Self-Organizing-Feature-Maps versus Statistical Clustering Methods: A Benchmark

A.Ultsch, C.Vetter
University of Marburg
FG Neuroinformatik & Künstliche Intelligenz
Research Report 0994

Abstract

In this paper we oppose some classical statistical clustering methods to a connectionist approach in the field of exploratory data analysis. Using an Artificial Neural network, namely Kohonen's Self-Organizing Feature Maps (SOFM), together with the Unified distance matrix or short U-matrix method we describe a new access to the domain of clustering algorithms.
We investigate the clustering capabilities of the different methods by choosing a sophisticated, the chainlink example and show how SOFM are superior to the described classical approaches.

1 Introduction

A main goal in the field of exploratory data analysis is to extract useful information out of large and usually high-dimensional data sets. In the past, clustering large datasets has been a domain of classical statistical methods. More recently a new approach, Kohonen's Self Organizing Feature Map (SOFM) has been proposed in order to classify high dimensional datasets, while having no analogous traditional method for unsupervised data analysis [Varfis/Versino93].
As new approaches from the field of connectionism were presented, those were compared to classical methods and similarities among them were said to be found [Demartines/Blayo91].

In particular the Kohonen algorithm was said to be similar to the k-means method, as it is stated, that the SOFM method *is* closely related or a partitioning method of the k-means type [Murthag/Hernandez94]. We can show, that SOFMs are not identical to the k-means type and in some cases outperform a variety of classical statistical approaches.

Since its introduction in 1982 [Kohonen82], SOFMs have been used for a wide range of applications, including robot programming, modeling of VLSI production, medical picture processing and others [Ritter/Schulten86], [Kohonen82], [Bertsch/Dengler87], [Keller/Fogelman88]. It's ability of mapping high-dimensional data to lower dimensions is applied to exploratory data analysis as well. Combing it with the U-Matrix method allows us to detect data clusters much easier than most classical methods [Ultsch/Siemon89].
In chapter 2 we provide a short description of a subset of some widely used statistical methods. We want to present two commonly employed traditional techniques, namely hierarchical clustering and the k-means algorithm.
An overview to Kohonen's SOFMs, the U-Matrix method and a guide to its use in cluster analysis is given in chapter 3. Extracting useful information form large or high-dimensional datasets often requires some preliminary reduction of their complexity. The field of data analysis encompasses various methods for clustering and representing information. Chapter 4 tries to compare the novel method to the classical approaches by introducing a sophisticated, non-trivial example.

2 Statistical Clustering Methods

The domination of statistical clustering methods in the field of complex data analysis has been very stable over the past. Many clustering algorithms have been developed in this domain, as described in man books and articles such as [Hartigan75], [Späth80], [Andernberg73] and [McLachlan/Basford88]. Here we present two of the most commonly used clustering algorithms, the hierarchical clustering and the k-means method. Further we want to give short descriptions how these methods work.

We first consider hierarchical clustering methods. These algorithms operate incrementally. Initially each data represent its own cluster. These techniques now try to link or combine different clusters. They operate on a matrix of distances between the data points rather than on the data points themselves. It is inherent to this methods, that once a cluster is allocated to a group, it cannot be reallocated as the number of groups or clusters decreases. The result of hierarchical clustering methods can be displayed as a tree-diagram or dendrogram, as shown

in figures 4.3 through 4.5. Starting with the single data points one can move along the tree as larger clusters are formed by combining different branches of the dendrogram. Finally you end up in one cluster.

