

CPE 722

Redes Neurais Não Supervisionadas e Agrupamento

Resumo, Visão Global

Agrupamentos (Clusterização)

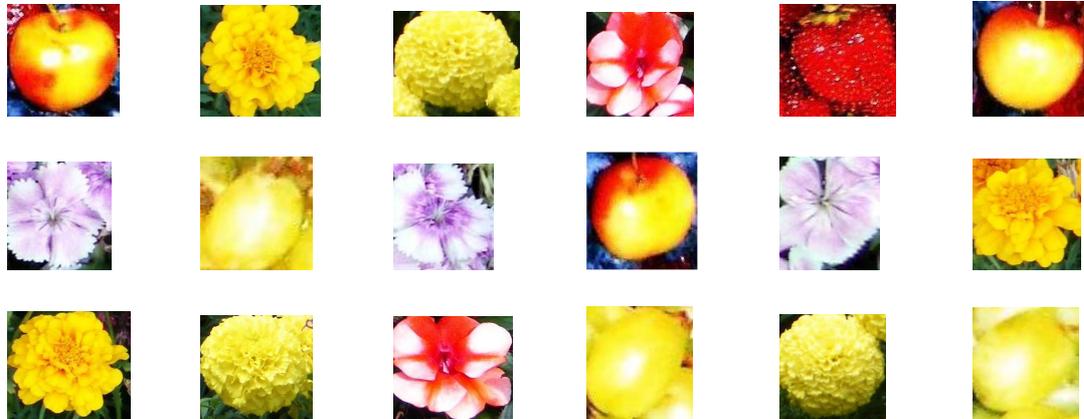
- Classificadores, Classificação por similaridade
- Aprendizado Supervisionado e Não Supervisionado
- Clusterização, processos clássicos
- Camada de Kohonen Simplificada
- Rede ART
- Counterpropagation
- Classes não esféricas
- SOM

Referências bibliográficas:

- *1 - Wasserman, P. – “Neural Computing”, Van Nostrand Reinhold, 1989, Cap 4, 8.
www.lps.ufrj.br/~caloba/wassermann.pdf
- *2 – Ivan Silva, I.; Spatti, D. e Flauzini, R. - "Redes Neurais Artificiais para Engenharia e Ciências Aplicadas", Artliber, 2010, cap 8-10.
- *3 - Duda, R.O., Hart, P.E., "Pattern Classification and Scene Anal.", Wiley, 1973, Cap 6.
www.lps.ufrj.br/~caloba/duda&hart1973cap6.pdf
- 4 - Haykin, S., “Neural Networks and Learning machines”, Pearson, 2009 ou “Redes Neurais, Teoria e Prática”, Bookman, 2001, Cap. 8.1, 8.2, Cap. 9.
<http://books.google.com.br/books?id=lBp0X5qfyjUC&pg=PA483&lpg=PP1> (parcial)
[http://rapidshare.com/files/127507178/Neural Networks - A Comprehensive Foundation - Simon Haykin.pdf](http://rapidshare.com/files/127507178/Neural_Networks_-_A_Comprehensive_Foundation_-_Simon_Haykin.pdf)
- 5 – Duda, R.O.; Hart, P.E.; Stork, D.G., "Pattern Classification", Wiley, 2001, Cap 10.
www.lps.ufrj.br/~caloba/duda&hart2001cap10.pdf
- 6 – Gersho, A., Gray, R.M. – “Vector Quantization and Signal Compression”, Kluwer, 1992, Part III.

Agrupamento, Clustering, Classificação por Similaridade

Classes agrupam “de forma natural”
conjuntos de dados com “similaridade interna”



Nível de agrupamento ?

Classificação por Similaridade

objetos físicos o_1 o_2

similaridade física



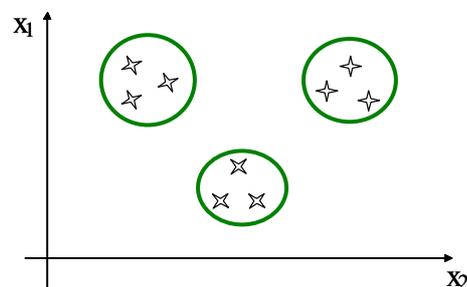
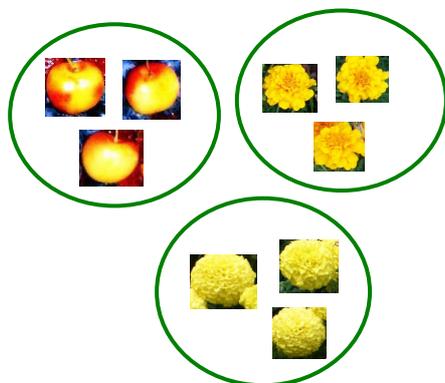
objetos matemáticos \underline{x}_1 \underline{x}_2

