Proceedings of the V Brazilian Conference on Neural Networks - V Congresso Brasileiro de Redes Neurais
pp. 667–672, April 2–5, 2001 - Rio de Janeiro - RJ - Brazil

A New Method for Data Clustering Using the Elastic (Neural) Net Algorithm

Rogerio L. Salvini, Luis Alfredo V. de Carvalho
Program of Systems Engineering and Computer Science - COPPE/UFRJ
Caixa Postal 68511 – CEP 21945-970, Rio de Janeiro, RJ, Brasil
rsalvini@cos.ufrj.br, alfredo@cos.ufrj.br

Abstract

This work proposes a new method for data clustering in a n-dimensional space using the Elastic Net Algorithm which is a variant of the Kohonen topographic map learning algorithm. The Elastic Net Algorithm is a mechanical metaphor in which an elastic ring is attracted by points in a bi-dimensional space while their internal elastic forces try to shun the elastic expansion. The different weights associated with these two kinds of forces lead the elastic to a gradual expansion in the direction of the bi-dimensional points. In this new method, the Elastic Net Algorithm is employed with the help of a heuristic framework that improves its performance for application in the n-dimensional space of cluster analysis. Tests were made with two types of data sets: (1) simulated data sets with up to 1000 points randomly generated in groups linearly separable with up to dimension 10 and (2) the Fisher Iris Plant database, a well-known database referred in the pattern recognition literature. The advantages of the method presented here are its simplicity, its fast and stable convergence, beyond efficiency in cluster analysis.

1. Introduction

Researchers from many areas often encounter situations best resolved by defining groups of homogeneous objects, whether they be individuals, firms, products, or even behaviors. Strategic decisions based on identifying groups within populations, such as segmentation and target marketing, would not be possible without an objective methodology. Also in Biology and Medicine, clustering is desirable, for example, in the creation of botanic taxonomy and in the analysis of psychiatric profiles [1]**.** In these instances, the analyst searches for "natural" structures in a set of experimental observations generally of a multivariate character.

The most commonly used technique for this purpose is the statistical cluster analysis which is a technique for grouping individuals or objects into clusters so that objects in the same cluster are more like to one another than objects in other clusters [2].

This work presents a new method for multivariate cluster analysis based on the Elastic Net Algorithm introduced by Durbin and Willshaw [3] using a heuristic framework that helps in the task of cluster analysis. In the next sections, this method will be detailed and some results will be shown.

2. Elastic Net Algorithm

The Elastic Neural Net Algorithm (ENA) is a heuristic algorithm, inherently parallel, developed from a hypothetical "tea-trade model" [4,5] for the establishment of topographically ordered, neighbor-preserving projections between neural structures with matching geometries. Like Kohonen's Self Organizing Feature Map (SOFM) [6], the ENA tries to find a topology preserving map between two spaces.

The algorithm was originally proposed to solve the Travelling Salesman Problem (TSP), a classical problem in the field of the Combinatorial Optimization [7]. The TSP can be described as: *"Given the position of* N *cities, what is the shortest hamiltonian tour (a closed path in which each city is visited once) joining these cities ?"*

The Elastic Net method considers an "elastic" ring with M points marked over it and that are subjected to two types of forces: internal stiffness forces E_j; j=1,2...M which induce the minimization of the length of the ring, and external forces C_{ij}; i=1,2...n and j=1,2...M which correspond to the attraction of the city i over the point j (See Figure 1).

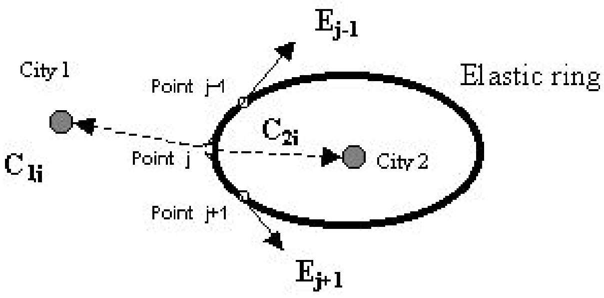


Figure 1: Forces that act on a point j in the elastic ring by two cities.

