Click to see slides of presentation

Unsupervised Learning Applied to Hydraulic Flow Unit Identification Based on Wireline Formation Pressure Data*

Jose Victor Contreras[1]

Search and Discovery Article #42260 (2018)**
Posted August 27, 2018

*Adapted from extended abstract based on oral presentation given at AAPG 2018 Southwest Section Annual Convention, El Paso, Texas, April 7-10, 2018

[1]Baker Hughes, a GE company, The Woodlands, Texas (jose.contreras2@bhge.com)

Abstract

Wireline formation testing (WFT) tools provide vital information about formation pressure in order to know the current energy available in reservoirs. In addition, the different pressure and mobility values delivered by WFT tools could be analyzed as a function of depth, and these results used to identify flow units. In some cases, there is a very large number of pressure and mobility values with depth, which makes it impractical and inefficient to do this process manually. In these cases, it is possible to develop a pattern recognition workflow, based on unsupervised learning, to classify pressure stations with similar characteristics and associate these pressure stations to a flow unit.

A workflow for pattern recognition based on unsupervised learning was developed to classify pressure stations with similar pressure, mobility and depth characteristics and assign those results to the corresponding flow units. A dendrogram or Self-Organizing Map algorithm was used to train and automatically classify pressure stations. This workflow based on unsupervised learning, assigns flow units to each pressure station, according to the similar characteristics described above.

During the workflow development, the data set was first processed to make it suitable to use in a dendrogram or Self-Organizing Map. Then, an unsupervised learning algorithm was applied to the resulting data in order to classify pressure stations with similar depth, pressure and mobility characteristics. Finally, these classification results were assigned to the identified flow units.

It is concluded that unsupervised learning is a powerful tool to identify flow units based on wireline formation pressure data, because of the high capacity for finding hidden patterns. This workflow adds more value to the WFT tool data, in terms of additional information that can be related to flow units. Moreover, this workflow reduces the error induced by human intervention during classification process and the time invested to do this task.

Introduction

Flow unit identification is a very important process performed during reservoir characterization studies in which it is possible to identify and estimate different petrophysical and geological characteristics that help to describe the fluid flow in porous media. Once these characteristics are quantified, it is relatively easy to classify geological units that share similar fluid flow characteristics. This assessment is very useful when studying vertical and lateral continuity in geological units. Wireline formation testing tools provide vital information that can be related directly with petrophysical characteristics, such as fluid mobility that supplies fluid flow characteristics from porous media and pressure information as a function of depth, which provides information about vertical continuity from contiguous rock volumes.

Flow unit identification assessment most of the time is performed by hand for human experts that use their knowledge and expertise to categorize geological units with similar characteristics by using cross plot techniques, histograms plots, etc. However, most of the time they have to deal with large amounts of observations and data that consists of more than three dimensions, which makes it very difficult to capture their characteristics and relationship in a single plot.

One method that can be used to overcome that kind of limitation is a pattern recognition technique based on unsupervised learning algorithms, with which, it is possible to classify large amounts of observations (pressure stations) in certain numbers of classes, where each class is comprised of observations that share similar attributes and those attributes are different compared with observations from other classes. Unsupervised learning algorithms has been applied effectively in many other Oil and Gas disciplines such as geophysics, petrophysics, as well as in other fields of studies like medical applications. Successful utilizations have been described in seismic facies classification (Roy et al., 2013) and image category detection (Seebock et al., 2017). Among unsupervised learning algorithms, the Self-Organizing Map (Kohonen, 1982) is commonly used to accomplish this task. The self-organizing map (SOM) is a topographic organization in which nearby locations in the map represent observations with similar properties.

This article proposes a workflow to identify and automatically classify hydraulic flow units based on unsupervised learning and WFT data attributes. The theory of SOM and the details of the workflow construction will be discussed. The results of three WFT data sets will be presented and compared with results from traditional processes to classify hydraulic flow units. This comparison demonstrates the potential of unsupervised learning algorithm to classify hydraulic flow units.

Unsupervised Learning

Unsupervised learning involves a process of auto association of information from the inputs in a set of classes with similar characteristics. Unsupervised learning algorithms are used to find patterns or features from data sets consisting of input data without labeled responses (Figure 1). The main objective is to explore the structure of the data sets to extract meaningful information without the guidance of a known target output, discovering hidden structures based on similarities. Unsupervised learning needs a criterion to terminate the process, otherwise the learning process continues even when a pattern has been found.

There are several categories of unsupervised learning techniques that allow us to discover hidden structures in data where we do not know the right answer in advance. One of these categories is called clustering which is a technique where the inputs are a set of elements, $\mathcal{X}$, and the distances function over these elements. That is, a function $d : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}_{+}$ that is symmetric, satisfies $d(x, x) = 0$ for all $x \in \mathcal{X}$ and often also satisfies the triangle inequality. Alternatively, the function could be a similarity function $s : \mathcal{X} \times \mathcal{X} \longrightarrow [0,1]$ that is symmetric and satisfies $s(x, x) = 1$ for all $x \in \mathcal{X}$ (Shalev-Shwartz and Ben-David, 2014). Additionally, some clustering algorithms also require an input parameter k (determining the number of required clusters). After this technique is applied, the result is a partition of the domain set $\mathcal{X}$ into subsets (Figure 2). That is, $C = (C_1, C_2, ..., C_k)$ where $U_{i=1}^{k} C_i = \mathcal{X}$ and for all $i \neq j, C_i \cap C_j = 0$. (Shalev-Shwartz and Ben-David, 2014).

In some situations the clustering is soft which mean the partition of $\mathcal{X}$ into the different clusters is probabilistic where the output is a function assigning to each domain point, $x \in \mathcal{X}$, a vector $p_1(x), p_2(x), \dots \dots \dots, p_k(x)$, where $p_i(x) = P[x \in C_i]$ is the probability that x belongs to cluster C_i.

In this article the attention will be focused on a class of unsupervised algorithm based on competitive learning in which the output neurons of a neural network compete between themselves for the right to respond resulting in only one neuron being activated at any one time. This activated neuron is called the winning neuron. This competition can be induced by having lateral inhibition connections (negative feedback paths) between the neurons. The result is the neurons are forced to organize themselves. This network is called a Self-Organizing Map (Kohonen, 1982). This article will be concentrated specifically on Kohonen SOM (Figure 3).

Self-Organizing Map

Kohonen SOM is a topographic organization in which nearby locations in the map represent inputs with similar properties. SOM is a group of neurons organized in a low dimension mesh where each neuron is represented by a weight vector of *m* dimensions in which *m* is equal to the input vector dimension. Neurons are connected to adjacent neurons by a vicinity relationship that dictates topographic organization to the map.

The self-organization process involves the following tasks:

Initialization: In this step, all connection weights (w_i) are initialized with small random values.