similaridade matemática



$$o_1 \approx o_2$$

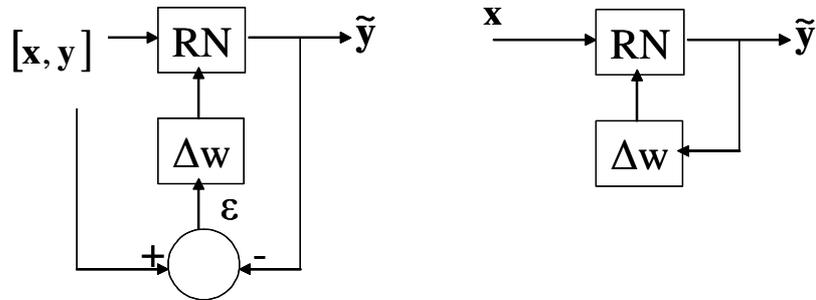
$$\underline{x}_1 \cong \underline{x}_2 \quad \text{ou} \quad |\underline{x}_1 - \underline{x}_2| \ll$$



Classificadores por similaridade.

***Aprendizado não supervisionado, cego.**

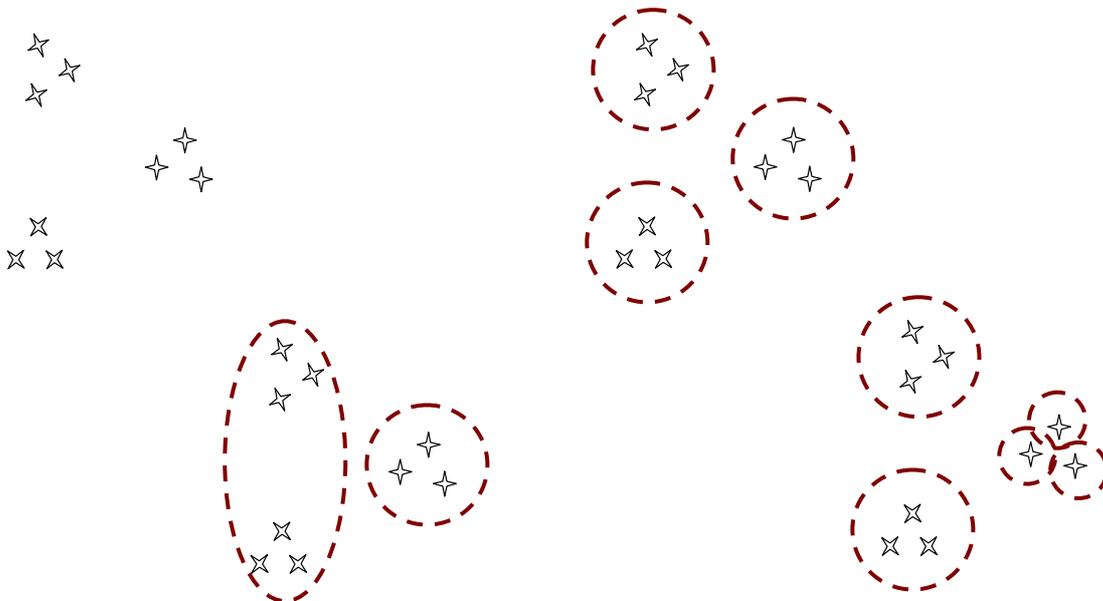
Treinamento supervisionado / não-supervisionado