Initially, the elastic ring is represented by a small circle centered in the centroid of the **n** cities. In an iterative procedure, the ring is gradually and non-uniformly elongated until it eventually passes sufficiently near to all the cities to define a tour around them (See Figure 2).

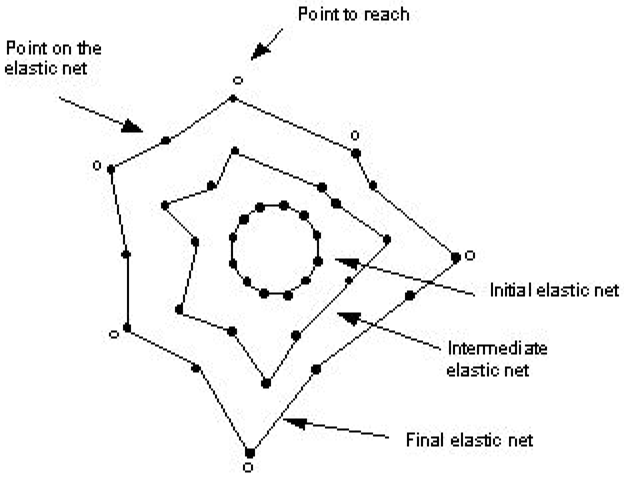


Figure 2: The process of expanding the elastic ring from the center to the cities.

Let the coordinates of a city i be denoted by the vector $\mathbf{x_i}$ and those of point j on the ring by $\mathbf{y_j}$. The forces of attraction of the cities over a point j in the ring are proportional to the distance between the city and the point, having the direction of the line that passes through the city and the point: $C_{ij} = w_{ij}\ (\mathbf{x_i} - \mathbf{y_j})$, where the parameters w_{ij} define the proportions. The two elastic forces that act upon a point j of the ring are proportional to the distance between the point j and its two neighboring points j–1 and j+1: $E_j = E_{j+1} + E_{j-1} = K\ (\mathbf{y_{j+1}} - \mathbf{y_j}) + K\ (\mathbf{y_{j-1}} - \mathbf{y_j})$ or $E_j = K\ (\mathbf{y_{j+1}} - \mathbf{2\ y_j} + \mathbf{y_{j-1}})$, where K is a control parameter of the algorithm. Summing the two types of forces leads us to the resultant that actuates on point j. The movement $\Delta\mathbf{y_j}$ of the point j of the elastic ring is assumed to be proportional to this resultant, or

$$\Delta \mathbf{y_j} = \alpha \sum_i w_{ij}(\mathbf{x_i} - \mathbf{y_j}) + \beta K(\mathbf{y_{j+1}} - 2\mathbf{y_j} + \mathbf{y_{j-1}}), \quad (1)$$

where α and β are weight coefficients. The control parameter K is a scale factor which decreases every $\aleph$ iterations, like the lowering of the "temperature" in the optimal Simulated Annealing Method [6]. The procedure of lowering K allows the ring to approximate the cities because the internal forces are reduced in relation to the external ones. The proportions w_{ij} are calculated with the help of a power function ϕ that defines an exponential decreasing influence of a city over a point as the distance between the two increases. The proportions are normalized with the division of the power function of one point by the total values of the power functions for all the points:

$$w_{ij} = \frac{\phi(|\mathbf{x_i} - \mathbf{y_j}|, K)}{\sum_k \phi(|\mathbf{x_i} - \mathbf{y_k}|, K)}. \quad (2)$$

The power function is given by $\phi(|\mathbf{x_i} - \mathbf{y_j}|^2, K) = \exp(-|\mathbf{x_i} - \mathbf{y_j}|^2 / 2K^2)$. Note that for an initially high value of K, the power function expresses an equally high influence of all the cities over all the points of the ring. As K decreases, the power function establishes a selective influence of a city over its nearest ring point.