Competition: In each pattern there are input units defined as $\boldsymbol{x} = \{x_i : i = 1, \dots, D\}$ and the connection weights between the input units i and the neurons j, written as $\boldsymbol{w_j} = \{w_{ji} : j = 1, \dots, N; i = 1, \dots, D\}$ where N is the total number of neurons. Neurons compute their respective values of a discriminant function which provides the basis for competition. The discriminant function can be defined as the Euclidean distance (Figure 4) between the input vector $\boldsymbol{x}$ and the weight vector $\mathbf{w_j}$ for each neuron j

$$d_{ij}(\mathrm{X}) = \sum_{i=1}^{D} \|x_i - w_{ji}\|$$

The particular neuron with the smallest value of the discriminant function is declared the winner. In other words, the neuron whose weight vector comes closest to the input vector is declared the winner.

Cooperation: The winning neuron determines the spatial location of a topological neighborhood of excited neurons thereby providing the basis for cooperation among neighboring neurons. When one neuron is stimulated, its closest neighbor tends to get excited more than those neurons located further away. The Gaussian topological neighborhood (Figure 5) for the neurons is defined as

$$T_{i,I(X)} = \exp\left(-S_{j,I(x)}^{2}/2\sigma^{2}\right)$$

Where $S_{i,I(x)}$ is the lateral distance between neurons *j* on the grid of neurons and the winning neuron, and $\sigma = \sigma(t)$ is a suitable decreasing time dependent function that contains information about the neighborhood radius.

Adaptation: In this process, excited neurons decrease their individual values of the discriminant function in relation to the input units through suitable adjustment of the associated connection weights. It is very important to the formation of ordered maps that not only the winning neuron gets its weights updated, but as topologically related subsets, its neighbors will have their weights updated as well, with a similar kind of correction imposed, although not as much as the winner itself. The appropriate weight update equation is

$$w_i(t+1) = \mathrm{w}_i(t) + \alpha(t)\, T_{i,I(X)}(t)[x_i - \mathrm{w}_i(t)]$$

Where $\alpha(t)$ is the time dependent learning rate and could be represented by a suitable decreasing time dependent function for example, a decreasing exponential function (Figure 6). The learning rate decreases as the learning process proceeds.

The weight vector $\boldsymbol{w_j}$ associated with the winner neuron is updated in such a way that it moves towards the input vector $\boldsymbol{x}$ (Figure 7).

Building the Workflow

Data Set

The wireline pressure test information that will be used in this study belongs to three vertical wells that were drilled in a sequence of interbedded shale sandstone formations. This data set comprises depth, formation pressure and fluid mobility information available for each pressure station with a number of pressure stations ranging from 34 to 59. A quality control process was performed on the raw data.

Data Preparation

The data set was organized in an input matrix in order to be used in the unsupervised learning algorithm. In the input matrix, each *ith* column have the elements that represent the measurements taken at each pressure station and each *ith* row represent the number of observations. For example, in well 1 the matrix has a size of 3 columns (characteristics) and 56 rows (number of pressure stations or observations).

Applying the Unsupervised Learning Algorithm

Creating the self-organizing maps:

The SOM will be created with a neural network that will learn how to classify pressure stations and assign those results to the corresponding hydraulic flow units. First step is to initialize the SOM where the map size, type of cell and initial weights with random values are specified. In this case, the SOM will have an arrangement of 2-dimension layer of 36 neurons (6x6) and hexagonal topology. During this process values for learning rate (α) = 0.1 and neighborhood radius (r) = 2 were specified.

Model selection:

Model selection process in this work consists of the selection of the optimum SOM size, which involves the map selection with the right numbers of cells that minimizes the average distance between each data vector and its best matching unit (BMU), and the topographic error. During the selection process, a sensitivity analysis was performed with different sizes of SOM of 6x6, 8x8, 10x10, 12x12 and 14x14, in order to study and analyze the impact of different map sizes in the pattern recognition task. Table 1 contains different SOM characteristics and its performance during the training process.

Table 1 shows that in a square map with random initialization vectors, as the map size increases the distance between data vector and its BMU decreases. However, the minimum value for topographic error was accomplished by the map size of 12x12 cells. In conclusion, the SOM that presents the best performance during the pattern recognition process is the map size 12x12.

The SOM selection process explained above corresponds to 59 pressure observation. Because the SOM size can vary according to the number of observations, another SOM selection process was carried out with a reduced number of pressure observations of 34 stations. Similar to the previous example, a sensitivity analysis was done with different sizes of SOM of 6x6, 8x8, 10x10, 12x12 and 14x14. Table 2 contains different SOM characteristics and its performance during the training process, for a reduced number of pressure observations.

Table 2 shows that in a square map with random initialization vectors, as the map size increases the distance between data vector and its BMU decreases while the topographic error increases. In this case, the best performance for the SOM was observed in the map sizes 10x10.

The sensitivity analysis developed with the number of observations demonstrate the SOM size must be proportional to the number of observations. For example, in this specific case for a number of observation around 60 the optimum SOM size is 12x12 neurons. Nevertheless, for a number of 34 observations the optimum SOM size is 10x10 neurons.

Identification of flow units:

Once the best size of the SOM was identified, the trained map was used to carry out the flow unit identification. An important characteristic in this type of SOM is that connection neurons with yellow, green and cyan colors indicate pressure stations with very different characteristics.

Example 1

The available wireline pressure information for well 1 includes depth, reservoir pressure and fluid mobility for a total of 56 pressure stations. The unsupervised algorithm was applied to this information and results are shown in Figure 8a.

Figure 8a shows three (03) zones with large differences in pressure trends or very different attributes. These three main zones are well defined and separated by several neurons with yellow and cyan colors. Inside these zones, several hydraulic flow units were identified which present small differences in pressure station attributes and are separated by 1 or 2 rows of neurons with a light blue color. In total 11 hydraulic flow units were recognized and are highlighted with red ovals. This information correlates very well with pressure trends recognized by using the pressure profile plot (Figure 8b).

This SOM map has the following characteristics: size 12x12, hexagonal topology, neighborhood radius (r) = 2, learning rate (α) = 0.1 and random weights initialization. The SOM performance parameters like average distance between each data vector and its BMU and topographic error are 12.8849 and 0.2857 respectively.

Example 2

For well 2, the wireline pressure information includes depth, reservoir pressure and fluid mobility for a total of 59 pressure stations. The unsupervised algorithm was applied to this information and results are shown in Figure 9a.

Figure 9a shows four (04) zones with huge differences in pressure trends. These four zones are well defined and separated by several neurons with yellow, green and cyan colors. Inside these zones, several hydraulic flow units were identified which present small differences in pressure station attributes and are separated by 1 or 2 rows of neurons with a light blue color. In total 20 hydraulic flow units were recognized by using SOM and are highlighted with red and orange ovals. This information correlates very well with pressure trends identified with the pressure profile plot (Figure 9b).

SOM map used for this example has the following characteristics: size 12x12, hexagonal topology, neighborhood radius (r) = 2, learning rate (α) = 0.1 and random weights initialization. The average distance between each data vector and its BMU and topographic error were 15.2526 and 0.2542 respectively.

Example 3

The available wireline pressure information for well 3 includes depth, reservoir pressure and fluid mobility for a total of 34 pressure stations. In this case, there are a reduced number of observation and pressure station attributes compared with examples 1 and 2. The unsupervised algorithm was applied to this information and results are shown in Figure 10a.

