

Mapas Auto Organizáveis
SOMs
Mapas de Kohonen

PSI 2886
Princípios de Neurocomputação
Edição de 2008

PSI 5886 – Prof. Emilio

Self Organizing Maps de Kohonen

- Estes slides foram usados nas aulas #8 e #9 na edição de 2008 da disciplina
- Vários slides aqui apresentados foram desenvolvidos conjuntamente com alguns de meus orientados ou usam imagens de seus trabalhos de pesquisa na EPUSP:
 - ✓ Clodis Boscarioli
 - ✓ Leandro Augusto da Silva
 - ✓ Renato Sassi(vide www.lsi.usp.br/icone para trabalhos relacionados)

Diversity of Neural Architectures

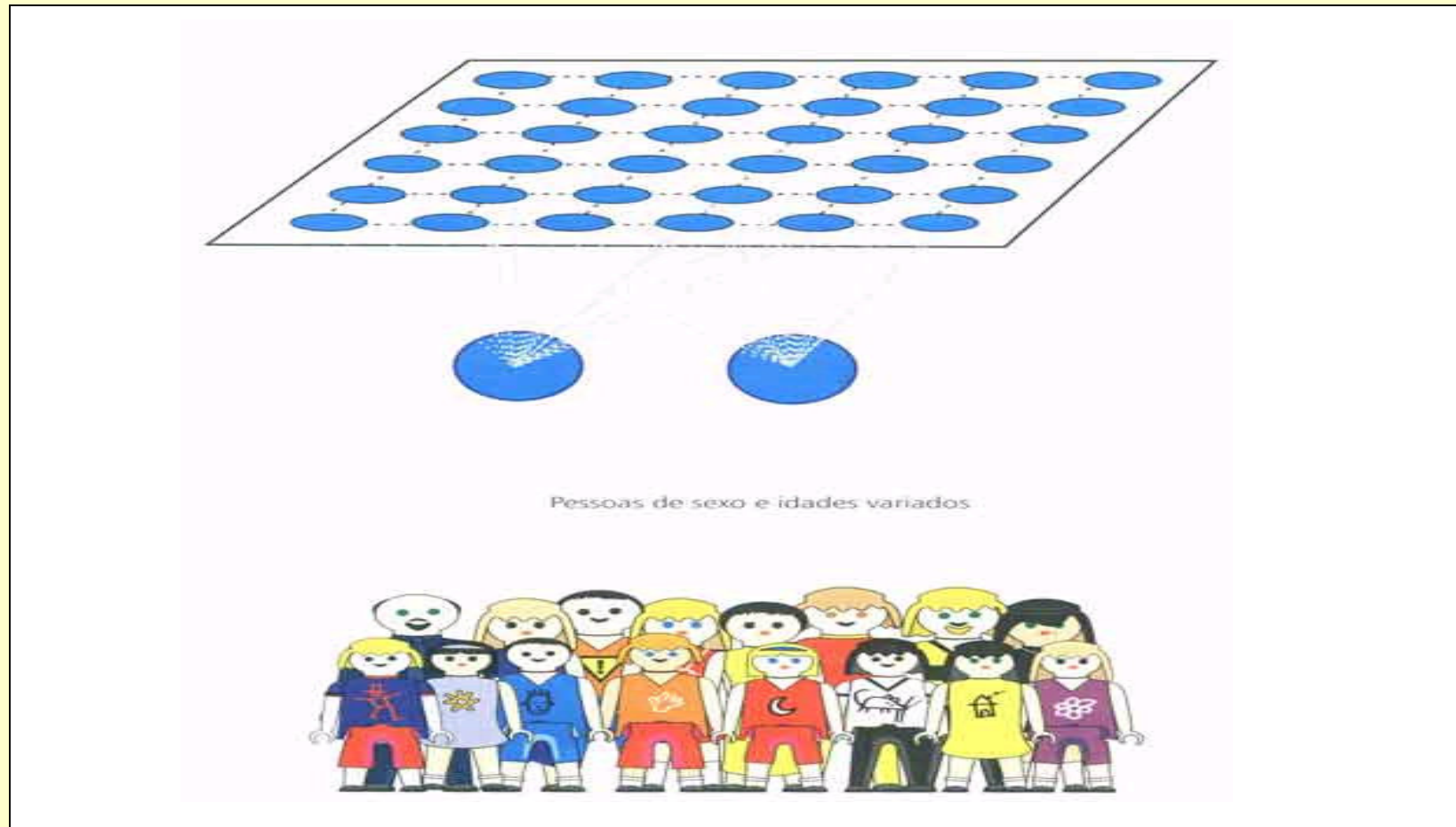
- Is “Perceptrons & Multi Layer Perceptrons (MPLs)” a synonym of Artificial Neural Networks?

No!!! There are many other very distinct topologies for arrangement of nodes

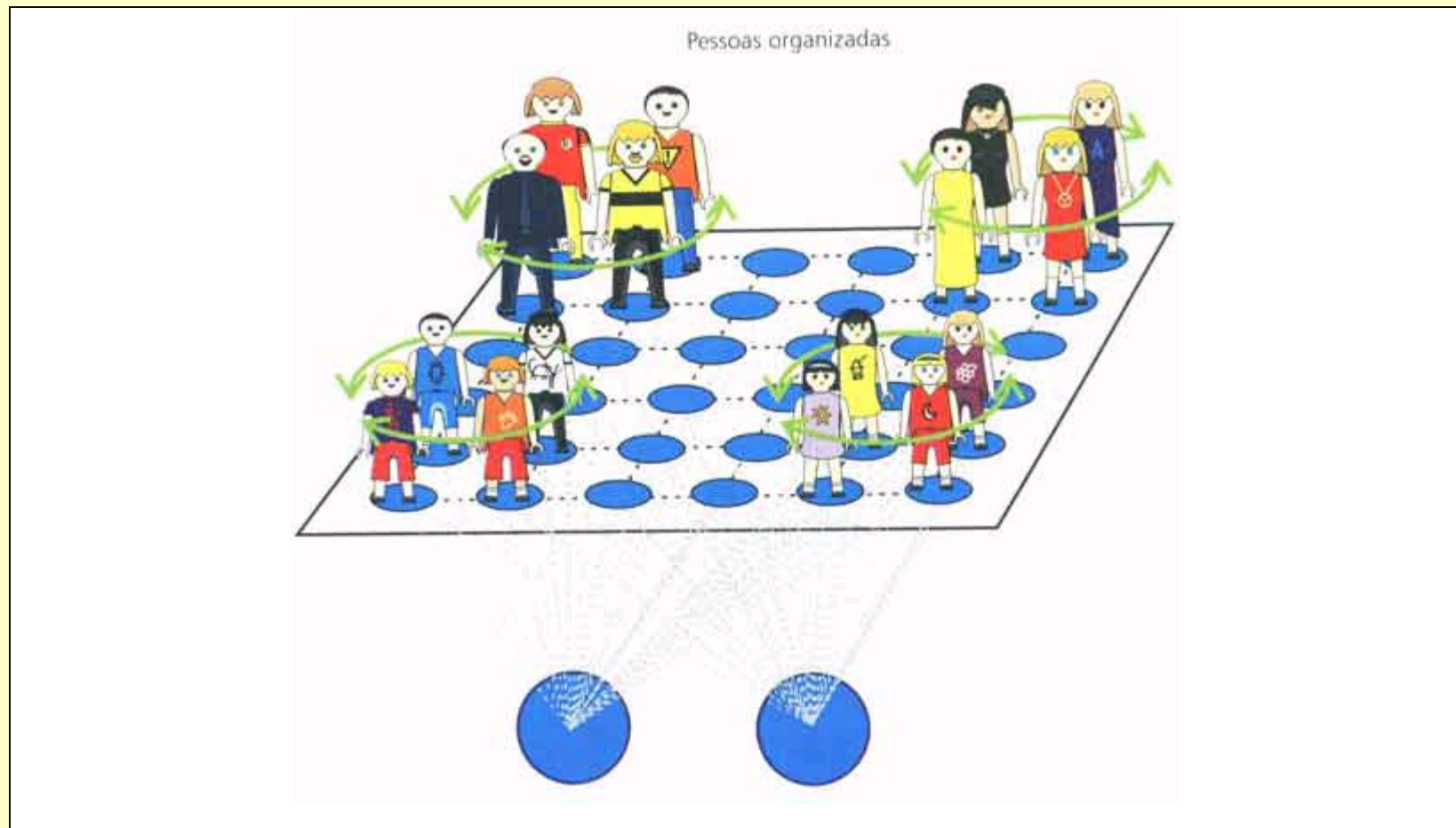
- Is Error Back Propagation a synonym of learning in artificial neural networks?

