NOVA IMS Information Management School

Data Mining

Self-Organizing Maps

24/11/2021

NOVA-IMS

Fernando Lucas Bação

bacao@isegi.unl.pt

http://www.isegi.unl.pt/fbacao

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

1

NOVA IMS Information Management School

AGENDA

- Cluster analysis
  - Clustering techniques
    - Self-organizing maps

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

2

NOVA IMS Information Management School

Clustering

3

NOVA IMS Information Management School

Clustering

- Grouping
  - Clustering
    - Hierarchical Methods
    - Partition Methods
    - Density-based
    - Self-Organizing Maps
  - Quartile-based clustering
    - Quartiles
      - Variables
      - Through time
    - RFM Analysis
      - Migration Matrices

4

NOVA IMS Information Management School

Self-Organizing Maps

5

NOVA IMS Information Management School

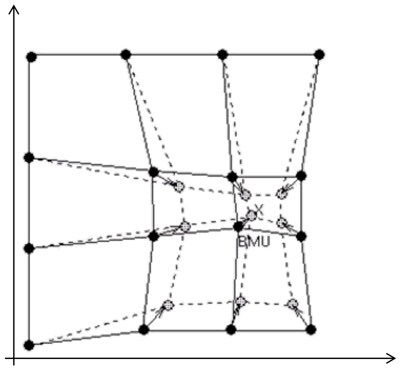
Self-Organizing Maps

- Unsupervised neural networks;
- Closely related to clustering;
- The inputs are connected to a two-dimensional (it may have several dimensions) matrix of units (neurons);
- Each unit is connected to its neighbors.
- What is its use?
  - Multidimensional data visualization;
  - Cluster detection;
  - Market segmentation;
  - Outlier detection;
  - Solve TSP, robot control, alarm detection, etc., etc., etc.

6

Self-Organizing Maps

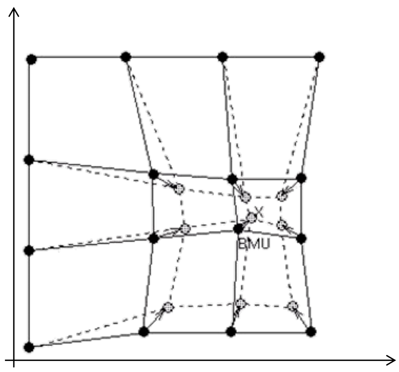
- Each neuron is a **vector in the input space**, just as the data patterns;
- During training, neurons are **pulled** to the positions of the input data, **dragging** with them their neighbors in the output space;
- The map can be seen as a **rubber sheet**, stretched and twisted, so that it passes in (or at least near) the data patterns.



7

Self-Organizing Maps

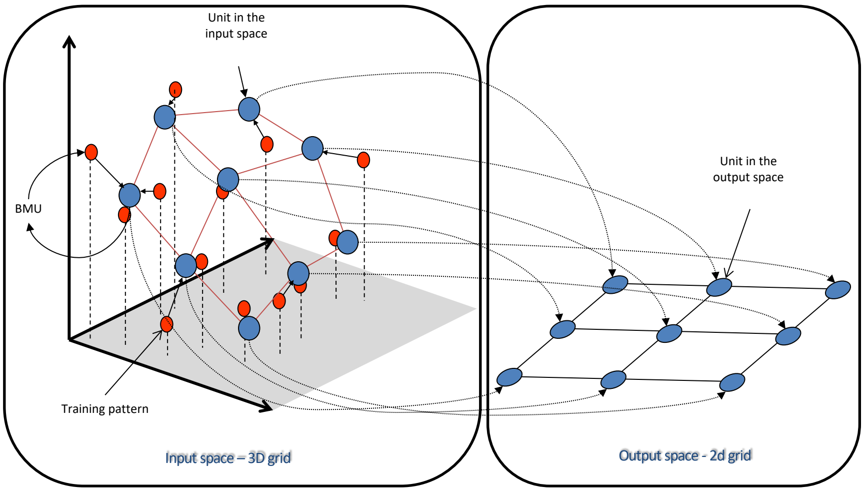
- Input patterns are compared with all neurons and the **closest is considered to be the winning neuron.**
- We consider that the input pattern is **mapped to the winning neuron**.
- The **winner is updated** (so that it resembles even more the data pattern that it represents), and its neighbors are also updated a little.
- There is always a slight difference between the data and the neurons that represent them. That difference is the **quantization error**.



