



CT 720 Tópicos em Aprendizagem de Máquina e  
Classificação de Padrões



## 4-Técnicas Não Paramétricas

# Conteúdo

1. Introdução
2. Estimação de densidades
3. Janelas de Parzen
4. Estimação  $k_n$ -vizinhos-mais próximos
5. Classificação fuzzy
6. Resumo

# 1-Introdução

- Técnicas não paramétricas de estimação
  - procedimentos para estimar distribuições arbitrárias
  - não assume forma analítica das densidade
  - estimação utiliza de dados de treinamento
- Este capítulo
  - apresenta métodos principais de estimação
  - estimação das densidades  $p(\mathbf{x}/\omega_i)$
  - estimação direta das probabilidades *a posteriori*  $P(\omega_i/\mathbf{x})$

## 2-Estimação de densidade

- Idéia básica

- probabilidade  $P$  que um vetor  $\mathbf{x}$  está em uma região  $R$

$$P = \int_R p(\mathbf{x}) d\mathbf{x} \quad (1)$$

- $P$  : versão média da função densidade  $p(\mathbf{x})$
- podemos estimar este valor médio de  $p$  estimando  $P$
- $\mathbf{x}_1, \dots, \mathbf{x}_n$  amostras independentes distribuídas de  $p(\mathbf{x})$
- probabilidade que  $k$  destas  $n$  amostras estejam em  $R$

$$P_k = \binom{n}{k} P^k (1-P)^{n-k} \quad (2)$$

binomial

$$E[k] = nP \quad (3)$$

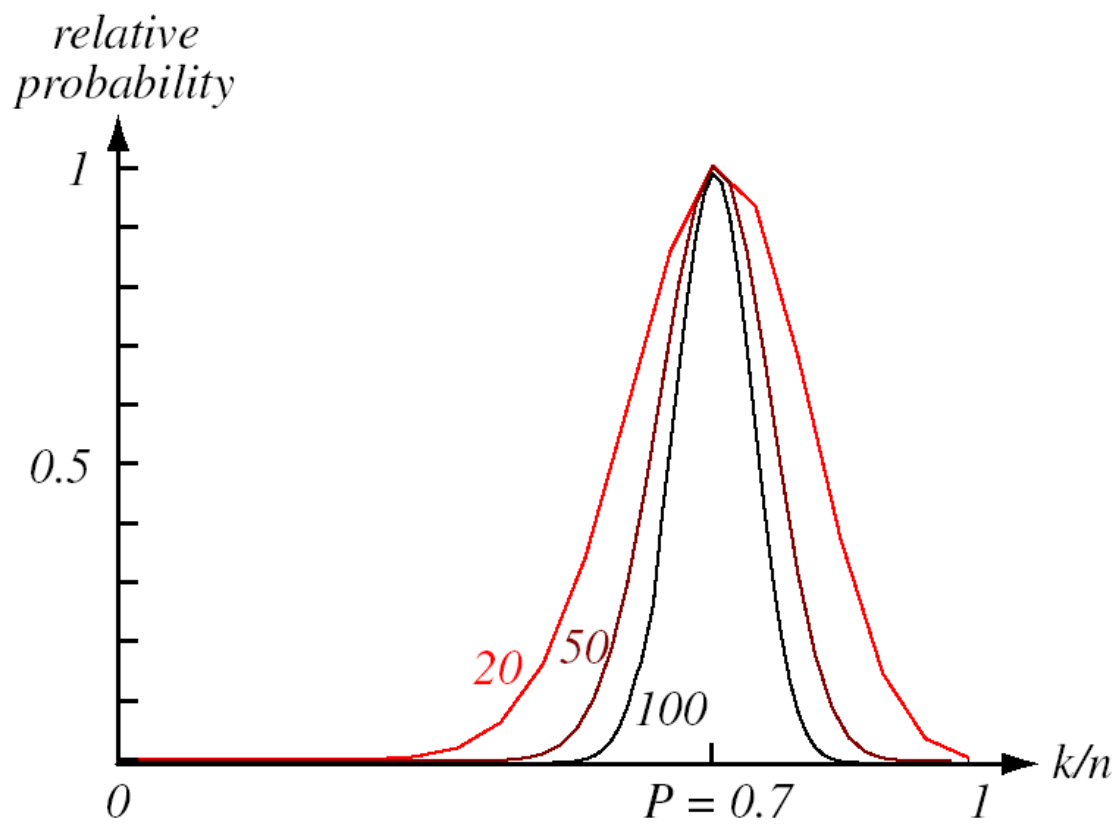
$$P = \frac{k}{n}$$

assumindo  $p(\mathbf{x})$  contínua,  $R$  pequena ( $p \sim \text{cte}$ )

$$\int_R p(\mathbf{x}) d\mathbf{x} \approx p(\mathbf{x})V \quad \mathbf{x} \in R, V \text{ volume de } R \quad (4)$$

combinando (1), (3) e (4)

$$p(\mathbf{x}) \approx \frac{k/n}{V}$$



## ■ Questões

1– se  $V$  é fixo, então  $k/n$  converge (em probabilidade) para  $P$  e

$$\frac{P}{V} = \frac{\int_R p(\mathbf{x}) d\mathbf{x}}{\int_R d\mathbf{x}} \quad \text{valor médio (espacial) de } p(\mathbf{x})$$

2– para obter  $p(\mathbf{x})$  ao invés do valor médio,  $V \rightarrow 0$

$V \rightarrow 0$  implica que  $R \rightarrow 0$  e pode não conter nenhuma amostra

isto é  $p(\mathbf{x}) \rightarrow 0$

## ■ Limitações

- número limitado de amostras
- $V$  não pode ser arbitrariamente pequena
- variância na razão  $k/n$  como estimativa de  $p(\mathbf{x})$

## ■ Questão

- podemos superar estas limitações m um número ilimitado de amostras ?