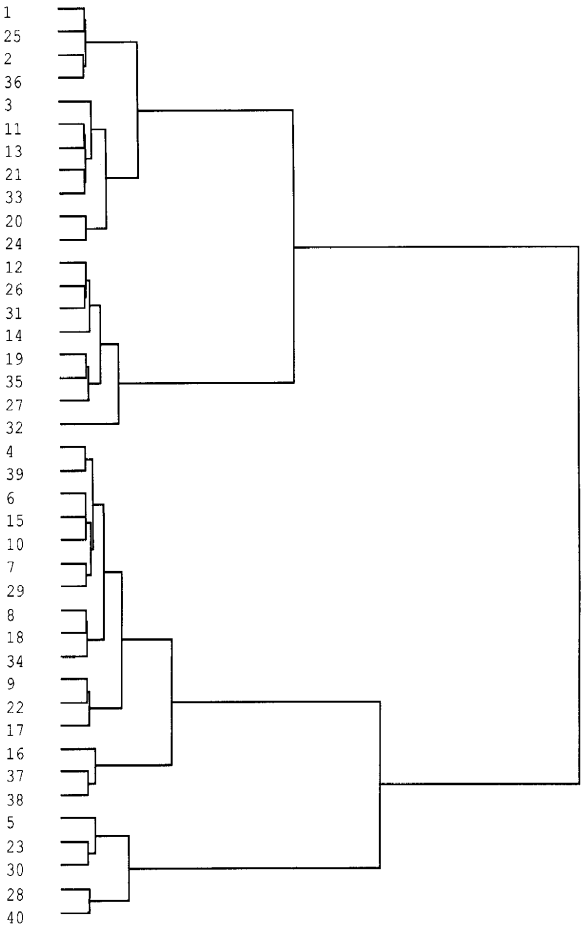


Figure x.x: Example of a dendrogram

The different hierarchical clustering methods distinguish in the way the distances between the data points or clusters are used to form larger subsets. Often used methods are the single linkage, the median linkage and the complete linkage cluster analysis, as explained below. The various methods are used to compute the distance of one object or cluster from another to determine whether the two should be merged or not.

While single linkage groups the points which are nearest to one another using the minimum distance between two points of a clusters, the complete linkage method defines the distance between two pairs of clusters as the largest distance between pairs of elements in each cluster. Median linkage averages all distances between pairs of objects using median in order to

decide how far they are apart.

The method SOFMs are often compared with is called the k-means algorithm. It's a partitioning algorithm and was presented among others by Hartigan in 1975 [Hartigan75] and improved in terms of speed by Hartigan and Wong in 1979 [Hartigan/Wong79]. Variations of this algorithm were considered by many other authors, e.g. [Späth85], [Darken/Moody90]. Its goal is it to minimize an reconstruction error, while assigning a nearest neighbor in order to compress the data vectors onto a smaller set of reference vectors. As a inherent feature it should be noted that it is absolutely necessary to define the number of final clusters in advance if using the k-means algorithm. Seeds for new clusters are chosen by splitting on the variable with the largest distance. K-means splits a data set into the selected number of clusters by maximizing between -relative to within - cluster variation. It' s an iterative procedure assigning the different data vectors to the specified number of non-overlapping clusters.

3 Self-Organizing-Feature-Maps & the U-Matrix Method

Kohonen's Self-Organizing Feature Maps represent one very well known and widely used Neural Network Model which employs unsupervised learning [Lipp87]. This means that the result is formed by the properties inherent to the data itself; there is no master which tells what is true or false. So no previous knowledge about the structure of the data is needed to be able to use the algorithm.

SOFM adapt to the training data in a way such that a high-dimensional-input space is reflected to a 2-dimensional grid on the generated Kohonen-map [Kohonen89]. By preserving the neighborhood on the topology of the map, structures of the input space can be discovered by exploring the feature mapIf the property of topology preservation holds, then similar input data should correspond to data on the Kohonen lattice with the same close relationship. . As cluster analysis groups similar objects into disjunct subsets called clusters [Deichsel/Trampitsch85], cluster in the $\Re^n$ should therefore also appear in the lattice of lower dimensionality. At this point, the SOFM arranges clusters into different regions of the lattice. As the distances among the data points are evenly distributed, no clustering would be detected without previous knowledge of the original membership. So it is clear, that Kohonen's SOFMs as a standalone are no clustering method.
In order to detect and display the structures we have developed a set of methods, called U-

matrix methods (UMMs). The idea of UMMs is it visulaize the Feature Map's topology. Analyzing the weights at each point of the grid with respect to their neighbors and then displaying the distance between two neighbor units as height , leads to an interpretable, 3-dimensional landscape of the Kohonen Map. This diagram we call Unified distance matrix or short U-matrix. The U-matrix therefore contains a geometrical approximation of the vector distribution in the Kohonen net. This display has valleys where the vectors in the lattice are close to each other and hills or walls where there are larger distances, indicating dissimilarities in the input data.

To make it more clear, we present a trivial example: Figure 3.1 shows the edge points of the used geometry. 20 random vectors added to each of them, generating 8 different clusters. A clustering algorithm should separate the data set into exactly 8 subsets and indicate, that the subsets are dissimilar. Figure 3.2 shows the U-Matrix of the dataset.

