Advances in Environmental Biology, 8(12) July 2014, Pages: 63-66

AENSI Journals

Advances in Environmental Biology

ISSN-1995-0756 EISSN-1998-1066

Journal home page: http://www.aensiweb.com/aeb.html

Nonlinear Principal Component Analysis (NLPCA) for Analysts' Equity Clustering

[1]Esmaeil Kamali Rad, [1]Saber Maleki[2], [3]Zinat Ansari

[1]*Professor in Department of Accounting, Science & Research Branch of Bushehr, Islamic Azad University, Bushehr, Iran*
[2]*Department of Accounting, Science & Research Branch of Bushehr, Islamic Azad University, Bushehr, Iran*
[3]*Department of Accounting, Shiraz Branch, Islamic Azad University, Shiraz, Iran*

ARTICLE INFO

Article history:
Received 15 April 2014
Received in revised form 22 May 2014
Accepted 25 May 2014
Available online 15 June 2014

Key words:
NLPCA, clustering, equity

ABSTRACT

Nonlinear principal (NLPCA) is used to identify and remove correlations among problem variables as an aid to dimensionality reduction, visualization, and exploratory data analysis. NLPCA uncovers both linear and nonlinear correlations, without restriction on the character of the nonlinearities present in the data. NLPCA operates by training a feed forward neural network to perform the identity mapping, where the network inputs are reproduced at the output layer. In the research were used 11 inputs that involve Cash‹ Short-Term Investments‹ Notes Receivable, Inventory‹ Spare Parts, Inventory Stock and Other Inventory, Advance Payment, Long-Term Assets, Notes Payable, Prepaid, Long-Term Liability that applied for clustering of equity by NLPCA method. The results show that there are 8 classes for the data.

To Cite This Article: Esmaeil Kamali Rad, Saber Maleki2, Zinat Ansari, Nonlinear Principal Component Analysis (NLPCA) for Analysts' Equity Clustering. ***Adv. Environ. Biol., 8(12),*** *63-66, 2014*

INTRODUCTION

The nonlinear principal component analysis (NLPCA) is a general purpose feature extraction algorithm producing features that retain the maximum possible amount of information from the original data set, for a given degree of data compression [1]. The coefficients of the linear transformation are such that if the feature transformation is applied to the data set and then reversed, there will be a minimum sum of squares difference between the original and reconstructed data [2]. The same criterion of optimality is adopted in NLPCA. If nonlinear correlations between variables exist, NLPCA will describe the data with greater accuracy and/or by fewer factors, provided that there are sufficient data to support the formulation of more complex mapping functions [3]. Artificial neural networks are not reviewed here; the reader is referred to Rumelhart *et al.* [4], Hoskins and Himmelblau [5], or Venkatasubramanian and Chan [6] for background information. The NLPCA method uses artificial neural network (ANN) training procedures to generate nonlinear features. The networks are of a conventional type, featuring feed forward connections and linear or sigmoidal nodal transfer functions, trained by back propagation. The particular network architecture used employs three hidden layers, including an internal "bottleneck" layer of smaller dimension than either input or output. The network is trained to perform the identity mapping, where the input is approximated at the output layer. Since there are fewer units in the bottleneck layer than the output, the bottleneck nodes must represent or encode the information in the inputs for the subsequent layers to reconstruct the input [7]. The concept of using a neural network with a bottleneck to concentrate information has been previously discussed in the context of "encoder/ decoder" problems [4, 7, 8]. In the paper we used 11 inputs that involve Cash ‹Short-Term Investments ‹Notes Receivable, Inventory‹ Spare Parts ‹Inventory Stock and Other Inventory‹ Advance Payment, Long-Term Assets, Notes Payable‹ Prepaid ‹Long-Term Liability for clustering of equity.

MATERIAL AND METHOD

Material:

In the study area used 12 charactristics that is following:

Corresponding Author: Saber Maleki, Department of Accounting, Science & Research Branch of Bushehr, Islamic Azad University, Bushehr, Iran,
Tel:+989144190785, E-mail: s.maleki@mehrpetrochem.com

Advances in Environmental Biology, 8(12) July 2014, Pages: 63-66

Table 1: Input data.