Figure 10a shows three (03) zones with large differences in pressure trends. These zones are well defined and separated by several neurons with yellow, green and cyan colors. Inside these zones, several hydraulic flow units were identified as a result of small differences in pressure station attributes. These pressure stations are separated by 1 or 2 rows of neurons with a light blue color. In total 7 hydraulic flow units were identified by using SOM and are highlighted with red ovals. This information correlates very well with different pressure trends recognized with the pressure profile plot (Figure 10b).

SOM map used for this example has the following characteristics: size 10x10, hexagonal topology, neighborhood radius (r) = 2, learning rate (α) = 0.1 and random weights initialization. In this case the average distance between each data vector and its BMU, and topographic error are 17.0162 and 0.5000 respectively.

Discussion

Results listed in Table 1 and Table 2 show that the number of observations significantly influences the SOM size, additionally, SOM features affect the classification performance. In all cases, results indicate that the classification process was significantly influenced by differences in depth and pressure values in pressure stations. It is worth noting that an additional improvement was observed when adding pressure gradient to this process. However, no additional improvement was detected when mobility was included in the classification process. These findings suggest that depth and pressure data have primary importance for the hydraulic flow unit identification, nevertheless other data sets have secondary importance.

Conclusions

Unsupervised learning by using SOM is an excellent tool to identify flow units based on wireline formation pressure data. The classification results show that SOMs were able to identify hydraulic flow units with large and small differences in its attributes.

Results indicate that the classification process was significantly influenced by depth and pressure information. No additional improvement was observed when using mobility in this process. A sensitivity analysis varying map size must be performed in order to select the optimum SOM size with the best performance.

Selected References

Bar-Joseph, Z. et al., 2001, Fast optimal leaf ordering for hierarchical clustering: Proceeding of the Ninth International Conference on Intelligent Systems for Molecular Biology.

Duda, Richard, Peter Hart, and David Stork, 2000, Pattern Classification, second edition: Wiley-Interscience, New York, USA.

Kohonen, T., 1995, Self-Organizing Map, second edition: Springer-Verlag, Berlin, p. 78-82.

Kohonen, T., 1982, Clustering, Taxonomy, and Topological Maps of Patterns: International Conference on Pattern Recognition (ICPR), p. 114-128.

Kiviluoto, K., 1996, Topology preservation in Self-Organizing Maps: Proceeding of International Conference on Neural Networks (ICNN), p. 294-299.

Roy et al., 2013, Active learning algorithms in seismic facies classification: SEGAM 2013-0769, presented at the SEG Houston 2013 Annual Meeting, Houston, Texas.

Seebock et al., 2017, Unsupervised learning for image category detection: 22nd Computer Vision Winter Workshop, Retz, Austria, February 6-8.

Shalev-Shwartz, S., and S. Ben-David, 2014, Understanding machine learning, from theory to algorithms: Cambridge University Press, New York, USA, p. 309-310.

Zhu, X., and Z. Ghahramani, 2002, Learning from labeled and unlabeled data with label propagation: Technical Report CMU-CALD-02-107, Carnegie Mellon University.

