Original Research Paper

Volume - 7 | Issue - 6 | June - 2017 | ISSN - 2249-555X | IF : 4.894 | IC Value : 79.96

Computer Science

Attribute Reduction to Enhance Classifier's Performance: a LD Case Study

Pooja Manghirmalani-Mishra Dept. of Computer Science, University of Mumbai, Mubai-98, India.

Sushil Kulkarni Dept. of Mathematics, Jai Hind College, Mumbai-20, India.

ABSTRACT The objective of this study is to reduce the overlapping attributes present in a data set and to impute missing values within the same dataset. For application purpose, a dataset of Learning Disability is taken where within Learning Disabilities; there are many different types as well as a variety of tests that may be done to diagnose the problem. This study proposes a Principle Component Analysis technique for dimensionality reduction along with Artificial Neural Network's Backpropagation algorithm to handle missing values of a dataset. This algorithm facilitates imputing the missing values in the preprocessing stage. The classification approach which is implemented by applying Winnow algorithm gives acceptable results that act as a tool for predicting the LD accurately amongst primary-grade school children.

KEYWORDS : Learning Disability, Missing values, Back-Propagation, Attribute Reduction, Principle Component Analysis, Classification, Winnow.

1. INTRODUCTION

Learning disability [LD] denotes to a neurological condition which disturbs an individual's ability to think and remember. It is established in disorders of listening, thinking, reading, writing, spelling or arithmetic [1]. These individuals are not attributed to medical, emotional or environmental causes despite having normal intellectual abilities [2].

Kirk [3] stated that, children with special learning disabilities exhibit a disorder in one or more of the basic psychological processes involved in understanding or in using spoken or written language. These may be manifested in disorders of listening, thinking, talking, reading, writing, spelling, or arithmetic. They do not include learning problems which are due primarily to visual, hearing, or motor handicaps, to mental retardation, emotional disturbance or to environmental deprivation.

For diagnosing LD, there does not exist a global method. Mostly detection is done using Wechsler Intelligence Scale for Children (WISC) test [4], conducted in the supervision of special educators and with the observation of parent and teachers. In this context, computational approach to detect LD is quite significant.

Computational prediction requires the data to be in a particular format. As this work deals with children, their mood swings and attitude needs to be considered while utilizing a live data. Keeping these circumstances in mind, data is bound to have missing values. And as already sated that there doesn't exists a single tool for LD diagnosis, a combinatory tool designed for the same purpose may have certain attributes which would overlap with others. Hence there falls a need to pre-process the data before applying any classifier.

This paper proposes a model for diagnosis and classification of LD. Section II of this paper explores in detail different computational methods and models applied for predicting LD. Having elaborately explored different approaches, we have found that there are still possible ways of approaching the given problem. Section III discusses the LD dataset. Section IV elaborates on the Data Pre-processing techniques used to impute missing values and to reduce the overlapping attributes Section V discusses the Winnow algorithm used for predicting LD. In Section VI, comparative results are shown and Section VII elaborates on the outcomes of the results whereas Section VIII discusses future objectives respectively.

2. TAXONOMY

Due to the hidden features of learning disabilities (LD), the classification or diagnosis of students with learning disabilities has always been a complex process. There is little agreement about what is the best procedure to diagnose a child with LD. Based on the survey done for this article; researchers have come across the following computational based methods which have been successful to diagnose LD at an early age. The literature discusses various including Artificial Intelligence and hybrid form of other techniques to diagnose LD.

Jain et al [5] proposed a simple Perceptron based artificial neural network (ANN) model for diagnosing LD using curriculum based test conducted by special educators. Bullinaria [6] applied a multi-layer feed forward Perceptron to diagnose dyslexia where letter strings were mapped to phoneme strings in multi- syllabic words. Wu et al [7] proved that multi-layer Perceptron with Backpropagation gave better results in diagnosing LD. They further attempted to diagnose LD using support vector machines (SVM) [8]. Salhi et al [9] used both wavelet transforms and ANN to diagnose LD from pathological voices. Novak et al [10] have calculated a set of features from signals of horizontal and vertical eye movement using self-organizing map and genetic algorithm (GA). They concluded that the reading speed increased with the probability of the patient being healthy.