No!!! The weights in other architectures (non MLPs) are defined through very different strategies which are nor guided by gradient descent

Kohonen Architecture (SOM) for datamining



Samples are grouped around prototypes in a 2-d grid



Rede de Kohonen (SOM)

Arquitetura

Aprendizado

Aplicações

Arquiteturas de Kohonen (SOM) para clustering



Dados são agrupados em torno de protótipos organizados em espaço bidimensional



Conceitos principais em SOMs de (Teuvo) Kohonen

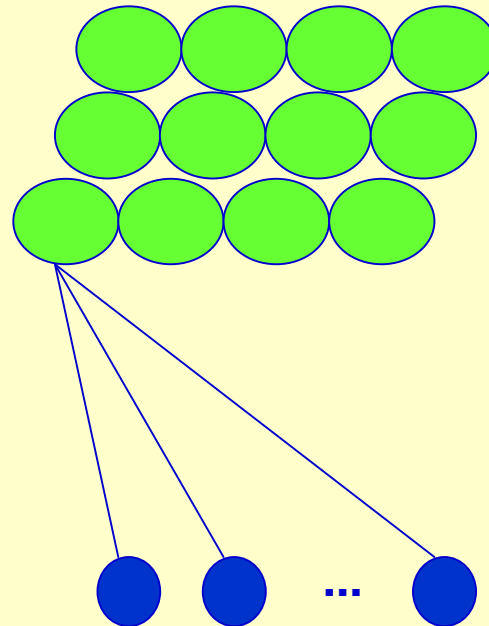
- Universo de população de dados massivos e multidimensionais (de alta dimensão) que se deseja visualizar e/ou organizar em classes / grupo
- Protótipos representando partes da população
- Protótipos organizados numa "grade", ou numa "fila", por semelhança
- Ordenação na grade
- Conceito de neurônio vencedor para definição do representante (aprendizado competitivo)

Arquiteturas de Kohonen (SOM) para clustering



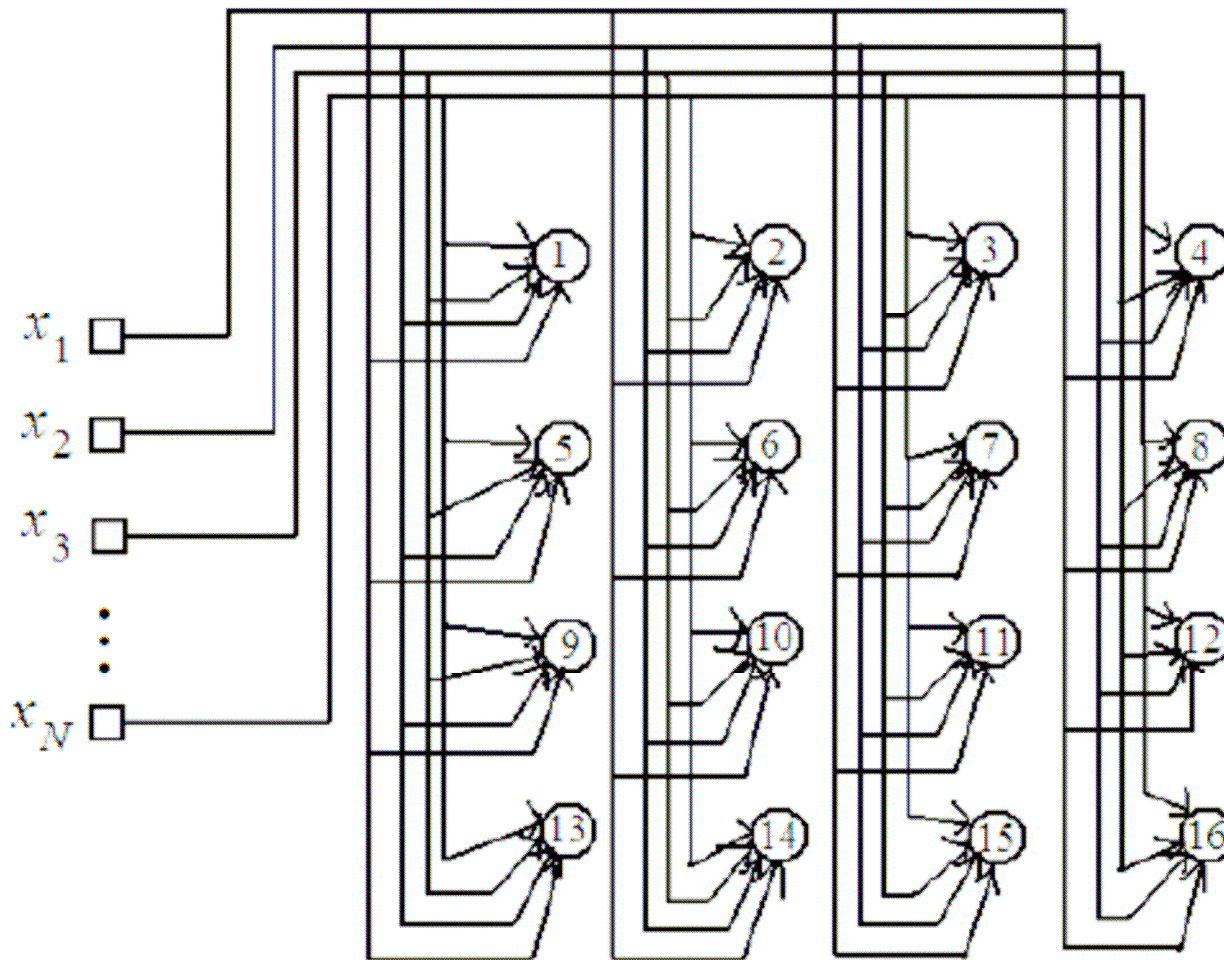
Arquitetura da rede SOM

**neurônios de saída
(cada neurônio tem N dimensões de entrada)**

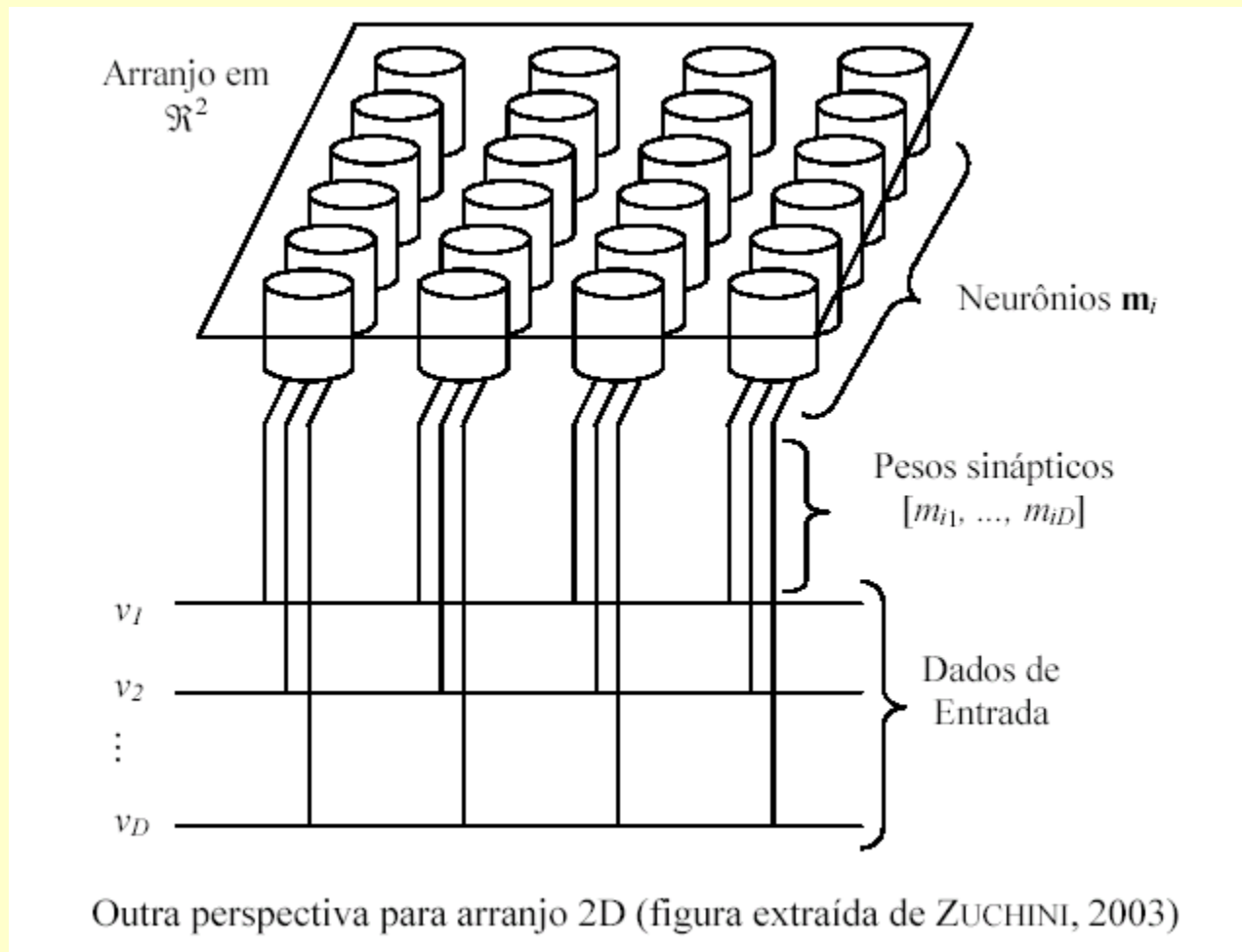


**padrão de entrada
(dimensão N)**

Arquitetura do SOM



Arrajo 2-D da Arquitetura SOM



Rede de Kohonen (SOM)

