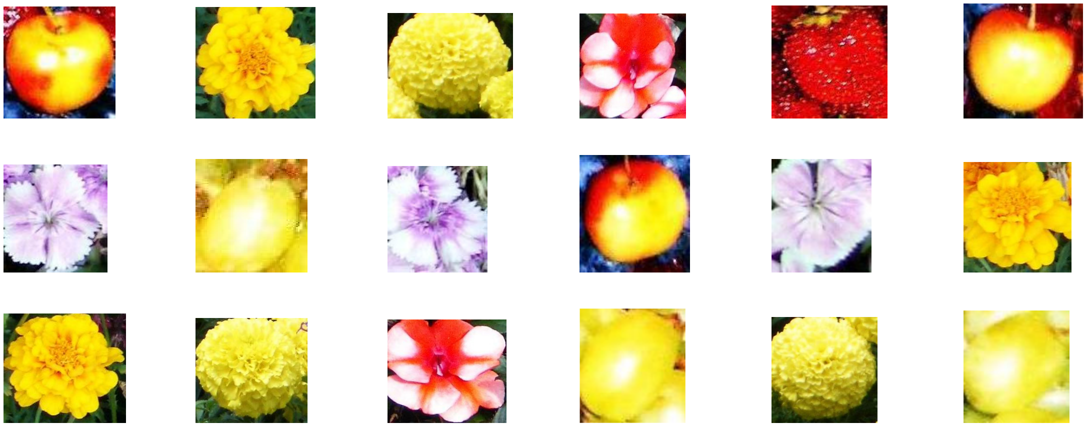
Camada de Kohonen

Classificação por Similaridade

Classes agrupam elementos `similares` entre si



versus classificadores arbitários

1 - Classificação por Similaridade – Critérios de pertinência

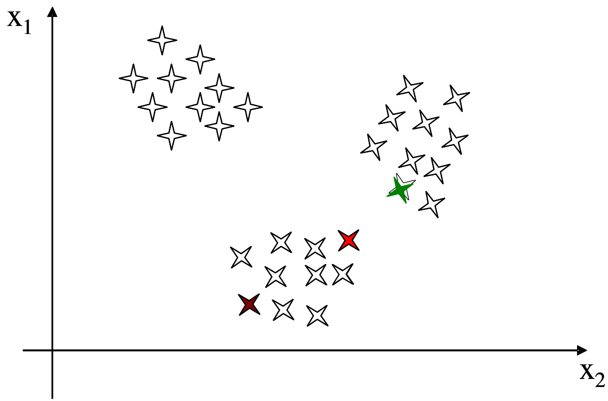
objetos físicos ○$_1$ ○$_2$ ⇒ objetos matemáticos $\underline{\mathbf{x}}_1$ $\underline{\mathbf{x}}_2$

similaridade física ⇒ similaridade matemática

○$_1$ $\approx$ ○$_2$ $\underline{x}_1 \cong \underline{x}_2$ ou $\left|\underline{x}_1 - \underline{x}_2\right| <<$

Classificadores por similaridade

Critérios de Pertinência à uma Classe



Critério básico

Dois elementos pertencem à mesma classe se estão próximos entre si

Critério: k vizinhos mais próximos.

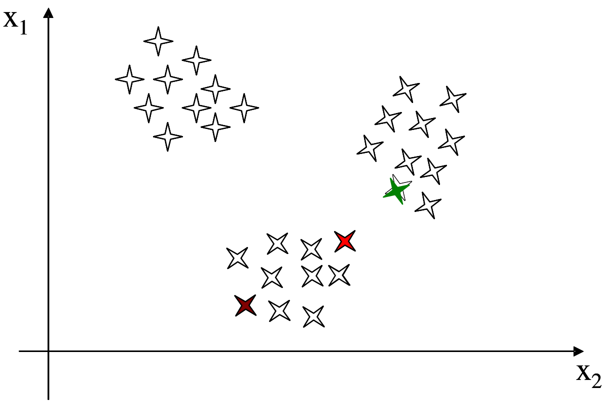
Considere os k elementos mais próximos da entrada que se pretende classificar. A entrada pertence à classe à qual pertencem a maioria dos k vizinhos.

Critério simplificado: vizinho mais próximo.

Se k=1 uma entrada pertence a uma classe se seu vizinho mais próximo é um elemento desta classe

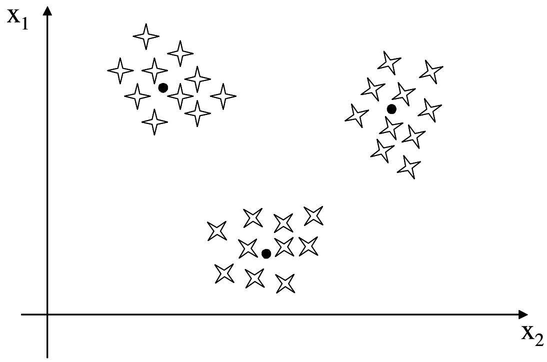
pouco prático

a descrição de uma classe exige a utilização de todos seus elementos (domínio da classe)



Padrão de classe:

Padrão $\underline{w}_i$ da classe C_i

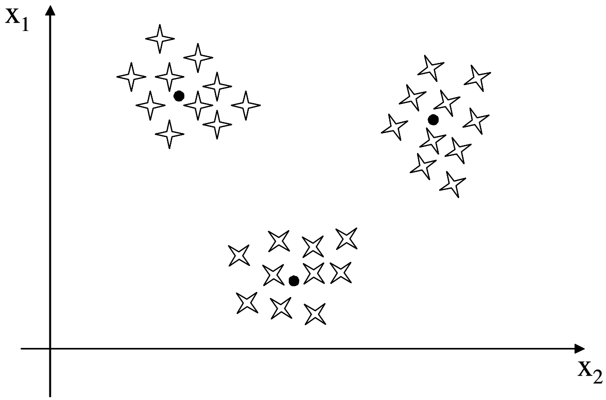


$$\underline{w}_i = \underset{\forall \underline{x}_j \in C_i}{E} \underline{x}_j = \frac{1}{N_i} \sum_{\forall \underline{x}_j \in C_i,\, j=1}^{N_i} \underline{x}_j$$

o baricentro é escolhido porque minimiza
a dissimilaridade
ou erro de representação
da classe

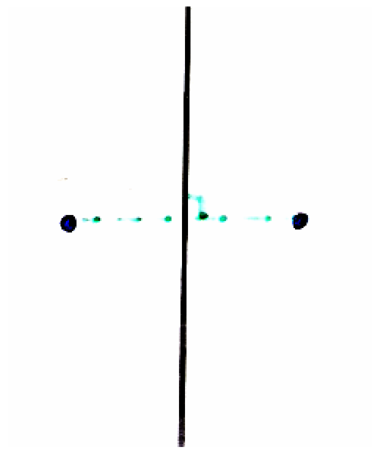
1.1 - Critério de pertinência 1 :

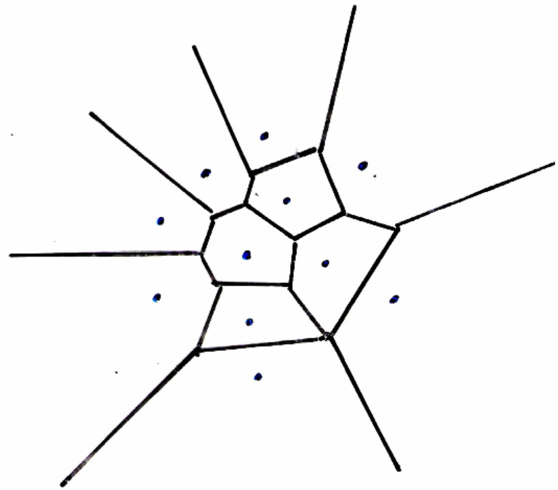
Padrão mais similar à entrada



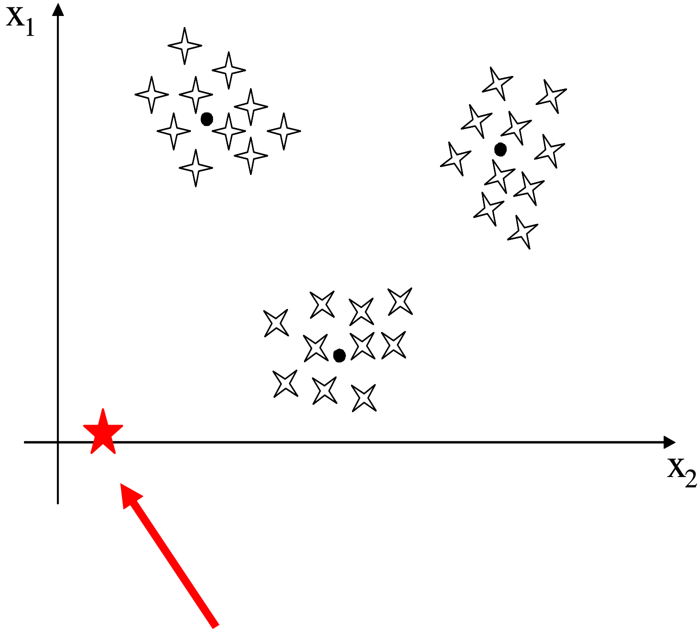
$$\underline{x}_i \in C_i \quad \text{sse} \quad \left|\underline{x} - \underline{w}_i\right|^2 < \left|\underline{x} - \underline{w}_j\right|^2 \quad \forall\, j \neq i \qquad \text{eq (1)}$$