Wu et al [11] later combined different feature selection algorithms like brute-force, greedy and GA along with ANN to improve the identification rate of LD. Georgopoulos et al [12] proposed that a hybridization of GA and fuzzy cognitive map was better equipped for accurate diagnosis. Macašet al [13] developed a system for extracting the features of eye movements from time and frequency domain. They concluded that Backpropagation based classification gave better results than that offered by Bayes and Kohonen network.

Manghirmalani et al proposed a soft computing technique called Learning Vector Quantization. The model classifies a child as learning disabled or non- learning disabled. Once diagnosed with learning disability, rule based approach is used further to classify them into types of learning disability that is dyslexia, dysgraphia and dyscalculia [14].

Manghirmalani et al further applied fuzzy logic to enhance the accuracy of LD classification [15]. They later improved large margin to accommodate the types of LD using semi-supervised learning algorithm (SVM) to obtain more accurate bifurcation of type of LD [16].

3. LD DATA SAMPLE

Mostly detection of LD is done using Wechsler Intelligence Scale for Children (WISC) test [17], conducted in the supervision of special educators and with the observation of parent and teachers.

The soft computing technique provides an alternative way to represent linguistic and subjective attributes of the real world in computing. It deals with the labeled and unlabeled data together. In this circumstance, computational approach to classify LD is quite significant.

3.1 Collection of Exhaustive Parameters

A curriculum-based test was designed with respect to the syllabus of

primary-level school going children. This test was conducted in schools for collecting LD datasets for testing. Historic data for LD cases were collected from LD Clinics of Government hospitals where the tests were conducted in real-time medical environments. The system was fed with 11 input units which correspond to 11 different sections of the curriculum-based test.

Table-1: Parameters of Curriculum-Based Test

Input Parameter	Weightage	Category of LD
Essay (ES)	10	Dysgraphia
Reading (RD)	10	Dyslexia
Comprehension (CP)	10	Dyslexia,Dysgraphia
Spelling (SP)	10	Dysgraphia
Perception (PP)	10	Dyslexia
Solve (SL)	10	Dysgraphia
Word Problem (WP)	10	Dyscalculia,Dyslexia
Mental Sums (MS)	10	Dyscalculia
Time (TM)	10	Dyscalculia
Calander (CL)	05	Dyscalculia
Money (MN)	05	Dyscalculia

Table-1 shows the initial 11 inputs corresponding to curriculum-based test. Dataset consists of 586 cases of LD children. The system was trained using 70% of the data items and the remaining was used to test the network [18, 19].

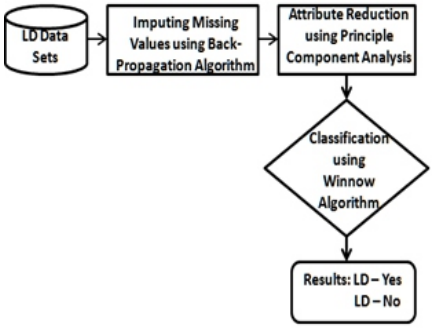


Figure 1: System flow chart

3.2 Proposed Methodology

The proposed methodology of this research work shown in the system flow chart given at Figure 1. Initially the data is pre-processed by imputing missing values followed by elimination of repetitive or overlapping attributes called as 'attribute reduction. Then this processed data is utilized for predicting the learning disability by applying Neural Network technique called as Winnow algorithm.

4. DATA PREPROCESSING

Before the data is used by prediction algorithm for classification, it has to be preprocessed in order to increase the classification accuracy and to facilitate the learning process of the neural network algorithm. This operation is crucial as neural network algorithms apply pattern matching on the data and hence this step makes the data more suitable for data mining [20]. Even for Data preprocessing, different processes exist in the preprocessing stage that are dimensionality reduction, feature subset selection, removal of noise from the data, and imputing the missing values [21, 22].

4.1 Imputing Missing Values

The aim of this study is to apply preprocess the data and make it more suitable for data mining. This study applies neural network model known as Backpropagation (BP) algorithm for imputing the missing values. Neural networks constitute a class of predictive modeling system that works by iterative parameter adjustment as stated by Chen [23].