## ■ Procedimento

- formar uma sequência de regiões  $R_1, R_2, \dots$  contendo  $\mathbf{x}$
- $R_1$  região a ser usada com uma amostra
- $R_2$  região a ser usada com duas amostras
- .....

$$p_n(\mathbf{x}) = \frac{k_n / n}{V_n} \quad (7)$$

- $V_n$  volume de  $R_n$
- $k_n$  número de amostras em  $R_n$
- $p_n(\mathbf{x})$   $n$ -ésima estimativa de  $p(\mathbf{x})$

- $p_n(\mathbf{x})$  converge para  $p(\mathbf{x})$  se

$$1 - \lim_{n \rightarrow \infty} V_n = 0$$

$$2 - \lim_{n \rightarrow \infty} k_n = \infty$$

$$3 - \lim_{n \rightarrow \infty} k_n / n = 0$$

- 1 – média espacial  $P/V$  converge para  $p(\mathbf{x})$  se as regiões diminuem uniformemente e  $p(\cdot)$  é contínua em  $\mathbf{x}$
  - 2 – frequência converge (em probabilidade) para  $p$ , se  $p(\mathbf{x}) \neq 0$
  - 3 – necessária para  $p_n(\mathbf{x})$  dada por (7) convergir; indica que, embora um número muito grande de amostras eventualmente estarão em uma pequena região  $R_n$ , elas formam uma fração pequena do número total de amostras.
- duas maneiras de obter sequências de regiões que satisfazem estas condições

1 – diminuir uma região inicial especificando  $V_n$  como função de  $n$ , e.g.

$$V_n = 1/\sqrt{n}$$

e mostrar que  $k_n$  e  $k_n/n$  se comportam apropriadamente  $p_n(\mathbf{x}) \rightarrow p(\mathbf{x})$

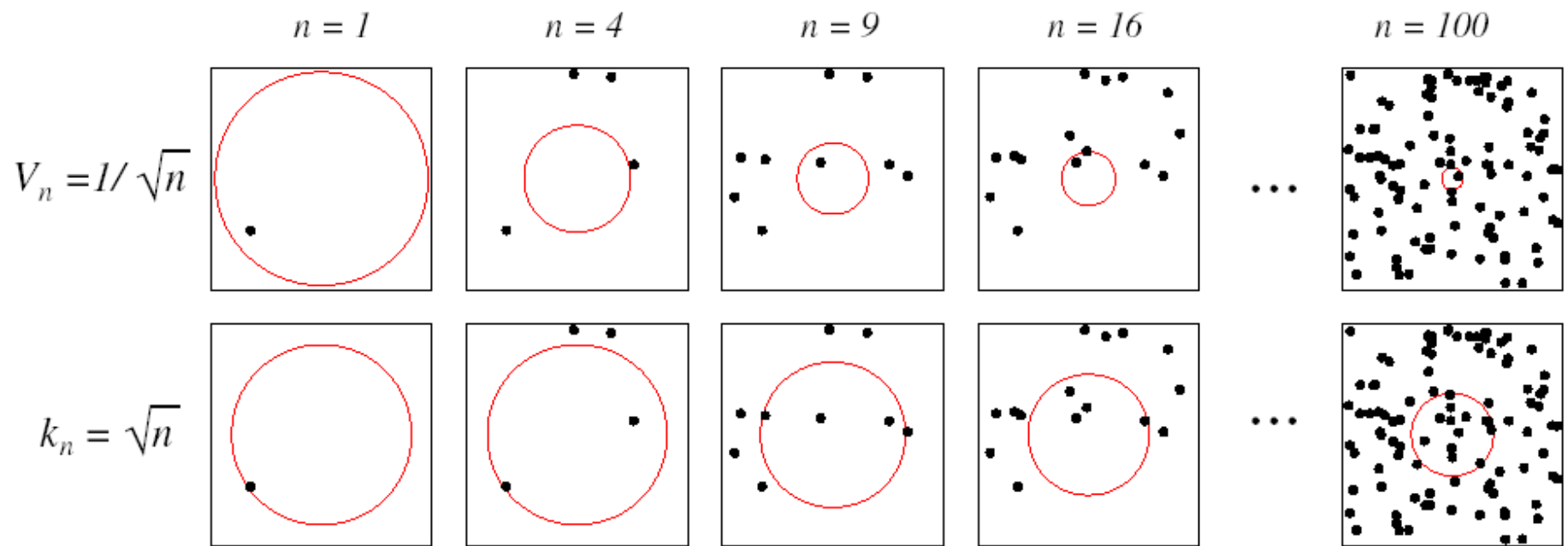
Janelas de Parzen

2 – especificar  $k_n$  como função de  $n$ , e.g.

$$k_n = \sqrt{n}$$

aqui o volume  $V_n$  cresce até que ele inclua  $k_n$  vizinhos de  $\mathbf{x}$

$k_n$ -vizinhos-mais-próximos



## 3-Janelas de Parzen

$$R_n \text{ hipercubo de dimensão } d \text{ com aresta } h_n \Rightarrow V_n = h_n^d \quad (8)$$

$k_n$  número de amostras que caem no hipercubo

$$\varphi(\mathbf{u}) = \begin{cases} 1 & |u_j| \leq 1/2 \quad j = 1, \dots, d \\ 0 & \text{caso contrário} \end{cases} \quad (9)$$

$\varphi(\mathbf{u})$  é um hipercubo centrado na origem

$\varphi((\mathbf{x} - \mathbf{x}_i)/h_n) = 1$  se  $\mathbf{x}_i$  cae dentro de  $V_n$  e zero caso contrário

$k_n$  número de amostras no hipercubo

$$k_n = \sum_{i=1}^n \varphi\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_n}\right) \quad (10)$$

substituindo em (7)

$$p_n(\mathbf{x}) = \frac{k_n / n}{V_n} \quad (7)$$

$$p_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{V_n} \varphi\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_n}\right) \quad (11)$$

– equação (11) sugere abordagem mais geral:

- usar classes mais gerais de funções ao invés do hipercubo
- estimativa de  $p(\mathbf{x})$  é a média de funções de  $\mathbf{x}$  e amostras  $\mathbf{x}_i$

– função janela é usada como um interpolador:

- cada amostra contribuindo para a estimativa de acordo com sua distância a  $\mathbf{x}$