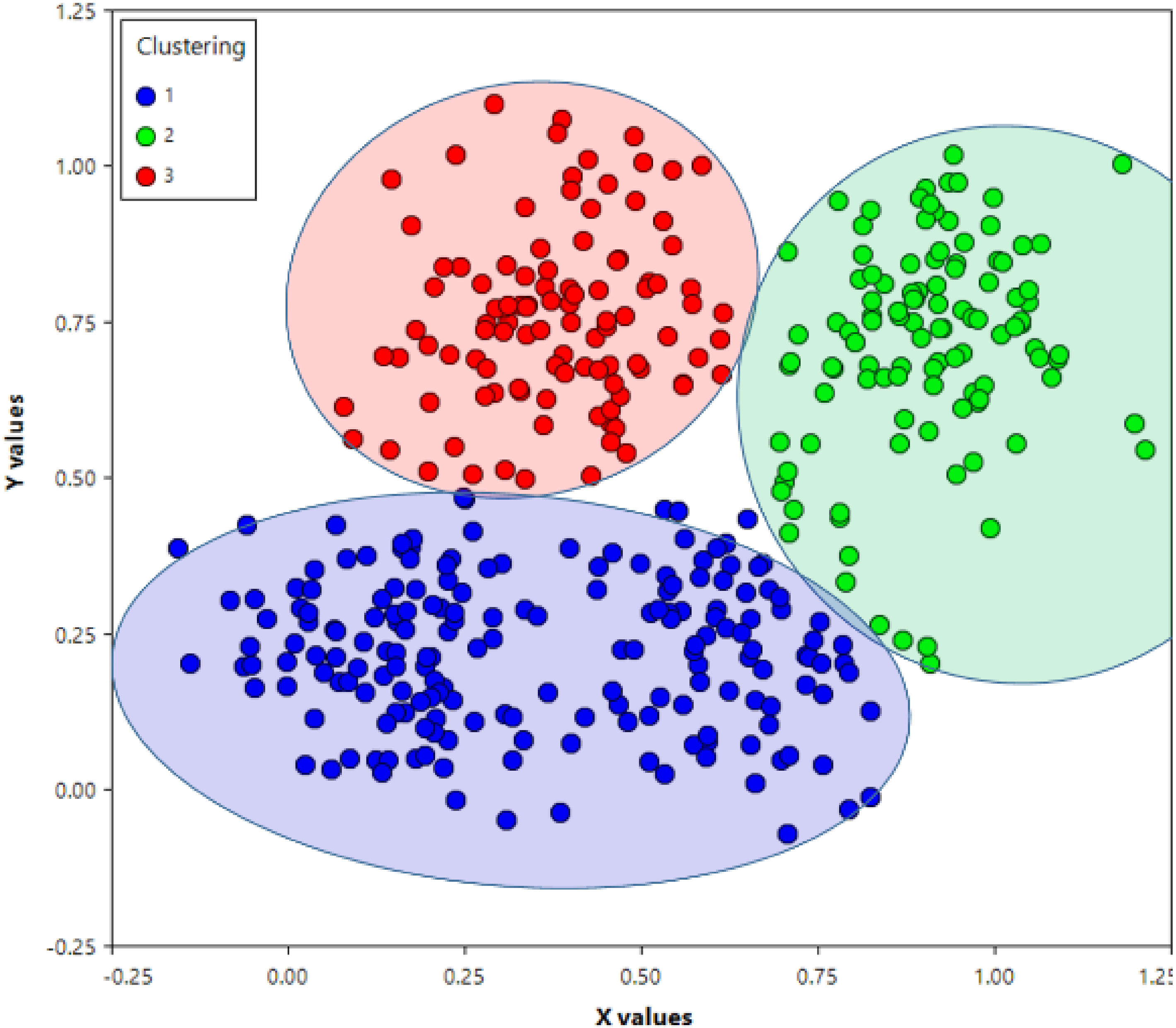

Figure 1. Pattern recognition of unlabeled data

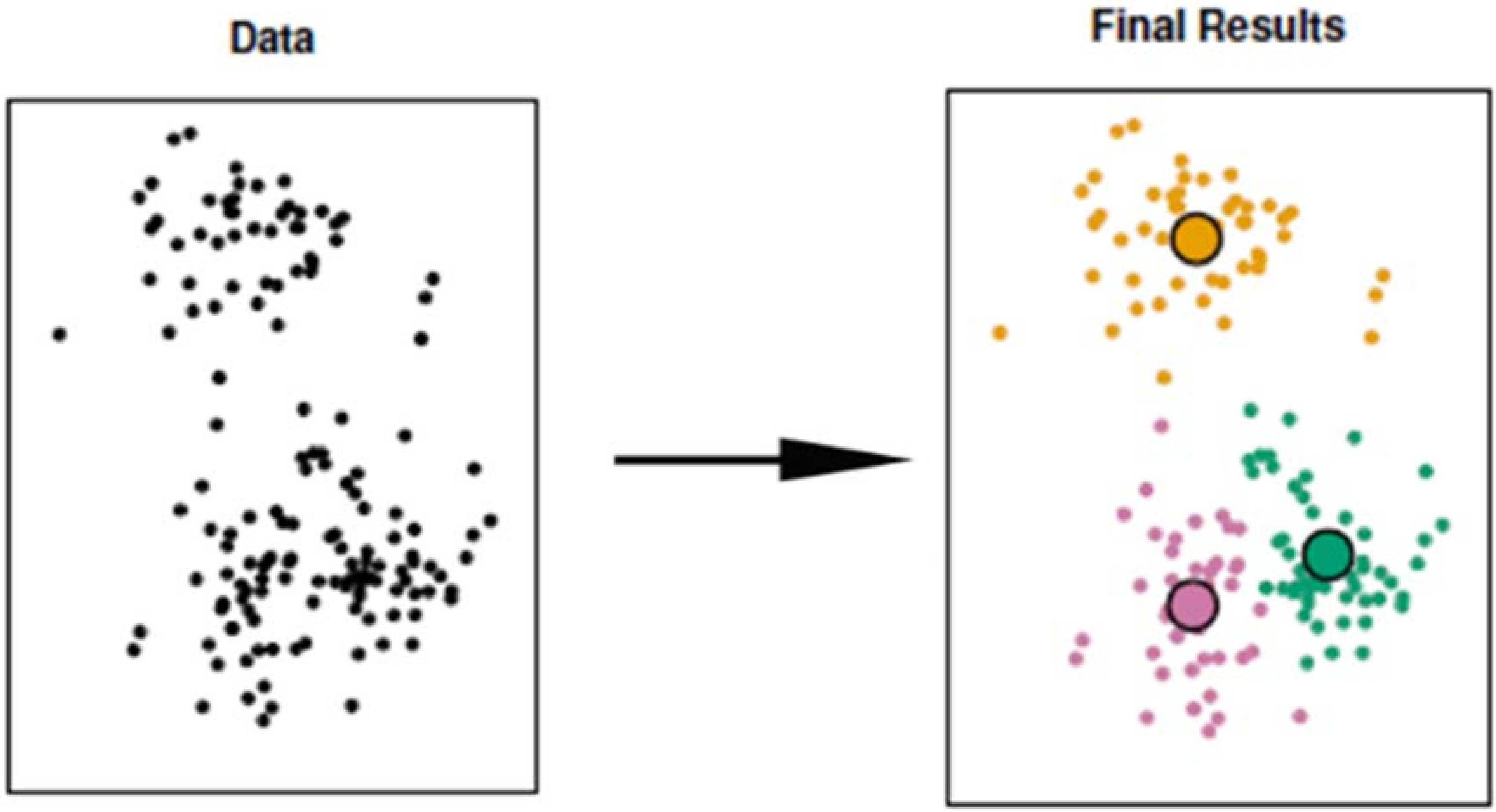

Figure 2. Classification process of unlabeled data