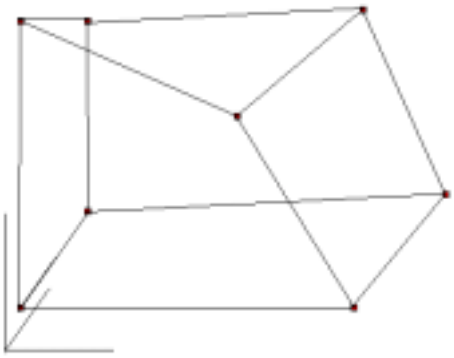


Figure 3.1: Example dataset

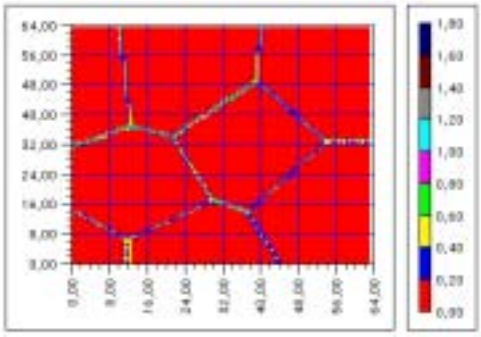


Figure 3.2: U-Matrix

To explore the properties of a n-dimensional data space we propose to interpret the landscape of the U-matrix. Given a n-dimensional data set the question is if there exists any structure in form of subsets i.e. clusters of data that are very similar to each other. The U-matrix should be interpreted as follows: If a subset of input data falls into a valley in the U-matrix then it's a cluster containing similar vectors. Neighbors among the U-matrix are close to each other in the $\Re^n$. If different subsets are separated by walls or hills then there is a larger dissimilarity among the different clusters.

As you can see, the example dataset is exactly devided into 8 different regions, with walls indicating the dissimilarity among the clusters. In order to separate a given data set it should be noted, that no previous information on the number of clusters is needed.

4 A Difficult Benchmark: The Chainlink example

In order to compare the different methods, a non-trivial example sould be found, which could serve as a benchmark for different clustering algorithms. Here we want to a present the chainlink example. As shown in figure 4.1 this benchmark consist of two intertwined 3-dimensional rings, of whom one is extended into x-y direction and the other one into x-z direction. The two rings can be thought of as a spatial chainlink with each one consisting of 500 datapoints. We now tried to figure out the abilities of the different methods in clustering this spatial benchmark problem.

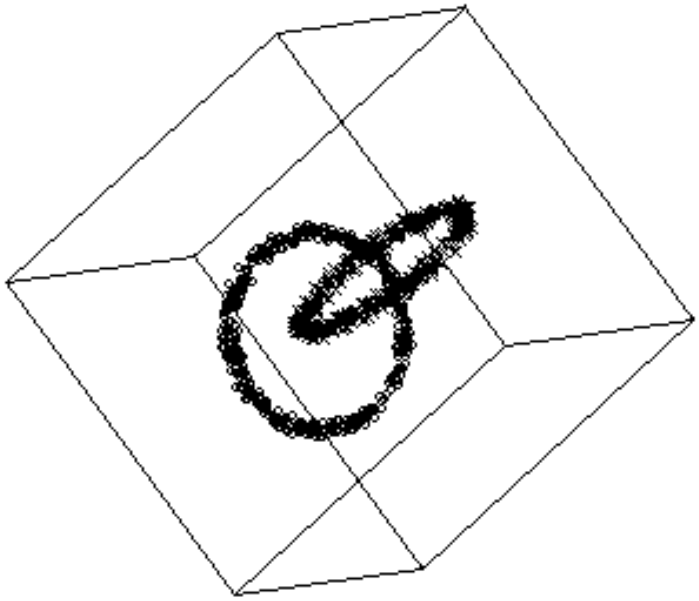


Figure 4.1: Spatial Chainlink

Statistical analysis is performed, in order to give us some initial information about the dataset. Using the Pearson Correlation Coefficient and the Principal Component Analysis1 we can see that the three input variables are in fact uncorrelated (see table 4.1). The scatterplots of figure 4.2 give some idea of the distribution of our data set, displaying each variable against the other and showing the views on the chainlink from the different directions.