8

NOVA IMS Information Management School

Self-Organizing Maps



Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

9

NOVA IMS Information Management School

Self-Organizing Maps

- **Comparisons with biology**
  - Biological systems have to use some kind of self-organization and adaptation.
  - There is evidence of:
    - Layered structure in the brain;
    - Those layers seem to spatially organize the information;
    - Similar “Concepts” are mapped to adjacent areas;
    - Experimental work with animals points to an organization similar to SOM in the visual cortex.

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

10

NOVA IMS Information Management School

The SOM Algorithm

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

11

NOVA IMS Information Management School

Self-Organizing Maps

- **SOM Algorithm:**
  - How does SOM processes data?

Video

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

12

Self-Organizing Maps

- **SOM Algorithm:**
  - What happens in the Input space?
  - If we only have two variables, optimization will look like this

SOM 1-dimensional

SOM 2-dimensional

NOVA IMS Information Management School

Self-Organizing Maps

- **SOM Algorithm:**
  - What happens in the output space

Color Demo

  - Suppose we want to group cells according to their rgb code (red, green and blue)
  - Each individual (color) represents a particular combination of rgb intensities
  - In this demo, we can see how the output space is transformed as individuals are presented to the network

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA IMS Information Management School

Self-Organizing Maps

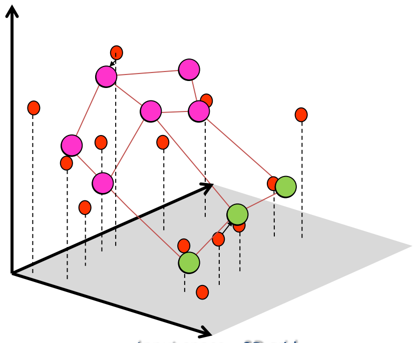
- **SOM Algorithm:**
  - Key outputs of the SOM:
    - U-Matrices;
    - Component plans;
    - Hit Plots.

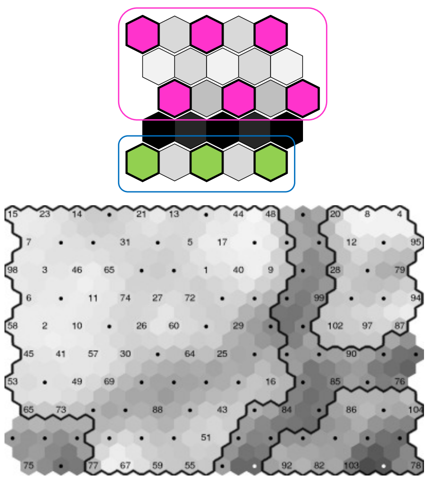
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