An energy function E that is always minimized (for constant K) as the algorithm progress can be defined:

$$E = -\alpha K \sum_i \ln \sum_j \phi(d, K) + \frac{\beta}{2} \sum_j |\mathbf{y_{j+1}} - \mathbf{y_j}|^2 \quad (3)$$

This energy function has the property that:

$$\Delta \mathbf{y_j} = -K \frac{\partial E}{\partial \mathbf{y_j}}. \quad (4)$$

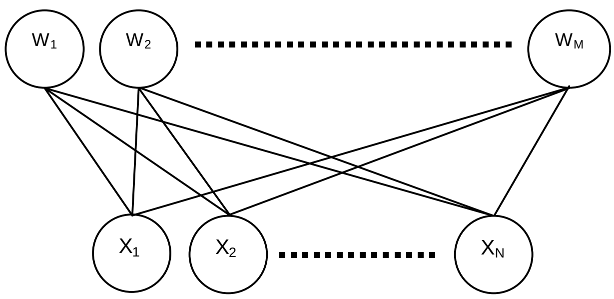
For fixed K, the path will converge to a (possibly local) minimum of E. At large values of K the energy function is smoothed and there is only one minimum (which do not correspond to a tour). At small values of K, the energy function contains many local minima corresponding to possible tours for the cities, and the deepest minimum is the shortest possible tour [8].

This is a special case of the general problem of best preserving neighborhood relations when mapping between different geometrical spaces. In our approach we extended these characteristics for clustering in n-dimensional spaces.

2.1. ANN interpretation of ENA

The Elastic Net Algorithm is a variation of the Kohonen topographic map learning algorithm that exhibits many of the same key qualitative features.

Consider an ANN system with M hidden units and N input units:



The training steps of the algorithm are:

- Step 0: Initialize weights $\mathbf{w_j}(0)$, j=1,...,M, to small values.
- Step 1: Compute response, $r_{ij}(t)$, of each hidden unit j to the ith stimulus pattern $\mathbf{x_i}$, i=1,...,N, over the input units at iteration t:

$$r_{ij}(t) = \frac{\exp(-|\mathbf{x_i} - \mathbf{w_j}(t)|^2 / 2K_t^2)}{\sum_{k=1}^{M} \exp(-|\mathbf{x_i} - \mathbf{w_k}(t)|^2 / 2K_t^2)} \tag{5}$$

- Step 2: Update weights. Let $\mathbf{x_i}$ be the ith element of the training set. For j=1,...,M:

$$\mathbf{w_j}(t+1) = \mathbf{w_j}(t) + \alpha \sum_{i}^{N} r_{ij}(t)[\mathbf{x_i} - \mathbf{w_j}(t)] + \\ + \beta K_t [\mathbf{w_{j+1}}(t) + 2\mathbf{w_j}(t) - \mathbf{w_{j-1}}(t)] \tag{6}$$

- Step 3: If $|\mathbf{w_j}(t) - \mathbf{w_j}(t\text{-}1)| < \varepsilon$ for all j=1,...,M then stop. Otherwise, let t=t+1 and return to step 1.

The weight update rule in the ENA possesses many of the same qualitative features as the original Kohonen learning rule. First, hidden units compete among themselves to select the hidden unit that is maximally responding. This competition is quantitatively instantiated by the denominator in (5). Second, units which are responding most strongly are trained with the stimulus pattern at a higher learning rate. This assumption is quantitatively instantiated by the $r_{ij}(t)[\mathbf{x_i} - \mathbf{w_j}(t)]$ term in the weight update rule in (6). Third, assuming that the hidden units are labeled such that hidden units j and j+1 are located physically next to one another, the second term of the weight update rule tends to make the average value of the weight vectors of the neighboring units j–1 and j+1 closer to the weight vector for unit j [9].

3. The Hierarchical Elastic Net Algorithm

In this section, a new hierarchical clustering method will be presented in which the ENA will be inserted. The main advantages of this approach are the decrease of the complexity of the calculations performed and, as a consequence, the decrease of the number of iterations necessary to the convergence of the method.