Arquitetura

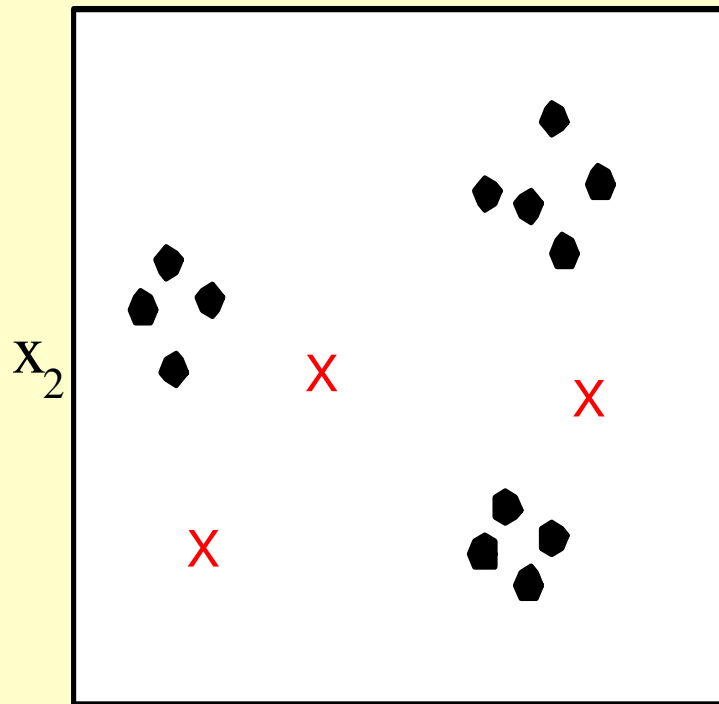
Aprendizado

Aplicações

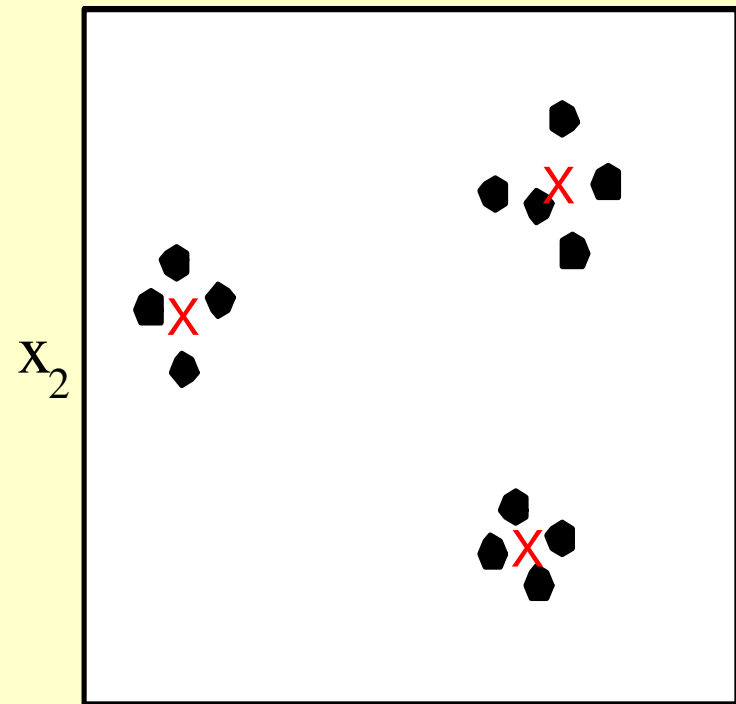
Redes SOM de Kohonen e o Aprendizado Não-Supervisionado

- O aprendizado não-supervisionado não requer saídas desejadas e por isso é conhecido pelo fato de não precisar de “professores” para o seu treinamento.
- Para o treinamento da rede, são usados apenas os valores de entrada. A rede trabalha essas entradas e se organiza de modo que acabe “classificando-as”, usando para isso, os seus próprios critérios.

Adaptation of the 2D-weights of prototypes in the Kohonen grid



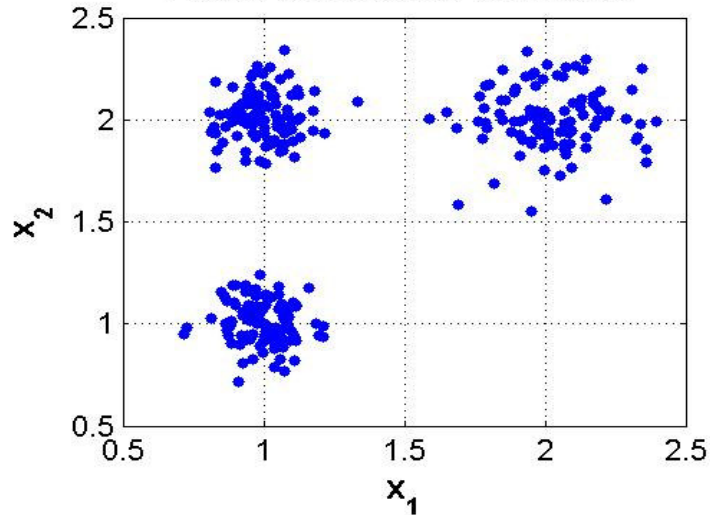
x_1
Início



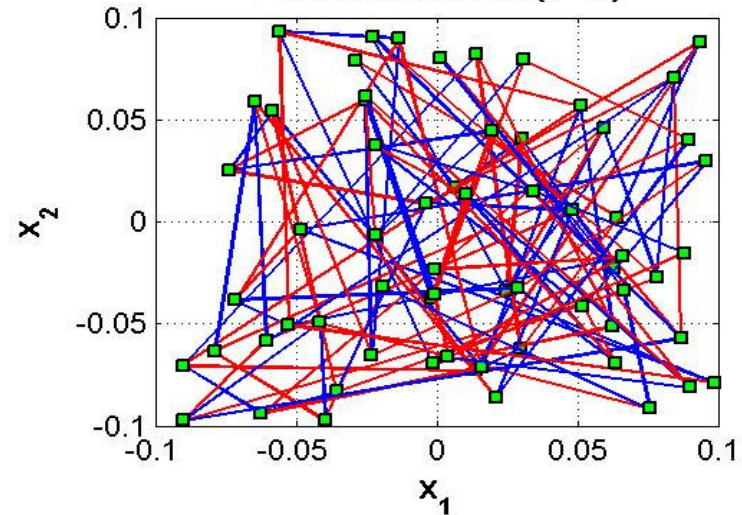
x_1
Fim

Aprendizado SOM

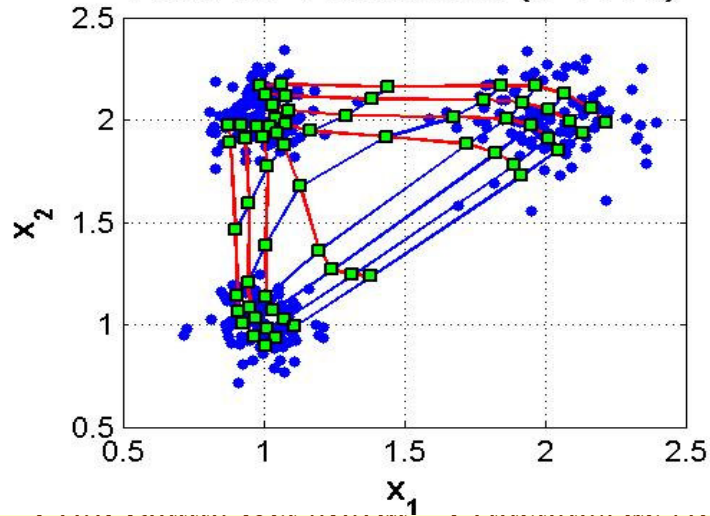
Base de dados artificial



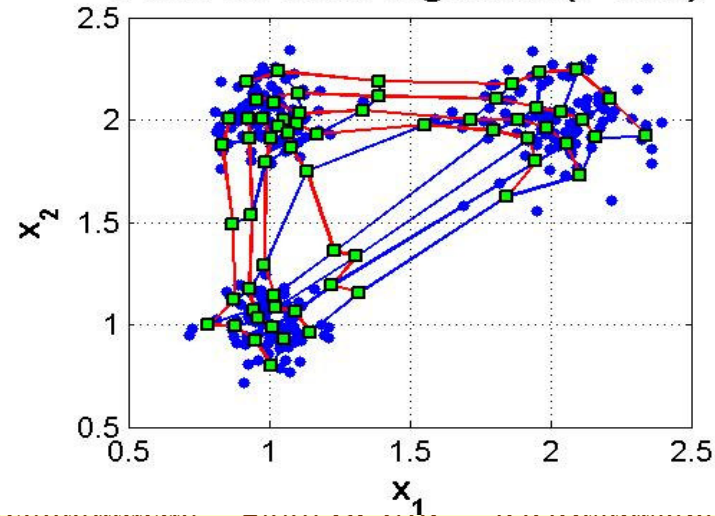
Pesos iniciais (t=0)



Fase de ordenacao (t=1000)



Fase de convergencia (t=500)



O Processo de Auto-organização

- Em linhas gerais:
 - Inicializa-se os vetores de pesos dos neurônios;
 - Apresenta-se dados de entrada à rede;
 - Seleciona-se o BMU para cada dado apresentado;
 - Atualiza-se o BMU e sua vizinhança;

Aprendizado SOM

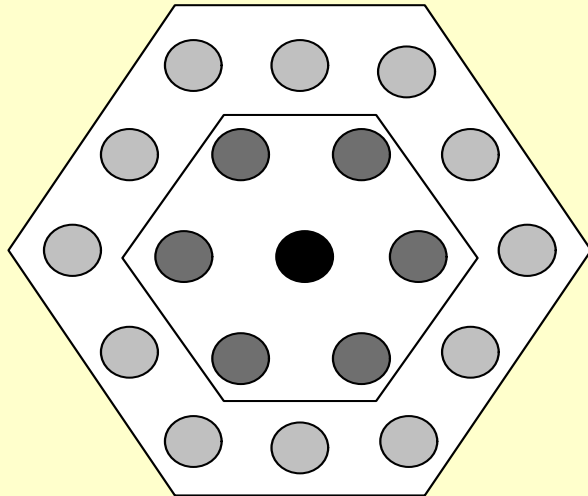
- **Inicialização:** defina o tamanho do mapa e seus respectivos pesos, \mathbf{w} , aleatoriamente. Defina a relação de vizinhança do mapa.
- **Ativação:** escolha um padrão de entrada, \mathbf{x} , do conjunto de treinamento.
- **Calcula a resposta do mapa:**

$$c = \underset{i}{\operatorname{argmin}} \{d(x(n), w_i)\}, \quad i \in \{1, \dots, M\}$$