– para que  $p_n(\mathbf{x})$  seja uma densidade legítima é necessário que

$$\varphi(\mathbf{x}) \geq 0 \quad (12)$$

$$\int \varphi(\mathbf{u}) d\mathbf{u} = 1 \quad (13)$$



- Efeito da largura da janela  $h_n$  em  $p_n(\mathbf{x})$

seja

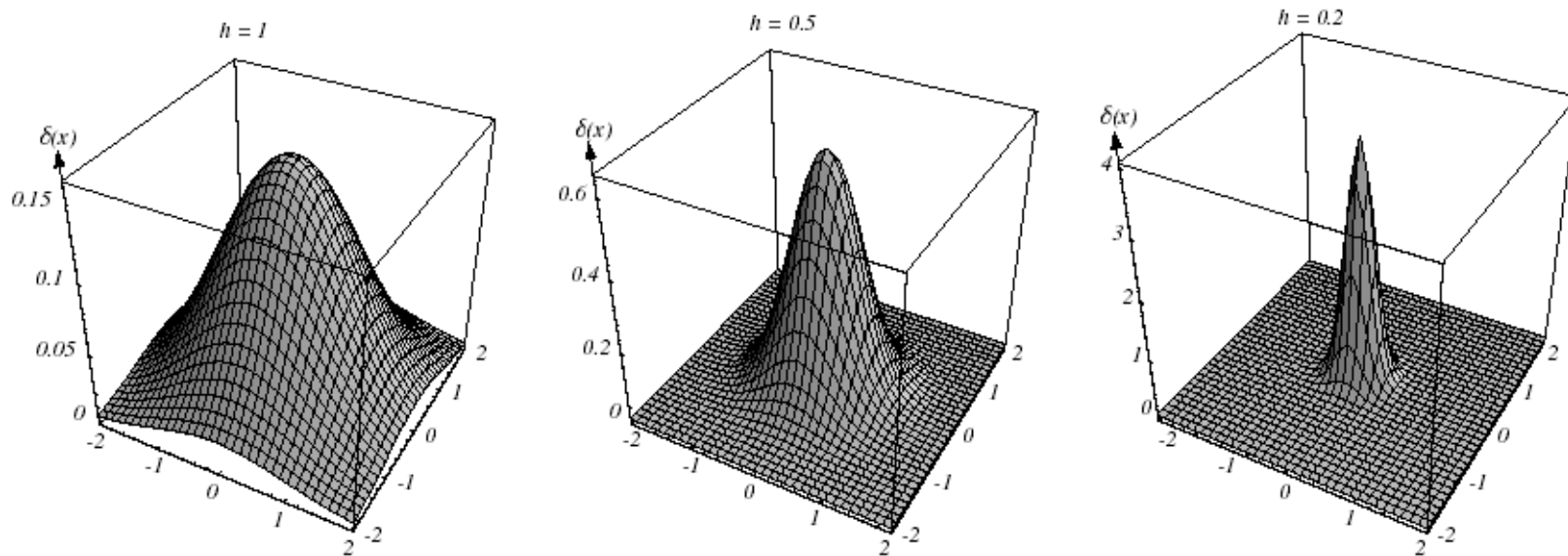
$$\delta_n(\mathbf{x}) = \frac{1}{V_n} \phi\left(\frac{\mathbf{x}}{h_n}\right) \quad (14)$$

então  $p_n(\mathbf{x})$  pode ser expressa como

$$p_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \delta_n(\mathbf{x} - \mathbf{x}_i) \quad (15)$$

$h_n(\mathbf{x})$  afeta tanto a amplitude quanto a dispersão de  $\delta_n(\mathbf{x})$

# Exemplo: janela de Parzen normal, simétrica, circular



– se  $h_n$  é muito grande

- amplitude de  $\delta_n$  é pequena
- $\mathbf{x}$  tem que estar longe de  $\mathbf{x}_i$  antes que  $\delta_n(\mathbf{x}-\mathbf{x}_i)$  afaste de  $\delta_n(0)$
- $p_n(\mathbf{x})$  superposição de  $n$  funções que variam suavemente

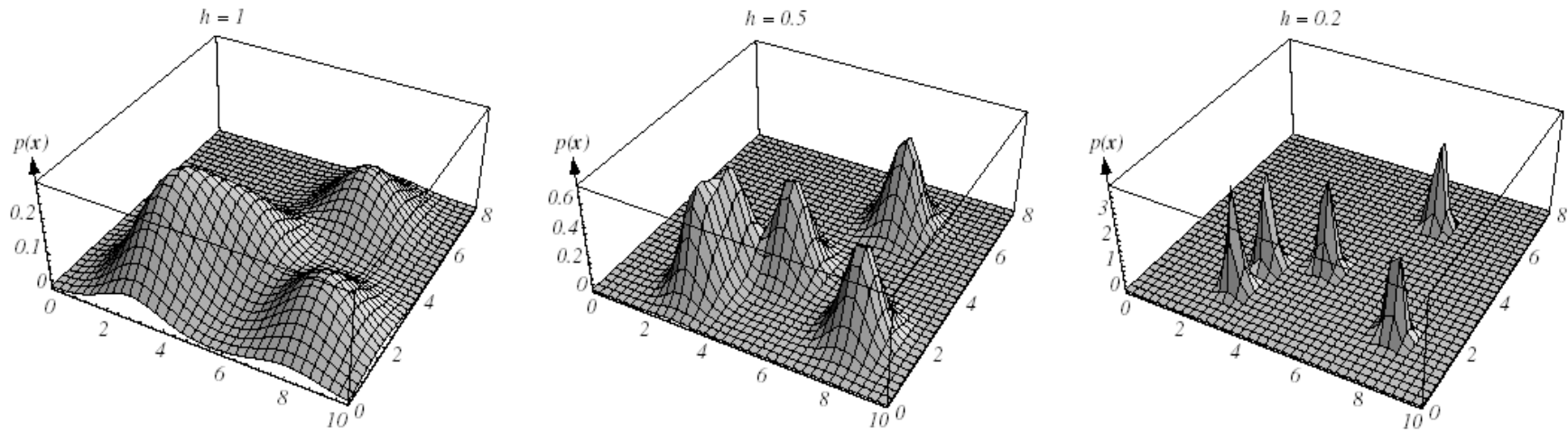
– se  $h_n$  é muito pequena

- pico de  $\delta_n(\mathbf{x}-\mathbf{x}_i)$  é grande e ocorre próximo a  $\mathbf{x} = \mathbf{x}_i$
- $p(\mathbf{x})$  é a superposição de  $n$  pulsos centrados nas amostras
- para qualquer  $h_n$  a distribuição é normalizada

$$\int \delta_n(\mathbf{x} - \mathbf{x}_i) d\mathbf{x} = \int \frac{1}{V_n} \varphi\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_n}\right) = \int \varphi(\mathbf{u}) d\mathbf{u} = 1 \quad (16)$$