This method has in two stages:

- *First stage*: the points in the space are reduced. This is done by matching the pair of points that are closer. The points matched are removed and a third point is generate at the middle of the distances among them. This process continues until only 3 points result.
- *Second stage*: application of the ENA in a hierarchical way. The points that remained in the previous phase are duplicated, i.e. they are unmatched in the inverse way. The ENA is then applied to converge for the points that generated these elastic ring points. When the algorithm converges, again the points of the elastic band are duplicated and the algorithm is applied for the points that generated them. This process continues until the algorithm converges for the last group of points that are the original points of the space.

The illustration below displays a graphic representation of the stages of the hierarchical method.

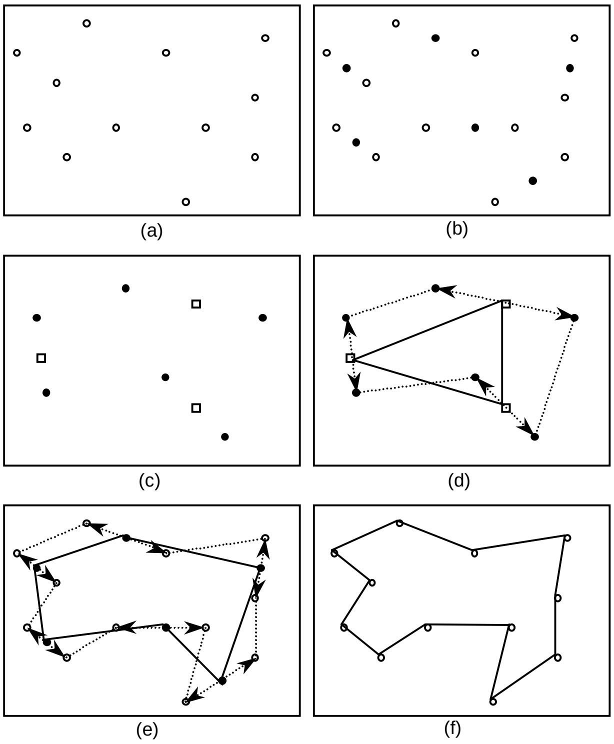


Figure 3: *First stage*: (a) Original points distributed in the space (b) The black points are generated by the matching of pairs of closest points (c) the generated points are matched again until the presence of only three of them. *Second stage*: (d) The three points of the early stage are duplicated and the ENA is applied only for its generators (e) The process is repeated until the generators are the original points (f) At the end of the process, all original points are associated with a point on the elastic band.

The final elastic ring will then form an unidimensional ordered mapping of the n-dimensional space points.

Some restrictions are made for better performance of the method. The pair of closer points are matched only if there is not some generated point whose distance for one of these points is smaller than the distance among them. In this case, the points are repeated so that in the next iteration one of them tries to match with this generated point. This measure tries to avoid possible crossings in the tour once it inhibits the matching of points that are distant. If there are 4 points at the end of the 1st phase, only the two closer are matched and the others are repeated in a way of always leaving 3 points to form the initial elastic band. The points that didn't match in the 1st stage are not duplicated in the 2nd stage.

In the original ENA, the number of points that form the initial elastic ring is at least twice the number of points of the space. Using the method developed above, the number of points of the elastic ring is the same as the number of points of the space. This reduces the number of forces between the space points and the points of the elastic band, and consequently it reduces the complexity of the algorithm. Moreover, due to the initial position of the points on the elastic ring (where each two points of the elastic band are positioned at stocking distance among two entrance points), they tend to reach very fast their nearest points of the space, again reducing the number of iterations.

4. Experimental Results

A line graph was used to visualize the results of the method. This graph traces the final sequence of the elastic points and the relative distances between each adjacent point. Every time a larger jump appears between two points in the graph, there is a group change. In this way, the process of definition of the groups is responsibility of the specialist that decides (as in the classical *dendogram* of the hierarchical techniques of clustering) the number and size of the clusters.

Two types of data sets were tested: (1) simulated data sets with up to 1000 points randomly generated in groups linearly separable with up to dimension 10 and (2) the Fisher Iris Plant database, a well-known database taken from the pattern recognition literature.