4.2 Steps of BP Algorithm

Following are the steps of a BP algorithm used for classification or prediction:

i Randomly choose the initial weights
ii While error is too large
iii For each training pattern (presented in random order)
  a Apply the inputs to the network
  b Calculate the output for every neuron from the input layer, through the hidden layer(s), to the output layer
  c Calculate the error at the outputs
  d Use the output error to compute error signals for pre-output layers
  e Use the error signals to compute weight adjustments
  f Apply the weight adjustments
  g Periodically evaluate the network performance

4.3 Steps of BP Algorithm for Imputing Missing Values

Gupta & Lam [24] introduced following procedure for reconstruction of missing values using multilayered networks.

i Collect all training cases without any missing value and call them the complete set.
ii Collect all training and test cases with at least one missing value and call them the incomplete set.
iii For each pattern of missing values, construct a multi-layered network with the number of input nodes in the input layer equal to the number of non-missing attributes, and the number of output nodes in the output layer equal to the number of missing attributes. Each input node is used to accept one non-missing attribute, and each output node to represent one missing attribute.
iv Use the complete set and the Backpropagation algorithm to train each network constructed in Step III. Since the complete set does not have missing values, different patterns of input-output pairs can be obtained from the complete set to satisfy the input-output requirements for different networks from Step III. As the output of a network is between 0 and 1, data have to be converted to values between 0 and 1 for this reconstruction procedure. The trained networks from Step IV calculate the missing values in the incomplete set.
v To construct such a Neural Network, a strategy is applied where one must start with a simple network and add extra nodes to the network until such addition does not improve the network performance. The system is implemented using Java.

4.4 Algorithm

Input: ProblemSize, InputPatterns, $iterations_{max}$, $learning_{rate}$
Output: Network

Network ← ConstructNetworkLayers()

Networkweights ← InitializeWeights(Network, ProblemSize)
For (i = 1 to iterationsmax)

Patterni ← SelectInputPattern(InputPatterns)

Outputi ← ForwardPropagate(Patterni, Network)

BackwardPropagateError(Patterni, Outputi, Network, learningrate)
UpdateWeights (Patterni, Outputi, Network, learningrate,)

End
Return (Network)

4.5 Attribute Reduction Using PCA

Principal Component Analysis (PCA) is a dimensionality reduction method. The data to be reduced consists of tuples described by n dimensions which are called PCA [25]. The PCA investigates for *k* n-dimensional orthogonal vectors that can be used to represent the data where k B n. The basic procedure behind PCA is:

(i) Input data is normalized so that each attribute comes within the similar range. This helps ensure that attributes with large domains will not dominate attributes with smaller domains.

(ii) PCA computes k orthogonal vectors that provide a basis for the normalized input data. These vectors which are perpendicular to others are referred to as the principal components.

(iii) The principal components are sorted in the order of decreasing strength.

In this study, these overlapping attributes are detached by applying the PCA and the number of attributes is reduced to eight. Dimensionality reduction is achieved by choosing sufficient eigenvectors to account for the variance in the original data.

Data is then filtered by changing to the principal component space, eliminating the poor eigenvectors and reverting back to the original form. For every eigenvector has a corresponding eigenvalue. An eigenvector is direction and eigenvalue is magnitude denoting the amount of variance in the data in that direction. The eigenvector with the highest eigenvalue is therefore the principal component. The amount of eigenvectors and eigenvalues that exist always equals to the number of dimensions the data set has. In any PCA algorithm, eigenvectors are predictably arranged so that the one with the largest eigenvalue is first which is corresponding to the largest variance being first. PCA is implemented in the environment of a mining tool Weka.

4.6 Steps of Applying PCA [26]

i Organize the data set

ii Calculate the empirical mean

iii Calculate the deviations from the mean

iv Find the covariance matrix

v Find the eigenvectors and eigenvalues of the covariance matrix

vi Rearrange the eigenvectors and eigenvalues

vii Compute the cumulative energy content for each eigenvector

viii Select a subset of the eigenvectors as basis vectors

ix Convert the source data to z-scores

x Project the z-scores of the data onto the new basis

4.7 Reduced Attributes

Table 2. Rank Reduced Attributes

Selected attributes: 1,2,3,4,5,6,7,9 : 8		
No.	Attribute	Rank
1	ES	0.6182
2	RD	0.5645
3	MS	0.5221
4	CP	0.4619
5	WP	0.3703
6	SL	0.3211
7	PP	0.2918
8	TM	0.1654

Table 2 shows the reduced attribute list (from eleven to 8) along with their ranks. The ranking is by the size of the eigenvalue. So the top ranked attribute is the Principle Component with the highest eigenvalue. The ranking output doesn't tell anything about the relative goodness of the original attributes. It can be analyzed by examining the size of the coefficients in the Principle Components.