	X	Y	Z
X	1,00000		
Y	0,01826	1,00000	
Z	0,00521	-0,00885	1,00000

Table 4.1: Pearson Correlation Coefficient

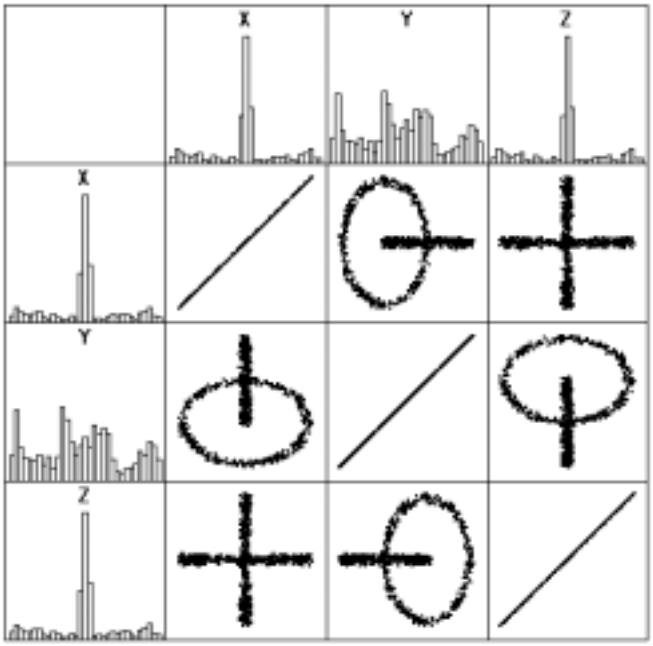
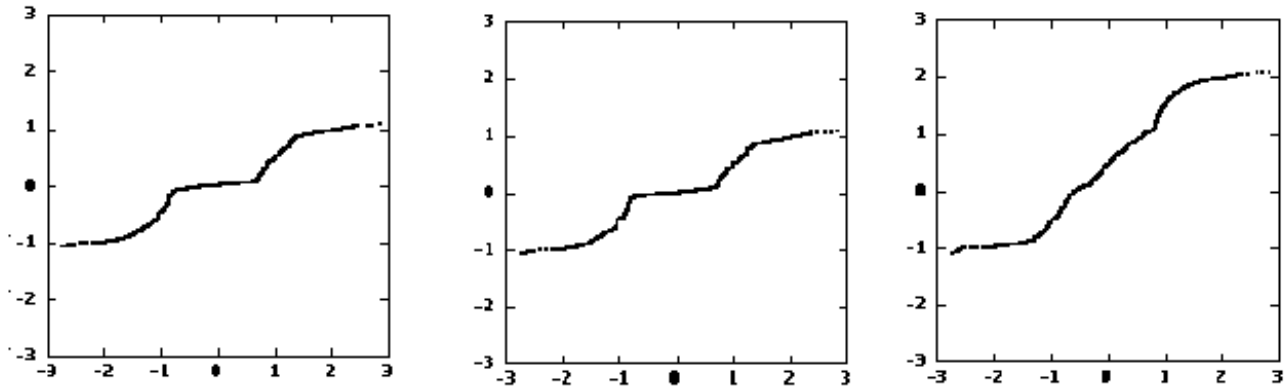


Figure 4.2: Scatterplots

Chi-Square test and the QQ-plots show, that the data is not normal distributed.

VARIABLE	N-OF-CASES	MAXDIF	PROBABILITY
X	1000,00000	0,23106	0,00000
Y	1000,00000	0,21000	0,00000
Z	1000,00000	0,23106	0,00000
NORMAL	1000,00000	0,02185	0,72625

Table 4.2: CHI-SQUARE TEST Using Standard Normal Distribution



Principal Component Analysis (PCA) could serve in order to see whether data sets of a high dimension could be described in fewer dimensions. Its main goal is it to summarize a multivariate dataset as accurately as possible using a few components. It consists of projecting the data onto vectorial subspaces that maximally preserve best the dataset variance. The first component will summarize a much of the joint variance as possible, while following components will do the same job on what is left as long as the left variance is negligible.

Using the information of table 4.2, we can see, that each of the variables collects about a third of the total variance. One can see that the 3-dimensional figure of the input space has to be described by also 3 dimension according to the PCA analysis.

1	2	3
33,95569	33,46597	32,57834

Table 4.2: Principal Component Analysis: Percentage of total variance explained

Keeping the above information in mind, it is clear that we found a non linear-seperable example, which has 2 evident clusters and cannot be reduced in its dimensionality by methods of projection, as PCA. It, therefore, serves as an adequate benchmark for the following research.

In the next section, we try to figure out the clustering capabilities of the different approaches:

Hierarchical Clustering

As a first step we worked with traditional hierarchical clustering methods, using single linkage, complete and median linkage clustering. Figures 4.3 through 4.5 show the dendrograms for the different linkage methods. As one can see single linkage tends to give the best results, while complete and median linkage fail to classify the data. At this point it must be stated, that there exists one main disadvantage on the interpretation of dendrograms: The fact of not having a previous knowledge of the original dataset, often leads to misinterpretation by the human expert [Gordon94]. The order of the branching point projection on the horizontal axis corresponds to the sequence of clustering steps.