From now on, the vectors $\mathbf{x_i}$ will represent the coordinates of the points that will be clustered in the *n*-dimensional space (and no more cities in a plane), and the vectors $\mathbf{y_j}$ are the coordinates of the points of the elastic band. All $\mathbf{x_i}$ were normalized in the range [0,1].

The chosen parameters for the ENA were: α=1.0, β=0.1, $K_{initial}$=0.1, rate of decrease of K at each iteration = 1%. These parameters were shown quite flexible being used for all databases tested.

4.1. Simulated Sets

The first groups of tests were a set of simulated data. Nine test groups were generated with 200, 300 (see Figure 4), 400, 500, 600, 700, 800, 900 and 1000 random points distributed respectively in the spaces R^2, R^3, R^4, R^5, R^6, R^7, R^8, R^9 and R^{10} and contained in 4, 3, 4, 5, 6, 7, 8, 9 and 10 clusters respectively.

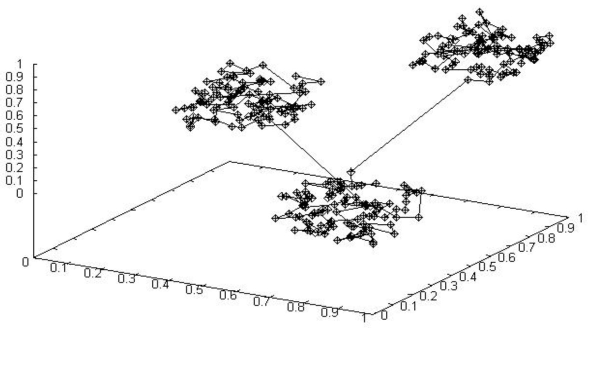


Figure 4: Example of a database with 300 random points distributed in 3 clusters in R^3

The information about the real number of clusters and the members of each class were left ignored in all the apprenticeships of the method and they were just used for comparison with the structure discovered by the process.

The results are shown in the table below:

Table 1: Simulated databases results

Base Id.	1	2	3	4	5	6	7	8	9
No. of points	200	300	400	500	600	700	800	900	1000
Dimension of the space	2	3	4	5	6	7	8	9	10
No. of clusters	4	3	4	5	6	7	8	9	10
Total no. of iterations	538	475	512	434	301	472	266	277	279
Classification rate (%)	100	100	100	100	100	100	100	100	100

Each point in the graph represented in Figures 5, 6, and 7 represents a point of the elastic band that is associated with an element of the data set. The value of the y-axis is the distance between the points $\mathbf{y_{j-1}}$ and $\mathbf{y_j}$. Therefore, each pick in the graph represents a larger relative distance between the points of the elastic band, meaning a cluster change. Examples of these graphs are shown below for the data sets with 300 (Figure 5) and 1000 elements (Figure 6).

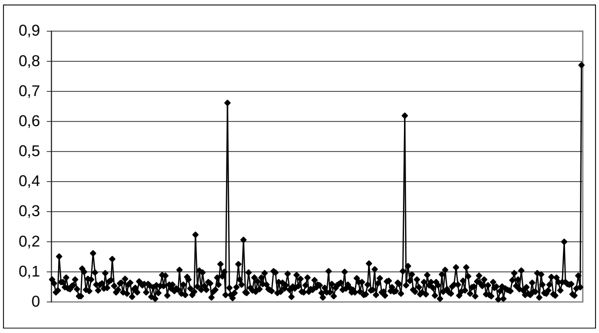


Figure 5: Line graph generated from the 300 data set. Picks represent jumps between possible clusters.