Separadores para o critério 1

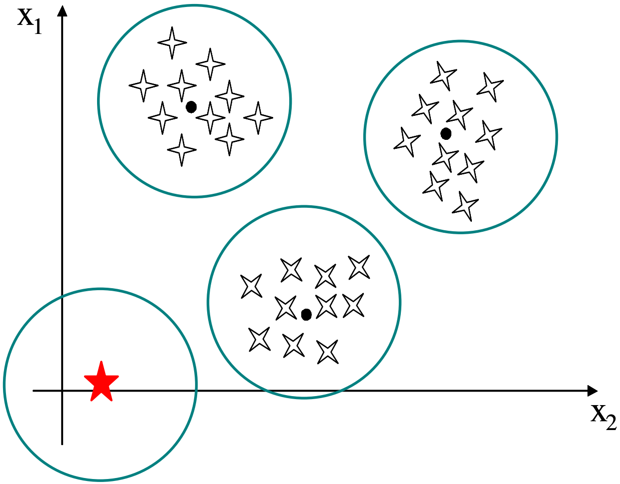


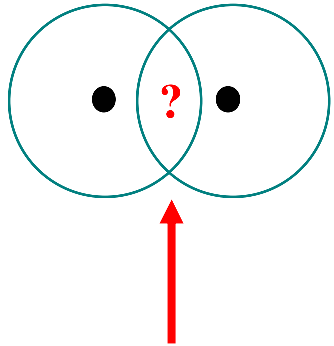


Tecelagem (diagrama) de Voronoi


x_1
x_2

1.2 - Critério de pertinêcia 2 – Padrão que satisfaz uma similaridade minima



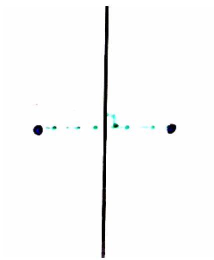


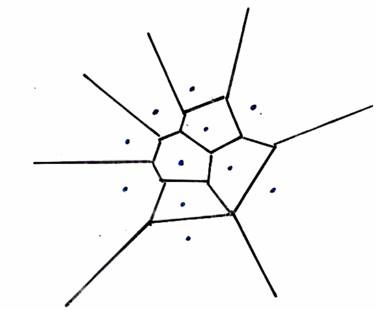
$$\underline{x}_i \in C_i \quad \text{sse} \quad \left|\underline{x} - \underline{w}_i\right|^2 < r_0^2 \qquad \text{eq (2)}$$

r_0 – raio de similaridade

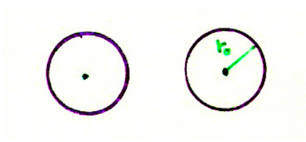
Separadores

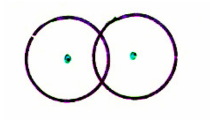
Critério 1 - padrão mais similar (mais próximo) a entrada   Eq(1)





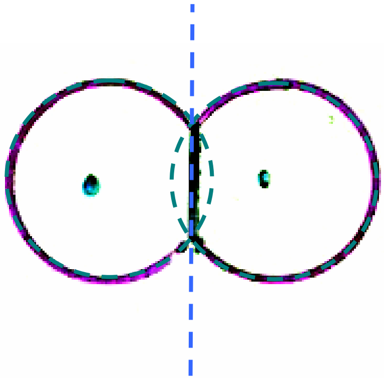
Critério 2 - mínima similaridade exigida Eq (2)

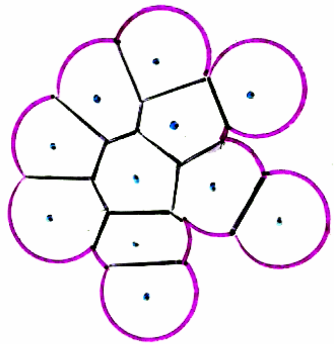




1.3 - Critérios 1 e 2, Eq(1) + Eq(2)

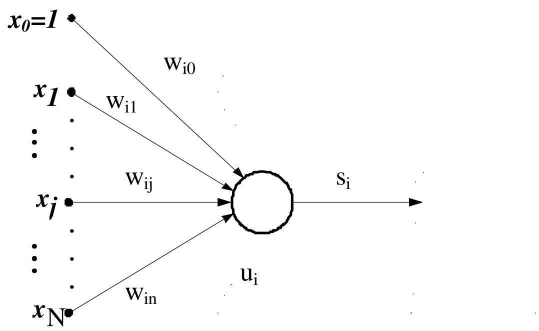
padrão mais similar & mínima similaridade atingida

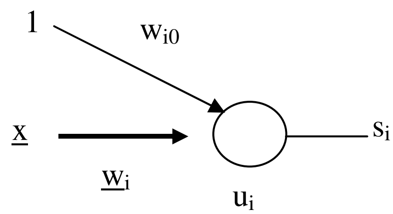




2 - Redes Neurais

Neurônio para redes feedforward





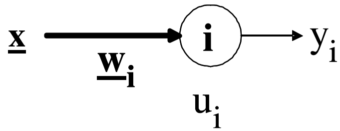
$$u_i = \sum_{j=1}^{N} w_{ij}\, x_j + w_{i0} = \underline{w}_i^t\, \underline{x} + w_{i0}$$

$$\tilde{y}_k = \varphi_k(u_k) = \begin{cases} u_k & \textit{neurônio linear} \\ tgh\,(u_k) & \textit{neurônio tgh} \end{cases}$$

