

# Memórias Associativas Recorrentes com Núcleo

**Aline Cristina de Souza**

**Orientador: Prof. Dr. Marcos E. Valle**

Instituto de Matemática, Estatística e Computação Científica - IMECC  
Universidade Estadual de Campinas - UNICAMP

14 de Maio de 2015

## **Face Recognition by Applying Wavelet Subband Representation and Kernel Associative Memory.**

Bai-Ling Zhang, Haihong Zang e Shuzhi Sam Ge, 2004.

135 citações / 55 citações (2010-2015)

Aplicações: memórias associativas; problemas de classificação; reconhecimento de faces;

## **Face Recognition by Applying Wavelet Subband Representation and Kernel Associative Memory.**

Bai-Ling Zhang, Haihong Zang e Shuzhi Sam Ge, 2004.

135 citações / 55 citações (2010-2015)

Aplicações: memórias associativas; problemas de classificação; reconhecimento de faces;

## **Recurrent Correlation Associative Memories: A Feature Space Perspective.**

Renzo Perfetti e Elisa Ricci (2008).

# Organização da Apresentação

- 1 Memórias Associativas Recorrentes por Correlação
- 2 Máquinas de Vetor de Suporte
- 3 Memórias Associativas Recorrentes com Núcleo

# Memórias Associativas com Núcleo - KAM

Seja  $\{(\mathbf{x}^1, \mathbf{y}^1), \dots, (\mathbf{x}^p, \mathbf{y}^p)\}$  o conjunto das memórias fundamentais.

A equação da KAM é dada por:

$$\mathbf{y} = \sum_{\xi=1}^p k(\mathbf{x}^{\xi}, \mathbf{x}) \mathbf{y}^{\xi}, \quad (1)$$

onde  $k(\mathbf{x}^{\xi}, \mathbf{x})$  é uma função núcleo.

Um exemplo de função núcleo é dado pela função de base radial gaussiana definida por:

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right). \quad (2)$$

# Memórias Associativas Recorrentes por Correlação - RCAM

- Memórias Associativas Recorrentes por Correlação foram introduzidas por Chiueh e Goodman, em 1991.

Seja  $\{\mathbf{x}^1, \dots, \mathbf{x}^p\}$  o conjunto das memórias fundamentais.

A equação de evolução da RCAM é definida por:

$$\mathbf{x}_i(t+1) = \operatorname{sgn} \left( \sum_{\xi=1}^p f_{\xi}(\langle \mathbf{x}^{\xi}, \mathbf{x}(t) \rangle) \mathbf{x}_i^{\xi} \right), \quad i = 1, \dots, n, \quad (3)$$

em que as funções  $f_{\xi} : [-n, n] \rightarrow \mathbb{R}$ ,  $\xi = 1, \dots, p$ , são chamadas funções peso.

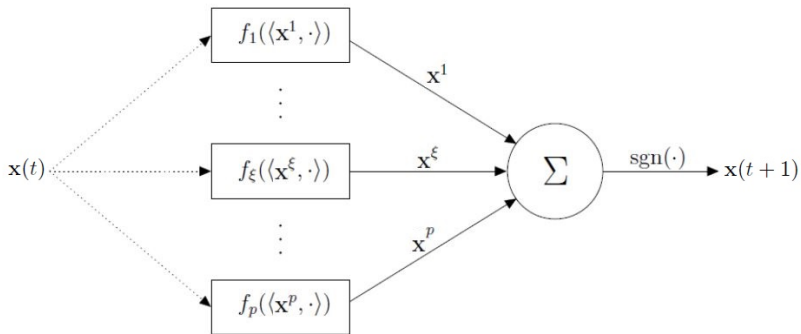


Figura: Diagrama de blocos da RCAM

# Memórias Associativas Recorrentes por Correlação Exponencial - ECAM

- A equação de evolução da ECAM é definida considerando, para todo  $k = 1, \dots, p$ ,  $f_k : [-n, n] \rightarrow \mathbb{R}$  dadas por

$$f_k(t) = e^{\alpha t}, \quad \alpha > 0. \quad (4)$$



# Memórias Associativas Recorrentes por Correlação - Exemplo

Consideremos, o conjunto das memórias fundamentais dado por:

$$\mathcal{A} = \left\{ \mathbf{x}^1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}^2 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \mathbf{x}^3 = \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

e a ECAM definida considerando  $\alpha = 1$ , isto é, consideremos que as funções peso  $f_\xi : [-n, n] \rightarrow \mathbb{R}$  são definidas por

$$f_\xi(t) = e^t, \quad \text{para } \xi = 1, 2, 3.$$

# Memórias Associativas Recorrentes por Correlação

Consideremos agora como padrão de entrada o vetor:

$$\mathbf{x}(0) = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \quad (5)$$