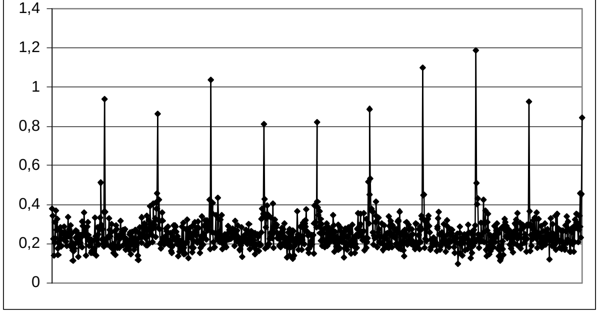


Figure 6: Line graph generated from the 1000 data set defined in R^{10}

4.2. Iris Plants Database

The Iris Plants Database, perhaps, is the best known database found in the literature of cluster analysis. Fisher's work [10] is a classic in the field and is referred frequently in cluster analysis texts. The data set contains 3 classes of iris plant with 50 instances each (iris setosa, iris versicolor and iris virginica). The first class (iris setosa) is linearly separable from the other two (iris versicolor and iris virginica); the two last groups are not linearly separable from one another. Each instance is formed by four real value inputs: sepal length, sepal width, petal length and petal width. The result obtained with the implemented model is seen in the graphic below (see Figure 7).

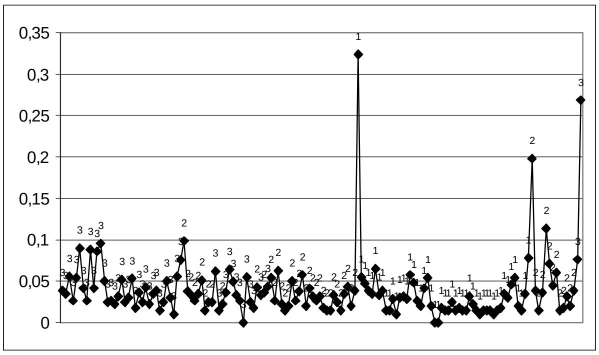


Figure 7: Graph showing results for the Iris plant database.

This graph shows a label on top of each point of the elastic band indicating to which class the element is associated (this information was not used in the clustering process). We can see a well-separated cluster labeled with 1 – this cluster is iris setosa plant type. The method also tried to separate the other two classes however it partially failed because these classes are overlapped.

Comparisons were made with traditional hierarchical methods of clustering (Single Linkage, Complete Linkage and Ward's Method) and also with the k-Means partition method. These methods were not also capable of separate correctly the types of plant of the Iris database, besides, the *dendograms* generated by the hierarchical methods are more confusing to be interpreted than the line graph presented in this work (see Figure 8). Moreover, the mistake percentage found in this method was almost the same of that found in the k-Means method.