Neurônio como medidor de similaridade

uma outra definição

$$u_i = -\left\| \underline{x} - \underline{w}_i \right\|^2 = -d_i^2 \leq 0$$

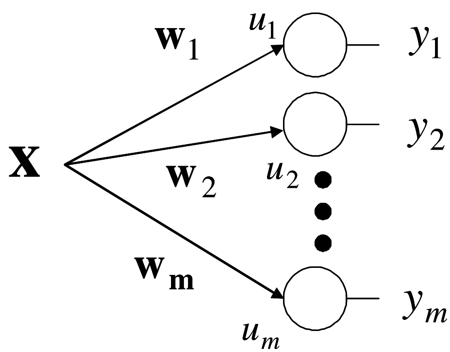


u_i - medida de similaridade entre $\underline{x}$ e $\underline{w}_i$

$u_i = 0$ distância nula = máxima similaridade

y_i – depende dos outros neurônios

3- Camada de Kohonen



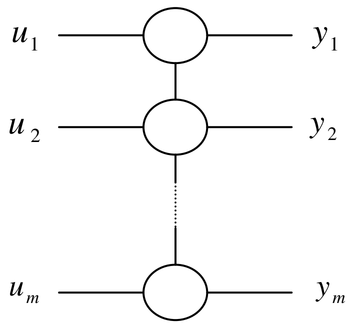
Template Matching

$$u_i = -\, d_i^2$$

maior u_i =
menor distância =
maior similaridade

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

Winner-takes-all



$$y_i = 1 \quad sse \quad u_i > u_j \quad \forall\, j \neq i$$

$$y_i = 0 \qquad caso\ contrário$$

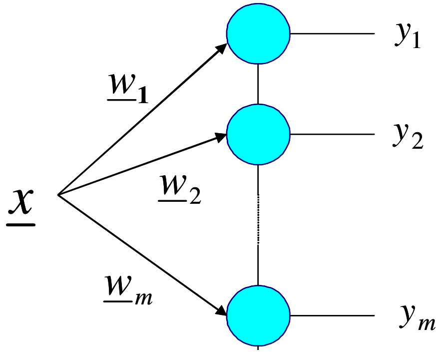
Camada de Kohonen

[versão unidimensional, simplificada (D=0)]

Classe C_i

Padrão $\underline{w}_i$

Indicador y_i

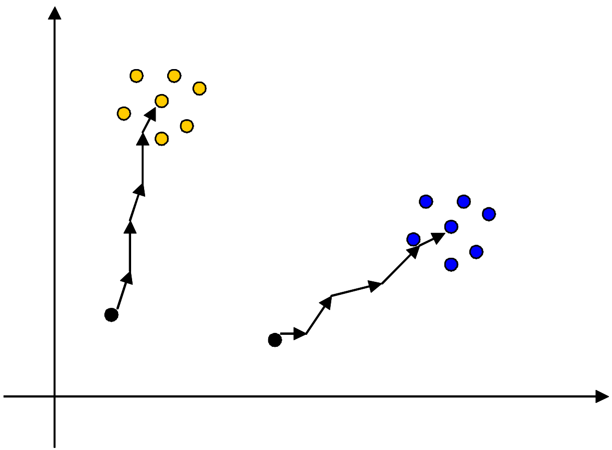


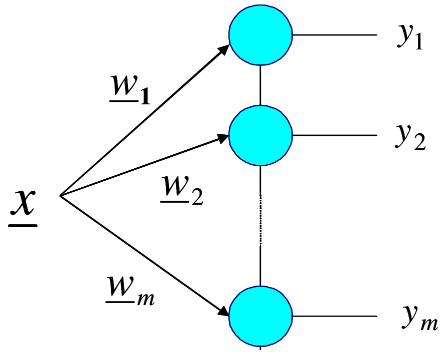
Se $y_i = 1$ então $\underline{x} \in C_i$

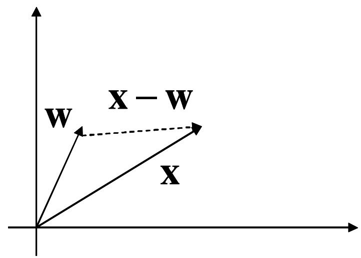
pelo critério 1 (padrão mais similar a entrada).

Camada de Kohonen

Treinamento Supervisionado

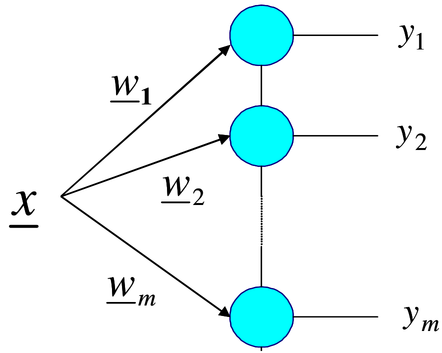






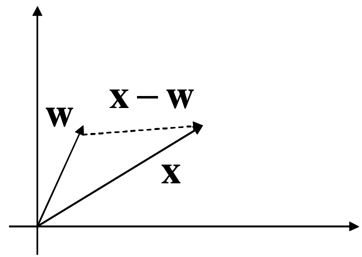
4 - Treinamento da Camada de Kohonen

$$\underline{x}(n) \in C_i \quad \Rightarrow \quad y_i = 1$$
$$y_j = 0 \;\; \forall\, j \neq i$$



$$\underline{w}_i(n+1) = \underline{w}_i(n) + \alpha[\,\underline{x}(n) - \underline{w}_i(n)\,]$$
$$= (1-\alpha)\,\underline{w}_i(n) + \alpha\,\underline{x}(n)$$

$$\underline{w}_j(n+1) = \underline{w}_j(n) \quad \forall\; j \neq i$$



Onde estabiliza, qual o valor final de <u>w</u> ?