Temos que  $\langle \mathbf{x}^1, \mathbf{x}(0) \rangle = -1$ ,  $\langle \mathbf{x}^2, \mathbf{x}(0) \rangle = 3$  e  $\langle \mathbf{x}^3, \mathbf{x}(0) \rangle = -3$ .

Dessa forma, obtemos:

$$\mathbf{x}(1) = \text{sgn} \left( \sum_{k=1}^p e^{\langle \mathbf{x}^k, \mathbf{x}(0) \rangle} \mathbf{x}^k \right) = \text{sgn} (e^{-1} \mathbf{x}^1 + e^3 \mathbf{x}^2 + e^{-3} \mathbf{x}^3) = \mathbf{x}^2.$$

Assim, a saída correspondente ao vetor  $\mathbf{x}(0)$  é  $\mathbf{x}^2$ .

# Memórias Associativas Recorrentes por Correlação

Temos que:

$\langle \mathbf{x}^1, \mathbf{x}(0) \rangle = -1$	$f_1(\langle \mathbf{x}^1, \mathbf{x}(0) \rangle) = e^{-1} \approx 0.3679$
$\langle \mathbf{x}^2, \mathbf{x}(0) \rangle = 3$	$f_2(\langle \mathbf{x}^2, \mathbf{x}(0) \rangle) = e^3 \approx 20.0855$
$\langle \mathbf{x}^3, \mathbf{x}(0) \rangle = -3$	$f_3(\langle \mathbf{x}^3, \mathbf{x}(0) \rangle) = e^{-3} \approx 0.0498$

Assim, podemos observar que, uma vez que as funções peso  $f_\xi$ ,  $\xi = 1, \dots, p$ , são não decrescentes, elas enfatizam os pesos das memórias fundamentais.

## Teorema (Convergência RCAM)

Seja  $f : [-n, n] \rightarrow \mathbb{R}$  uma função contínua e monótona não decrescente. Então a RCAM com equação de evolução:

$$\mathbf{x}(t + 1) = \operatorname{sgn} \left( \sum_{\xi=1}^p f(\langle \mathbf{x}^{\xi}, \mathbf{x}(t) \rangle) \mathbf{x}^{\xi} \right) \quad (6)$$

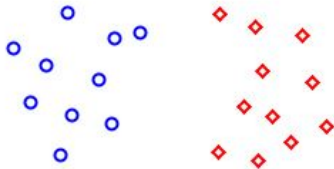
é assintoticamente estável em ambos os modos síncrono e assíncrono de atualização.

- Capacidade de Armazenamento

A capacidade de armazenamento da Memória Associativa por Correlação Exponencial (ECAM), cresce exponencialmente com o comprimento dos vetores  $\mathbf{x}^\xi$ ,  $\xi = 1, \dots, p$ .

# Máquinas de Vetor de Suporte

Sejam  $\{\mathbf{x}^\xi, y_\xi\} \subset \mathbb{R}^n \times \{-1, 1\}$ ,  $\xi = 1, \dots, p$  o conjunto de exemplos de treinamento para um problema de classificação e suponha que as classes representadas pelo 1 e  $-1$  são linearmente separáveis.



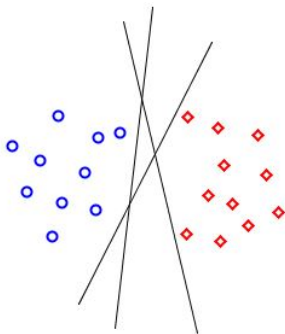
A equação do hiperplano de separação é dada por:

$$\mathbf{w}^t \mathbf{x} + b = 0, \quad (7)$$

em que  $\mathbf{w}$  é um vetor de pesos e  $b$  o bias.