15

NOVA IMS Information Management School

Self-Organizing Maps – U-Matrix



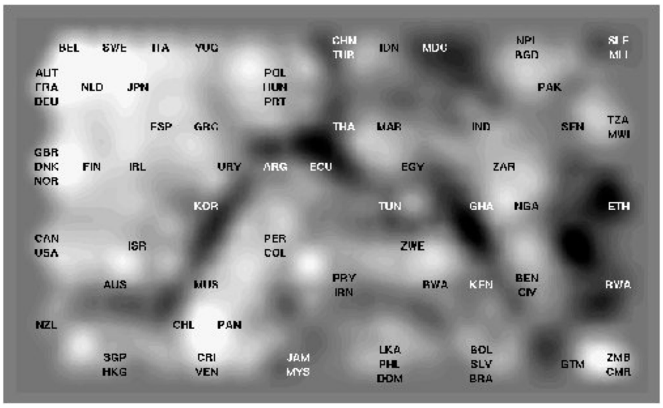


Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

16

NOVA IMS Information Management School

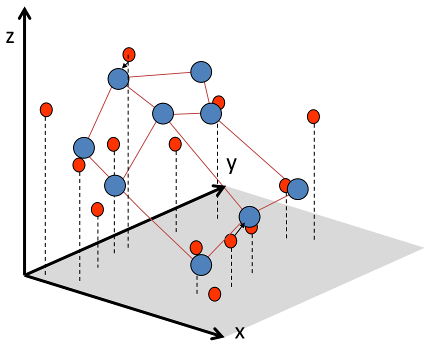
Self-Organizing Maps – U-Matrix



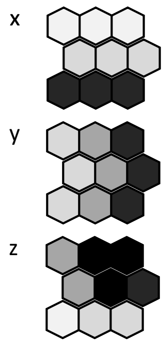
Kaski, S. and Kohonen, T. 1996. Exploratory Data Analysis by the Self-Organizing Map: Structures of Welfare and Poverty in the World. In: Apostolos, P. N., Refenes, Y. A., Moody, J. and Weigend, A. (eds.) Neural Networks in Financial Engineering. Singapore: World Scientific, 498-507.