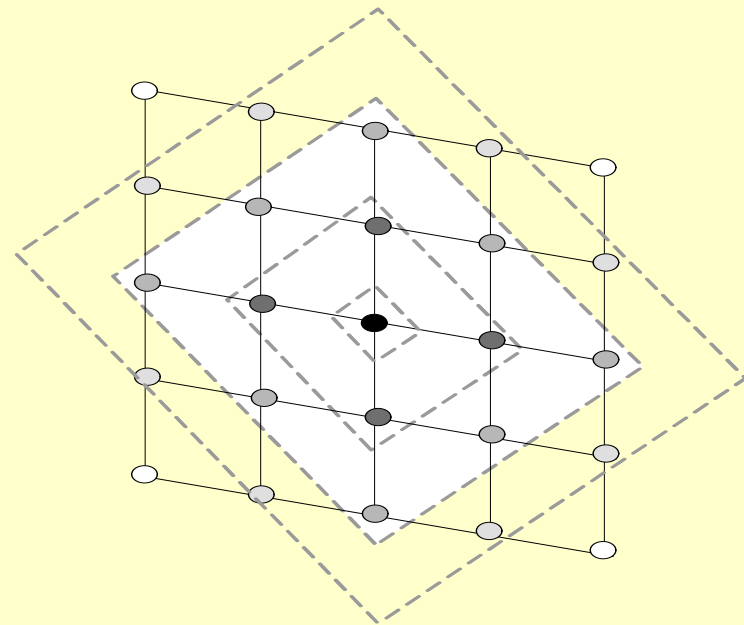
- **Adaptação:** $h_{ci}(t) = \exp^{-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}}$, $i \in N_c$