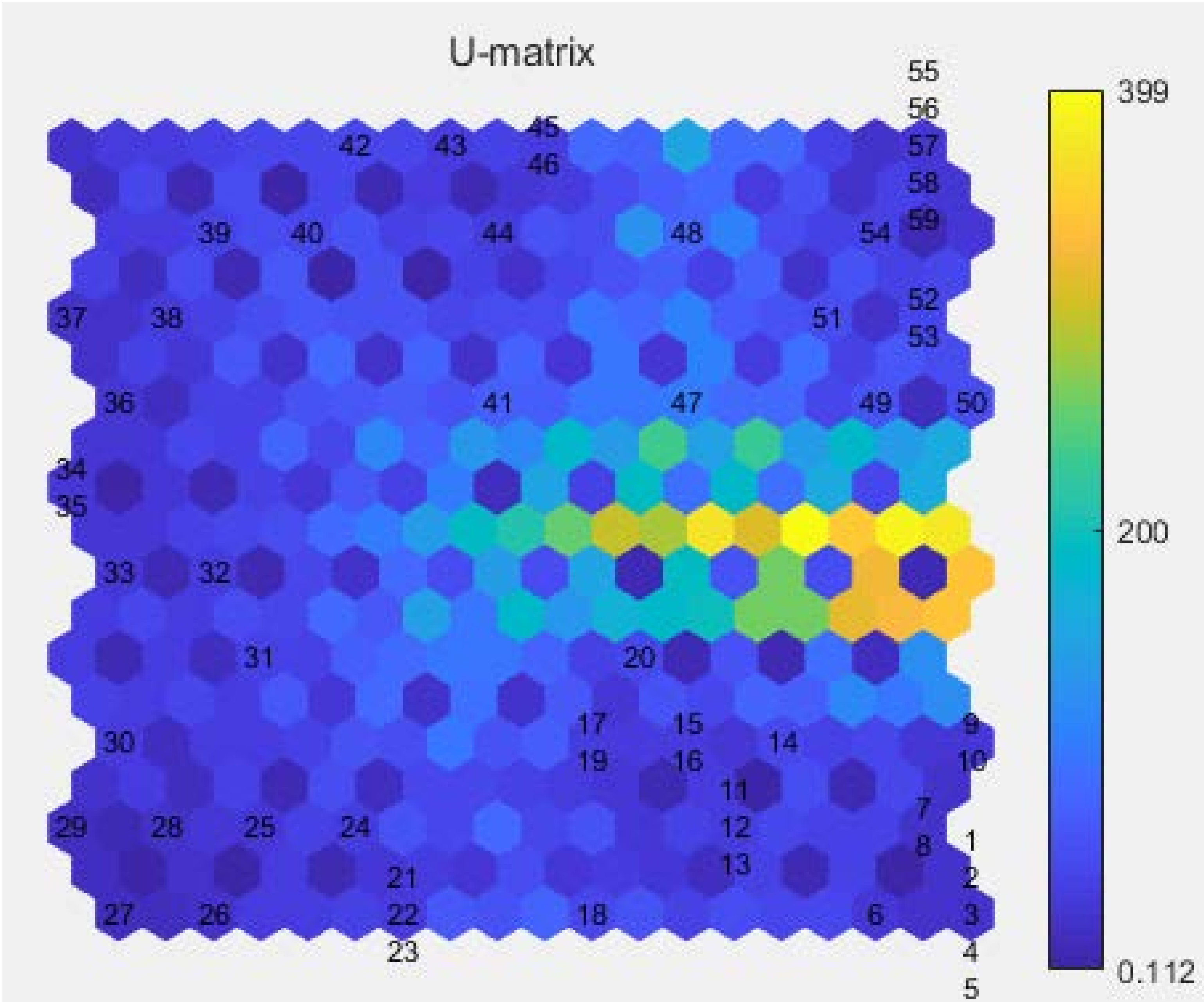

Figure 3. Classification with Self-Organizing Map

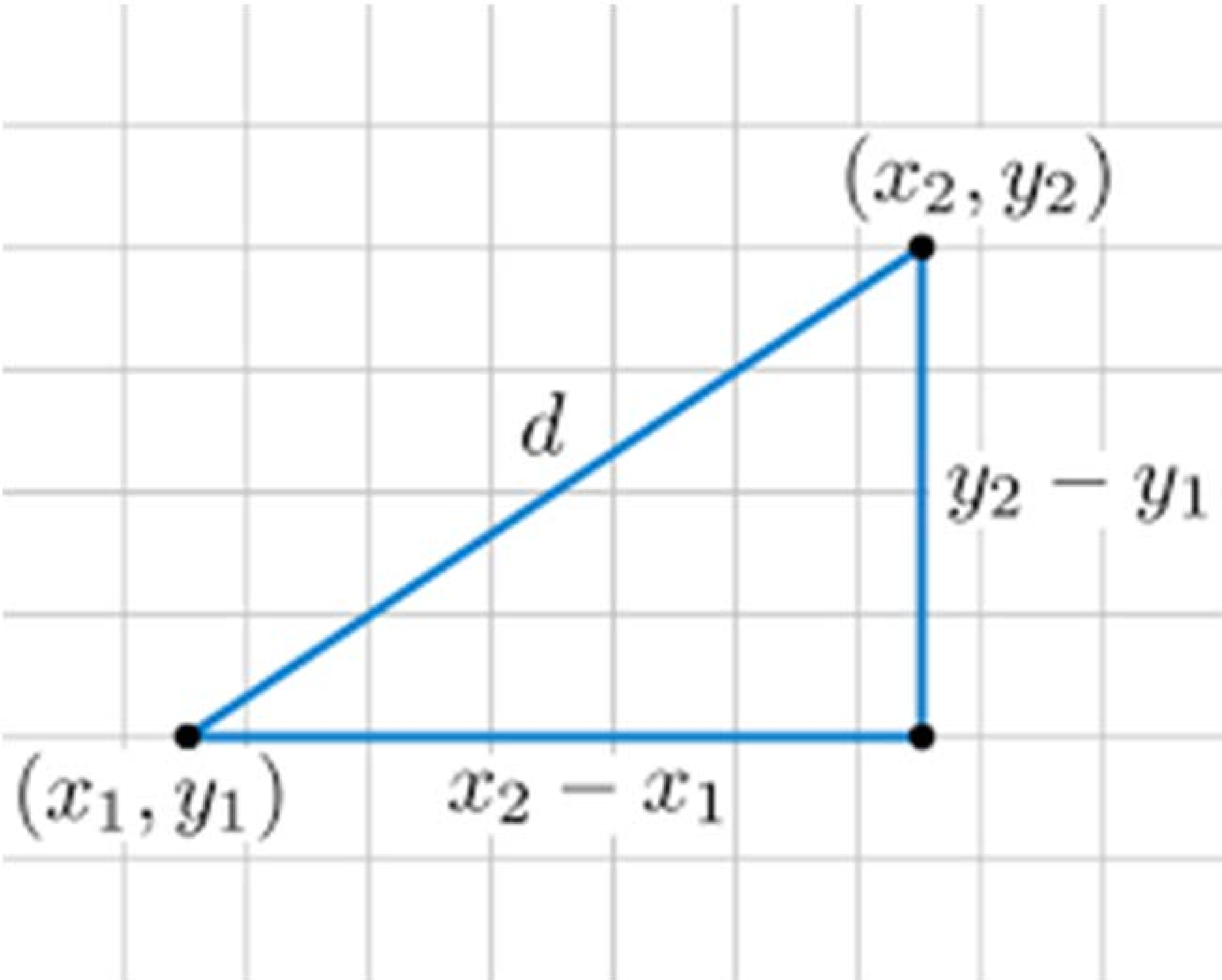

Figure 4. Euclidean distance