5. CLASSIFICATION OF LD

In the previous work [5], Perceptron algorithm was implemented on the same dataset using 11 attributes. Perceptron deals with learning problems by updating the weight function based on the history of mistakes done by the system. In particular, the weight function is updated additively:

$$\vec{w}_{t+1} = \vec{w}_t + \vec{y}_t$$

This work applies the Winnow algorithm, which is similar to Perceptron algorithm. However, Winnow algorithm updates the weight function multiplicatively [27]. Winnow algorithm maintains a weight vector $\vec{w}_t$. Let $w_{i,t}$= the weight on feature i at round t, and $x_{i,t}$ the i^{th} component of $\vec{x}_t$.

Suppose $y_t \in \{0,1\}, and\ \vec{x}_t \in \{0,1\}^n$ is a binary string.

Below are the steps of Winnow algorithm:

a. Initialize all the weights to one:

$$\mathbf{w_{1,1} = w_{2,1} = \cdots = w_{n,1} = 1}$$

b. For each example $\vec{x}_t$,

i. output *1, if* $\vec{w}_t \cdot \vec{x}_t \geq n$

ii. output *0, otherwise.*

c. If the algorithm makes a mistake,

i. *If* $y_t = 1$, then $\forall i$ such that $x_{i,t} = 1, w_{i,t+1} \leftarrow w_{i,t}\ (1 + \epsilon)$

ii. *If* $y_t = 0$, then $\forall i$ such that $x_{i,t} = 1, w_{i,t+1} \leftarrow \frac{w_{i,t}}{1+\epsilon}$

iii. *In both cases* $y_t = 1\ or\ 0$, $\forall\ x_{i,t} = 0, w_{i,1+1} \leftarrow w_{i,t}$

d. If there is no mistake, then

$$\mathbf{\vec{w}_{t+1} \leftarrow \vec{w}_t}$$

In the Winnow algorithm, Є is a parameter. If mistake is made on a positive example, increase the weights of the features for $x_{i,t}$ =1 . Similarly, if mistake is made on a negative example, decrease the weights of all the $x_{i,t}$ that contributes to:

$$\vec{w}_{i,t} \cdot \vec{x}_{i,t}$$

5.1 Training of the network

The system is trained using the training pair:

$\{x^1, d^1\}, \{x^2, d^2\} \ldots \{x^m, d^m\}$

where x^k is the k^{th} input vector and $d^k = \{1, 0\}$,
1 being 'Normal' and
0 being 'Learning Disabled'.

The system is trained by adjusting the weight vector to make the output y^k corresponding to the input x^k match with the desired output d^k. The system is implemented using Java.

5.2 Detection measures

The performance of the various classifiers may be presented in terms of accuracy, correctness and coverage:

$$\text{accuracy} = \frac{\text{No. of cases correctly classified}}{\text{Number of Cases}} \times 100\%$$

$$\text{correctness} = \frac{\text{No. of cases correctly classified}}{\text{Number of cases classified}} \times 100\%$$

$$\text{coverage} = \frac{\text{No. of cases classified}}{\text{Number of cases}} \times 100\%$$

Accuracy may be measured on the training set or on a test set. A high figure for training set accuracy does not mean that the performance of the classifier will be good in practice. In this paper, accuracy results on the test set since these give a better indication of how well the classifier is able to generalize on new examples.

6. RESULTS

This paper shows that the Winnow algorithm results when compared with Perceptron algorithm which is already implemented on the same data. Their results are compared in the tables 2 and 3 given below. Later the accuracy of the system when the dataset is preprocessed is given in Table 4.

Graph 1 represents the accuracy of the algorithm without attribute reduction and considering only those entries whose data is complete (i.e. removing the missing value data) where as Graph 2 represents the accuracy of the algorithm where the dataset is with imputed data in case of missing values and post-attribute reduction.

Table 3 represents the accuracy measures of a Perceptron Algorithm implemented on the same data set [5]. Table 4 represents the implementation of Winnow algorithm which is 1st made to run on the same data which is not pre-processed. Table 5 shows the accuracy of Winnow algorithm on the processed data.