Evolução do $\underline{w}$ de uma classe

Passo n: $\underline{x}(n)$ $\underline{w}(n)$

$x(n) \in C_i$

$$\begin{cases} \underline{w}_i(n+1) = (1-\alpha)\,\underline{w}_i(n) + \alpha\,\underline{x}(n) \\ \\ \underline{w}_i(n+1) = \underline{w}_i(n) \qquad \forall j \neq i \end{cases}$$

$\underline{w}(0)$

$\underline{w}(1) = (1-\alpha)\ \underline{w}(0) + \alpha\ \underline{x}(1)$

$$\begin{aligned}
\underline{w}(2) &= (1-\alpha)\ \underline{w}(1) + \alpha\ \underline{x}(2) \\
&= (1-\alpha)[(1-\alpha)\ \underline{w}(0) + \alpha\ \underline{x}(1)] + \alpha\ \underline{x}(2) \\
&= (1-\alpha)^2\ \underline{w}(1) + \alpha(1-\alpha)\ \underline{x}(1) + \alpha\ \underline{x}(2)
\end{aligned}$$

$$\begin{aligned}
\underline{w}(3) &= (1-\alpha)\ \underline{w}(2) + \alpha\ \underline{x}(3) \\
&= (1-\alpha)[(1-\alpha)^2\ \underline{w}(0) + \alpha(1-\alpha)\ \underline{x}(1) + \alpha\ \underline{x}(2)] + \alpha\ \underline{x}(3) \\
&= (1-\alpha)^3\ \underline{w}(0) + \alpha(1-\alpha)^2\ \underline{x}(1) + \alpha(1-\alpha)\ \underline{x}(2) + \alpha\ x(3)
\end{aligned}$$

$\vdots$

$$\begin{aligned}
w(n) &= (1-\alpha)^n\ \underline{w}(0) + \alpha[(1-\alpha)^{n-1}\ \underline{x}(1) + (1-\alpha)^{n-2}\ \underline{x}(2) + \ldots + \underline{x}_n] \\
&= (1-\alpha)^n\ \underline{w}(0) + \alpha \sum_{i=1}^{n} (1-\alpha)^{n-i}\ \underline{x}(i)
\end{aligned}$$

algumas aproximações:

$$0 < \alpha < 1 \qquad 0 < (1-\alpha) < 1 \qquad n >> 1 \qquad (1-\alpha)^n \to 0$$

$$1 + (1-\alpha) + (1-\alpha)^2 + (1-\alpha)^3 + \ldots = \sum_{i=0}^{\infty} (1-\alpha)^i = \frac{1}{1-(1-\alpha)} = \frac{1}{\alpha}$$

$$n >> 1 \qquad \alpha \approx \frac{1}{\sum_{i=0}^{n-1} (1-\alpha)^i}$$

$$\underline{w}(n) = \overset{\mathbf{0}}{(1-\alpha)^n} \, \underline{w}(0) + \alpha \sum_{i=1}^{n} (1-\alpha)^{n-i} \, \underline{x}(i)$$

$$\cong \frac{\sum_{i=1}^{n} (1-\alpha)^{n-i} \vec{x}(i)}{\sum_{i=0}^{n-1} (1-\alpha)^i} = \frac{\sum_{i=0}^{n-1} (1-\alpha)^i \vec{x}(n-i)}{\sum_{i=0}^{n-1} (1-\alpha)^i}$$

Média das entradas que pertencem a classe

ponderada geometricamente pelo tempo !

$$\underline{w}_n \cong \frac{\sum_{i=1}^{n} (1-\alpha)^{n-i}\vec{x}(i)}{\sum_{i=1}^{n} (1-\alpha)^{n-i}} = \left(\frac{1}{\sum_{i=0}^{n-1} (1-\alpha)^{i}} \right) \sum_{i=0}^{n-1} (1-\alpha)^{i}\vec{x}(n-i)$$

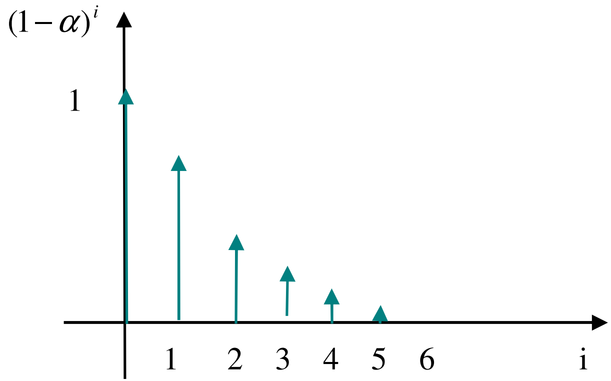
instante atual = n

$\underline{x}$(n-í) = entrada atrasada

de i intervalos de tempo

$(1-\alpha)^i$ = ponderador

para a entrada $\underline{x}$(n-í)



Note que se

$$\alpha \cong 1$$

e a estatística de $\vec{x}$ for invariante no tempo

$$\vec{w}_n \cong \left(\frac{1}{\sum_{i=0}^{n-1} (1 - \alpha)^i} \right) \sum_{\mathrm{i=0}}^{\mathrm{n-1}} (1 - \alpha)^i \vec{x}(n-i) \cong E[\vec{x}]$$

$$\vec{w}_n \cong E[\vec{x}]$$

4.1 – Tempo de medida

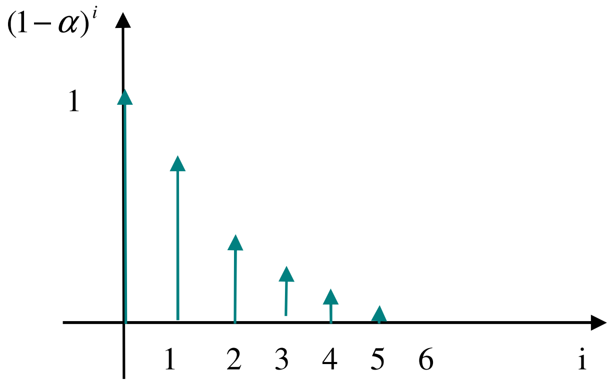
Fim (prático) da soma ponderada (tempo de medida)