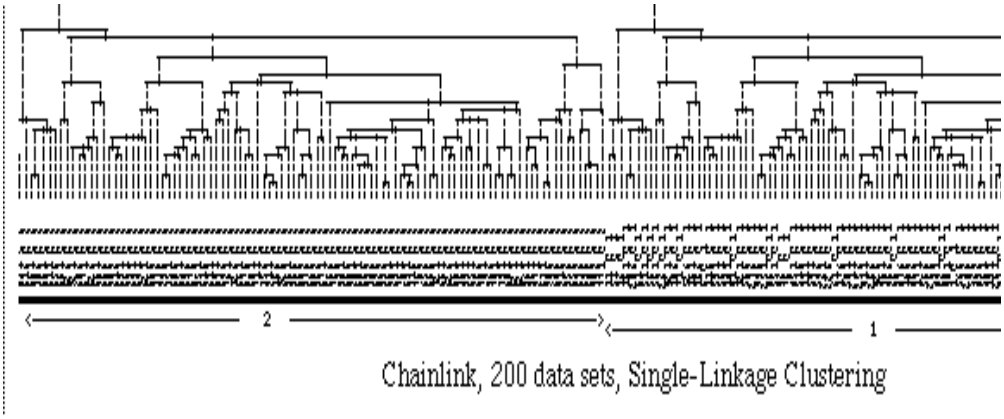


Figure 4.3

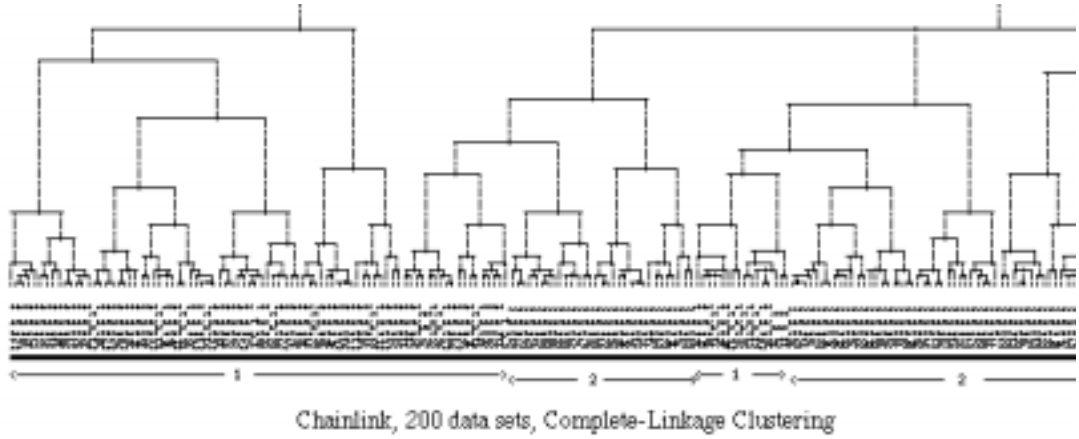


Figure 4.4

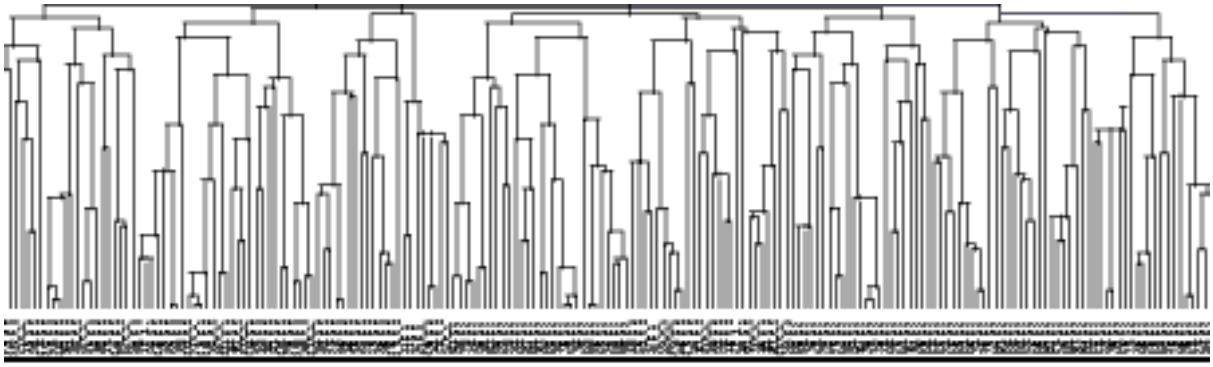


Figure 4.5

K-Means

Second, we presented the dataset to the classical k-means approach. It was found that this clustering approach wasn‘t able to distinguish between the two different rings.. Figure 4.6 and 4.7 show the results of k-means with 2 respective 4 clusters. Having a closer look at figure 4.6 one can see, that k-means devides the spatial construction of the rings following the x-z coordinate plane, and therefor fails to classify the data correctly. In figure 4.7 the datapoints of the two rings are clustered according to the given k-means algorithm into 4 different regions. But as ther are datapoints of one cluster , which belong to different rings, it also performs very badly.