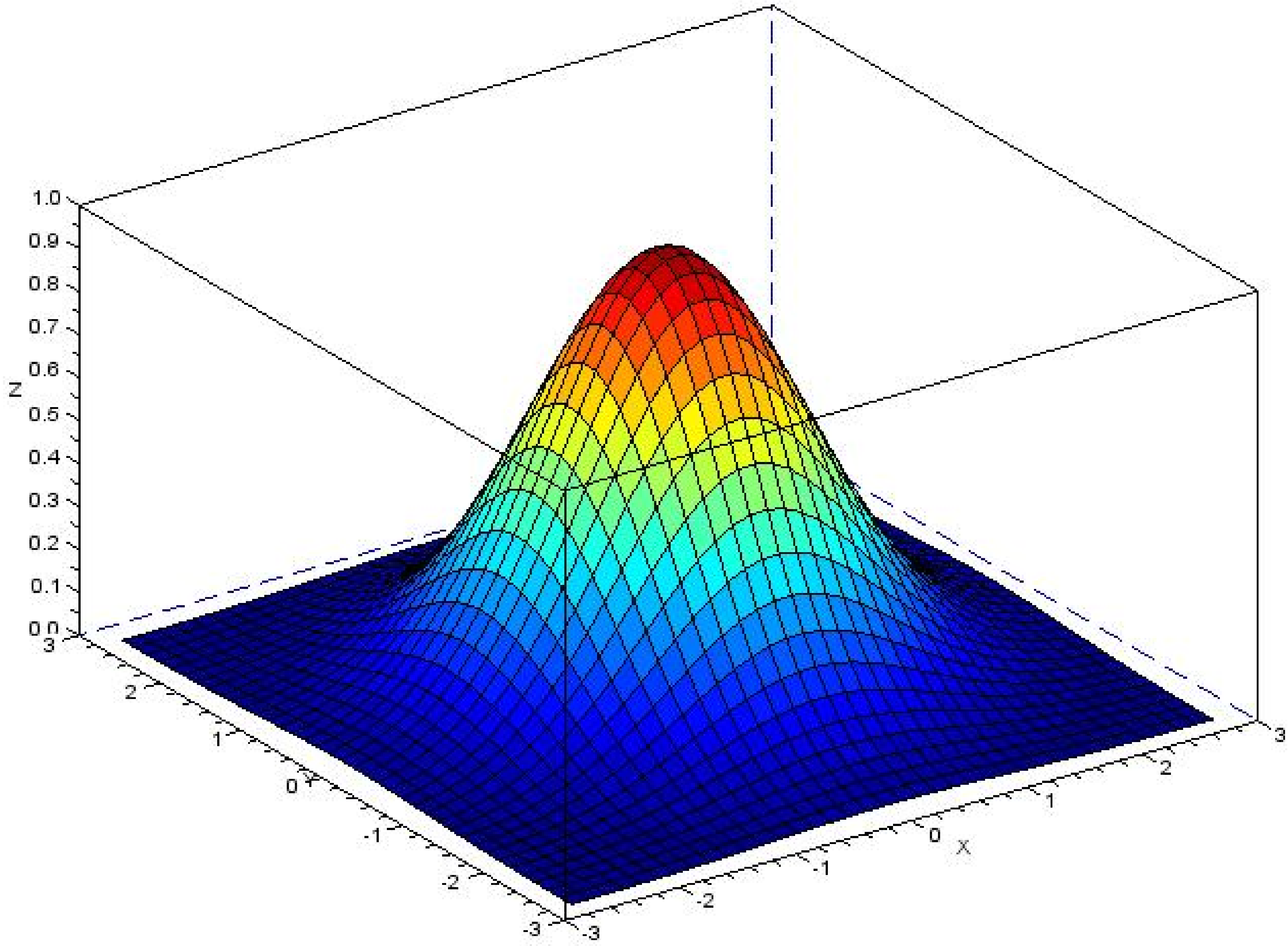

Figure 5. Gauss function

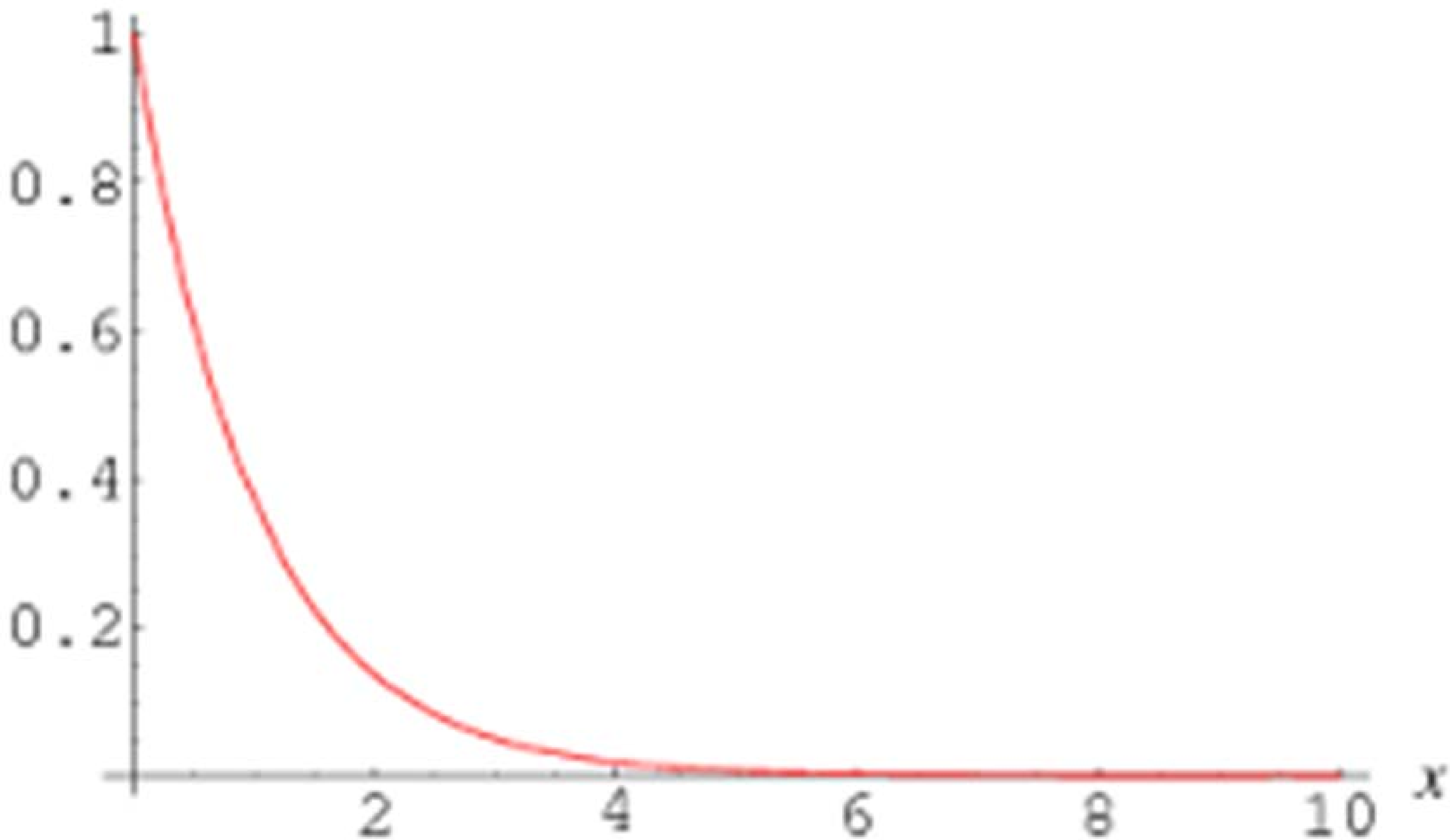

Figure 6. Learning rate

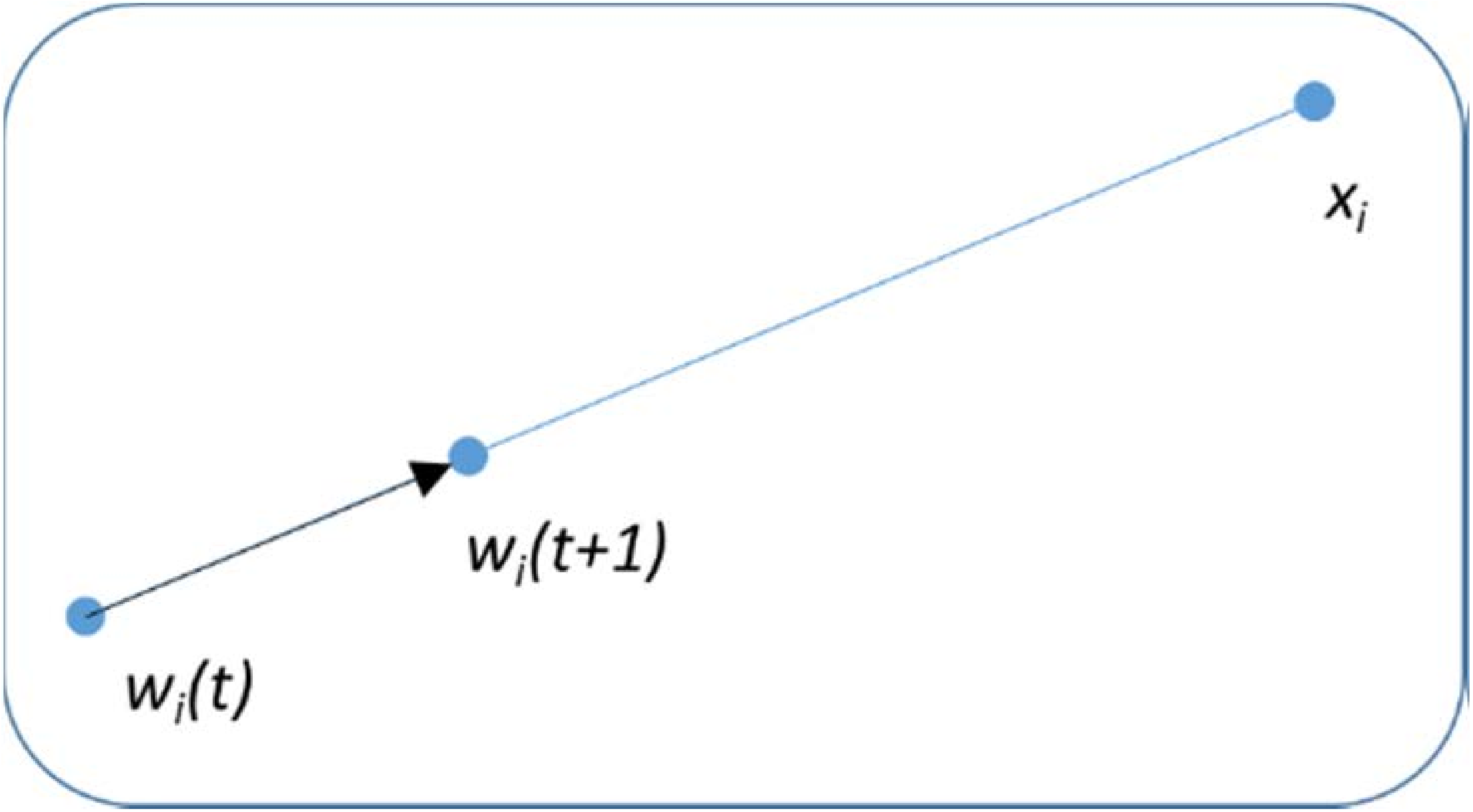