Atraso zero **Ponderador = $(1-\alpha)^0 = 1$**

Último atraso à ser considerado: **Ponderador = $.02 = (1-\alpha)^i$**

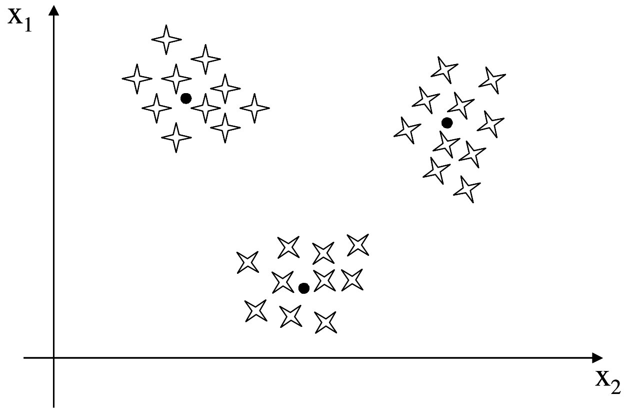
$$i = \frac{\ln .02}{\ln(1-\alpha)} \approx \frac{-4}{-\alpha}\Big|_{\alpha\to 0}$$

$$i \approx \frac{4}{\alpha}$$



4.2 - Erro na determinação do baricentro:

Se em uma classificação por similaridade cada elemento $\underline{x}$ de uma classe C_k pode ser representado pelo padrão $\underline{w}_k$ da classe adicionado de um vetor de ruído $\underline{r}$ com média nula.



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

Cada componente x_j de $\underline{x}$ é então representada pela componente correspondente de $\underline{w}_k$, w_j, adicionada de um ruído r_j de média nula e variância σ_{xj}^2

$$x_j = w_j + r_j$$

Com que precisão componente w_j é calculada pela camada de Kohonen, qual a sua variância σ_{wj}^{2} ? A componente j de $\underline{w}$, w_j, em um instante n >> 1 é dada por

$$w_j \cong \left(\frac{1}{\sum_{i=0}^{n} (1-\alpha)^i} \right) \sum_{i=0}^{n} (1-\alpha)^i x_j(n-i)$$

Sendo uma soma ponderada sua variância será dada por

$$\sigma_{wj}^2 \cong \left(\frac{1}{\sum_{i=0}^{\infty} (1-\alpha)^i} \right)^2 \sum_{i=0}^{\infty} (1-\alpha)^{2i} \sigma_{xj}^2 \qquad ou \qquad \sigma_{wj}^2 \cong \frac{\alpha}{2} \sigma_{xj}^2$$

4. 3 – Compromisso entre Erro vs. Tempo de Medida

Quanto menor α mais precisamente o baricentro será determinado. Mas, como esperado, maior o tempo (número de passos) necessário para calculá-lo

$$\sigma_{wj}^2 \cong \frac{\alpha}{2}\,\sigma_{xj}^2 \qquad\qquad \#\,passos = i \cong \frac{4}{\alpha}$$

Exemplo:

σ_x = .05 (5%)

σ_w = .01 (1%) requerido $\quad \alpha = \dfrac{2\sigma_{wj}^2}{\sigma_{xj}^2} = .08 \qquad \#\,passos = \dfrac{4}{\alpha} = 50\ passos$

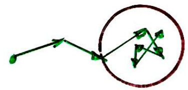
Mas se:

σ_w = .001 (.1%) requerido $\quad \alpha = \dfrac{2\sigma_{wj}^2}{\sigma_{xj}^2} = .0008 \qquad \#\,passos = \dfrac{4}{\alpha} = 5.000\ passos$!!

4.4 Considerações:

4.4.1. Fim do treinamento ?

$$E[\Delta \underline{w}] = \underline{0}$$



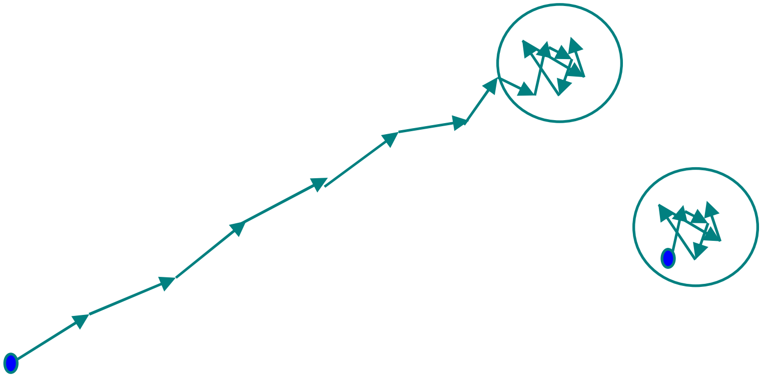
$$\Delta\vec{w} = \alpha\left(\vec{x} - \vec{w}\right)$$

$$E(\Delta\vec{w}) = E[\alpha\left(\vec{x} - \vec{w}\right)] = \alpha[E(\vec{x}) - \vec{w}] = 0$$

$$\vec{w} = E(\vec{x})$$

4.4.2. Valores iniciais das sinapses

Irrelevantes ! Mas escolher o valor inicial como o primeiro elemento da classe acelera o treinamento, porque a sinapse já começa dentro da classe.