$$w_i(t + 1) = w_i(t) + \alpha(t) h_{ci}(t) \{d(x(n), w_i)\}$$

Vizinhança SOM

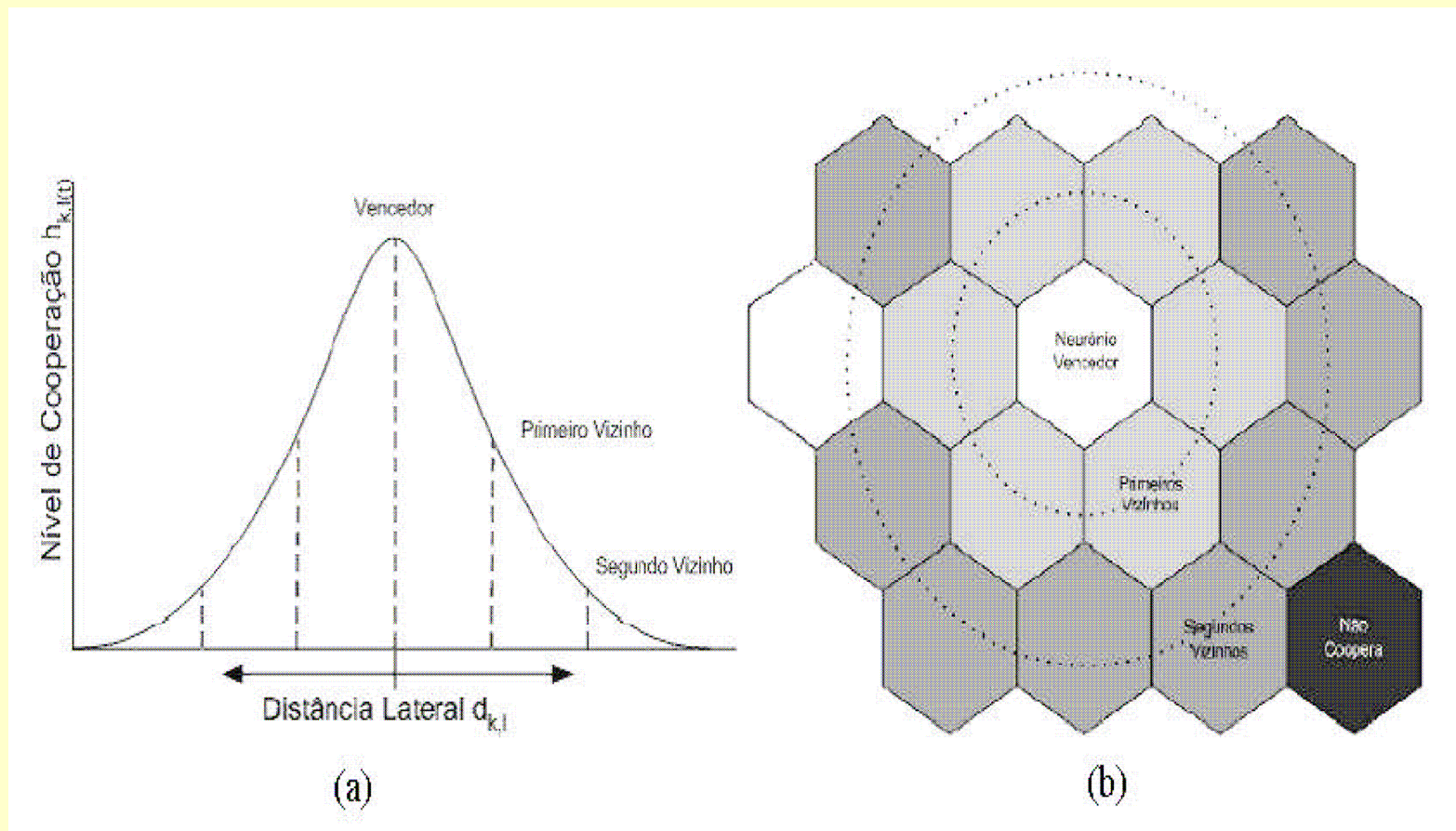


Vizinhança hexagonal



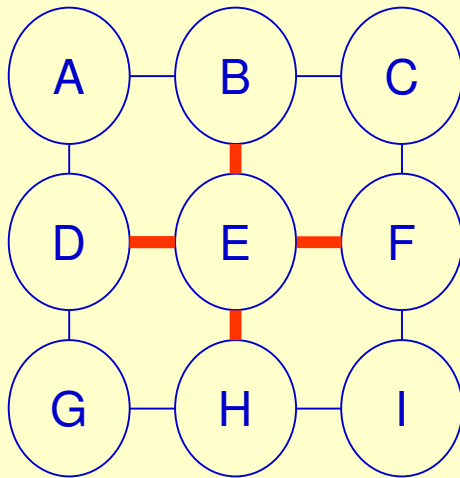
Vizinhança retangular

Vizinhança do SOM

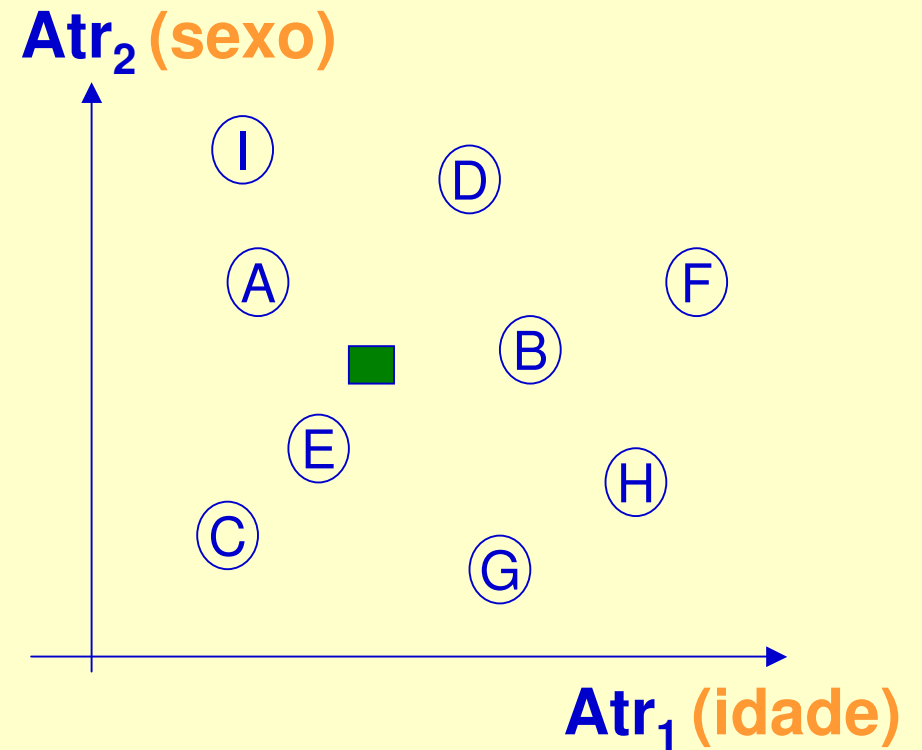


Aprendizado SOM

Espaço de protótipos



Espaço Atributos



Exemplo de adaptação em protótipos 2-D com inicialização não randômica / linear

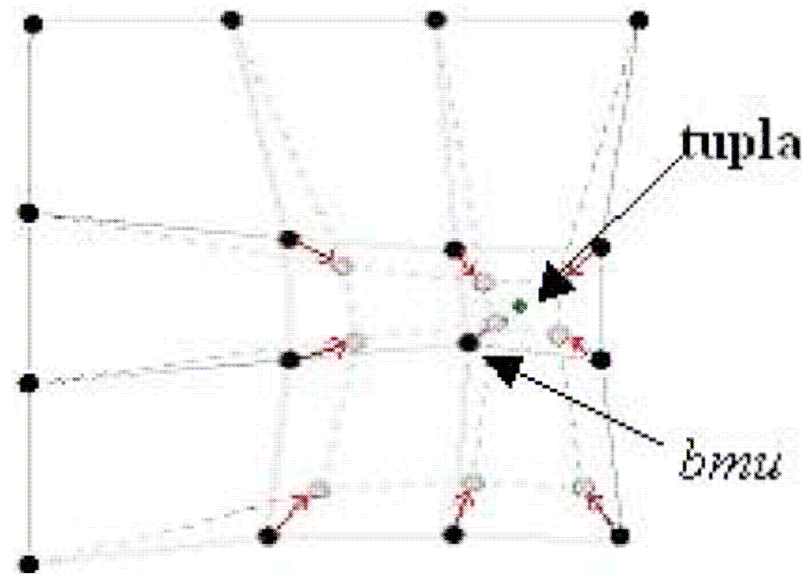
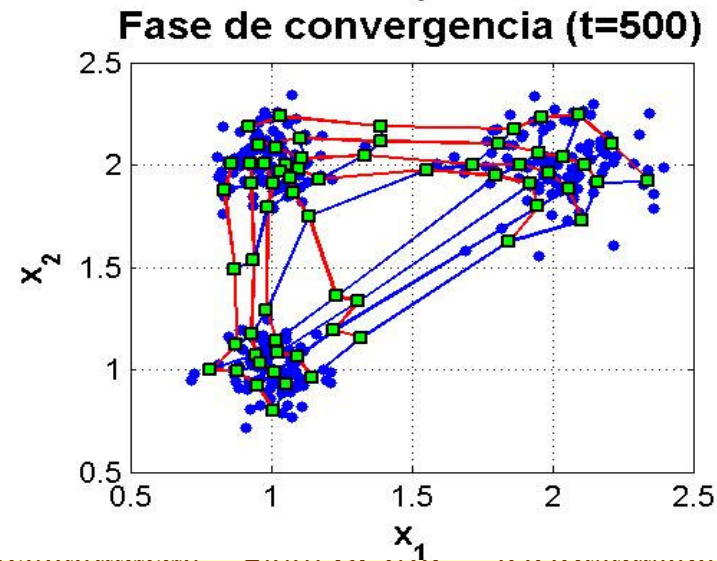
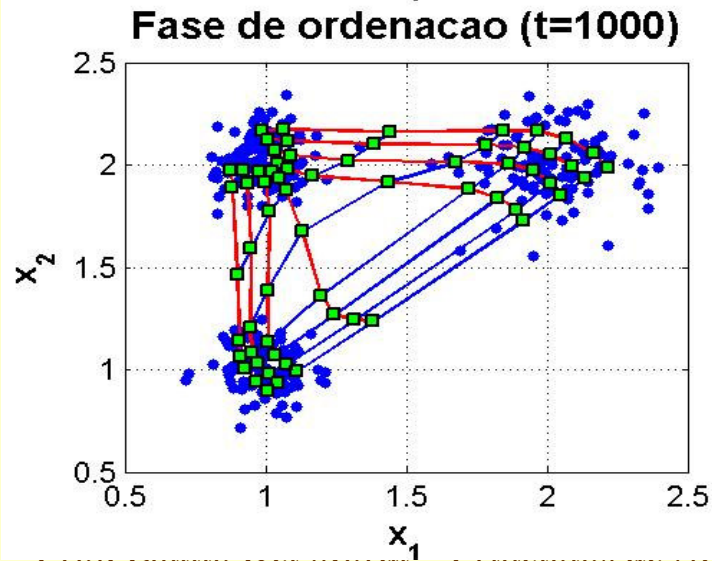
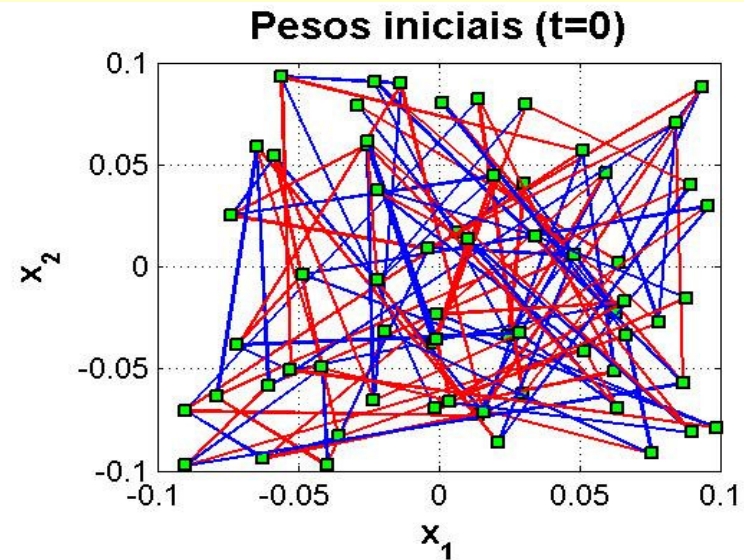
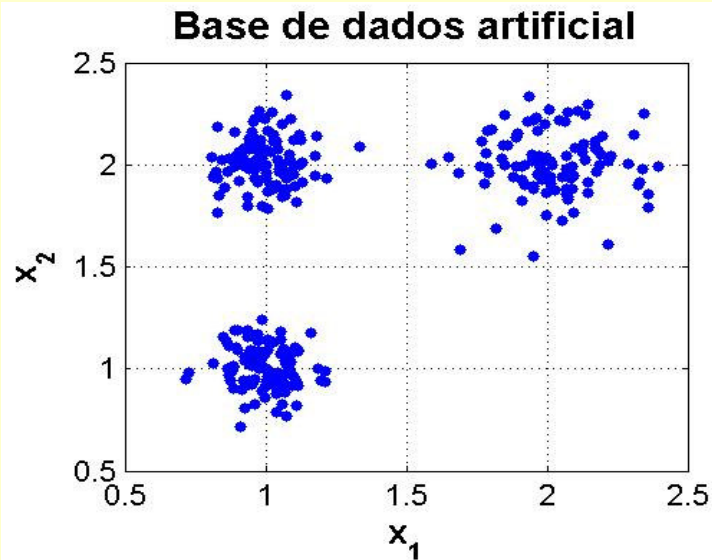


Figura 3.4: Atualização do neurônio vencedor.
(VESANTO, 1997)

Aprendizado SOM



*Da redução de população a
conjunto de protótipos à tarefa
de identificação de grupos
relevantes*

Clustering com SOM - Exemplo

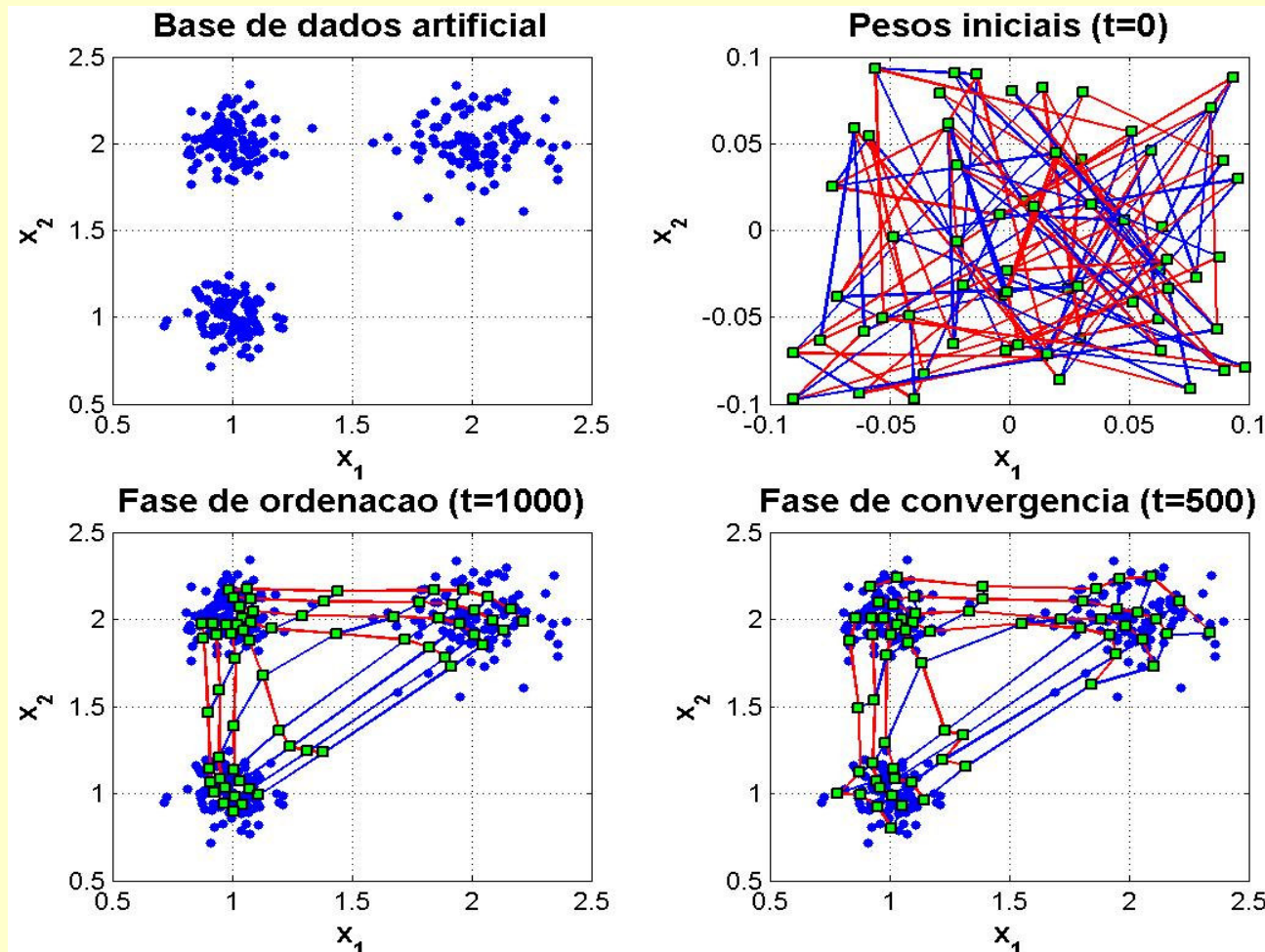


Dados são agrupados em torno de protótipos organizados em espaço bidimensional.