Table 3: System Accuracy of Perceptron Algorithm [5]

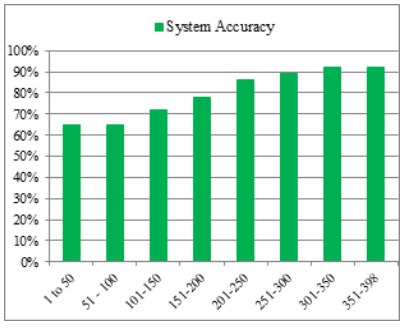
Detection Measure	Percentage
Accuracy	90%
Correctness	80%
Coverage	80%

Table 4: System Accuracy of Winnow Algorithm before Data Pre-Processing

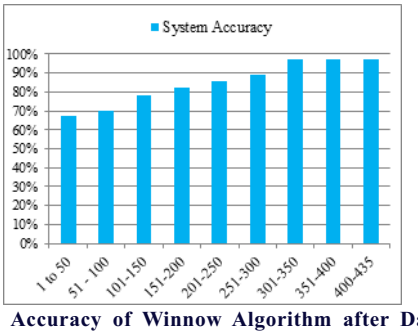
Detection Measure	Percentage
Accuracy	93%
Correctness	89%
Coverage	88%

Table 5: System Accuracy of Winnow Algorithm after Data Pre-Processing

Detection Measure	Percentage
Accuracy	97%
Correctness	96%
Coverage	96%



Graph 1. Accuracy measures of Winnow Algorithm without Data Pre-processing.



Graph 2. Accuracy of Winnow Algorithm after Data Pre-processing.

7. CONCLUSION

This paper shows that the Winnow algorithm updates faster than Perceptron algorithm and Winnow algorithm will make fewer mistakes than the Perceptron algorithm when the number of relevant feature is considerably less than the total number of features. Even a minor increment in the accuracy of system where the data is of sensitive nature plays an important role.

The goal is yet to reach to an accuracy level such that a system can very precisely diagnose a child from LD with minimum processing time and computing errors. Such a system will then at large help a lot of people including children facing the LD problem, their parents, their teachers and the doctors under whose medical supervision these children are.

8. DISCUSSIONS

After never-ending discussions with doctors, special educators, teachers and parents of children with LD, the process of diagnosing LD in India could understood and monitored. The traditional procedure for diagnosing LD is very burdensome and time consuming. The system doesn't allow the commencement of remedial education for these children unless diagnosis is not complete. Due to this, children find it difficult to adjust with the regular education pattern.

This paper demonstrate an attempt to accurately classify LD by first applying data imputing and attribute reduction technique and later using that data for LD classification. This method is not only simple and easy to replicate in huge volumes but gives good results based on accepted benchmarks. However, there is scope for further enhancement of system by finding a combination of algorithms so as to build up a model that is satisfactorily more accurate.

This study investigates the possibility of parameter classification in order to distinguish irrelevant and superfluous variables which lead to decrease in diagnosis process time and increase in accuracy. This preprocessing method is beneficial for the special educators, doctors and teachers by providing suggestions that lead to the elimination of superfluous tests and remarkably reducing of time needed for diagnosing LD.

On the whole, the focus of this research is to identify early diagnosis and proper classification of LD and to support special education community in their quest to be with mainstream.

