ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

Ensembled PCA-ELM for the Prediction of Diseases in Rice Leaves

Gayatri Parasa

Department of Computer Science & Engineering, Annamalai University, Tamilnadu.

gayathriparasa20@gmail.com

M. Arulselvi

Department of Computer Science & Engineering, Annamalai University, Tamilnadu.

marulcse.au@gmail.com

Abstract:

The conventional wisdom acknowledges the universal approximation capabilities of single-hidden-layer feedforward networks (SLFNs) employing additive models. Nevertheless, the efficiency of training such models was notably sluggish until the advent of the extreme learning machine (ELM) introduced by Huang et al. Pre-ELM, gradient-based algorithms were the go-to for efficiently training SLFNs, necessitating iterative application until a satisfactory model was achieved. This slow convergence hindered the widespread adoption of SLFNs, despite their generally commendable performance. The introduction of ELM transformed SLFNs into a viable choice for rapidly classifying a large number of patterns. Traditionally, hidden nodes were randomly initiated and fine-tuned (though not universally in all approaches). This paper proposes a deterministic algorithm that initiates each hidden node with an additive activation function, optimizing it for training with ELM. The algorithm leverages information obtained from principal components analysis to tailor the hidden nodes, significantly reducing computational costs compared to subsequent ELM improvements while surpassing their performance.

Keywords: Principal Component Analysis, Extreme Learning Machine, Classification

Introduction:

Artificial neural networks (ANNs) primarily revolve around single-hidden-layer feedforward networks (SLFNs). SLFNs, characterized by lacking side or back connections between nodes, have prompted the development of numerous training techniques for adjusting their parameters

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

and structure. Among the well-known architectures is the multilayer perceptron (MLP), typically consisting of sigmoid nodes and commonly trained using the backpropagation algorithm (BP). While this architecture can be trained using various algorithms, the most prevalent ones are categorized as either gradient-based or heuristic. Both types share common challenges, including difficulty in handling large datasets and sluggish convergence under such circumstances. These characteristics contribute to the slow construction of SLFNs, resulting from the need to adjust numerous parameters using time-consuming algorithms that must be iteratively applied to attain a suitable model. Consequently, despite their overall commendable performance, SLFNs are not as widely used as they could be. The extreme learning machine (ELM), an algorithm that significantly reduces the traditional computational time required for SLFN training using gradient-based approaches. ELM streamlines the training process into two steps: randomly configuring the hidden layer and fitting a linear combination using the Moore-Penrose generalized inverse matrix. This algorithm's rapidity and validated performance make it a noteworthy advancement. Following the introduction of ELM, various approaches aimed to enhance its original version's performance, particularly in selecting the number of hidden nodes and rapidly fitting their parameters. Incremental extreme learning machine (I-ELM) variants which randomly initializes hidden node weights and ranks features resulting from applying hidden node transformations on the training set.

Building on these advancements, this paper introduces the robust principal component analysis extreme learning machine (PCA-ELM) algorithm. Unique features include deterministically determining the number of hidden nodes and their weights based on information retrieved from a PCA analysis of the training set. Notably, the proposed algorithm is distinct from applying the original ELM over covariates obtained from a PCA analysis. In the realm of high dimensionality, where the curse of dimensionality poses a challenge, the objective of Principal Component Analysis (PCA) is to mitigate this issue by diminishing the dimensionality of the data while retaining a substantial portion of the original dataset's variation. PCA facilitates the computation of a linear transformation that effectively maps data from a space with high dimensions to one with fewer dimensions. The goal is to minimize the

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

following expression to ensure the preservation of as much information as possible:

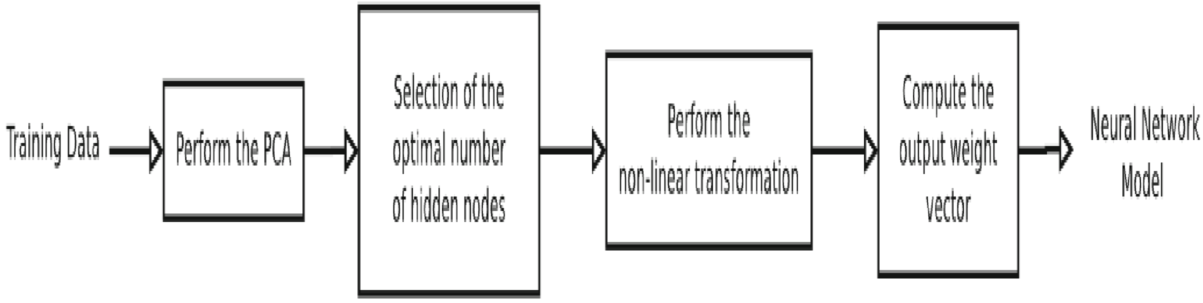

Fig 1.1: Working process of Ensembled PCA-ELM

The PCA method lends itself to a geometric interpretation by projecting data along directions where the data exhibits the most variation. These directions are determined by the eigenvectors of the covariance matrix corresponding to the largest eigenvalues. The magnitude of these eigenvalues reflects the variance of the data along the associated eigenvector directions. Key properties of PCA include:

- The new variables resulting from PCA are uncorrelated.
- The covariance matrix encapsulates only second-order statistics among vector values.
- As the new variables are linear combinations of the original variables, interpreting their meaning can be challenging.