4.4.3. Passo de treinamento

reduzir ao longo do tempo $\alpha(n) = \alpha_0 \, e^{-\frac{n}{N_0}}$

$$\alpha(0) = \alpha_0 \qquad \alpha(n+1) = k\,\alpha(n) \qquad \text{onde} \qquad k = 1 - \frac{1}{N_0}$$

o processo acaba ($\alpha(n) << \alpha_0$) para $n > 4N_0$

e as sinapses pouco ativadas (classes pouco populosas) ?

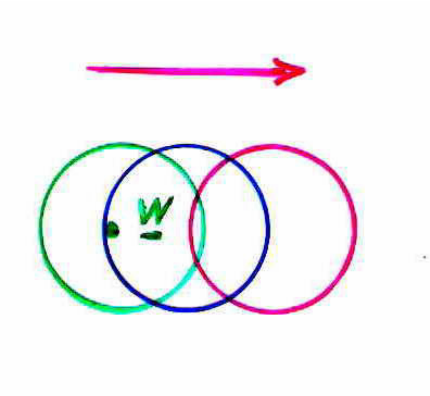
usar α diferenciado por sinapse

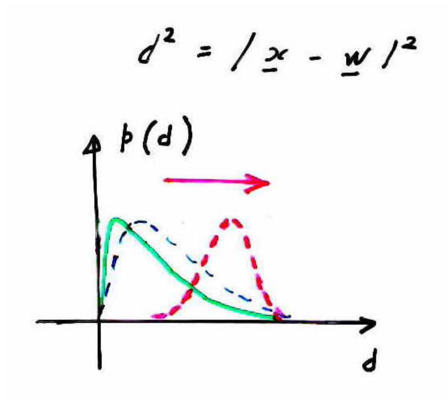
$$\alpha_{w_j}(0) = \alpha_0 \qquad \alpha_{w_j}(n_j + 1) = k\,\alpha_{w_j}(n_j)$$

onde n_j é o número de vezes que a sinapse $\underline{w}_j$ foi treinada.

4.4.4 Treinamento dinâmico, adaptativo

Uma classe varia de posição no tempo. Como saber ?

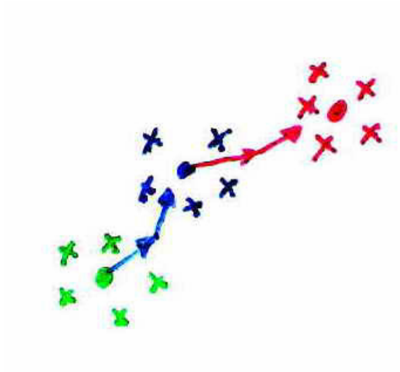


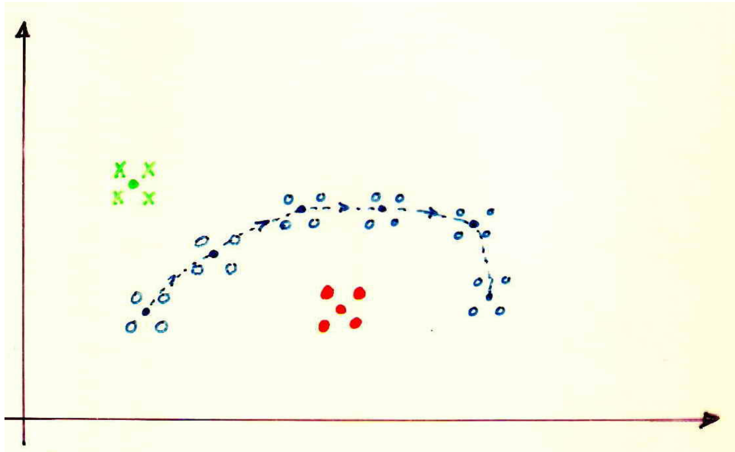


O valor médio de $u_i = - d^2_i$ diminui

(o valor da distância média (ou a da moda) aumenta)

Como corrigir ? Ligar o treinamento até que d^2_i volte a seus valores normais.

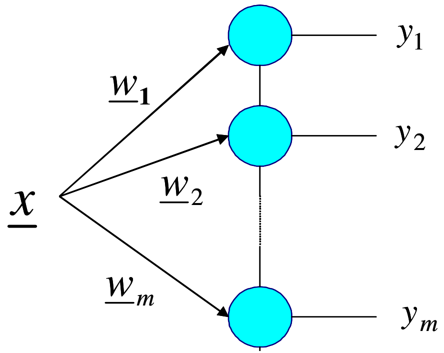




Se a variação do baricentro for lenta e o treinamento estiver ligado o padrão segue o baricentro

5 – Um método alternativo

LVQ – Learning Vector Quantization



$\vec{x}(n) \in C_i$ aproxima o padrão vencedor de $\vec{x}(n)$ e afasta os demais padrões de $\vec{x}(n)$

$$\vec{x}(n) \in C_i \qquad \begin{cases} y_i = 1 \\ y_j = 0 \quad \forall\, j \neq i \end{cases}$$

$$\vec{w}_i(n+1) = \vec{w}_i(n) + \alpha\left[\vec{x}(n) - \vec{w}_i(n)\right]$$

$$\vec{w}_j(n+1) = \vec{w}_j(n) - \alpha\left[\vec{x}(n) - \vec{w}_j(n)\right] \quad \forall\, j \neq i$$

ou, usando uma fórmula única

$$\vec{x}(n) \in C_i \qquad \begin{cases} y_i = 1 \\ y_j = 0 \quad \forall\, j \neq i \end{cases}$$

$$\vec{w}_j(n+1) = \vec{w}_j(n) + (2y_i - 1)\,\alpha\left[\vec{x}(n) - \vec{w}_j(n)\right] \qquad \forall\, j$$