Assim, temos que:

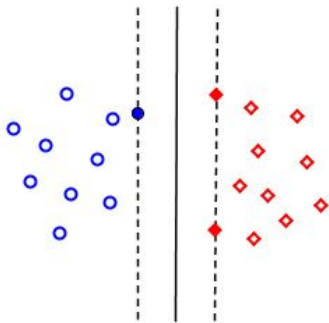
$$\begin{cases} \mathbf{w}^t \mathbf{x}^\xi + b \geq 0, & \text{se } y_\xi = 1 \\ \mathbf{w}^t \mathbf{x}^\xi + b < 0, & \text{se } y_\xi = -1 \end{cases}$$



Para um dado vetor  $\mathbf{w}$  e bias  $b$ , a distância entre um hiperplano de separação e o vetor  $\mathbf{x}^\xi$  mais próximo é chamada **margin de separação** e será denotada por  $\rho$ .



Uma Máquina de Vetor de Suporte identifica o hiperplano de separação ótimo, o qual maximiza a margem de separação. Os vetores  $\mathbf{x}^{\xi}$  mais próximos do hiperplano de separação ótimo são chamados **vetores de suporte**.



Suponha que o hiperplano ótimo seja definido por:

$$\mathbf{w}_0^t \mathbf{x} + b_0 = 0 \quad (8)$$

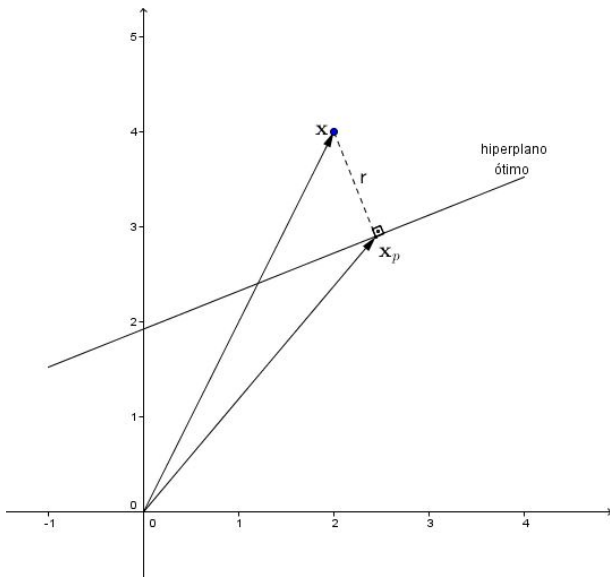
e considere a função  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  definida por

$$g(\mathbf{x}) = \mathbf{w}_0^t \mathbf{x} + b_0.$$

Dado  $\mathbf{x} \in \mathbb{R}^n$ , podemos escrever

$$\mathbf{x} = \mathbf{x}_p + r \frac{\mathbf{w}_0}{\|\mathbf{w}_0\|},$$

onde  $\mathbf{x}_p$  é a projeção ortogonal de  $\mathbf{x}$  sobre o hiperplano ótimo e  $|r|$  é a distância de  $\mathbf{x}$  ao hiperplano.



Como  $\mathbf{x}_p$  pertence ao hiperplano, temos que  $g(\mathbf{x}_p) = 0$ .

Assim,

$$g(\mathbf{x}) = \mathbf{w}_0^t \mathbf{x} + b_0 = \mathbf{w}_0^t \left( \mathbf{x}_p + r \frac{\mathbf{w}_0}{\|\mathbf{w}_0\|} \right) + b_0 = g(\mathbf{x}_p) + r \|\mathbf{w}_0\|.$$

Portanto,

$$r = \frac{g(\mathbf{x})}{\|\mathbf{w}_0\|}.$$

Podemos reescalar  $\mathbf{w}_0$  e  $b_0$  para obter:

$$\begin{cases} \mathbf{w}^t \mathbf{x}^\xi + b \geq 1, & \text{se } y_\xi = 1 \\ \mathbf{w}^t \mathbf{x}^\xi + b \leq -1, & \text{se } y_\xi = -1 \end{cases}$$

Observemos que para os vetores de suporte é válida a igualdade.