Elements	Notes receivable	Inventory	Inventory stock and other inventory	Advance payment	Long-term assets	Notes payable	prepaid	spare parts	Cash	Long-term liability	Short-term investments	equity
Maximum	5253206	688701	2542277	2521124	17363330	5695291	3139402	9726510	1182705	3001470	3564611	5253206
Minimum	3885	0	0	39	54030	16278	3068	0	900	0	-3868050	3885
Average	1261087	120072	636372	326940	4689123	1478423	704671	1653372	214590	368339	270927	1261087
STDEV	1399529	163515	623930	655760	4879249	1763770	791798	2447987	308493	759892	1794244	1399529

Method:

NLPCA generalizes the principal components from straight lines to curves (nonlinear). Thus, the subspace in the original data space which is described by all nonlinear components is also curved [9].

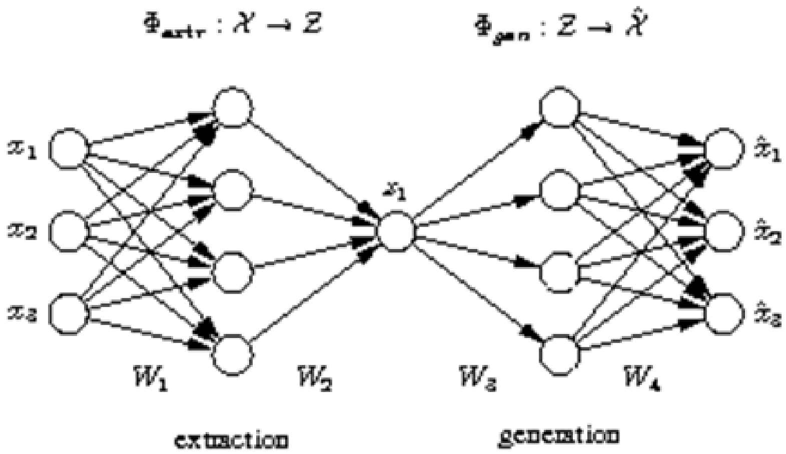

Fig. 1: Auto-associative neural network (Autoencoder) [9]

Nonlinear PCA (NLPCA) is used to approximate this trajectory by a curve referred to as nonlinear component. Which, in order to analysis cyclic phenomena, must be a closed curve hence a circular component. Here, a neural network with circular units is used to generate circular components [10].

Fig. 2: Nonlinear PCA [9]

RESULTS AND DISCUSSION

The results of the recearch is show in Figure 3 and Figure 4.

Advances in Environmental Biology, 8(12) July 2014, Pages: 63-66

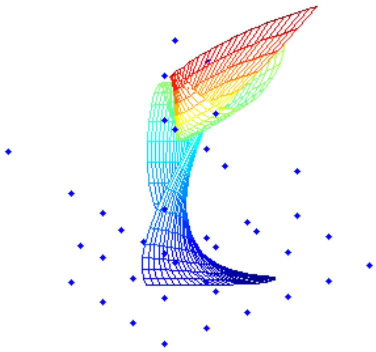

Fig. 2: Nonlinear PCA for the data in the research.

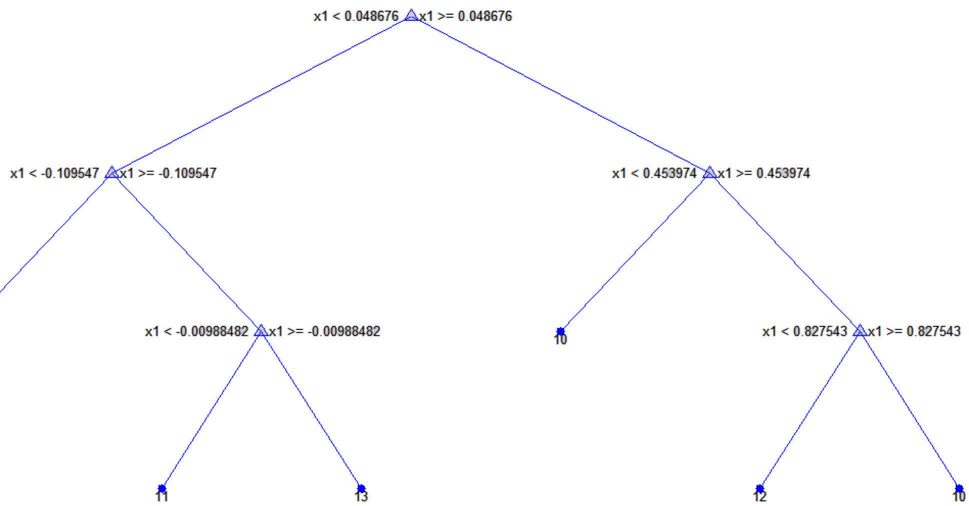

Fig. 4: clustring of the datas.