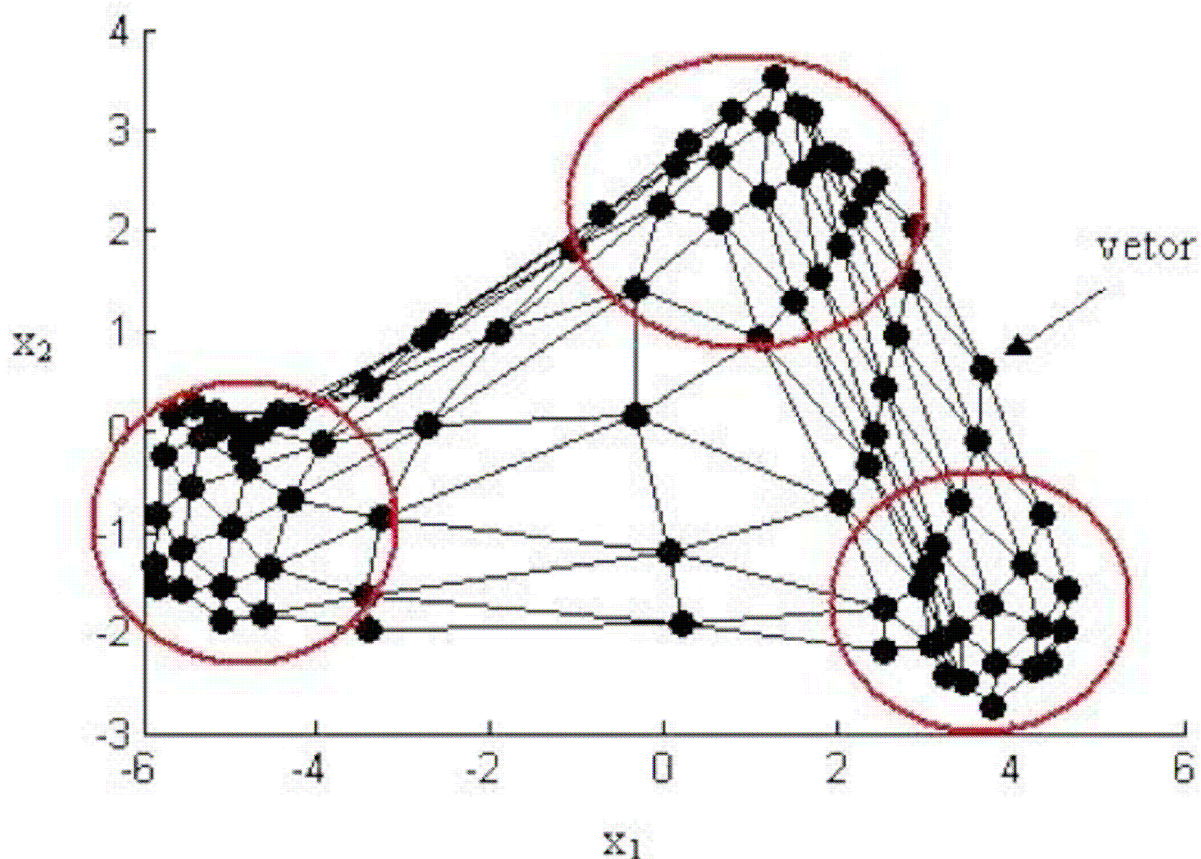
Agrupamento SOM não faz tudo ...

- Uma vez representada a população por protótipos, resta ainda particionar estes em grupo
- E se os grupo de protótipos não estiverem tão separados assim como na ilustração?
- E se uma inspeção visual dos protótipos (de por exemplo 50 dimensões) não for possível para a definição imediata dos grupos?

Uma primeira tentativa ... Histograma de frequencias para identificação de picos e vales na grade de representantes



Histograma / Protótipos mais votados indicam partes do mapa mais significativas



Note que o histograma seria visível na grade de protótipos

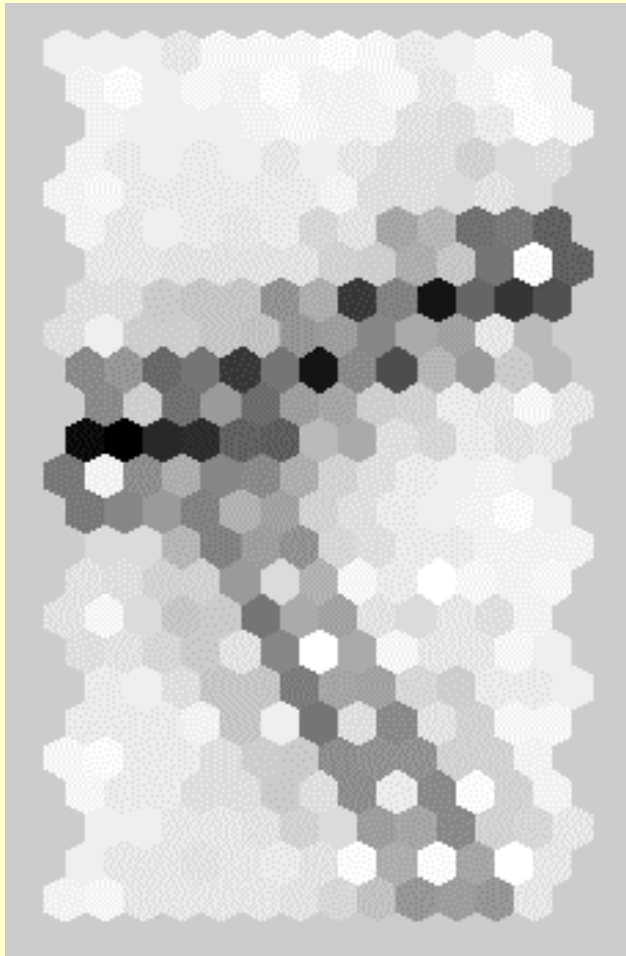


Dados são agrupados em torno de protótipos organizados em espaço bidimensional.

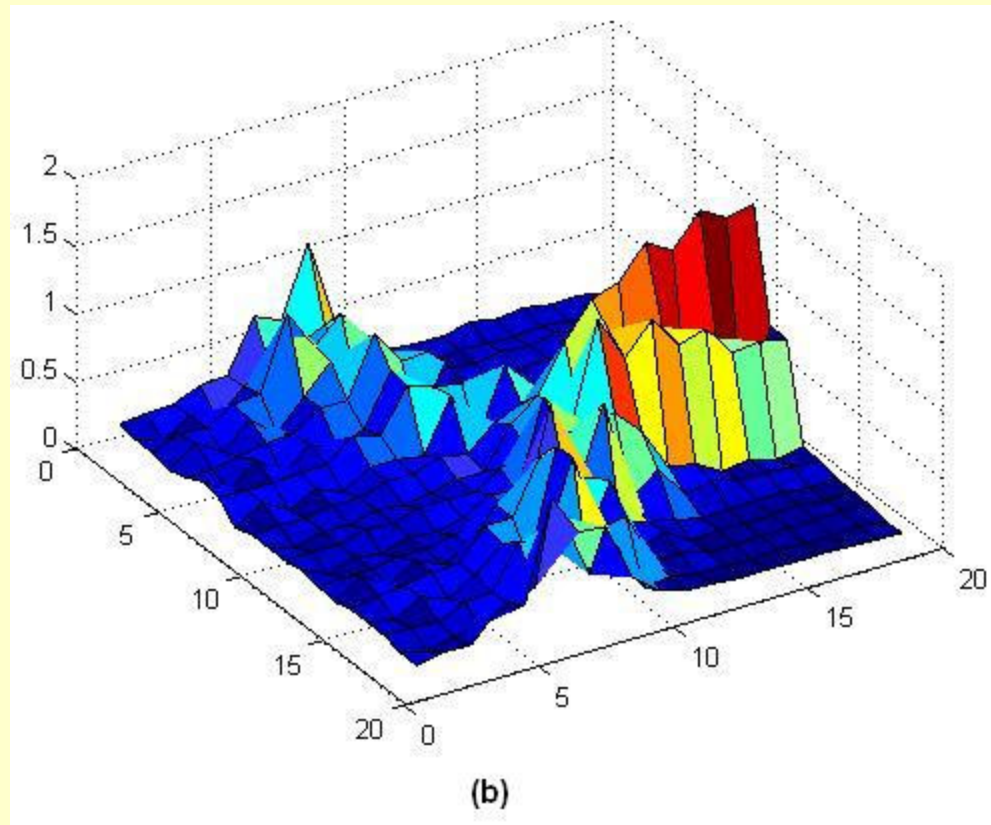
Visualização → Matriz U

- Para efetuar análise de agrupamentos, apenas mapeamento topologicamente ordenado do SOM não é suficiente, pois a informação de distâncias entre os neurônios é perdida.
- Matriz de distâncias unificadas (*U-matrix*) é um método de visualização de um SOM treinado, que permite a detecção visual das relações topológicas dos neurônios.
- A idéia básica é usar a mesma métrica que foi utilizada durante o treinamento para calcular distâncias entre pesos sinápticos de neurônios adjacentes.

Matriz U - Exemplos

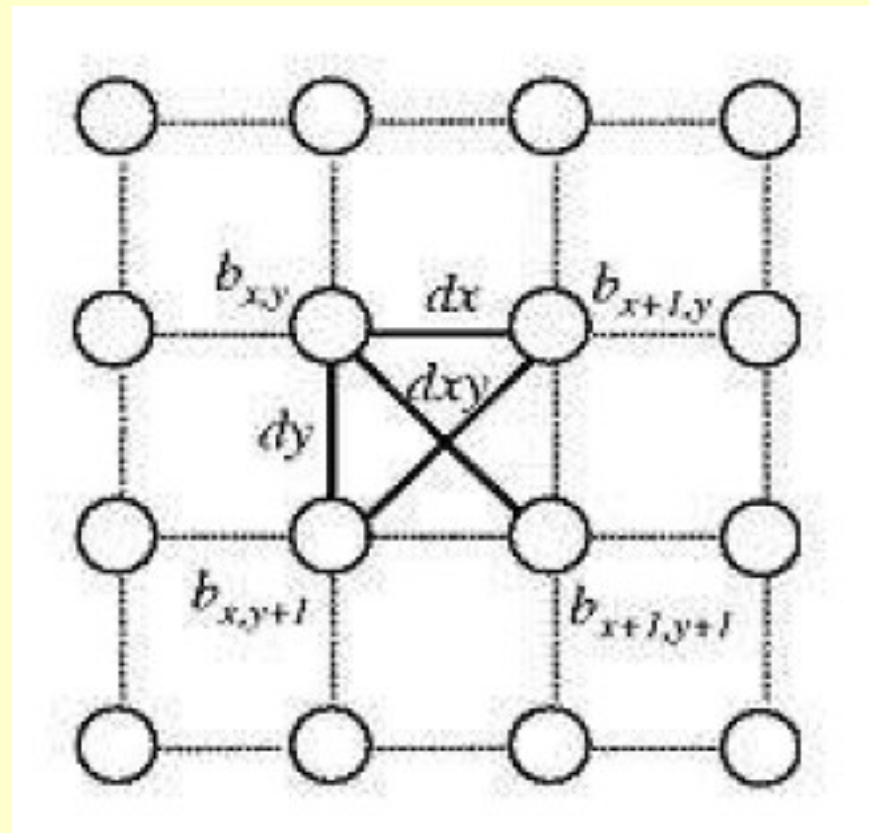


2-D



3-D

Matriz U

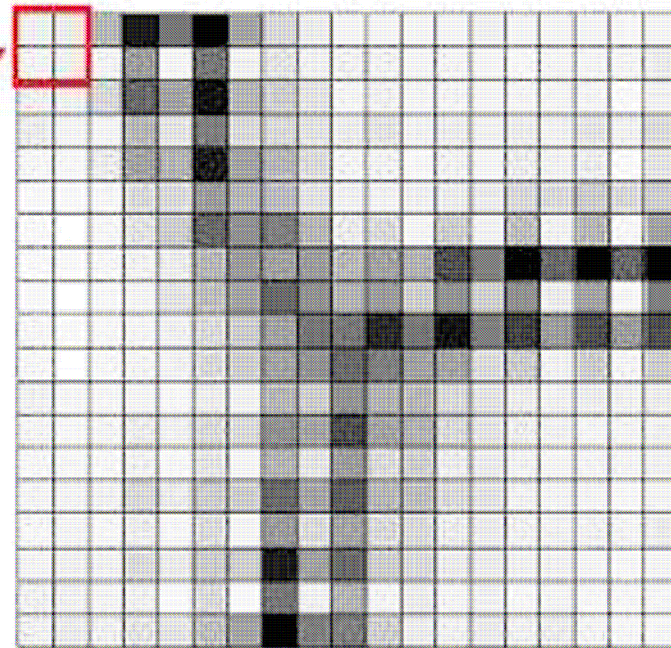


Matriz U

Células da Matriz-U referentes à uma unidade de mapeamento.