Considere um vetor de suporte  $\mathbf{x}^{(s)}$ . Então, temos

$$g(\mathbf{x}^{(s)}) = \mathbf{w}_0^t \mathbf{x}^{(s)} + b_0 = \pm 1, \quad \text{para } y_{(s)} = \pm 1.$$

Assim, temos

$$r^{(s)} = \frac{g(\mathbf{x}^{(s)})}{\|\mathbf{w}_0\|} = \begin{cases} \frac{1}{\|\mathbf{w}_0\|}, & \text{se } y_{(s)} = 1 \\ -\frac{1}{\|\mathbf{w}_0\|}, & \text{se } y_{(s)} = -1 \end{cases}$$

Portanto, denotando a margem de separação por  $\rho$ , temos:

$$\rho = |r| = \frac{1}{\|\mathbf{w}_0\|}.$$

Logo, maximizar  $\rho$  é equivalente a minimizar  $\|\mathbf{w}_0\|$ .

## Problema Primal

Dado o conjunto de dados de treinamento  $\{(\mathbf{x}^\xi, y_\xi)\}_{\xi=1}^p$  linearmente separáveis, determine  $\mathbf{w}$  e  $b$  que minimizam a função

$$\Phi(\mathbf{w}) = \frac{1}{2} \mathbf{w}^t \mathbf{w},$$

sujeito à restrição

$$y_\xi(\mathbf{w}_0^t \mathbf{x}^\xi + b) \geq 1, \quad \forall \xi = 1, \dots, p.$$

Este problema pode ser resolvido utilizando o **Método dos Multiplicadores de Lagrange**, para transformar o problema primal no problema dual correspondente.

Consideremos a função Lagrangeana do problema primal :

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w}^t \mathbf{w} - \sum_{i=1}^p \alpha_i (y_i (\mathbf{w}^t \mathbf{x}^i + b) - 1),$$

onde  $\alpha_i \geq 0$ , são chamados multiplicadores de Lagrange.

Condições de Otimização:

$$1) \frac{\partial J(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = \mathbf{0} \Rightarrow \mathbf{w} - \sum_{i=1}^p \alpha_i y_i \mathbf{x}^i = \mathbf{0} \Rightarrow \mathbf{w} = \sum_{i=1}^p \alpha_i y_i \mathbf{x}^i$$

$$2) \frac{\partial J(\mathbf{w}, b, \alpha)}{\partial b} = 0 \Rightarrow \sum_{i=1}^p \alpha_i y_i = 0$$



## Problema dual

Dado o conjunto de dados de treinamento  $\{(\mathbf{x}^\xi, y_\xi)\}_{\xi=1}^p$ ,  
encontre os multiplicadores de Lagrange  $\{\alpha_\xi\}_{\xi=1}^p$  que  
minimizam a função

$$Q(\alpha) = \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \alpha_i \alpha_j y_i y_j (\mathbf{x}^i)^t \mathbf{x}^j - \sum_{i=1}^p \alpha_i$$

sujeito às restrições

$$\sum_{i=1}^p \alpha_i y_i = 0,$$

$$\alpha_i \geq 0,$$

para todo  $i = 1, \dots, p$ .

Além disso, temos que

$$\alpha_{0i}[y_i(\mathbf{w}_0^t \mathbf{x}^i + b_0) - 1] = 0,$$

para  $i = 1, \dots, p$ .

Assim, se  $\mathbf{x}^i$  não for vetor de suporte, então  $\alpha_{0i} = 0$ .

- O problema dual é formulado inteiramente em termos dos dados de treinamento;
- Após calcular os valores de  $\alpha_{0i}$  ótimos, podemos calcular  $\mathbf{w}_0$  por meio da equação:

$$\mathbf{w}_0 = \sum_{i=1}^{N_s} \alpha_{0i} y_i \mathbf{x}^i,$$

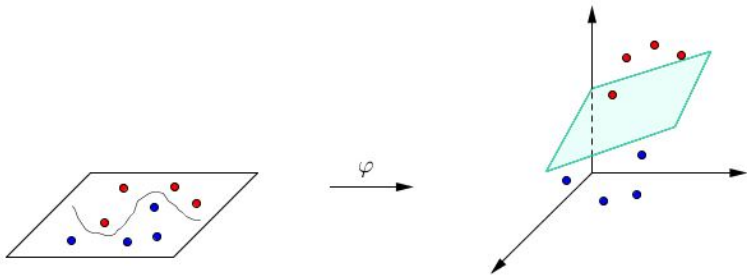
onde  $N_s$  é o número de vetores de suporte.