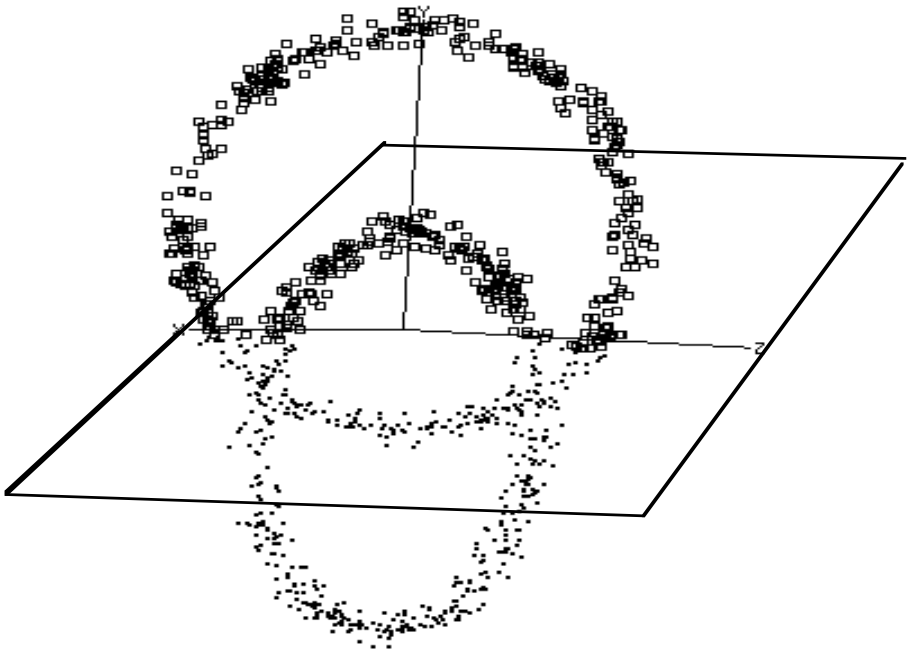


Figure 4.6: k-means: 2 clusters

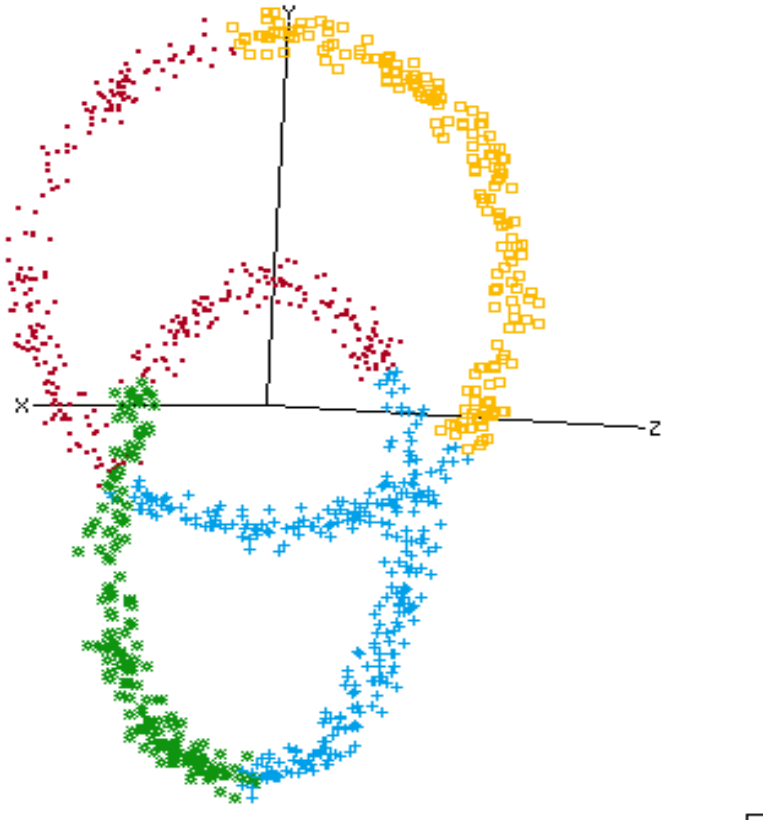


Figure 4.7: k-means: 4 clusters

SOFM & U-Matrix

Next we tried to classify the data using a Kohonen Feature Map with a 64x64 two dimensional grid. After training the SOFM, it was able to cluster the data correctly. In Figure 4.8 the location of the data on the 64x64 grid are displayed. This type of diagram is called coordinate matrix in the sequel. As we can see the grid is clearly devided into two different regions, each on representing the datapoints of one ring. Looking a the U-Matrix picture in figure 4.9 we notice that the different data points are devided by walls, so that we easily can figure out, which group a certain data point belongs to. For a better understanding we displayed the different data points of the two rings with distinguishable makers (one with the crosses, one with the circles). But even without that information we could distinguish the two rings by superposing the coordinate of the SOFM