Valor de dx	Valor de du
Valor de dy	Valor de dxy

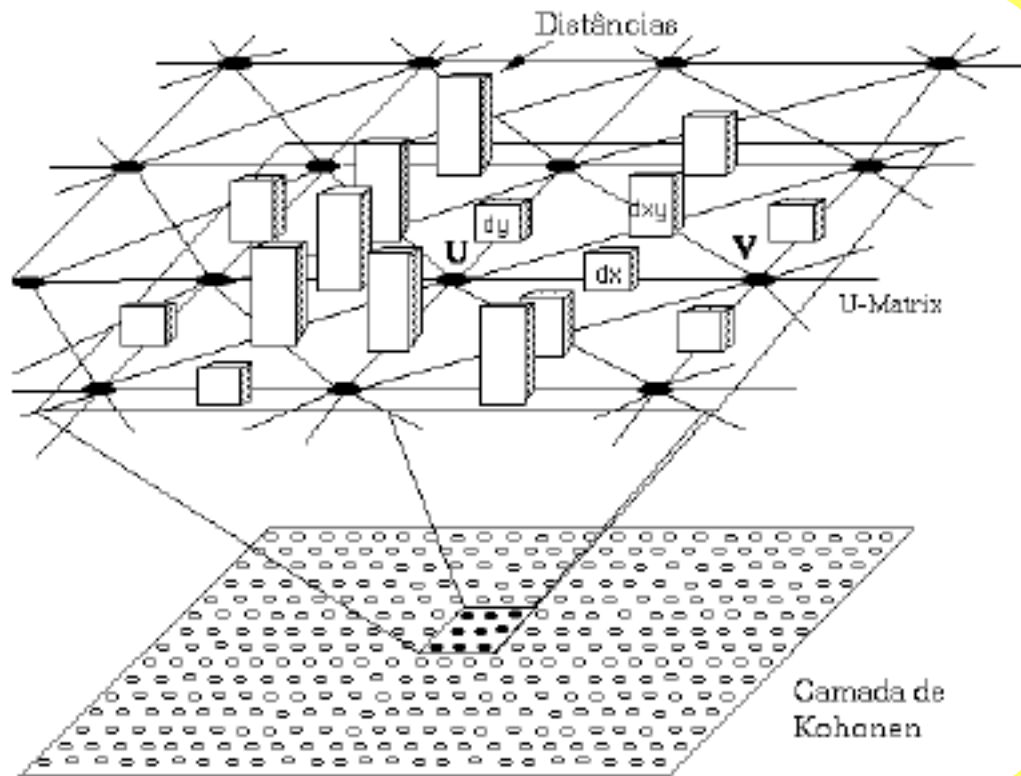
du é um valor médio de dx , dy e dxy .



Matriz U

- O resultado é uma imagem $f(x, y)$, na qual as coordenadas de cada pixel (x, y) são derivadas das coordenadas dos neurônios no grid do mapa (por exemplo, $(1,1), (1, 2) \dots (X, Y) \rightarrow (1,1), (1, 2) \dots (2*X-1, 2*Y-1)$), e a intensidade de cada pixel na imagem $f(x,y)$ corresponde a uma distância calculada.
- A matriz U é geralmente complexa em casos reais, tornando seu uso restrito à visualização, como ferramenta de auxílio na separação manual dos agrupamentos de um SOM.

Matriz U



Representação
3-D

Medidas de Avaliação

- **Erro de Quantização:** É uma medida sobre a resolução do mapa. Corresponde à distância entre o vetor de características e seu BMU.
- **Erro Topológico:** Avalia o quanto a estrutura do mapa aproxima padrões próximos do espaço de entrada, a partir do primeiro e do segundo BMUs.
- **Produto Topológico:** Objetiva quantizar a preservação de relações de vizinhança em mapeamentos.

Erro de Quantização

$$E_q = \frac{\sum_{k=1}^n \|v_k - w_{bmu}\|}{n}$$

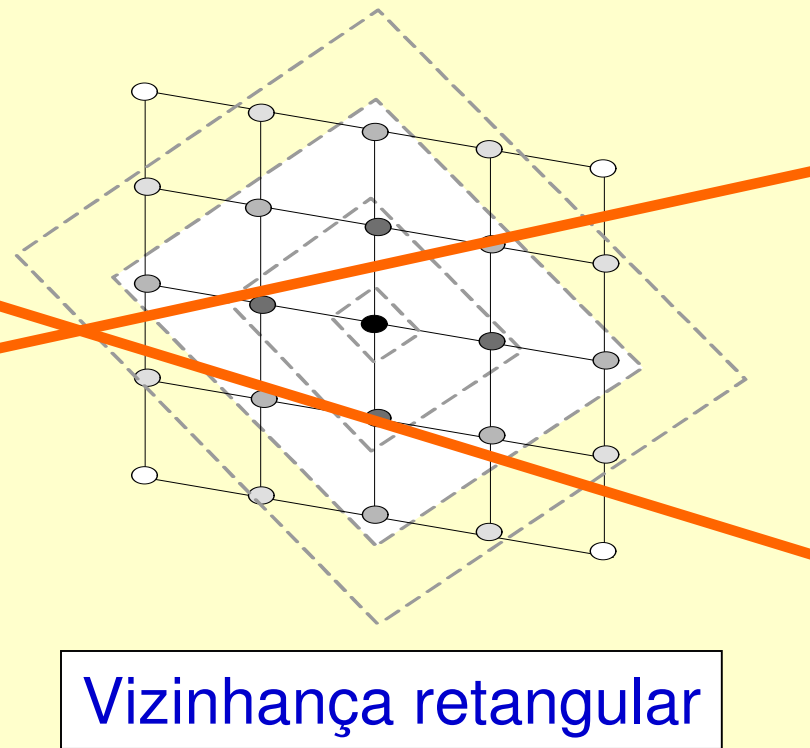
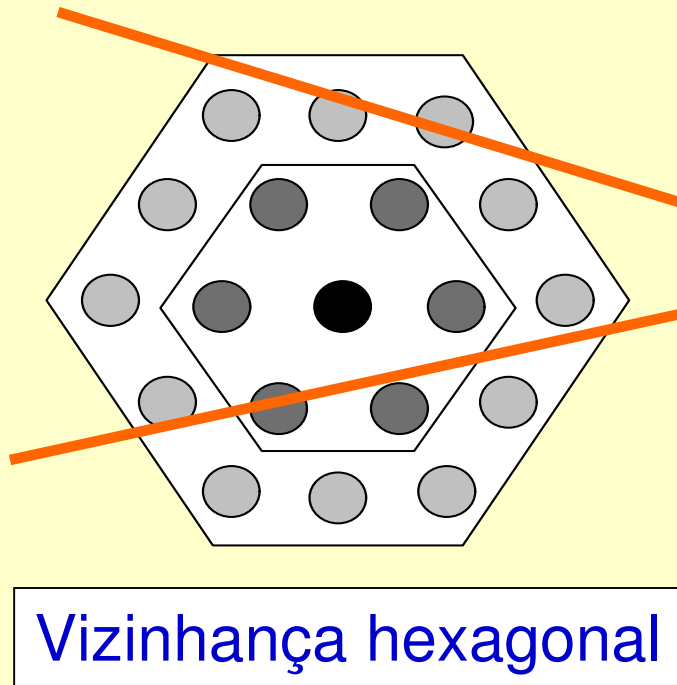
Linear Vector Quantization

- “Primo do SOM”, sem aprendizado dos vizinhos
- O conceito de Protótipos representando a população de dados é mantido
- A vizinhança de aprendizado é eliminada, de forma que a grade ordenada não mais existe