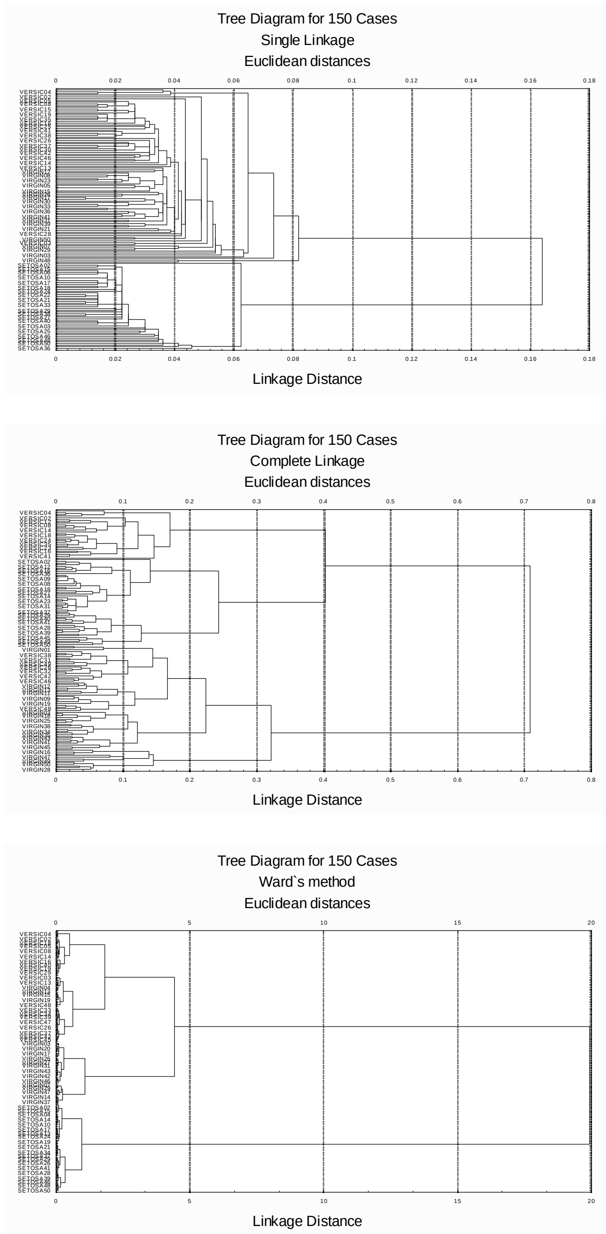


Figure 8: Dendograms generated by the hierarchical methods (Single Linkage, Complete Linkage, Ward's Method) for the Iris Plant database

5. Conclusions

This work proposes a new method for data clustering in a *n*-dimensional space based on the Elastic Net Algorithm (ENA). The ENA is a variation of the Kohonen topographic map learning algorithm that is much easier to analyze also exhibiting many of the same key qualitative features. In this approach, the ENA is employed within a hierarchical framework that improves the performance of the algorithm. A new method of visualization was also introduced which uses a line graph showing relative distances of the elastic points for the interpretation of the results. Tests were made with two types of data sets: (1) simulated datasets with up to 1000 points randomly generated in groups linearly separable with up to dimension 10 and (2) the Fisher Iris Plant database, a well-known database from the pattern recognition literature. Excellent results (with 100% of success) were reached for the first group of data, however the method differentiated only one class of Iris database because the other two are overlapped. An increase of fuzziness or overlap of the clusters will degrade the interpretation of the results because groups tend to get closer, leading to the attenuation of the nonlinearities between classes. It must be noted that the decrease of performance also occurs with other methods. The advantages of the method presented here are the simplicity of the algorithm, the fast and stable convergence, and the efficiency in cluster analysis when we have well-defined clusters.

References

[1] ROZENTHALM, M., CARVALHO, L.A.V., LAKS, J., ALMEIDA, A.F.C., MACIEL, C., ENGELHARDT, E., "Neuropsicologia da esquizofrenia e redes neurais artificiais", *Rev. Bras. de Neurologia*, v. 34, n. 4, pp. 139-144, 1998.

[2] HAIR JR., J.F. *et al*., "Cluster Analysis". In: Borkowsky, D. (ed), *Multivariate data analysis: with readings*, 4 ed., chapter 8, New Jersey, USA, Prentice-Hall, Inc., 1995.

[3] DURBIN, R., WILLSHAW, D., "An analogue approach to the travelling salesman problem using an elastic net method", *Nature*, v. 326, n. 16, pp. 689-691, 1987.

[4] VON DER MALSBURG, CH., WILLSHAW, D.J., "How to label nerve cells so that they can interconnect in an ordered fashion", *Proc. natn. Acad. Sci*., v. 74, n. 11, pp. 5176-5178, 1977.

[5] WILLSHAW, D.J., VON DER MALSBURG, C.H., "A marker induction mechanism for the establishment of ordered neural mappings: Its application to the retinotectal problem", *Phil. Trans. R. Soc. Ser. B*, v. 287, pp. 203-243, 1979.

[6] KOHONEN, T., "Self-organized formation of topologically correct feature maps", *Biological Cybernetics*, v. 43, pp. 59-69, 1982.

[7] LAWLER, E.L. et al, *The Traveling Salesman Problem*. 1 ed. New York, John Wiley & Sons, 1985.

[8] DURBIN, R., SZELISKI, R., YUILLE, A., "An Analysis of the Elastic Net Approach to the Travelling Salesman Problem", *Neural Computation*, v. 1, pp. 348-358, 1989.

[9] GOLDEN, R.M., "ANN Dynamical Systems". In: *Mathematical Methods for Neural Network Analysis and Design*, 1 ed., chapter 2, London, England, MIT Press, 1996.

[10] FISHER,R.A., "The use of multiple measurements in taxonomic problems", *Annual Eugenics*, v. 7, Part II, pp. 179-188, 1936.