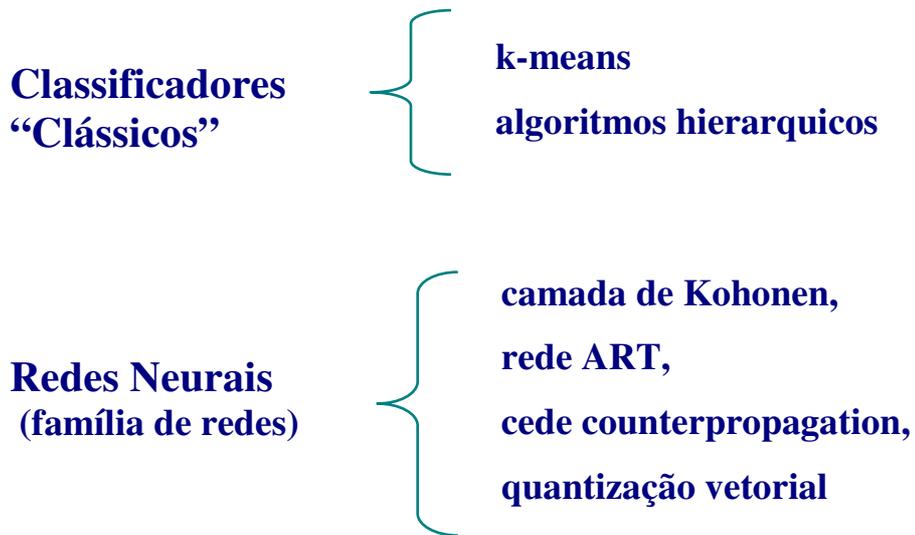
***Plasticidade.**

Classificadores por similaridade.

Critério: **minimizar a dispersão intra-classe e maximizar o espalhamento inter-classes**

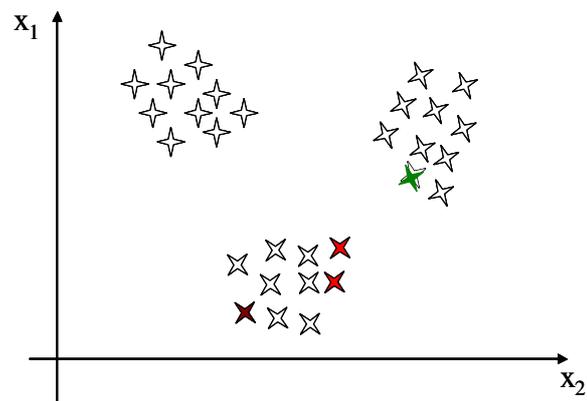


Classificadores por similaridade.



Classificadores por similaridade

pertinência à uma classe



Critério básico: Dois elementos pertencem à mesma classe se estão próximos entre si

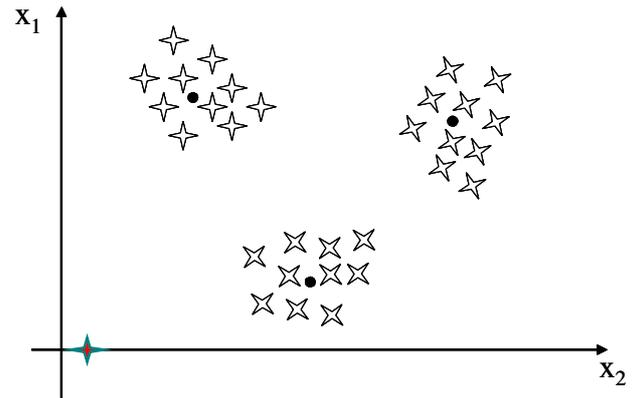
$$| \underline{x}_1 - \underline{x}_2 | \ll \text{ pouco prático}$$

Padrão de classe:

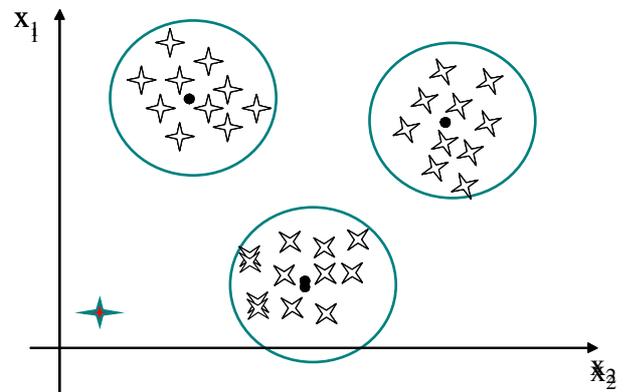
$$\underline{w}_i = \underset{\forall \underline{x}_j \in C_i}{E} \underline{x}_j = \frac{1}{N_i} \sum_{\forall \underline{x}_j \in C_i, j=1}^{N_i} \underline{x}_j$$

Critério 1 :**Padrão mais similar à entrada**

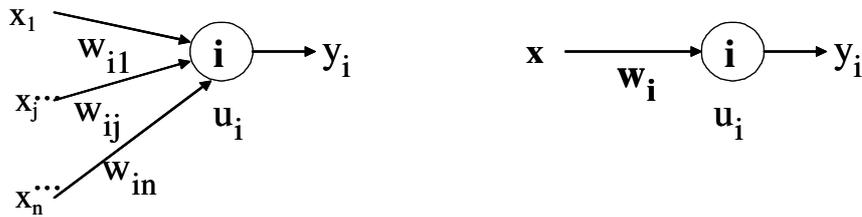
$$\underline{x}_i \in C_i \quad \text{sse} \quad |\underline{x} - \underline{w}_i|^2 < |\underline{x} - \underline{w}_j|^2 \quad \forall j \neq i$$