Figure 7. Adaptation process

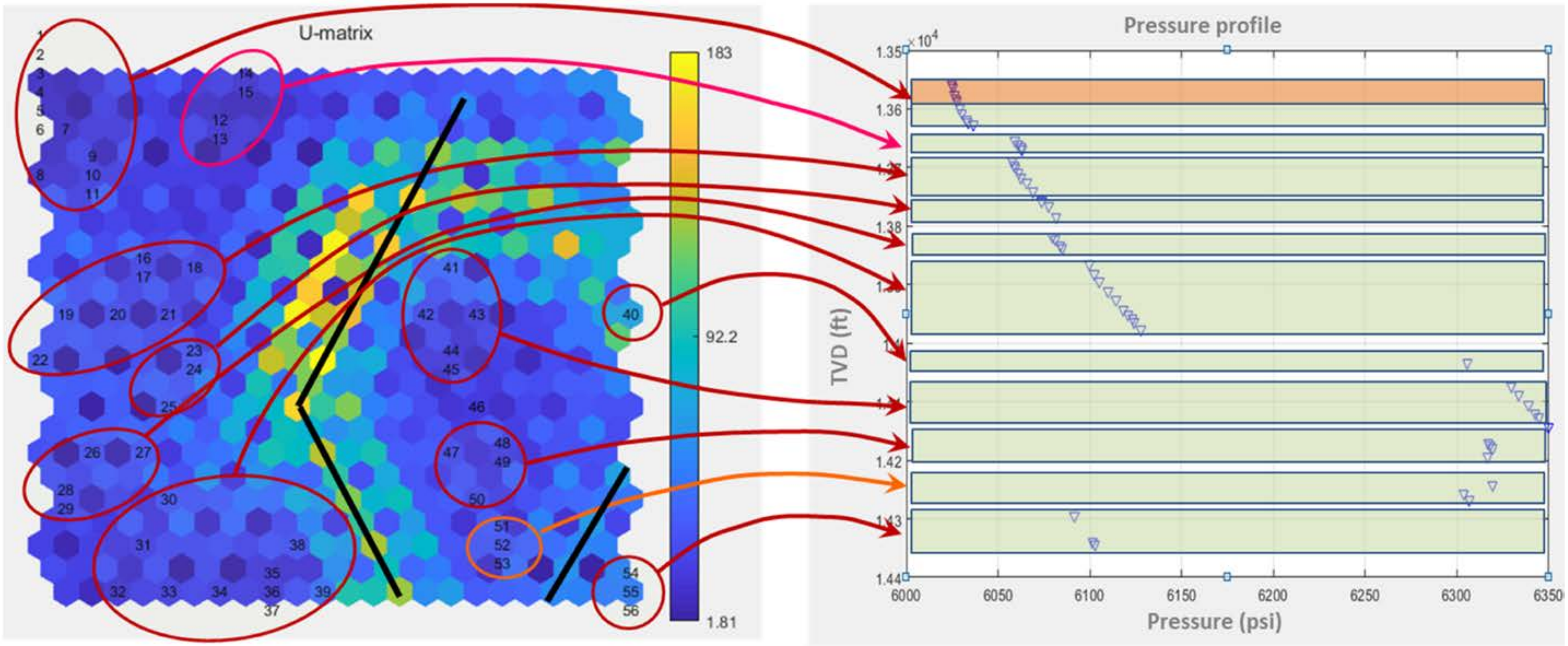

Figure 8. (a) SOM; (b) Pressure profile. Hydraulic flow unit identification in example 1