with the generated U-Matrix landscape.

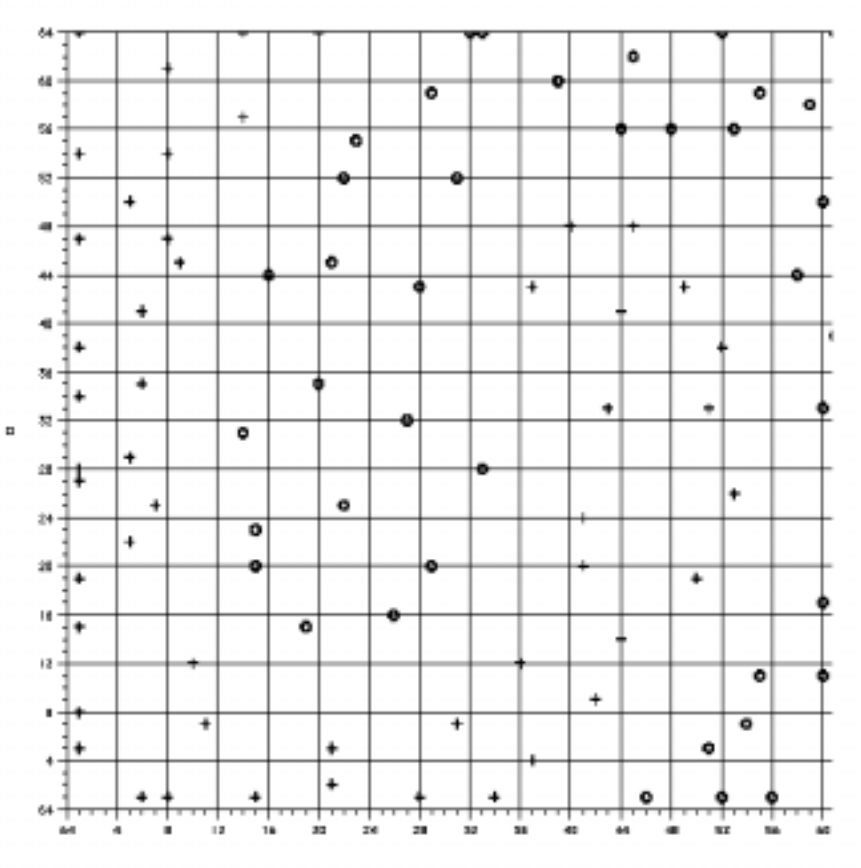


Figure 4.8: Coordinate Matrix of the data

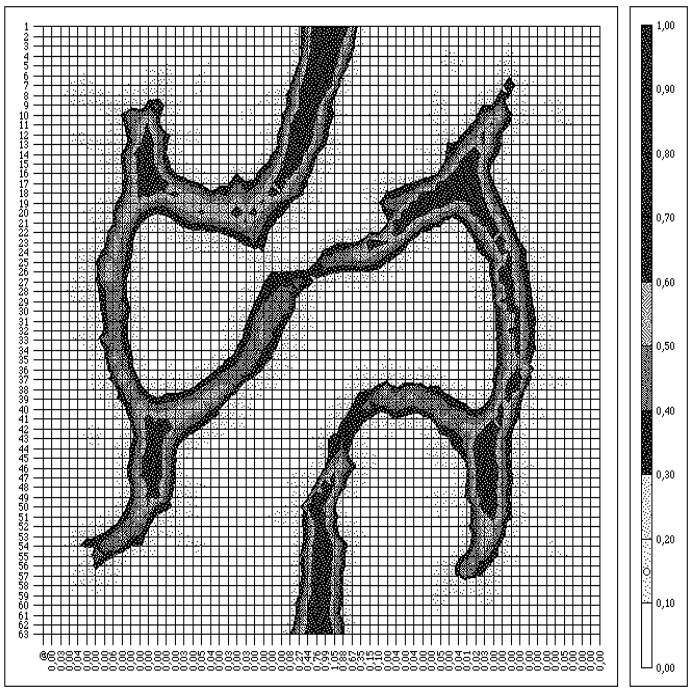


Figure 4.9: U-Matrix of the data

5 Conclusion

Although it imposes some difficulties to compare different clustering algorithms, which partially rely on different principles, we have shown that SOFM together with the U-Matrix method can prove to be superficient to other well known statistical clustering methods. It has been capable of classifying a difficult, artificially generated dataset using unsupervised learning , while other methods need previous information on the dataset, or even fail to classify the data correctly. Moreover other experiments tend to encourage us, to use this method for wide range of analysis, even from such domain as medical data.