**Similaridade mínima****Critério 2 –****Padrão que satisfaz
uma similaridade mínima
com a entrada**

$$\underline{x}_i \in C_i \quad \text{sse} \quad |\underline{x} - \underline{w}_i|^2 < r_0^2$$

 r_0 – raio de similaridade

Neurônio como medidor de similaridade



$$u = -|\vec{x} - \vec{w}_i|^2$$

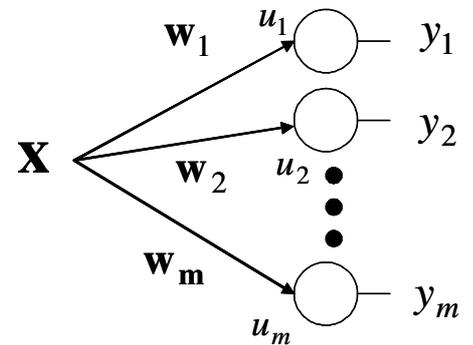
u_i - medida de similaridade entre \underline{x} e \underline{w}_i

Template Matching

$$u_i = -d_i^2$$

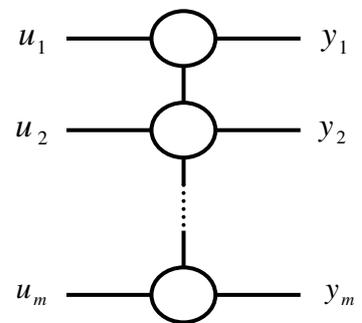
maior u_i =
menor distância =
maior similaridade

u_i é uma medida de similaridade entre \underline{x} e \underline{w}_i



Winner-takes-all

$y_i = 1$ sse $u_i > u_j \quad \forall j \neq i$
 $y_i = 0$ caso contrário

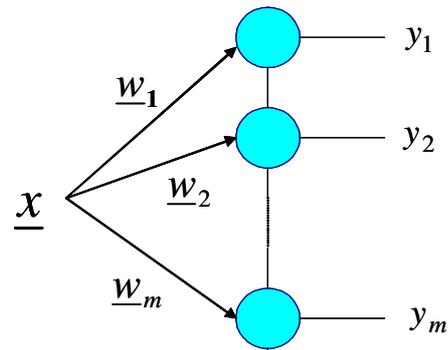


Camada de Kohonen (simplificada, $D=1$)

Classe C_i

Padrao \underline{w}_i

Indicador y_i

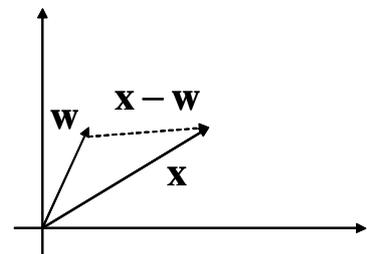
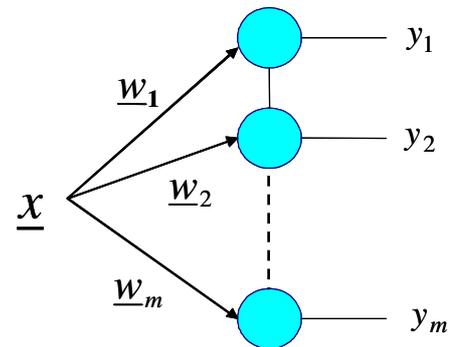
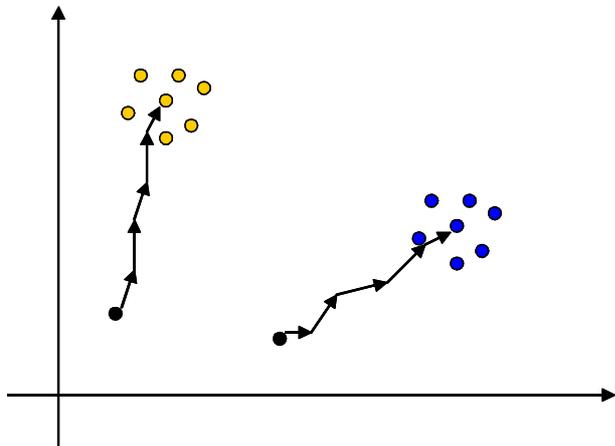


Se $y_i = 1$ então $\underline{x} \in C_i$

pele critério 1 (padrão mais similar a entrada).