- $h_n \rightarrow 0 \Rightarrow \delta_n(\mathbf{x}-\mathbf{x}_i) \rightarrow$  função delta de Dirac centrada em  $\mathbf{x}_i$

## Exemplo: estimativa de densidades com janelas de Parzen



estimativa com 5 pontos, janela normal, simétrica, circular

## ■ Análise

- $V_n$  muito grande: estimativa tem pouca resolução
- $V_n$  muito grande: estimativa tem grande variabilidade estatística
- número limitado de amostras: temos que aceitar um compromisso
- número infinito de amostras: se  $V_n \rightarrow 0$  quando  $n \rightarrow \infty$  então

$$p_n(\mathbf{x}) \rightarrow p(\mathbf{x})$$

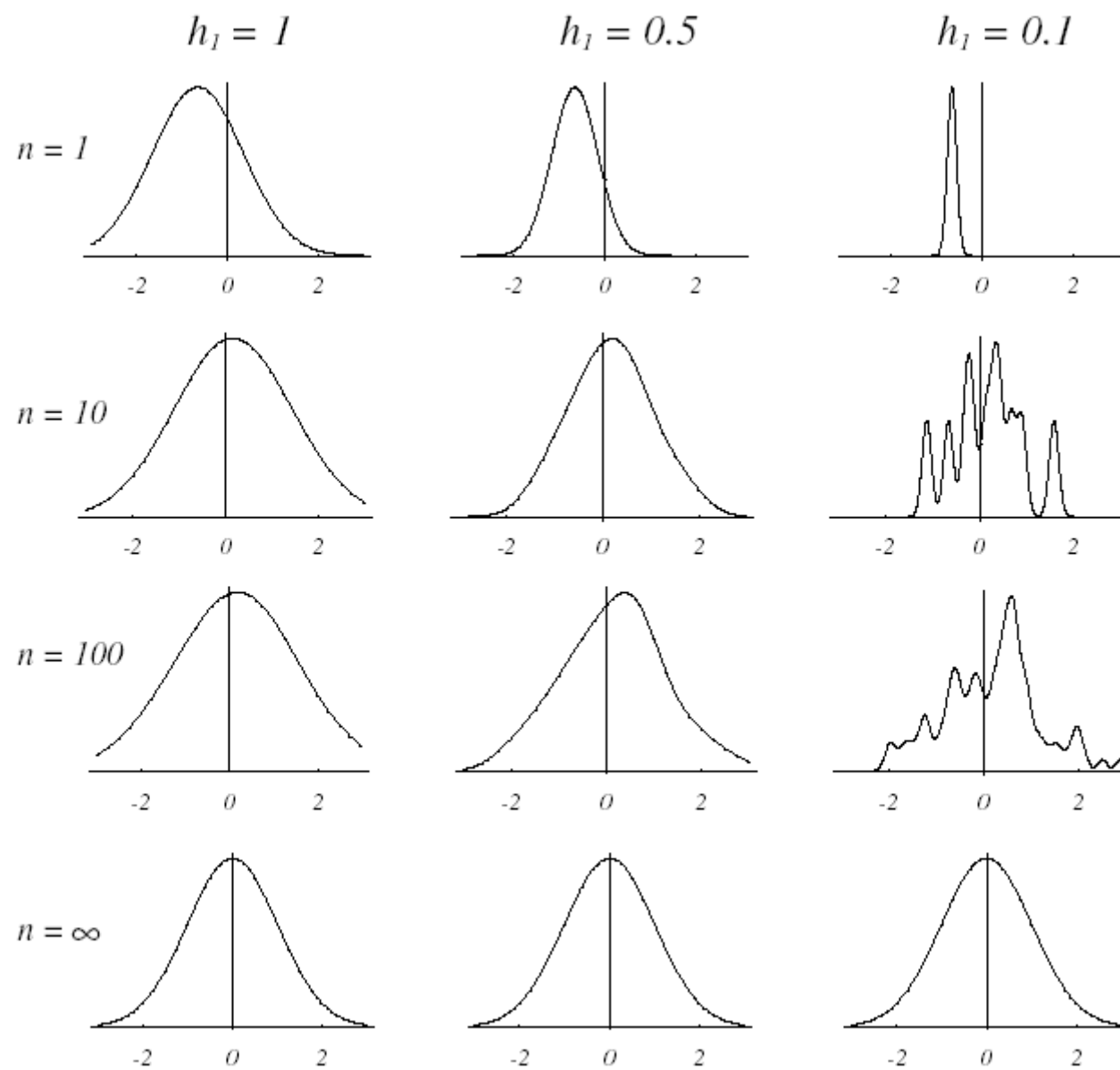
## ■ Exemplo

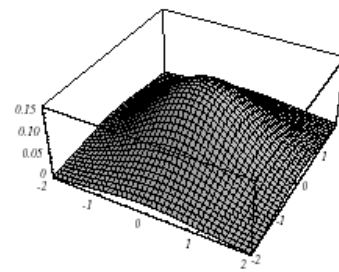
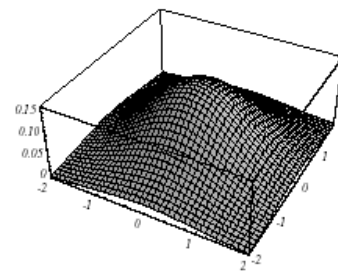
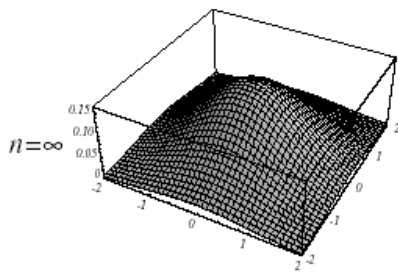
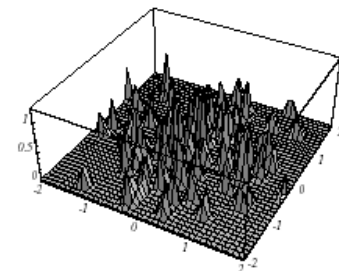
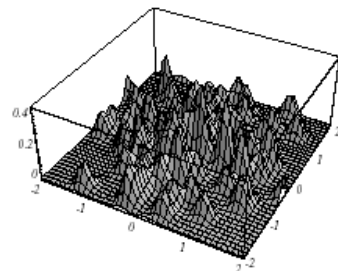
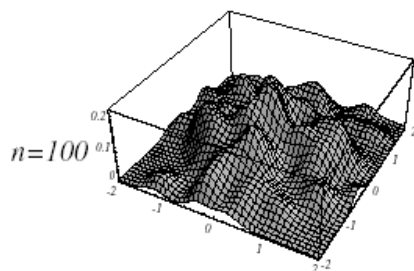
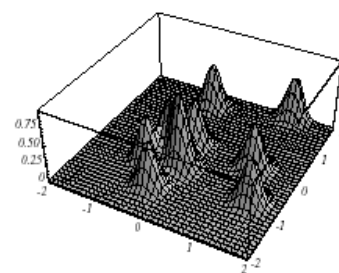
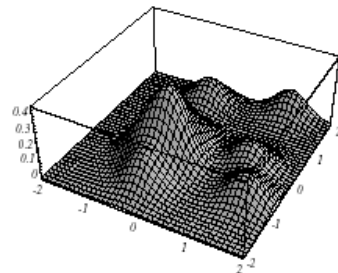
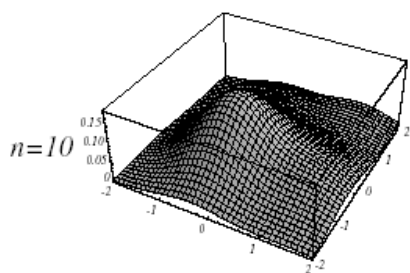
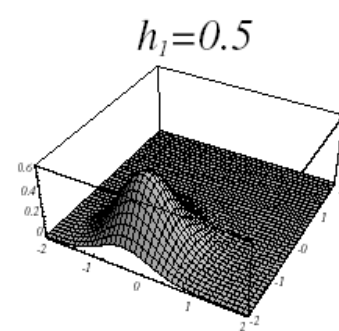
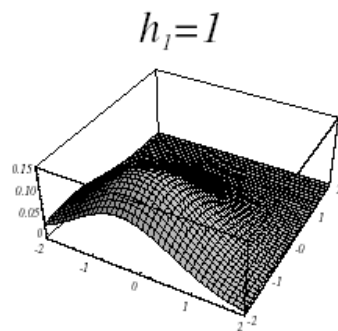
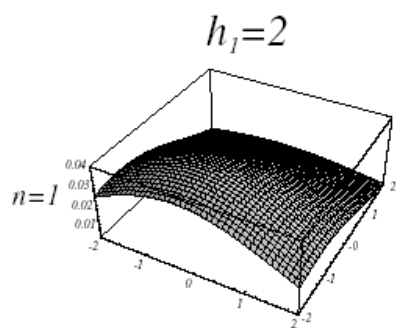
$$p(\mathbf{x}) = N(0,1)$$