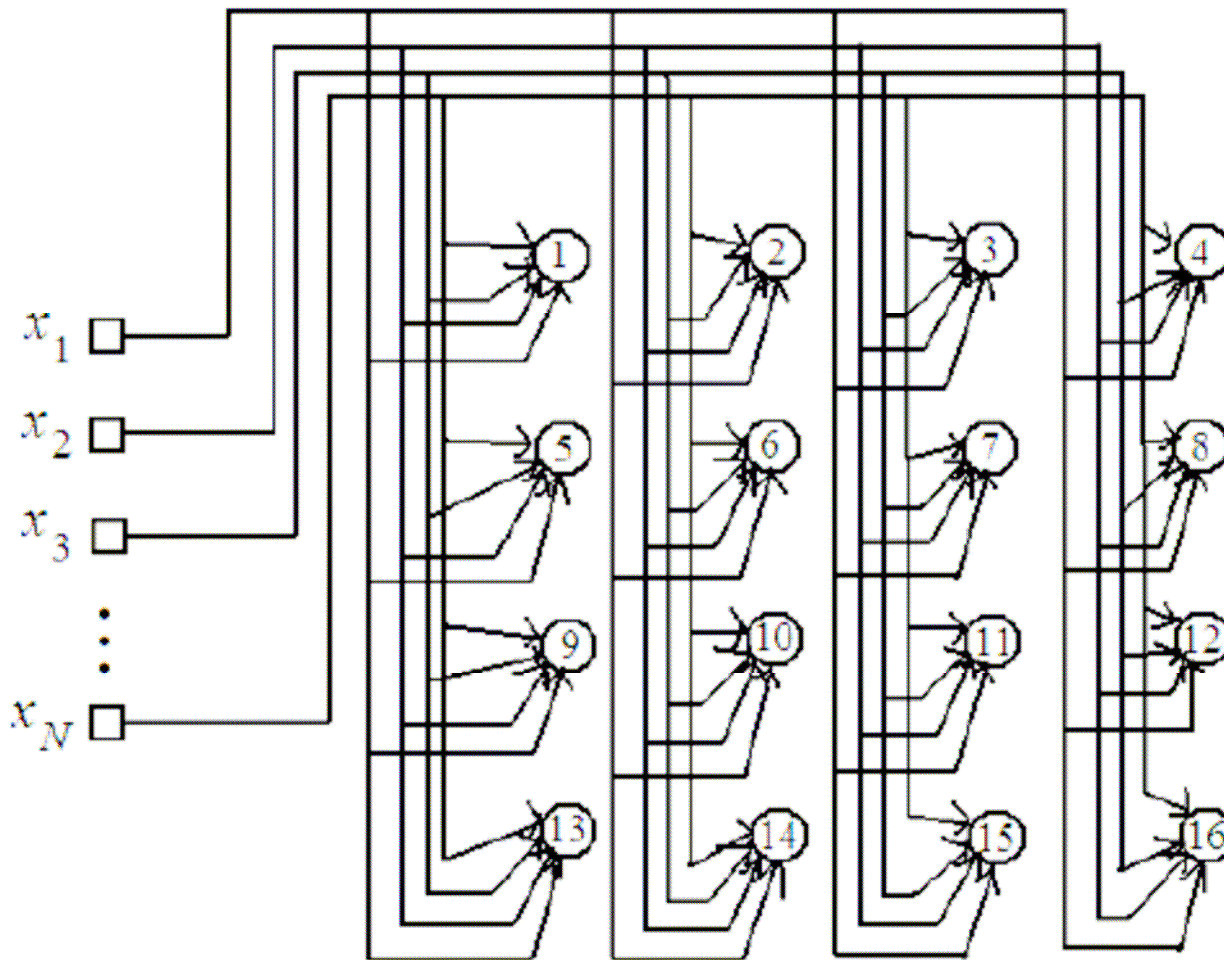
Arquiteturas de Kohonen (SOM) para clustering



Vizinhança SOM



Arquitetura do SOM



Erro Topológico

O **Erro Topológico** (E_t) avalia o quanto a estrutura do mapa aproxima vetores de dados de entrada próximos, a partir do primeiro e do segundo neurônios, de forma que corresponderá ao percentual de vetores de dados cujos primeiro e segundo BMUs não são vizinhos no mapa.

$$E_t = \frac{1}{n} \sum_{k=1}^n u(v_k) \quad (3.7)$$

$$\text{em que } u(v_k) = \begin{cases} 1, & \text{se o primeiro e segundo BMUs não forem adjacentes;} \\ 0, & \text{caso contrário.} \end{cases}$$

Rede SOM

Aplicações

SOM - Aplicações

- Visualização e análise de dados de dimensionalidade elevada;
- Projeção não-linear do espaço de dados de entrada, em \mathfrak{R}^M , para o espaço de dados do arranjo, \mathfrak{R}^N , executando uma redução de dimensionalidade $M < N$. Normalmente, N para 2 ou 1;
- Algoritmo de Classificação (com outras arquiteturas neurais, ...);
- Amostragem, extração de características;
- Algoritmo de Clustering (detecção de agrupamentos) - utilizado em conjunto com “*U-Matrices*”.

Applications of the Self Organizing Maps (SOM) of Kohonen

- Partition of the universe of multidimensional data in sub-groups of similar elements
- Data mining (clustering + visualization)
- Multidimensional data pre processing aiming facilitated classification, within more specific sub universes

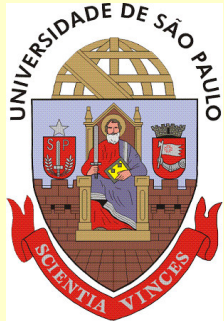
Algumas aplicações de mapas auto organizáveis de Kohonen (SOM)

- Divisão de universos de dados multidimensionais em sub-grupos de elementos similares
(Congresso ABAR 2003 – segmentação de empresas do setor elétrico – trabalho de Virginia Parente et al.)
- Data mining (R.J.Sassi, L.A.Silva e C.Boscarioli)
- Pré processamento de dados multidimensionais para posterior classificação, de padrões dentro de sub universos mais específicos (C.Boscarioli e L.A.Silva)
- Recuperação de imagens baseado em conteúdo (L.A.Silva)

Algumas aplicações de mapas auto organizáveis de Kohonen (SOM)

Descoberta de conhecimento
em base de dados de
marketing

(R.J.Sassi e L.A.Silva)



LSI – *Laboratório de Sistemas Integráveis*

ICONE – *Grupo de Inteligência Computacional,
Modelagem e Neurocomputação*



Análise de Agrupamentos por Meio de Mapas Auto-organizáveis

Clodis Boscarioli

Data Mining and Clustering of Multidimensional data

Clodis Boscarioli. PhD. Student.

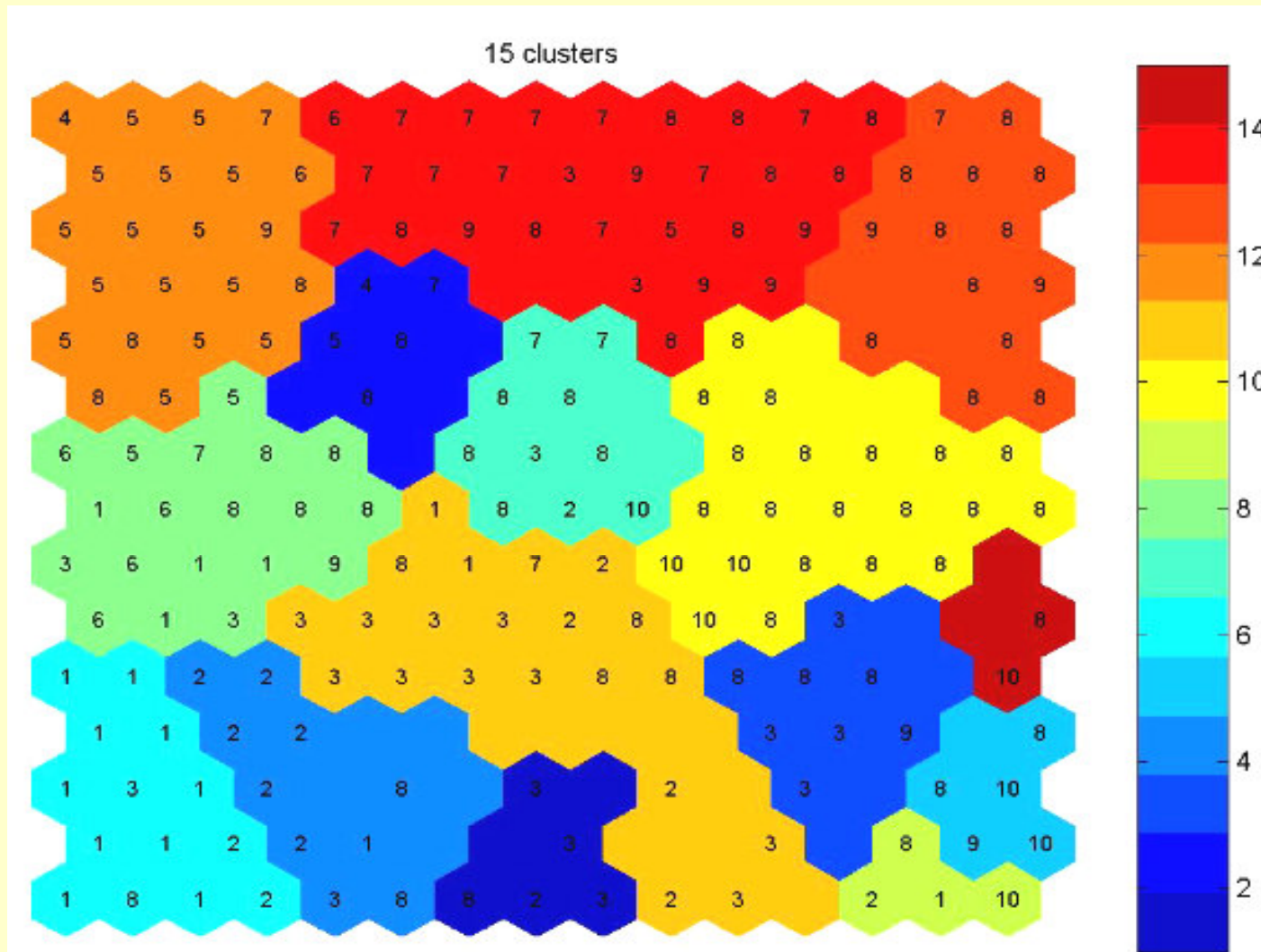
Research area: The application of clustering techniques to Data Mining. More specifically working in the development and evaluation of Self-Organizing Maps algorithms for this task.

Interests: Database systems, data mining and non-supervised neural networks.

*Mais aplicações de SOM feitas
pelos pesquisadores do grupo
ICONE-EPUSP podem ser
conhecidas através do site
www.lsi.usp.br/icones
e também através das
publicações lá relacionadas*

Ferramentas desenvolvidas por Vesanto (Finlândia)

Example of clustering / visualization of clustering in a Kohonen SOM



My coordinates ...

Emilio Del-Moral-Hernandez

Polytechnic School - University of São
Paulo (EPUSP)

Department of Electronic Systems
Engineering

emilio_del_moral@ieee.org

www.lsi.usp.br/icone



USP

Universidade de São Paulo
B R A S I L