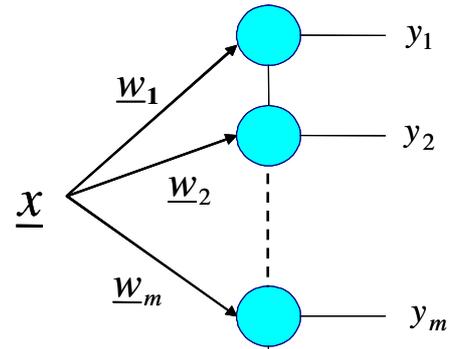
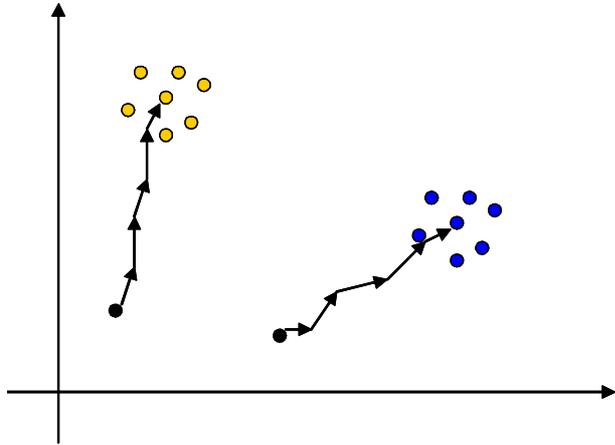
Camada de Kohonen

Treinamento Supervisionado



Camada de Kohonen

Treinamento não supervisionado



E a similaridade mínima ?

Camada de Kohonen (simplificada, aumentada)

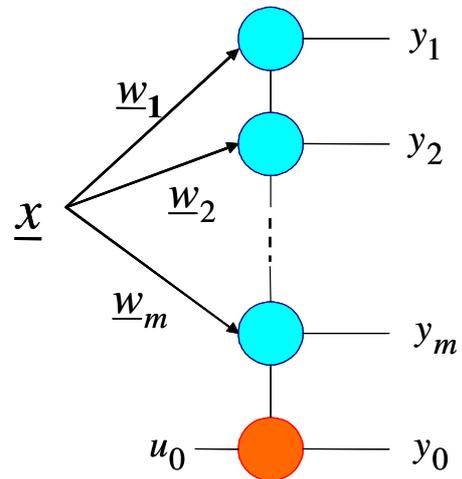
$$u_0 = -r_0^2$$

Se $y_i = 1$ então

$$\underline{x} \in C_i$$

Se $y_0 = 1$ então

$$\underline{x} \notin C_i \quad \forall i$$



Rede ART (modificada)

Condições iniciais:

$$\mathbf{u}_0 = -\mathbf{r}_0^2, \text{ neurônios desativados}$$

Treinamento:

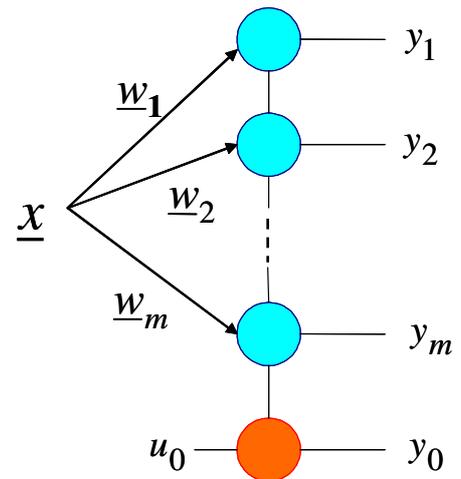
Apresentar entrada $\underline{x}(n)$

$$\text{Se } y_0(n) = 1$$

Ativar um neurônio i desativado

$$\text{Se } y_i(n) = 1$$

Treinar o neurônio vencedor, i



Mecanismos auxiliares

Consciência

Neurônio muito treinado “abre mão” do treinamento

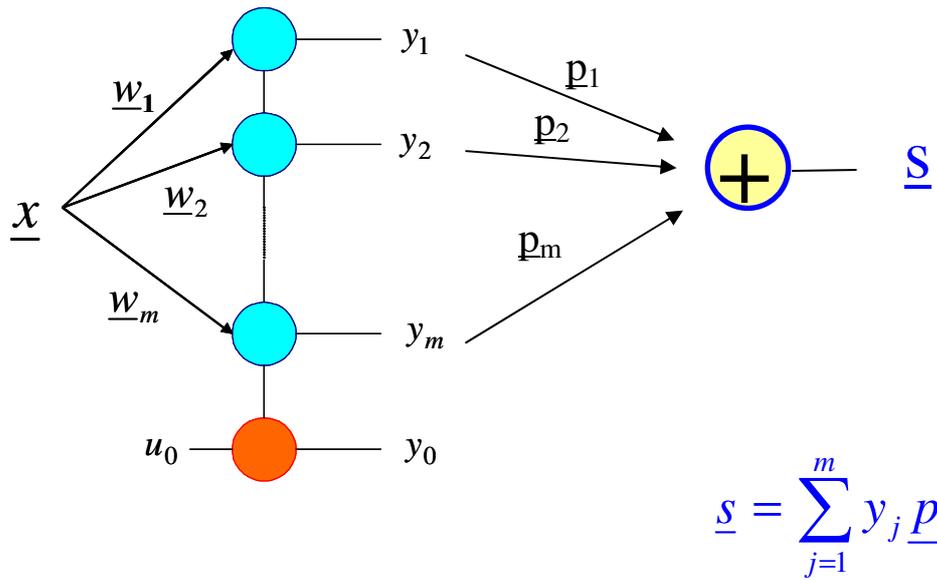
Esquecimento – total ou parcial

Padrão não usado por N entradas consecutivas é desativado
(mas pode ser guardado como possível “lembrança”)

Lembrança

Se $y_0 = 0$ verifique se alguma “lembrança” serve. Se sim, ative-a. Se não, ative novo neurônio.