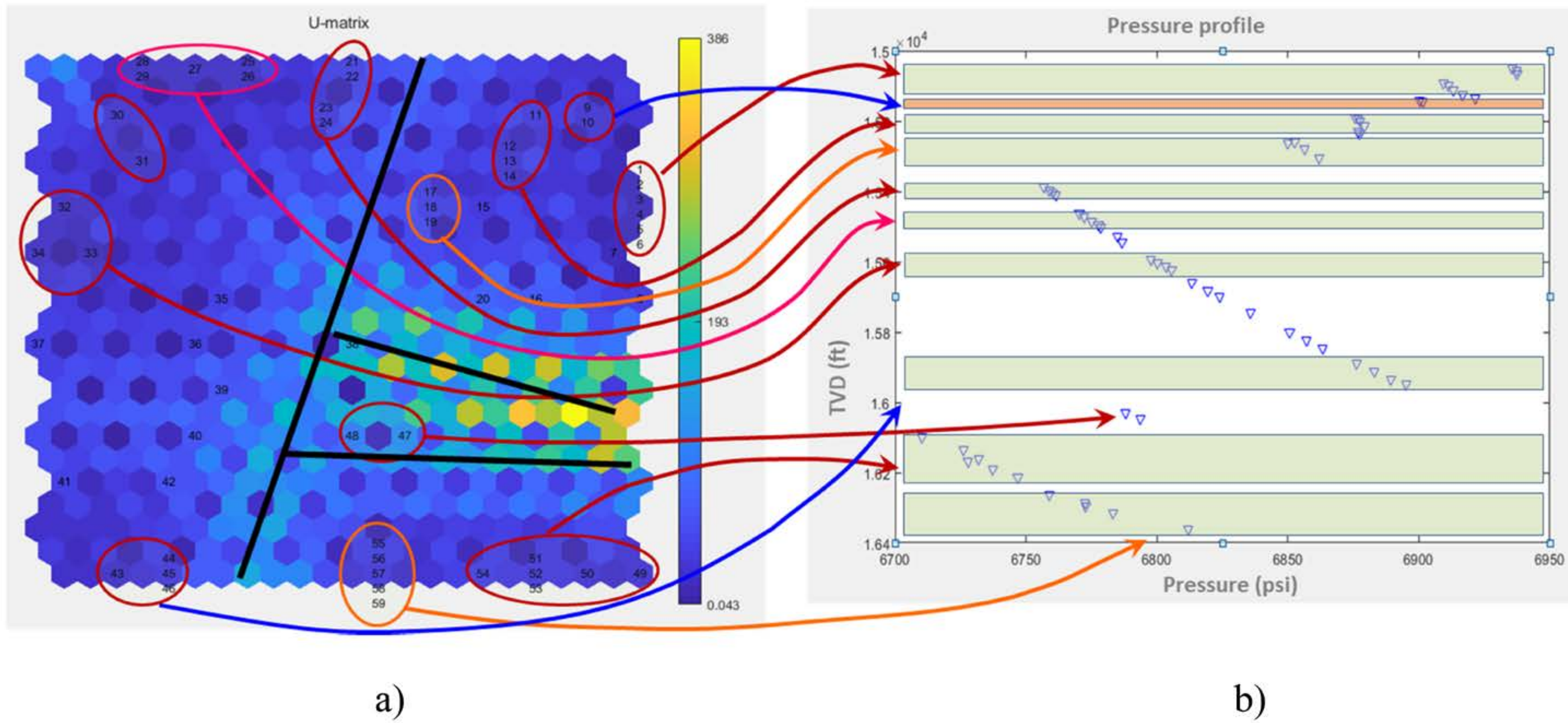

Figure 9. (a) SOM; (b) Pressure profile. Hydraulic flow unit identification in example 2.

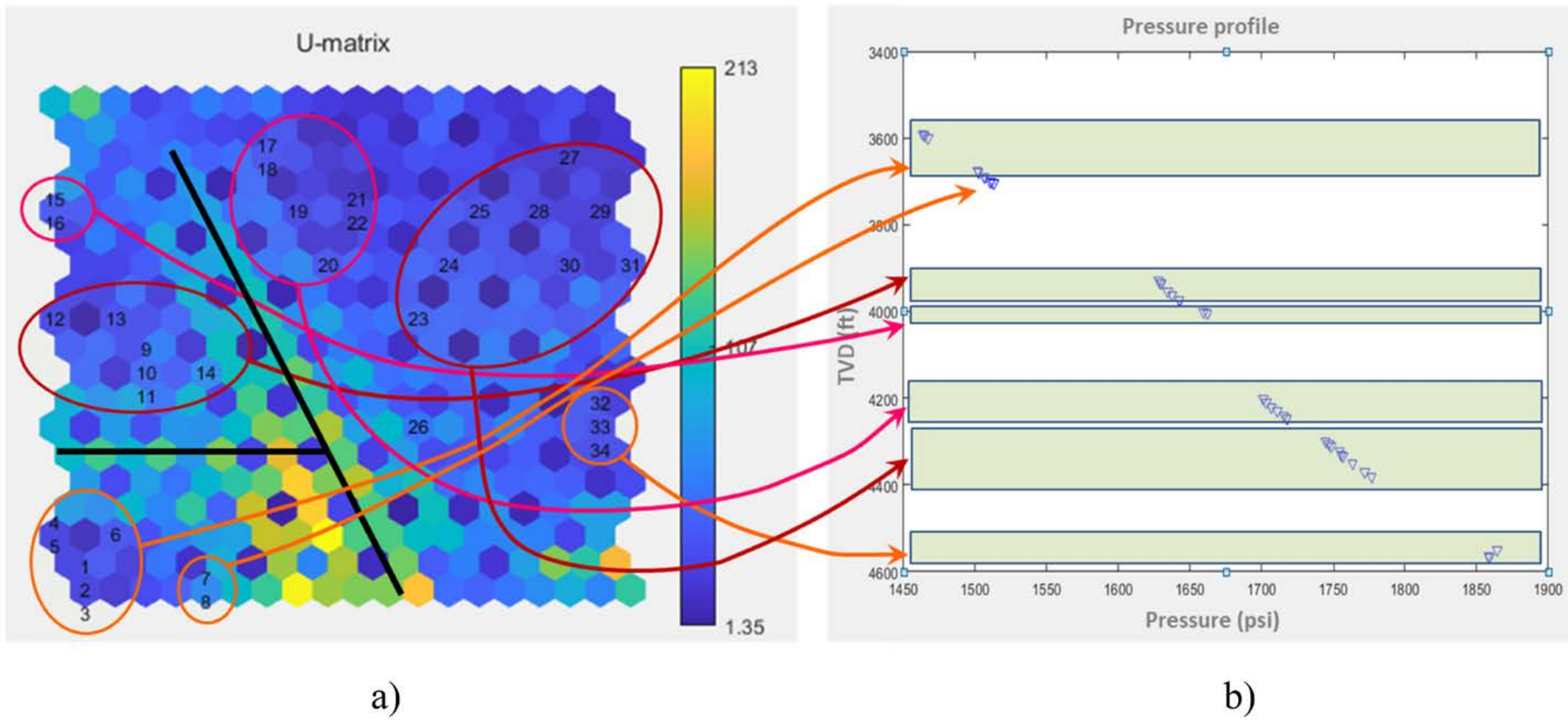

Figure 10. (a) SOM; (b) Pressure profile. Hydraulic flow unit identification in example 3.

SOM	α	r	Size	Initialization of SOM	Topology	Average distance to BMU	Topographic error
1	0.1	2	6x6	Random weight	Hexagonal	39.6858	0.2203
2	0.1	2	8x8	Random weight	Hexagonal	27.0665	0.3220
3	0.1	2	10x10	Random weight	Hexagonal	22.2155	0.1864
4	0.1	2	12x12	Random weight	Hexagonal	15.2526	0.1542
5	0.1	2	14x14	Random weight	Hexagonal	14.8312	0.3729
6	0.1	1	10x10	Random weight	Hexagonal	23.1025	0.5085
7	0.2	2	10x10	Random weight	Hexagonal	16.5107	0.2203

Table 1. Comparison results for different SOM characteristics.

SOM	α	r	Size	Initialization of SOM	Topology	Average distance to BMU	Topographic error
1	0.1	2	6x6	Random weight	Hexagonal	32.9814	0.2647
2	0.1	2	8x8	Random weight	Hexagonal	21.9935	0.3529
3	0.1	2	10x10	Random weight	Hexagonal	17.0162	0.5000
4	0.1	2	12x12	Random weight	Hexagonal	16.6751	0.6765
5	0.1	2	14x14	Random weight	Hexagonal	15.2521	0.8824
6	0.1	1	10x10	Random weight	Hexagonal	21.2789	0.2647

Table 2. Comparison results for different SOM characteristics with reduced number of observations.

Click to view slides of presentation.