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

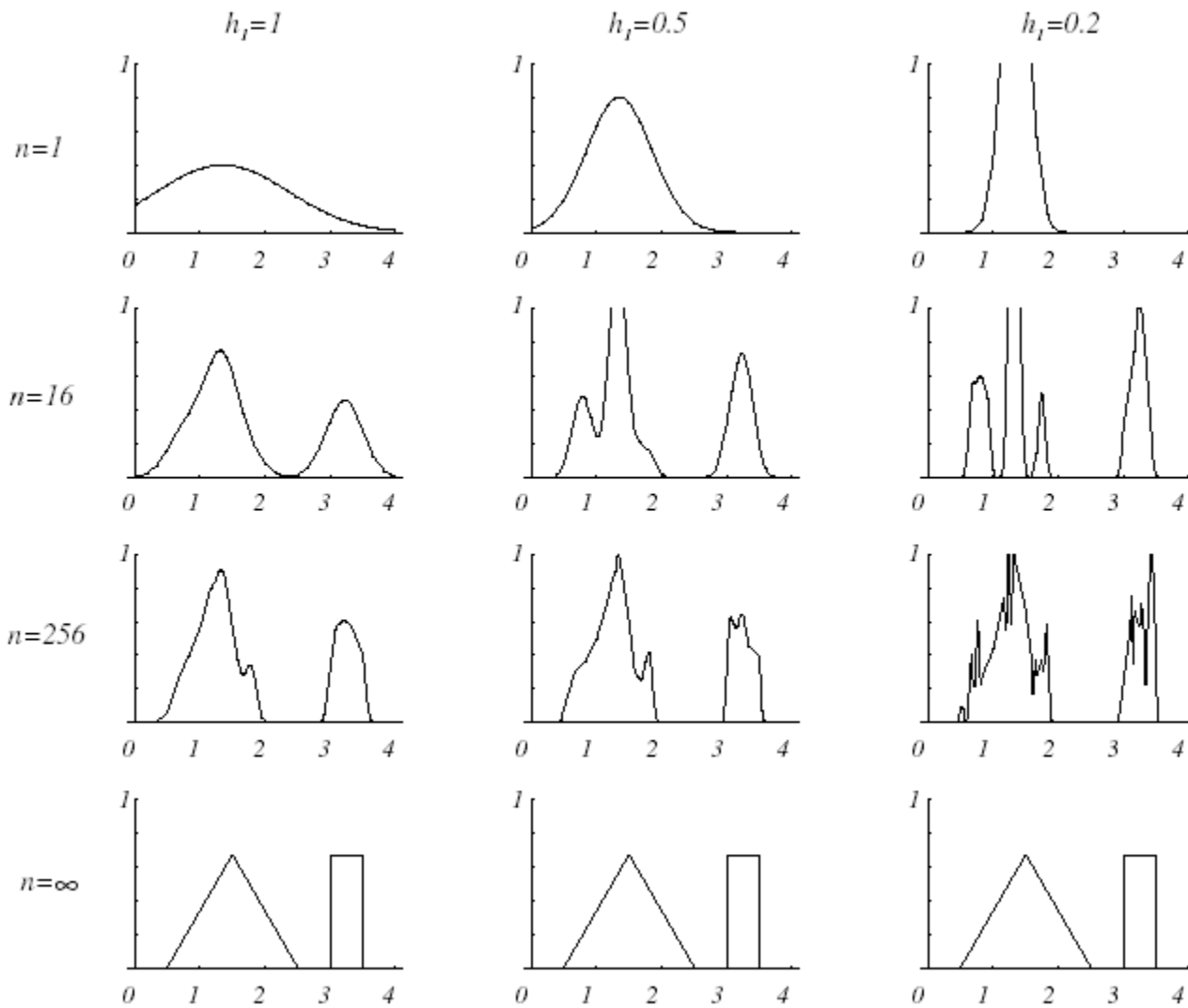
$$h_n = \frac{h_1}{\sqrt{n}}$$

$$p_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} \varphi\left(\frac{x - x_i}{h_n}\right)$$



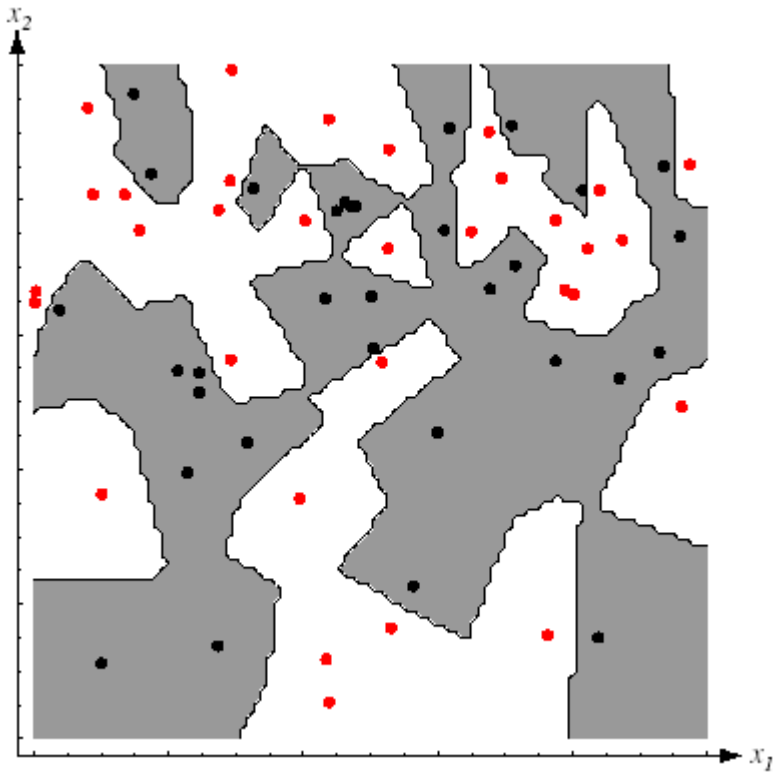




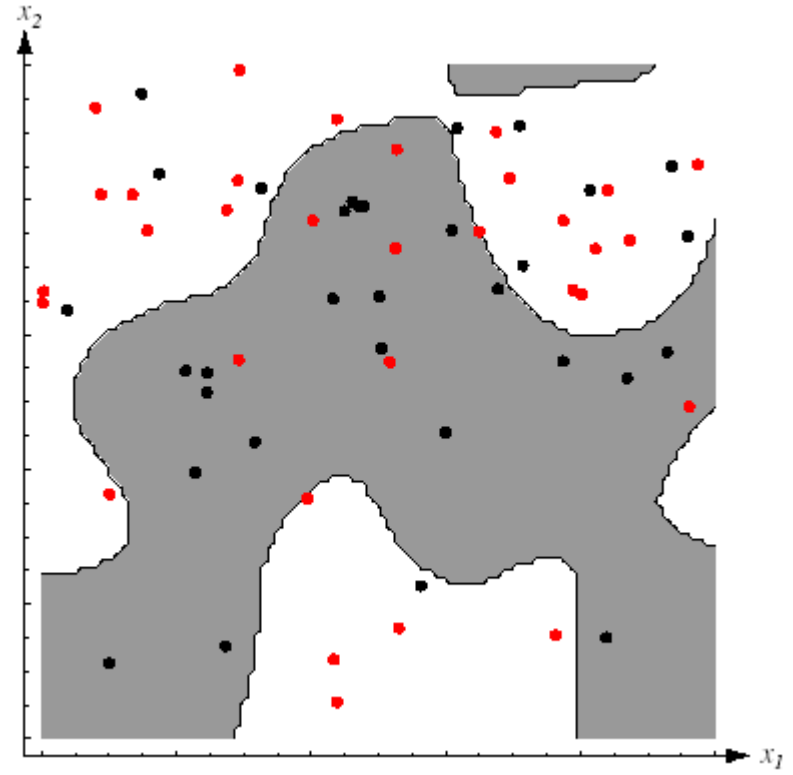


## ■ Exemplo classificação

- estimar densidades para cada categoria
- classificar pela classe com maior *a posteriori*
- regiões de decisão dependem da função janela
- erro de treinamento arbitrariamente pequeno se janela é suficientemente pequena
- erro treinamento pequeno não significa erro de teste pequeno
- formas Gaussianas são plausíveis na ausência de informação
- difícil justificar teoricamente a escolha da largura da janela
- vantagem: generalidade
- problemas: número de amostras pode ser grande  
cresce exponencialmente com a dimensão  
*curse of dimensionality*