Counterpropagation

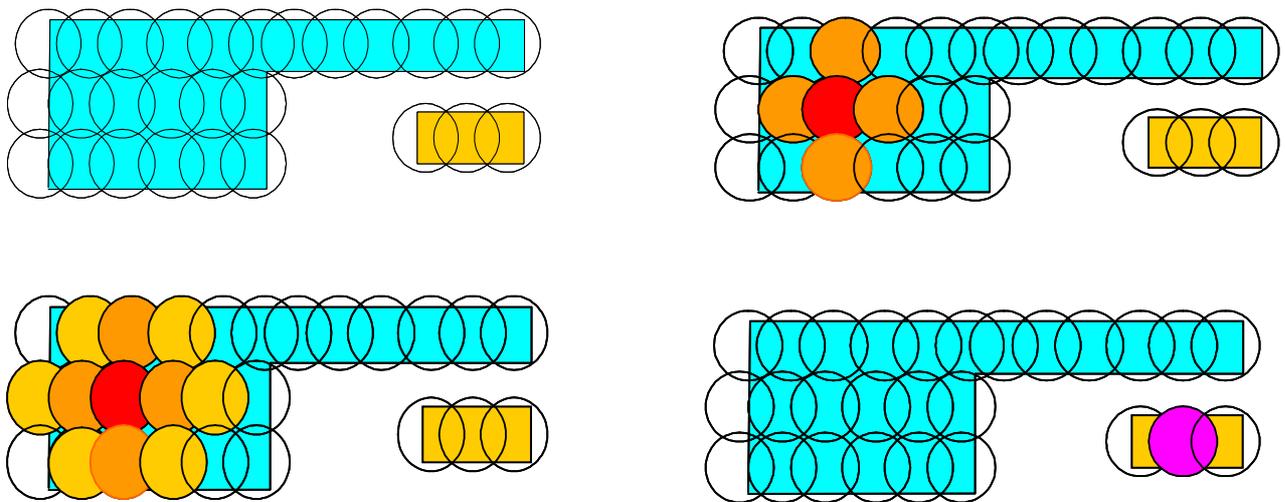


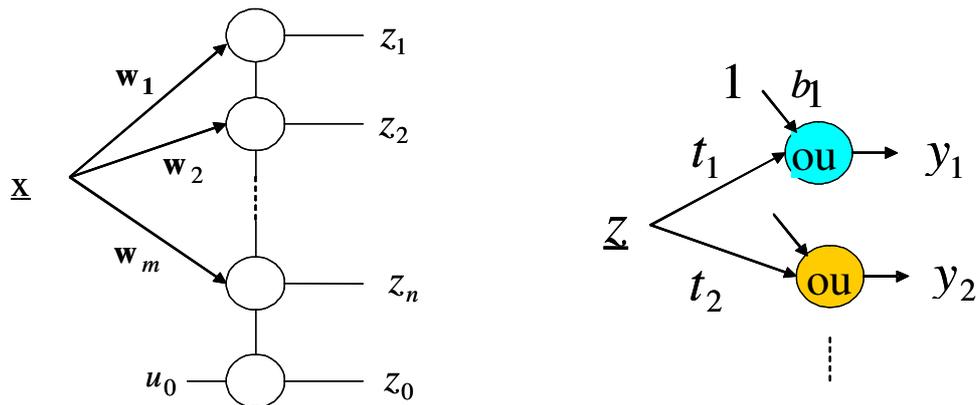
Memória hetero associativa

$$\underline{s} = \underline{p}_i \quad (y_i=1)$$

Separadores esféricos.

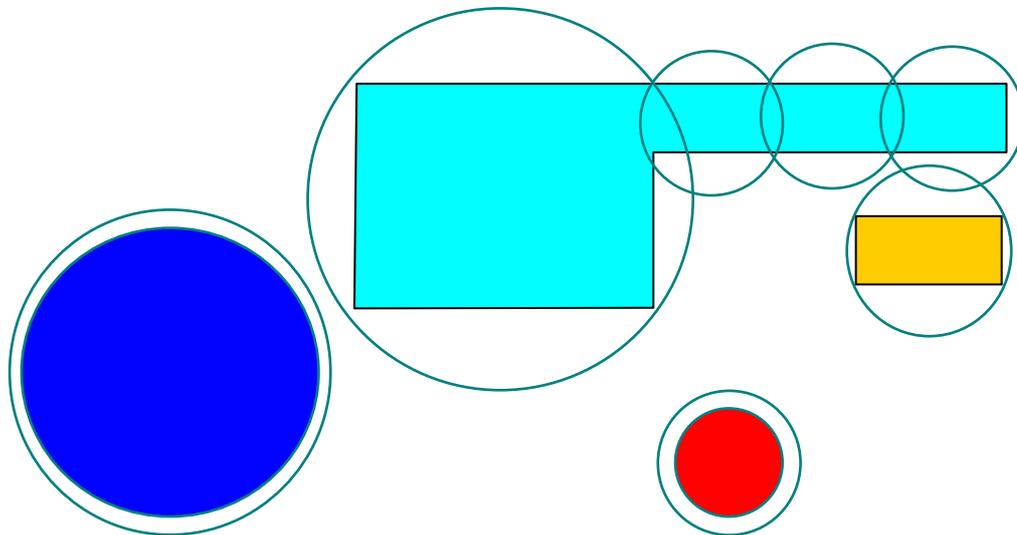
E as Classes Não-Esféricas ?





Classes com dimensões variáveis

ART com raio variável



Comparação de Classificadores ART vs. BP

Treinamento não supervisionado		ART
Plasticidade		ART
Velocidade de treinamento		ART / BP
Mínimos locais		ART
Velocidade de operação		ART / BP
Precisão	Classes esféricas	ART
	Caso geral	BP

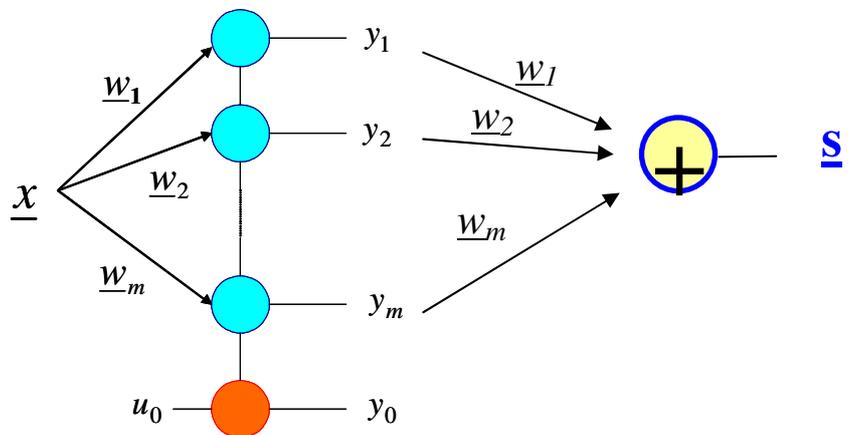
Counterpropagation – de novo

$$\text{Se } \underline{p}_i = \underline{w}_i$$

$$\underline{x} = \underline{w}_i + \underline{r}$$



$$\underline{s} = \underline{w}_i$$



Memória endereçável por conteúdo (ou parte)