REFERENCES:

[Andernberg73] M.R.Andernberg: *Cluster Analysis for Applications*, New York, Academic Press, 1973.

[Bertsch/Dengler78]: H.Bertsch, J.Dengler: *Klassifizierung und Segmentierung medizinischer Bilder mit Hilfe der selbstlernenden topologischen Karte*, in: Paulus, E.(ed.):*DAGM-Symposium Mustererkennung*, Springer Informatik Fachberichte 149, 1978, 16 - 170.

[Darken/Moody90] C.Darken, J.Moody*: Fast adaptive K-Means Clustering: some Empirical Results*, Proc. International Joint Conference on Neural Networks Vol II, San Diego, New York, IEEE Computer Scienc Press, pp.233-238, 1990.

[Deichsel/Trampisch85]: G.Deichsel, H.J.Trampisch: *Clusteranalyse und Diskriminanzanalyse*, Gustav Fischer Verlag, Stuttgart, New York, 1985.

[Demartines/Blayo91]: P.Demartines, F.Blayo: Data Analysis*: How to compare Kohonen neural networks to other techniques*, in: Proc. of IWANN 91, Prieto (ed.*): Lecture Notes in Computer Science*, Springer Verlag, pp. 469-476, 1991.

[Gordon94] A.D.Gordon: *Identifying Genuin Clusters in a Classification*, Computational Statistics & Data Analysis 18, pp. 561-581, 1994.

[Hartigan75] J.A.Hartigan: *Clustering Algorithms*, Wiley and Sons, New York, 1975

[Hartigan/Wong79] J.A.Hartigan, M.A.Wong: *A K-means Clustering Algorithm*, Appplied Statistics, 28, pp.100-108., 1979.

[Keller/Fogelman88] M.Keller, F. Fogelman-Soulie: *Topological Maps and their Applications to Electrophoresis image recognition*, Proc. Neuro-Nimes, Nimes, France, November 1988, pp 403-413.

[Kohonen82]: T.Kohonen: *Self Organized Formation of Topological Correct Feature Maps* , Biol.Cyberbetics. Vol 43, 1982, pp.59-96.

[Kohonen89]: T.Kohonen: *Self Organization and Associative Memory*, Springer Verlag, Berlin, 1989.

[Lipp87]: R.P. Lippmann: *An Introduction to Computing with Neural Nets*, IEEE ASSP Magazine, April 1987, pp. 4-22.

[McLachlan/Basford] G.J.McLachlan, K.E.Basford: *Mixture Models*, New York: Marcel Dekker, Inc., 1988.

[Murtagh/Hernandez94] F.Murtagh, M.Hernandez-Pajares: *The Kohonen Self-Organizing Map Method: An Assessment*., Journal of Classification, in press

[Ritter/Schulten]: H.Ritter, K.Schulten*: Kohonen's Self-Organizing Maps: Exploring their Computational Capabilities*, Biol. Cybernetics Vol 54, 1986, pp.99-106.

[Sarle94] W.S.Sarle: *Neural Networks and Statistical Models*, Proc. of the 19th Annual SAS Users Group Inernational Conference, April 1994.

[Späth80] H.Späth: *Cluster Analysis Algorithms*, Chichester, UK, 1980.

[Späth85] H. Späth: *Cluster Dissection and Analysis*, Chichester, UK, 1985.

[Ultsch/Siemon89]: A.Ultsch, H.P.Siemon*: Exploratory Data Analysis: Using Kohonen Networks on Transputers*, Research Report No.329, University ofDortmund, 1989.

[Ultsch/Siemon90]: A.Ultsch, H.P.Siemon: *Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis*, Proc. of the intl. Neural Networks Conference, Kluwer Academic Press, Dordrecht, pp.305-308, 1990

[Varfis/Versino92] A.Varfis, C.Versino: *Clustering of Socio-Economic Data with Kohonen Maps*, IDG VSP, Ispra, Italy, pp. 813-833, 1992.