It's essential to recognize certain assumptions underlying PCA:

- Linearity: Patterns are assumed to be linear combinations of a basis. Non-linear methods like kernel PCA address this assumption.
- Principal components with larger associated variances signify meaningful structures, while those with lower variances denote noise.
- The principal components are orthogonal, allowing PCA to be solvable using linear algebra decomposition techniques such as singular value decomposition (SVD).

The original ELM algorithm advocates for and validates the use of neural networks with sigmoid nodes in a randomly fitted hidden layer, transforming the feature space into a new one. In contrast, PCA represents an orthogonal transformation of initial axes into new ones,

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

maximizing variance. To illustrate PCA analysis, consider two well-defined clusters in a 2D space. These clusters represent two classes in a classification problem, with patterns near the cluster boundaries overlapping in their projections over the two axes. The initial axes can be rotated using PCA, resulting in two new clusters. This rotation optimally captures the variance in the data and aids in the differentiation of patterns, particularly those near the cluster boundaries.

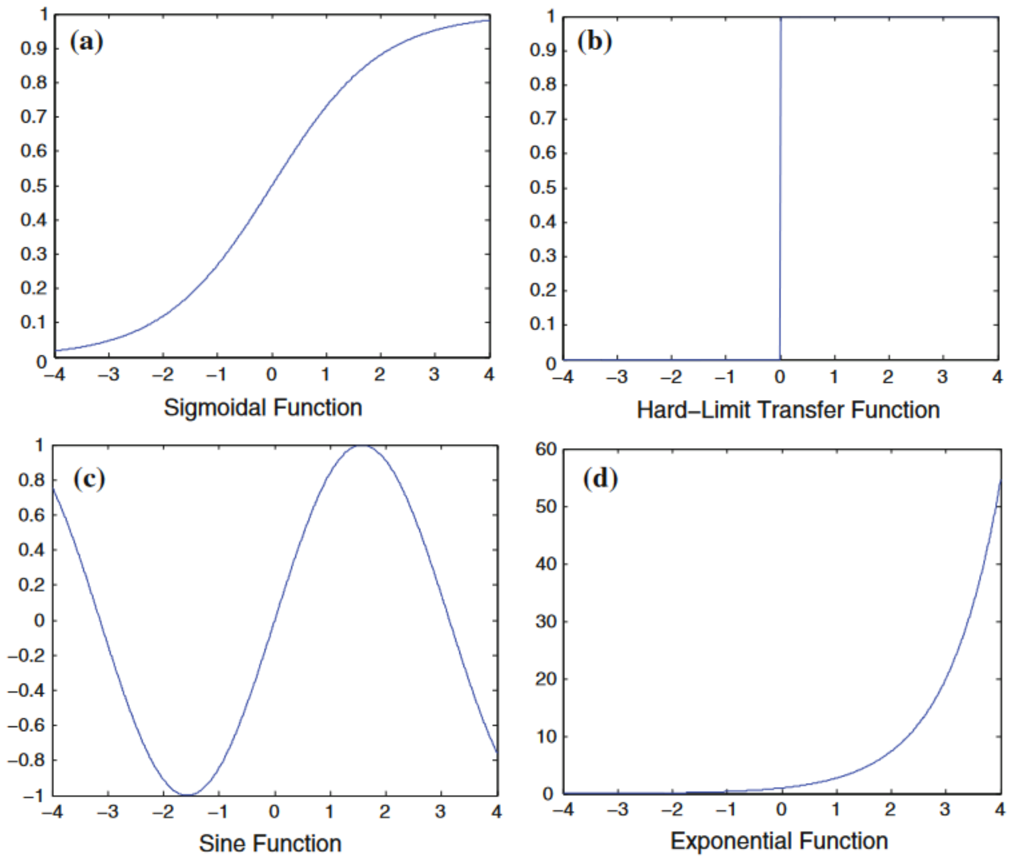

Fig 1.2: Graph of different activation functions used for ensembled procedure.