Memória auto associativa

Filtragem não linear

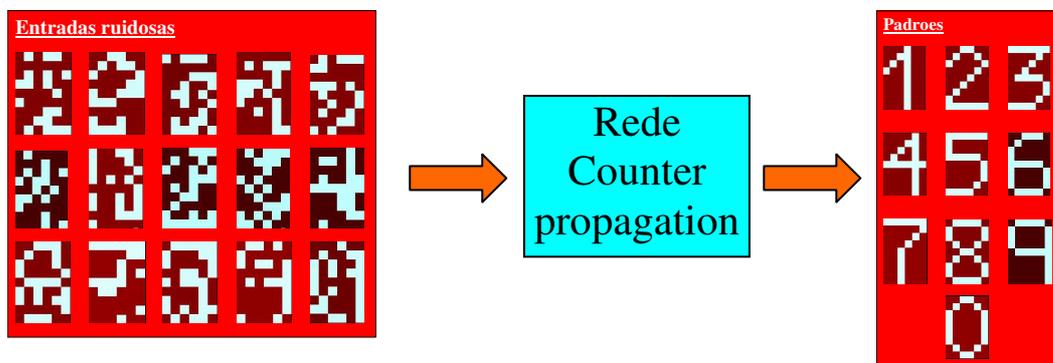
Compressão e restauração da informação

Detecção de padrões desconhecidos:

Dolphin Ecolocation



Filtragem não linear: OCR



Compressão da Informação: Imagens

(1x)



(11x)



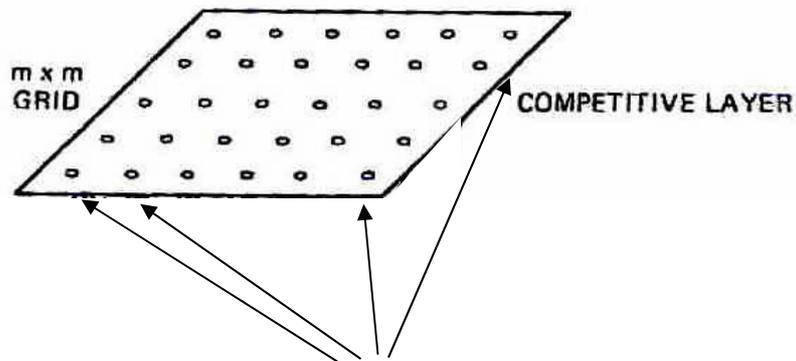
(15x)



(28x)



SOM - Mapas Auto-Organizáveis de Kohonen



Identificação cega de padrões

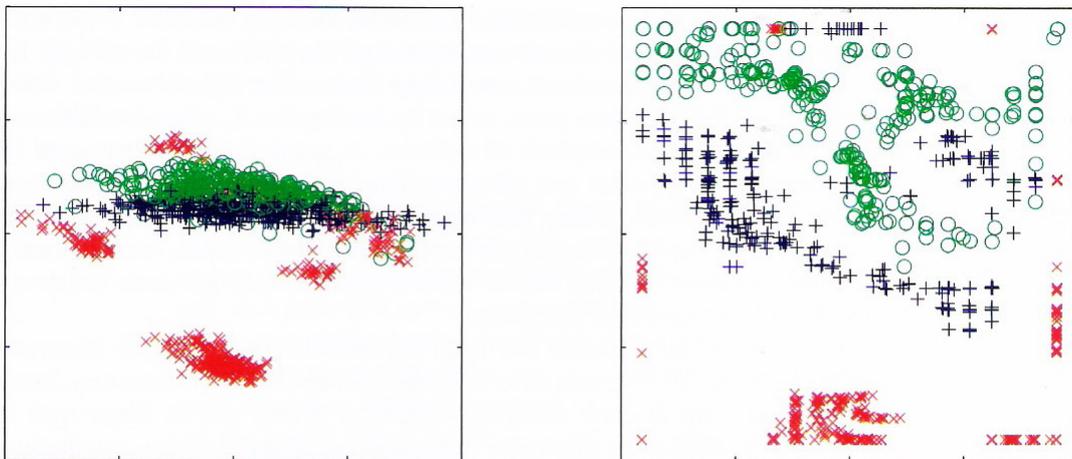


Figure 12.21 Plot of the oil flow data set visualized using PCA on the left and GTM on the right. For the GTM model, each data point is plotted at the mean of its posterior distribution in latent space. The nonlinearity of the GTM model allows the separation between the groups of data points to be seen more clearly.

Pré-Processamento

e Pós Processamento