniques, Artificial Intelligence Review, Volume 26, Issue 3, pp. 159-190, USA, November 2006, DOI: [10.1007/s10462-007-9052-3](https://doi.org/10.1007/s10462-007-9052-3).
- [6] Magules, G.D., and Vrahatis, M.N., adaptive algorithms for neural networks supervised learning : a deterministic optimization approach, Int. J. Bifurcation Chaos, Volume 16, Issue 07, July 2006, DOI: [10.1142/S0218127406015805](https://doi.org/10.1142/S0218127406015805).
- [7] Gacquer, D., Delcroix, V., Delmotte, F., Piechowiak, S., comparative study of supervised classification algorithm for the detection of atmospheric pollution, Engineering Applications of Artificial Intelligence, Volume 24, Issue 6, pp. 1070-1083, ISSN 0952-1976, September 2011.
- [8] Chen, H., Marny, M.T., and Chaudresekaran, H., A neural network training algorithm utilizing multiple sets of linear equation, Neurocomputing, Volume 25, Issues 1-3, pp. 55-72, ISSN 0925-2312, April 1999.
- [9] Kotsiantis, S. B., Supervised machine learning: a review of classification techniques, Informatica 31, pp. 249-268, 2007.
- [10] Liu, Q., and Wu, Y., supervised learning, Web Data Mining, Data-Centric Systems and Applications, pp. 63-132, 2011, DOI: [10.1007/978-3-642-19460-3_3](https://doi.org/10.1007/978-3-642-19460-3_3).
- [11] Diene, O., and Bhaya, A., perceptron training algorithm designed using discrete-time control Liapunov functions, Intelligent Control, ISIC 2007. IEEE 22nd International Symposium on , vol., no., pp.608-613, 1-3 Oct. 2007, DOI: [10.1109/ISIC.2007.4450955](https://doi.org/10.1109/ISIC.2007.4450955).
- [12] Maimon, O., and Rokach, L., Data Mining by Attribute Decomposition with semiconductors manufacturing case study, in D. Bracha, Editor, Data Mining for Design and Manufacturing: Methods and Applications, pp 311-336, Kluwer Academic Publishers, 2001, DOI: [10.1007/978-1-4757-4911-3_13](https://doi.org/10.1007/978-1-4757-4911-3_13).
- [13] Livieris, I. E., and Pintelas, P., A survey on algorithms for training artificial neural networks, Technical Report TR08-01, Department of Mathematics, University of Patras, 2008.
- [14] Jain, A. K., Mao, J., and Mohiuddin, K.M., Artificial Neural Networks: A Tutorial, Computer, pp. 31-44, 1996.
- [15] Bharsar, H., and Ganatra, A., A comparative study of training algorithm for supervised machine learning, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, September 2012.
- [16] Sharma, K., and Rajpoot, M., comparative study of supervised learning algorithm for sensor data, International Journal of Advanced Technology & Engineering Research (IJATER), Volume 2, Issue 4, July 2012.
- [17] Riedmiller, M., Machine Learning: Multilayer Perceptrons, available online : https://www.google.co.in/search?newwindow=1&q=ml.informatik.uni-freiburg.de%2F_media%2Fteaching%2Fss10%2F05_mlps.printer.pdf%E2%80%8E&oq=ml.informatik.uni-freiburg.de%2F_media%2Fteaching%2Fss10%2F05_mlps.printer.pdf%E2%80%8E&gs_l=serp.12...77260.77260.0.78848.1.1.0.0.0.301.301.3-1.1.0...0.0.1c.1.17.serp.t9fWz3Rv-_1.
- [18] Pfeifer, R., Damain, D., and Fuchslin, R., Neural Networks, available online : <https://www.google.co.in/search?newwindow=1&q=R.+Pfeifer,+DO.+Domain+and+R.+F%C3%BCchslin+Neural+Net+works&spell=1&sa=X&ei=86HjUduzK4WFrAft1YGgDw&ved=0CCsQBSgA&biw=1066&bih=523>.



- [19] Han, J., and Kamker, M., Data Mining: Concepts and Techniques, Second Edition, Elsevier, 2011.
- [20] Rojas, R., Neural network a systematic introduction, Springer-Verlag, Berlin, Germany, 1996.
- [21] Dietterich, T. G., : An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization”, Machine Learning, 1999 Kluwer Academic Publishers, Boston, Manufactured, Netherlands, pp.1-22, 1999.
- [22] Banfield, R.E., Hall, L.O., Bowyer, K.W., and Kegelmeyer, W.P., A Comparison of Decision Tree Ensemble Creation Techniques, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol. 29 (1) , no.1, pp.173,180, Jan. 2007, doi: 10.1109/TPAMI.2007.250609.
- [23] Bouckaert, R. R., Choosing between Two Learning Algorithms Based on Calibrated Tests, In ICML'03, Washington DC, 2003.

Biography Authors:



Dr. Rafiqul Zaman Khan: Dr. Rafiqul Zaman Khan, is presently working as a Associate Professor in the Department of Computer Science at Aligarh Muslim University, Aligarh, India. He received his B.Sc Degree from M.J.P Rohilkhand University, Bareilly, M.Sc and M.C.A from A.M.U. and Ph.D (Computer Science) from Jamia Hamdard University. He has 18 years of Teaching Experience of various reputed International and National Universities viz King Fahad University of Petroleum & Minerals (KFUPM), K.S.A, Ittihad University, U.A.E, Pune University, Jamia Hamdard University and AMU, Aligarh. He worked as a Head of the Department of Computer Science at Poona College, University of Pune. He also worked as a Chairman of the Department of Computer Science, AMU, Aligarh. His Research Interest includes Parallel & Distributed Computing, Gesture Recognition, Expert Systems and Artificial Intelligence. Presently 04 students are doing PhD under

his supervision. He has published about 35 research papers in International Journals/Conferences. Names of some Journals of repute in which recently his articles have been published are International Journal of Computer Applications (ISSN: 0975-8887), U.S.A, Journal of Computer and Information Science (ISSN: 1913-8989), Canada, International Journal of Human Computer Interaction (ISSN: 2180-1347), Malaysia, and Malaysian Journal of Computer Science(ISSN: 0127-9084), Malaysia. He is the Member of Advisory Board of International Journal of Emerging Technology and Advanced Engineering (IJETA), Editorial Board of International Journal of Advances in Engineering & Technology (IJAET), International Journal of Computer Science Engineering and Technology (IJCSET), International Journal in Foundations of Computer Science & technology (IJFCST) and Journal of Information Technology, and Organizations (JITO).



Haider Khalaf Jabbar Allamy: Haider allamy, is a research scholar in the Department of Computer Science; Aligarh Muslim University, Aligarh, India. He joined to his Ph.D course in 28-09-2012. His research interest includes Artificial Intelligence and Expert system. He did B.Sc Degree in computer science from University of Basrah, College of Science, Iraq, M.Sc in computer science from Jamia Hamdard University, Delhi, India. He is envoy by the University of Maysan/Iraq for complete his Ph.D course in Aligarh Muslim University, Aligarh, India.