The proposed PCA-ELM method is subjected to a comparative analysis with other ELM algorithms employing various artificial neural network (ANN) models. Specifically, our approach is contrasted with the following:

Original Extreme Learning Machine (ELM):

For ELM (RBF), centers are randomly selected from data points, and widths are randomly drawn between the 20th and 80th percentiles of the distance distribution of the input space. ELM differs from ELM with the Radial basis function by measuring the distance of each

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

pattern to its centroid and weighting the final output by its radius. The number of nodes (m) in the hidden layer is determined through cross-validation on the training set, exploring values.

Optimally Pruned Extreme Learning Machine (OP-ELM):

In OP-ELM (RBF), centre and width values are initialized similarly to the ELM (RBF) algorithm. The number of nodes (m) in the hidden layer is fixed at 100, as OP-ELM automatically prunes useless neurons.

Conclusion:

The PCA-ELM algorithm, which we propose, stands out as a swift and resilient ELM-based methodology. Our innovation involves estimating hidden node parameters by leveraging information extracted from PCA applied to the training set, and the determination of output node parameters is achieved using the Moore-Penrose generalized inverse. Through rigorous experimentation conducted on fifteen widely recognized datasets, our algorithm has been thoroughly validated. The obtained results underwent meticulous statistical scrutiny employing Bonferroni–Dunn, Nemenyi, and Friedman tests. This rigorous statistical analysis unequivocally demonstrates the superior performance of our approach compared to previous methodologies. Notably, our algorithm introduces a crucial enhancement by eliminating the random initiation of hidden neurons, contributing to its robustness and reliability.

References:

1. Asuncion A, Newman D (2007) UCI machine learning repository. http://www.ics.uci.edu/~mlearn/ MLRepository.html. Accessed 8 Sept 2007

2. Cao J, Lin Z, Huang G (2011) Composite function wavelet neural networks with differential evolution and extreme learning machine. Neural Process Lett 33(3):251–265

3. Chen L, Zhou L, Pung HK (2008) Universal approximation and qos violation application of extreme learning machine. Neural Process Lett 28(2):81–95

4. Dunn OJ (1961) Multiple comparisons among means. J Am Stat Assoc 56:52–56

5. Feng G, Huang GB, Lin Q, Gay R (2009) Error minimized extreme learning machine with growth of hidden nodes and incremental learning. IEEE Trans Neural Netw 20(8):1352–1357

6. Friedman M (1940) A comparison of alternative tests of significance for the problem of m rankings. Ann Math Stat 11(1):86–92

7. Hochberg Y, Tamhane A (1987) Multiple comparison procedures. Wiley, New York

ISSN PRINT 2319 1775 Online 2320 7876

  UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

8. Huang GB, Chen L (2007) Convex incremental extreme learning machine. Neurocomputing 70(16–18):3056–3062

9. Huang GB, Chen L (2008) Enhanced random search based incremental extreme learning machine. Neurocomputing 71(16–18):3460–3468

10. Huang GB, Chen L, Siew CK (2006) Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans Neural Netw 17:4

11. Huang GB, Li MB, Chen L, Siew CK (2008) Incremental extreme learning machine with fully complex hidden nodes. Neurocomputing 71(4–6):576–583

12. Huang GB, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern B 42(2):513–529

13. Huang GB, Zhu Q, Siew C (2006) Extreme learning machine: theory and applications. Neurocomputing70(1–3):489–501

14. Huang GB, Zhu QY, Siew CK (2004) Extreme learning machine: a new learning scheme of feedforward neural networks. IEEE Int Conf Neural Netw Conf Proc 2:985–990

15. Jaeger H (2001) The “echo state” approach to analysing and training recurrent neural networks. GMD Rep 148:1435–2702

16. Kim J, Shin H, Lee Y, Lee M (2007) Algorithm for classifying arrhythmia using extreme learning machine and principal component analysis. In: 29th Annual international conference of the IEEE, engineering in medicine and biology society, 2007. EMBS, New York, pp 3257–3260

17. Mich Y, Sorjamaa A, Bas P, Simula O, Jutten C, Lendasse A (2010) OP-ELM: optimally pruned extreme learning machine. IEEE Trans Neural Netw 21(1):158–162

18. Miche Y, Sorjamaa A, Lendasse A (2008) Op-elm: theory, experiments and a toolbox. In: Artificial neural networks—ICANN 2008, lecture notes in computer science, vol 5163. Springer, Berlin, pp 145–154

19. Rong HJ, Ong YS, Tan AH, Zhu Z (2008) A fast pruned-extreme learning machine for classification problem. Neurocomputing 72(1–3):359–366

20. Sánchez-Monedero J, Gutiérrez PA, Fernández-Navarro F, Hervás-Martínez C (2011) Weighting efficient accuracy and minimum sensitivity for evolving multi-class classifiers. Neural Process Lett 34(2):101–116

21. Schlkopf B, Smola AJ, Müller KR (1999) Kernel principal component analysis. In: Advances in kernel methods: support vector learning. MIT Press, Cambridge, pp 327–352