O passo de aprendizagem deve decrescer com o tempo

$$\alpha = \alpha(n) = \alpha_0 \exp\left(-\frac{n}{\tau_\alpha}\right) \qquad \alpha_0 = 0{,}1 \qquad \tau_\alpha = 1000$$

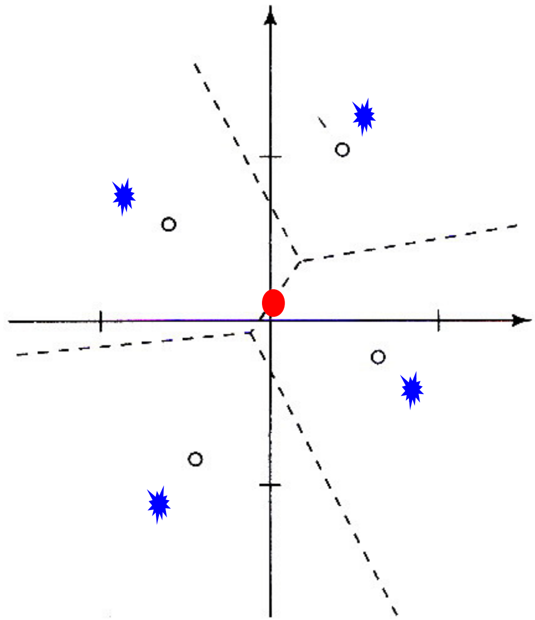
Este processo melhora as fronteiras de classificação (Haykin).

O equilíbrio é atingido quando $E[\Delta\vec{w}_i]=\vec{0}$, o que ocorre quando

$$\vec{w}_i = p_i \underset{\forall\vec{x}\in C_i}{E}(\vec{x}) - (1-p_i)\underset{\forall\vec{x}\notin C_i}{E}(\vec{x}) =$$

$$= 2p_i \underset{\forall\vec{x}\in C_i}{E}(\vec{x}) - \underset{\forall\vec{x}}{E}(\vec{x}) =$$

$$= 2p_i\vec{m}_i - \vec{m}$$



onde p_i é a probabilidade de uma entrada $\vec{x}$ pertencer à classe C_i , $\vec{m}_i$ é o baricentro de C_i e $\vec{m}$ é o baricentro de todas as entradas, normalmente $\vec{m}=\vec{0}$. O efeito é deslocar os padrões das classes ( ) para longe do centro de todas as entradas ( ) mas também dos baricentros (O) das mesmas.

Este processo corresponde a otimizar a função objetivo

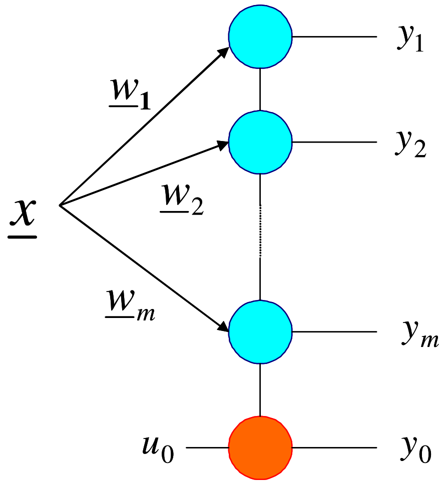
$$F(\vec{w}_i) = \sum_{\forall \vec{x} \in C_i} \left| \vec{x} - \vec{w}_i \right|^2 - \sum_{\forall \vec{x} \notin C_i} \left| \vec{x} - \vec{w}_i \right|^2$$

isto é, minimizar a dispersão intra classe de C_i e maximizar a distância de w_i aos elementos das outras classes.

6 - E o segundo critério (similaridade mínima) ?

Camada de Kohonen (aumentada)

$$u_0 = -r_0^2$$



Se $y_i = 1$ entao

$\underline{x} \in C_i$

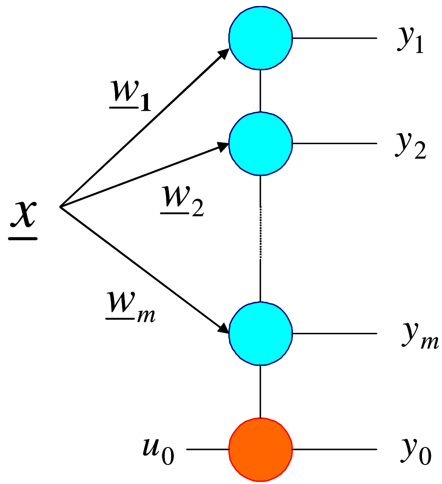
pelos critérios 1 (padrao mais similar a entrada) e

2 (satisfaz a similaridade minima exigida).

Se $\mathbf{y_i = 1}$ então

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

pelos critérios
1 (centro de classe mais similar à entrada) e
2 (satisfaz à similaridade mínima)



Se $\mathbf{y_0 = 1}$ então

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

$\underline{\mathbf{x}}$ não satisfaz ao critério 2
para nenhuma classe

7 – HVQ – Hierarquical Vector Quantization