- Para calcular  $b_0$ , utilizamos  $\mathbf{w}_0$  e a equação:

$$\mathbf{w}_0^t \mathbf{x}^{(s)} + b_0 = 1, \quad \text{para } y^{(s)} = 1.$$

Assim, temos  $b_0 = 1 - \sum_{i=1}^{N_s} \alpha_{0i} y_i (\mathbf{x}^i)^t \mathbf{x}^{(s)}$ .

Pontos não linearmente separáveis podem ser mapeados por uma função não linear  $\varphi$  em um espaço  $\mathcal{H}$  de dimensão maior, chamado *espaço de características*, de forma que as imagens sejam linearmente separáveis.



A função de decisão ótima é dada por:

$$g(\mathbf{x}) = \sum_{i=1}^p \alpha_0 y_i K(\mathbf{x}^i, \mathbf{x}) + b_0, \quad (9)$$

onde  $K(\mathbf{x}^i, \mathbf{x}) = \langle \varphi(\mathbf{x}^i), \varphi(\mathbf{x}) \rangle$  é chamada *função núcleo*.

Somente os multiplicadores de Lagrange associados aos vetores de suporte são não nulos, então a função de decisão depende somente deles.

A maximização da margem de separação em  $\mathcal{H}$  pode ser obtida por meio da resolução do seguinte problema na forma dual

## Problema dual

Dado o conjunto de dados de treinamento  $\{(\mathbf{x}^\xi, y_\xi)\}_{\xi=1}^p$ , encontre os multiplicadores de Lagrange  $\{\alpha_\xi\}_{\xi=1}^p$  que minimizam a função

$$Q(\alpha) = \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \alpha_i \alpha_j y_i y_j K(\mathbf{x}^i, \mathbf{x}^j) - \sum_{i=1}^p \alpha_i$$

sujeito às restrições

$$\sum_{i=1}^p \alpha_i y_i = 0,$$

$$\alpha_i \geq 0,$$

para todo  $i = 1, \dots, p$

Uma função  $K$  é um núcleo válido para uma máquina de vetor de suporte se a matriz  $K = (k_{ij})_{p \times p}$  definida por

$$k_{ij} = K(\mathbf{x}^i, \mathbf{x}^j),$$

for simétrica semidefinida positiva.

## Exemplos

- 1 Função de Base Radial Gaussiana

$$K(\mathbf{x}^i, \mathbf{x}^j) = \exp(-\|\mathbf{x}^i - \mathbf{x}^j\|^2 / \sigma^2).$$

- 2 Função Núcleo Polinomial

$$K(\mathbf{x}^i, \mathbf{x}^j) = (\langle \mathbf{x}^i, \mathbf{x}^j \rangle + R)^d,$$

onde  $d$  é inteiro positivo e  $R > 0$ .

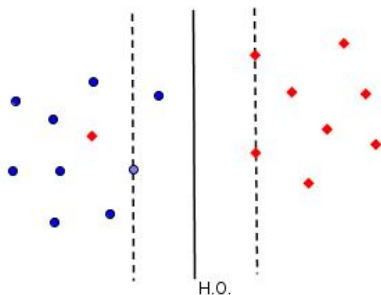
# Hiperplano ótimo para padrões não linearmente separáveis

O objetivo é determinar um hiperplano ótimo que minimize a probabilidade de erro de classificação.

A margem de separação é dita *suave* se um dado de treinamento  $(\mathbf{x}^i, y_i)$  violar a condição

$$y_i(\mathbf{w}^t \mathbf{x}^i + b) \geq 1.$$

Temos dois casos:





Introduzimos variáveis de folga  $\xi_i$ ,  $i = 1, \dots, N$ , na definição do hiperplano de separação da seguinte forma:

$$y_i(\mathbf{w}_i^t \mathbf{x}^i + b) \geq 1 - \xi_i, \quad i = 1, \dots, N.$$

## Problema Primal

Dado o conjunto de treinamento  $\{(\mathbf{x}^i, y_i)\}_{i=1}^N$  encontre os valores ótimos de  $\mathbf{w}$  e  $b$  que satisfazem as restrições

$$y_i(\mathbf{w}_i^t \mathbf{x}^i + b) \geq 1 - \xi_i, \quad i = 1, \dots, N$$

$$\xi_i \geq 0,$$

de modo que  $\mathbf{w}$  e as variáveis de folga minimizem a função

$$\Phi(\mathbf{w}, \xi) = \frac{1}{2} \mathbf{w}^t \mathbf{w} + c \sum_{i=1}^N \xi_i,$$

onde  $c > 0$  é constante.