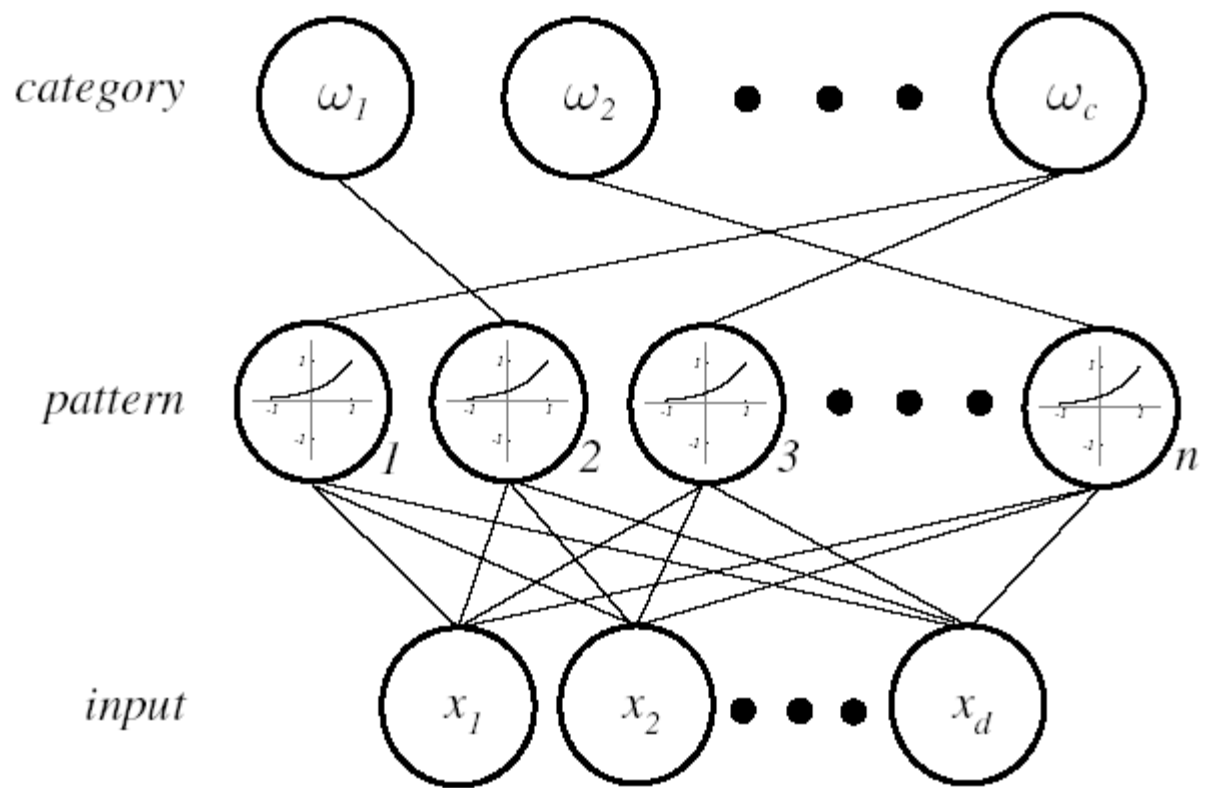
$h$  pequeno



$h$  grande

## ■ Redes probabilísticas

- implementação de janelas de Parzen com redes neurais
- hipóteses: desejamos obter estimador de Parzen com  $n$  padrões  
cada padrão tem dimensão  $d$   
padrões amostrados aleatoriamente de  $c$  classes
- características da rede
  - $d$  unidades de entrada (*input units*)
  - $n$  unidades intermediárias (*pattern units*)
  - $c$  unidades de saída (*category units*)
- arquitetura *feedforward*



## ■ Treinamento de redes probabilísticas

Algoritmo 1: Treinamento de PNN

```
initialize  $j \leftarrow 0$ ,  $k \leftarrow 0$ ,  $a_{ji} \leftarrow 0$ 
begin
  do  $j \leftarrow j + 1$ 
    do  $k \leftarrow k + 1$ 
       $x_{jk} \leftarrow x_{jk} / \|\mathbf{x}_j\|$    normalize
       $w_{jk} \leftarrow x_{jk}$            train
    until  $k = d$ 
  if  $\mathbf{x} \in \omega_i$  then  $a_{ji}$ 
until  $j = n$ 
end
```

- Classificação com redes probabilísticas

$$net_j = \mathbf{w}_j^t \mathbf{x}$$

$$\begin{aligned} \varphi\left(\frac{\mathbf{x} - \mathbf{w}_j}{h_n}\right) &\propto e^{-(\mathbf{x} - \mathbf{w}_j)^t (\mathbf{x} - \mathbf{w}_j) / 2\sigma^2} = \\ &= e^{-(\mathbf{x}^t \mathbf{x} - \mathbf{w}_j^t \mathbf{w}_j - 2\mathbf{w}_j^t \mathbf{x}) / 2\sigma^2} \\ &= e^{(net_j - 1) / \sigma^2} \end{aligned}$$

categoria =  $\max g_i(\mathbf{x})$

$g_i(\mathbf{x})$  = soma estimativas locais

## Algoritmo 2: Classificação com PNN

```
initialize  $j \leftarrow 0$ ,  $\mathbf{x} \leftarrow$  test pattern  
begin  
  do  $j \leftarrow j + 1$   
     $net_j \leftarrow \mathbf{w}_j^t \mathbf{x}$   
    if  $a_{ji} = 1$  then  $g_i \leftarrow g_i + \exp(net_j - 1) / \sigma^2$   
  until  $j = n$   
  return  $class \leftarrow \arg \max_i g_i(\mathbf{x})$   
end
```



## 4- Estimação $k_n$ -vizinhos-mais-próximos

- alternativa para o problema da melhor janela
- $V$  função dos dados de treinamento ao invés de uma função arbitrária do número de amostras
- para estimar  $p(\mathbf{x})$ : centrar uma célula em  $\mathbf{x}$  e deixá-la crescer até incorporar  $k_n$  amostras ( $k_n$ -vizinhos de  $\mathbf{x}$ )
- $k_n$  é uma função de  $n$