9. REFERENCES

[1]. Lisa L. Weyandt; The Physiological Bases of Cognitive and Behavioural Disorders, Routledge: Blausen Medical Communications, 1st Edition, United States, 2005.
[2]. Lerner, Janet W, Learning Disabilities: Theories, Diagnosis, and Teaching Strategies; Houghton Mifflin (T); 6th edition; 1993
[3]. S.A Kirk, Educating Exceptional Children Book, Wadsworth Publishing, 14th Edition, 2014.
[4]. Kaplan, Robert M.; Saccuzzo, Dennis P., Psychological Testing: Principles, Applications, and Issues, Belmont (CA): Wadsworth. pp. 262, 7th Edition, 2009
[5]. Kavita Jain, Pooja Manghirmalani, Jyotshna Dongardive, Siby Abraham, Computational Diagnosis of Learning Disability, International Journal of Recent Trends in Engineering, pp. 64-66, Vol. 2, No. 3, 2009.
[6]. John A Bullinaria, Neural Network Models of Reading Multi- Syllabic Words, International Joint Conference on Neural Networks, pp. 283-286, 1993.
[7]. Tung-Kuang Wu, Ying-Ru Meng, Shian-Chang Huang, Application of ANN to the identification of students with LD,ICAI, pp.162-168. 9, 2006.
[8]. Tung-Kuang Wu, Ying-Ru Meng and Shian-Chang Huang, Identifying & Diagnosing Students with LD Using ANN & SVM, IEEE International Joint Conference on Neural Networks, pp. 4387-4394, Vancouver, BC, 2006.
[9]. L. Salhi, M. Talbi, A. Cherif, Voice Disorders Identification using Hybrid Approach: Wavelet Analysis and Multilayer Neural Networks, Proceedings of World Academy of Science, Engineering and Technology Volume 35, pp. 3003-3012, 2008.
[10]. D. Novak, P. Kordk, M. Macas, M. Vyhnalek, R. Brzezny, L. Lhotska, School Children Dyslexia Analysis using Self-Organizing Maps, IEEE 26th Annual International Conference, pp. 1-4 , San Francisco, 2004.
[11]. Tung-Kuang Wu, Ying-Ru Meng and Shian-Chang Huang, Effects of Feature Selection on the Identification of Students with LD Using ANN, Springer-Verlag Heidelberg, pp. 565-574 2006.
[12]. Voula Georgopoulos and Chrysostomos Stylios, Genetic Algorithm enhanced fuzzy cognitive maps for medical diagnosis, IEEE International conference on Fuzzy Systems, pp. 2123-2128, Hong Kong, 2008.
[13]. Martin Macaš, Lenka Lhotská and Daniel Novák, Bio-inspired methods for analysis and classification of reading eye movements of dyslexic children , Department of Cybernetics, Czech Technical University in Prague, Czech Republic NiSIs Symposium, pp. 1-5, 2005.
[14]. Pooja Manghirmalani, Zenobia Panthaky, Kavita Jain; Learning Disability Diagnosis and Classification-A Soft Computing Approach; IEEE World Congress on Information and Communication Technologies (WICT), pp. 483-488, Mumbai, 2011.
[15]. Pooja Manghirmalani, Darshana More, Kavita Jain; A Fuzzy Approach to Classify Learning Disability; International Journal of Advanced Research in Artificial Intelligence, pp. 1-7, 2012.
[16]. Pooja Manghirmalani, Sushil Kulkarni, Classification Of Data Using Semi-Supervised Learning (A Learning Disability Case Study), International Journal of Computer Engineering and Technology, pp. 432-440, 2013.
[17]. Kavita Jain, Pooja Manghirmalani Mishra, Sushil Kulkarni. A Neuro-Fuzzy System to Diagnose Learning Disability; IEEE International Conference on Radar, Communication and Computing, pp. 645-657, Chennai, 2012
[18]. Pooja Manghirmalani, Sushil Kulkarni, Developing Prognosis Tools To Identify LD In Children Using Machine Learning Technologies, National Conference on Spectrum of Research Perspectives, pp. 87-95, Mumbai, 2013
[19]. Pooja Manghirmalani Mishra, Sunita Magre Sushil Kulkarni, A Computational Based study for Diagnosing Learning Disability amongst Primary Students, 5th National Conference on Revisiting Teacher Education, pp. 20-26, Mumbai, 2015
[20]. U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, The KDD process for extracting useful knowledge from volumes of data, Communications of the ACM, vol. 39, no. 11, pp. 27-34, 1996.
[21]. Han J, Kamber M, Data mining: concepts and techniques, 2nd edition, Morgan Kaufmann, Elsevier Publishers. 2008.
[22]. D. Tomar and S. A. Agarwal, Survey on Data Mining approaches for Healthcare, International Journal of Bio-Science and Bio-Technology, vol. 5, no. 5, pp. 241-266. 2013
[23]. Chen, Z., Data Mining and Uncertain Reasoning: An Integrated Approach. Wiley, 2001
[24]. Gupta, A. & Lam, M., The weight decay Backpropagation for generalizations with missing values, Annals of Operations Research 78, pp. 165-187, 1998
[25]. Jolliffe I.T., Principal Component Analysis, Springer Series in Statistics, Springer-Verlag, Berlin, 2002.
[26]. Jolliffe I.T. Principal Component Analysis Series: Springer Series in Statistics 2nd edition, Springer, NY, 2002,
[27]. M. Schmitt, Identification criteria and lower bounds for Perceptron-like learning rules, Neural Computations edition 10, pp. 235–250, 1998