## Problema dual

Dado o conjunto de dados de treinamento  $\{(\mathbf{x}^i, y_i)\}$ , encontre os multiplicadores de Lagrange  $\{\alpha_i\}_{i=1}^N$  que minimizam a função

$$Q(\alpha) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (\mathbf{x}^i)^t \mathbf{x}^j - \sum_{i=1}^N \alpha_i$$

sujeito às restrições

$$\sum_{i=1}^N \alpha_i d_i = 0,$$

$$0 \leq \alpha_i \leq c.$$

Seja  $\mathcal{A} = \{\mathbf{x}^1, \dots, \mathbf{x}^p\} \subset \{-1, 1\}^n$  o conjunto das memórias fundamentais. A equação de evolução de uma Memória Associativa Recorrente com Núcleo é definida por:

$$x_i(t+1) = \operatorname{sgn} \left( \sum_{\xi=1}^p \alpha_i^\xi x_i^\xi K(\mathbf{x}^\xi, \mathbf{x}(t)) \right), \quad \forall i = 1, \dots, n, \quad (10)$$

onde  $K$  é uma função núcleo válida do produto interno entre  $\mathbf{x}^\xi$  e  $\mathbf{x}(t)$  e  $\alpha_i^\xi$ ,  $i = 1, \dots, n$ ,  $\xi = 1, \dots, p$  são coeficientes não negativos.

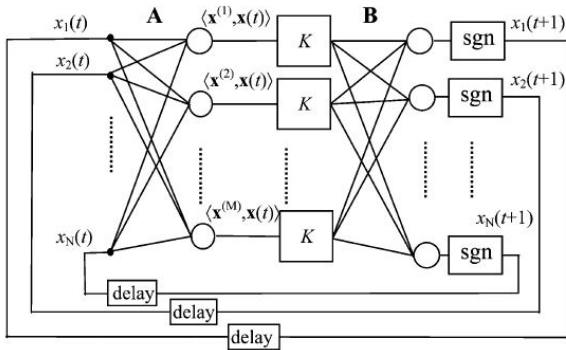


Figura: Diagrama de blocos da RKAM <sup>1</sup>.

<sup>1</sup>Retirada do artigo: Recurrent Correlation Associative Memories: A Feature space Perspective. Renzo Perfetti e Elisa Ricci (2008).

Comparando a função de decisão  $g$  dada por (9) com a equação de evolução da RKAM (10)

$$g(\mathbf{x}) = \sum_{i=1}^p \alpha_{0i} y_i K(\mathbf{x}^i, \mathbf{x}) + b_0,$$

$$x_i(t+1) = \operatorname{sgn} \left( \sum_{\xi=1}^p \alpha_i^{\xi} x_i^{\xi} K(\mathbf{x}^{\xi}, \mathbf{x}(t)) \right)$$

e considerando que os neurônios são atualizados de forma assíncrona, cada passo de atualização pode ser interpretado como uma operação de classificação, que atribui o rótulo  $x_i(t+1)$  ao vetor  $\mathbf{x}(t)$ .

A síntese da RKAM pode ser formulada como a solução de  $n$  problemas de classificação, em que os conjuntos de treinamento são dados por  $\{(\mathbf{x}^\xi, x_i^\xi)\} \subset \mathcal{A} \times \{-1, 1\}$ ,  $\xi = 1, \dots, p$  e  $i = 1, \dots, n$ .

## Propriedade

A RKAM se torna equivalente a RCAM com  $f = K$ , se a matriz  $K = (k_{ij})_{p \times p}$  definida por

$$k_{ij} = K(\mathbf{x}^i, \mathbf{x}^j),$$

e tal que  $k_{ii} = k_0$ , para todo  $i = 1, \dots, p$  e satisfaz a condição

$$k_{ii} \gg |k_{ij}|,$$

para  $i, j = 1, \dots, p, i \neq j$ .