NOVA IMS Information Management School

Self-Organizing Maps – Component Planes

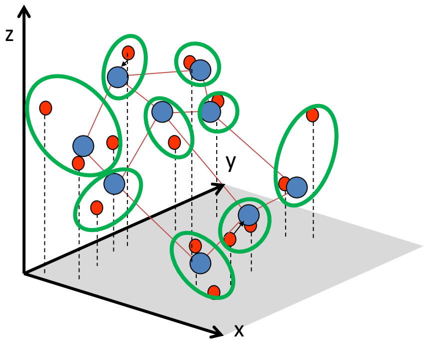


Input space – 3D grid

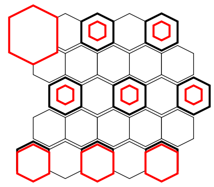


NOVA IMS Information Management School

Self-Organizing Maps – Hit Maps



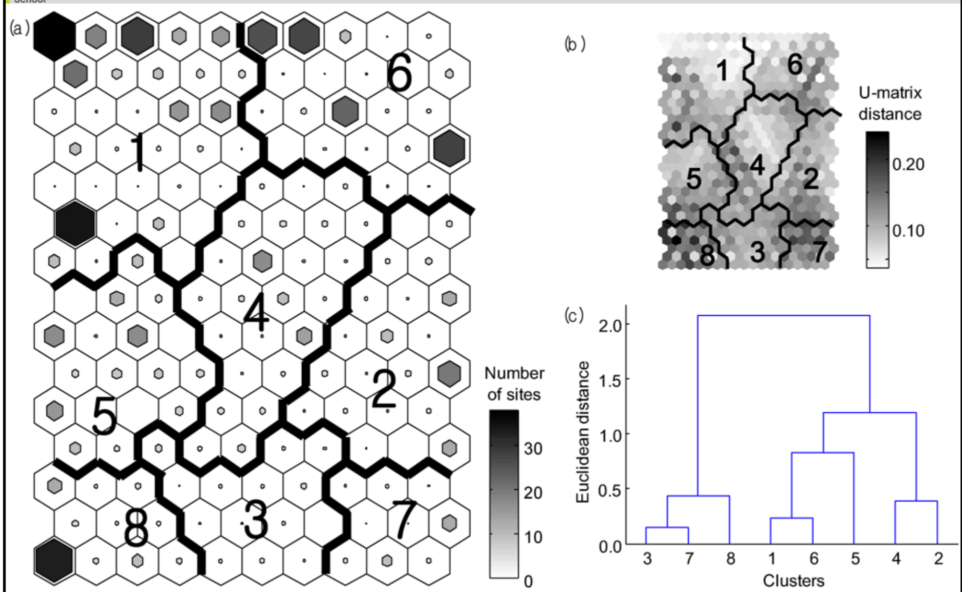
Input space – 3D grid



19

NOVA IMS Information Management School

Self-Organizing Maps – Interpreting the U-Matrix



20

NOVA IMS Information Management School

Self-Organizing Maps

- **SOM Algorithm:**

Step 0: Randomly initialize the weights w_{ij}
Set the neighborhood topological parameters
Set the learning rate

Step 1: While stop condition false, do steps 2-7

Step 2: For each input vector x, do steps 3-5

Step 3: For each j, execute:

$$D(j) = \Sigma (w_{ij} - x_i)^2$$

Step 4: Find the unit that minimizes D(j)

Step 5: For every j unit within the predefined and for all the i:

$$w_{iJ}\,(new) = w_{iJ}\,(old) + \alpha\,[x_i - w_{iJ}\,(old)]$$

Step 6: Update the learning rate

Step 7: Update (reduce) the radius of the topological neighborhood

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

21

NOVA IMS Information Management School

Self-Organizing Maps

- **SOM Algorithm:**

Vectors to classify:

(1, 1, 0, 0); (0, 0, 0, 1); (1, 0, 0, 0); (0, 0, 1, 1)

Maximum number of clusters to form:

m = 2

Learning rate:

$\alpha(0) = .6,$ $\alpha(t+1) = .5\,\alpha(t)$

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

22

NOVA IMS Information Management School

Self-Organizing Maps

- **SOM Algorithm:**

Step 0: Initial matrix of weights

(0.2, 0.6, 0.5, 0.9);

(0.8, 0.4, 0.7, 0.3);

Initial radius: R = 0

Initial learning rate:

$\alpha(0) = .6$

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

23

NOVA IMS Information Management School

Self-Organizing Maps

- **SOM Algorithm:**

Step 1: Initialize training

Step 2: first vector (1, 1, 0, 0);

Step 3:

$D(1) = (.2-1)^2 + (.6-1)^2 + (.5-0)^2 + (.9-0)^2$

$= 1.86;$

$D(2) = (.8-1)^2 + (.4-1)^2 + (.7-0)^2 + (.3-0)^2$

$= 0.98$

Vectors of weights:

(0.2, 0.6, 0.5, 0.9)

(0.8, 0.4, 0.7, 0.3)

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

24

NOVA IMS Information Management School

Self-Organizing Maps

- **SOM Algorithm:**

Step 4: The input vector is closest to unit 2, therefore

$$j = 2 = (0.8, 0.4, 0.7, 0.3)$$

Step 5: The weights of the winning unit are adjusted

$$w_{i2}\text{ (new)} = w_{i2}\text{ (old)} + .6\,[x_i - w_{i2}\text{ (old)}]$$

$$= .4\,w_{i2}\text{ (old)} + .6x_i$$

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

25

NOVA IMS Information Management School

Self-Organizing Maps

- **SOM Algorithm:**

Step 4: The input vector is closest to unit 2, therefore

$$j = 2 = (0.8, 0.4, 0.7, 0.3)$$

Step 5: The weights of the winning unit are adjusted

$$w_{i2}\text{ (new)} = w_{i2}\text{ (old)} + .6\,[x_i - w_{i2}\text{ (old)}]$$

$$= .4\,w_{i2}\text{ (old)} + .6x_i$$

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

26

NOVA IMS Information Management School

Self-Organizing Maps

- **SOM Algorithm:**

Thus, the weight matrix is adjusted:

.2	.92
.6	.76
.5	.28
.9	.12

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

27

NOVA IMS Information Management School

Self-Organizing Maps

- **SOM Algorithm:**

Thus, the weight matrix is adjusted:

.08	.92
.24	.76
.20	.28
.96	.12

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

28