According to the Figure 4, there are 13 classes that show in Table 1.

Table 2: Eight classes of the data.

Features	Classes						
	1	2	3	4	5	6	7
Notes receivable	<0	0-500000	500000-1000000	1000000-2000000	2000000-3000000	3000000-4000000	>4000000
Inventory	<0	0-50000	50000-100000	100000-200000	200000-300000	300000-400000	>400000
Inventory stock and other inventory	<0	0-500000	500000-1000000	1000000-2000000	2000000-3000000	3000000-4000000	>4000000
Advance payment	<0	0-50000	50000-100000	100000-200000	200000-300000	300000-400000	>400000
Long-term assets	<0	0-500000	500000-1000000	1000000-2000000	2000000-3000000	3000000-4000000	>4000000
Notes payable	<0	0-50000	50000-100000	100000-200000	200000-300000	300000-400000	>400000
prepaid	<0	0-50000	50000-100000	100000-200000	200000-300000	300000-400000	>400000
spare parts	<0	0-50000	50000-100000	100000-200000	200000-300000	300000-400000	>400000
Cash	<0	0-50000	50000-100000	100000-200000	200000-300000	300000-400000	>400000
Long-term liability	<0	0-50000	50000-100000	100000-200000	200000-300000	300000-400000	>400000
Short-term investments	<0	0-50000	50000-100000	100000-200000	200000-300000	300000-400000	>400000
equity	<0	0-50000	50000-100000	100000-200000	200000-300000	300000-400000	>400000

Advances in Environmental Biology, 8(12) July 2014, Pages: 63-66

Conclusion:

In the research were used 11 inputs that involve Cash ‹Short-Term Investments ‹Notes Receivable, Inventory‹ Spare Parts ‹Inventory Stock and Other Inventory‹ Advance Payment, Long-Term Assets, Notes Payable‹ Prepaid‹ Long-Term Liability for clustering of equity. For clustering of data used NLPCA method. According to results, there are 13 classes for the data.

REFERENCES

[1] Ebrahimi, E., N. S.F. Astan, M. Mousavi, Heydari, 2014. ANFIS-Aided Grading of Almonds by Using Sound Signals, Advances in Environmental Biology, 8(4): 1046-1053.

[2] Farhadi, M., E. Farshadfar, 2014. Stability Analysis of wheat-Agropyron Disomic Addition Lines Using Univariate Stability Parameters, Advances in Environmental Biology, 8(7): 2399-2406.

[3] Osouli, N., E. Khoshbakht, Z. Ansari, M. Kazemi. 2014. Using Self-Organizing Map (SOM) algorithm for Analysts' equities clustering, Advances in Environmental Biology, 8(1): 185-189.

[4] Rumelhart, D.E., G.E. Hinton, R.J. Williams, 1986. "Learning Internal Representations by Error Propagation," Parallel Distributed Processing, 1, D. E. Rumelhart and J. L. McClelland, eds., MIT Press, Cambridge, MA.

[5] Hoskins, J.C., D.M. Himmelblau, 1988. "Artificial Neural Network Models of Knowledge Representation in Chemical Engineering,"Comput. Chem. Eng., 12: 881.

[6] Venkatasubramanian, V., K. Chan, 1989. "A Neural Network Methodology for Process Fault Diagnosis," AIChE J., pp: 35.

[7] Sanger, T.D., 1989. "Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network,"Neural Networks, 2: 459.

[8] Ackley, D.H., G.E. Hinton, T.J. Sejnowski, 1985. "A Learning Algorithm for Boltzmann Machines," Cognitive Sci., 9: 147.

[9] Scholz. M., 2012. Validation of nonlinear PCA. *Neural Processing Letters*, 36(1): 21-30.

[10] Scholz. M., 2007. Analysing Periodic Phenomena by Circular PCA. Bioinformatics Research and Development. Lecture Notes in Computer Science, 4414: 38-47.