A condição

$$k_{ii} \gg |k_{ij}|,$$

para  $i, j = 1, \dots, p$ ,  $i \neq j$  corresponde a aproximar a ortogonalidade de qualquer par de memórias fundamentais distintas no espaço de características. De fato, denotando o ângulo entre  $\mathbf{x}^i$  e  $\mathbf{x}^j$  no espaço de características por  $\theta_{ij}$ , temos:

$$\cos \theta_{ij} = \frac{\langle \varphi(\mathbf{x}^i), \varphi(\mathbf{x}^j) \rangle}{\|\varphi(\mathbf{x}^i)\| \|\varphi(\mathbf{x}^j)\|} = \frac{k_{ij}}{\sqrt{k_{ii} k_{jj}}} \approx 0.$$

# Núcleo Exponencial

Consideremos o núcleo exponencial definido por

$$K(\mathbf{x}^i, \mathbf{x}^j) = a^{\langle \mathbf{x}^i, \mathbf{x}^j \rangle},$$

com  $a > 1$ .

A equação de evolução da RKAM se torna:

$$x_i(t+1) = \operatorname{sgn} \left( \sum_{j=1}^p \alpha_i^j x_i^j a^{\langle \mathbf{x}^i, \mathbf{x}(t) \rangle} \right), \quad i = 1, \dots, n \quad (11)$$

# Relação entre a função de base radial Gaussiana e o núcleo exponencial

Observemos que  $\|\mathbf{x}^i - \mathbf{x}^j\|^2 = 2n - 2\langle \mathbf{x}^i, \mathbf{x}^j \rangle$ .

Temos que:

$$\exp(-\|\mathbf{x}^i - \mathbf{x}^j\|^2/\sigma^2) = \exp(-2n/\sigma^2)\exp(2\langle \mathbf{x}^i, \mathbf{x}^j \rangle/\sigma^2) = Ga^{\langle \mathbf{x}^i, \mathbf{x}^j \rangle},$$

com  $a = \exp(2/\sigma^2)$  e  $G = \exp(-2n/\sigma^2)$ .

Este fator  $G$  faz com que os multiplicadores de Lagrange fiquem multiplicados por  $1/G$ , o que não influencia no comportamento do sistema descrito pela equação (11).

Portanto o núcleo Gaussiano é equivalente ao núcleo exponencial.

Temos que  $\langle \mathbf{x}^i, \mathbf{x}^j \rangle = n - 2d_H(\mathbf{x}^i, \mathbf{x}^j)$ , onde  $d_H(\mathbf{x}^i, \mathbf{x}^j)$  denota a distância de Hamming entre  $\mathbf{x}^i$  e  $\mathbf{x}^j$ .

Assim, para o núcleo exponencial, a matriz  $K = (k_{ij})$  é dada por:

$$k_{ii} = a^n, \quad k_{ij} = a^n a^{-2d_H(\mathbf{x}^i, \mathbf{x}^j)}.$$

Portanto, temos:

$$\cos \theta_{ij} = \frac{\langle \varphi(\mathbf{x}^i), \varphi(\mathbf{x}^j) \rangle}{\|\varphi(\mathbf{x}^i)\| \|\varphi(\mathbf{x}^j)\|} = \frac{k_{ij}}{\sqrt{k_{ii} k_{jj}}} = a^{-2d_H(\mathbf{x}^i, \mathbf{x}^j)}.$$

Dessa forma, o núcleo exponencial ortogonaliza os padrões de entrada se a distância entre eles é suficientemente grande. Desde que  $\cos \theta_{ij} \rightarrow 0$  quando  $a \rightarrow \infty$ , o núcleo exponencial pode ortogonalizar padrões de entrada arbitrários se  $a$  é suficientemente grande.

A equação de evolução da ECAM é definida por:

$$\mathbf{x}_i(t+1) = \operatorname{sgn} \left( \sum_{\xi=1}^p a^{\langle \mathbf{x}^\xi, \mathbf{x}(t) \rangle} \mathbf{x}_i^\xi \right), \quad i = 1, \dots, n, \quad (12)$$

Portanto a ECAM é equivalente a RKAM com núcleo exponencial se  $\alpha_i^j = 1$  para todo  $i = 1, \dots, n$ , para todo  $j = 1, \dots, p$ .

Obrigada!