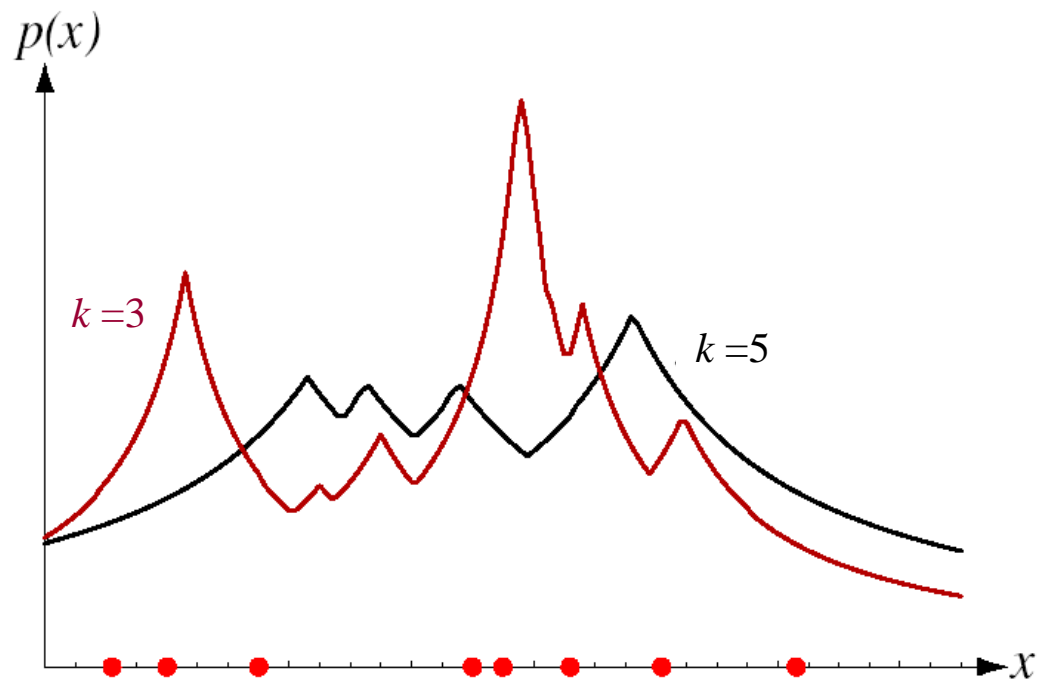
- Estimativa de  $p(\mathbf{x})$

$$p_n(\mathbf{x}) = \frac{k_n / n}{V_n} \quad (30)$$

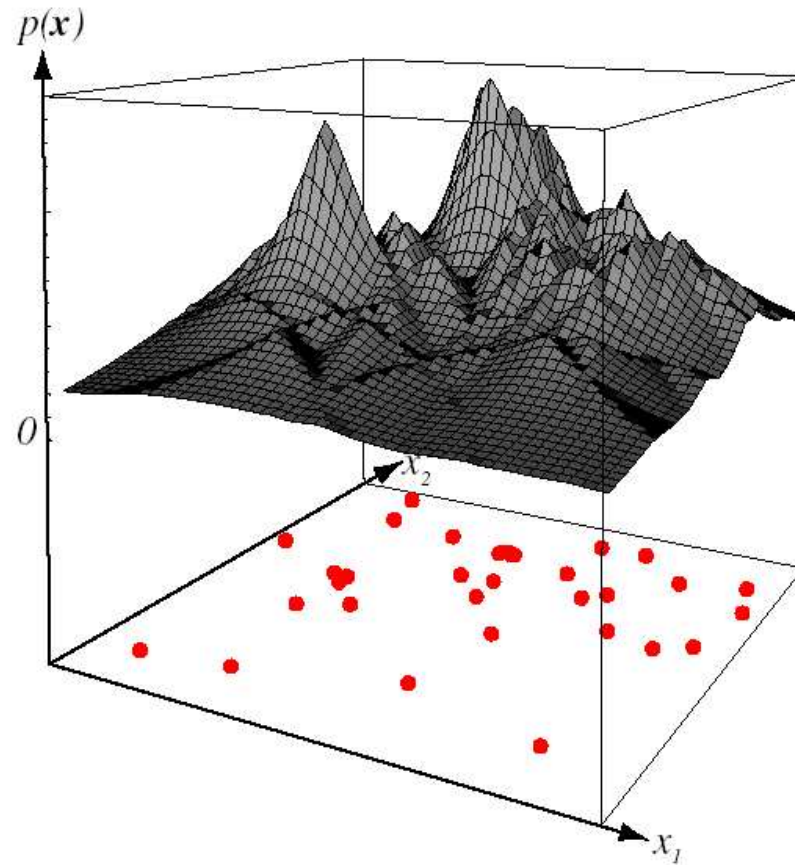
condições de convergência

$$\lim_{n \rightarrow \infty} k_n = \infty$$

$$\lim_{n \rightarrow \infty} k_n / n = 0$$



$p(\mathbf{x})$  é contínua mas  $\partial p(\mathbf{x})/\partial \mathbf{x}$  não



$k = 5$

$p(\mathbf{x})$  é contínua mas  $\partial p(\mathbf{x})/\partial \mathbf{x}$  não

■ Estimação de probabilidades *a posteriori*

- $n$  amostras rotuladas
- célula de volume  $V$  no entorno de  $\mathbf{x}$
- $k$  amostras,  $k_i$  rotuladas  $\omega_i$  em  $V$
- estimativa de  $p(\mathbf{x}, \omega_i)$

$$p_n(\mathbf{x}) = \frac{k_n / n}{V_n}$$

- estimativa de  $P(\omega_i, \mathbf{x})$

$$P_n(\omega_i | \mathbf{x}) = \frac{p_n(x, \omega_i)}{\sum_{i=1}^c p_n(\mathbf{x}, \omega_j)} = \frac{k_i}{k}$$

# 5-Classificação fuzzy

## ■ Conjuntos

Classificam objetos em conceitos gerais

- números pares
- cidades que são capitais
- carros esportes
- números ímpares
- times de futebol
- .....

## Conjuntos ?????

- *grandes* cidades da América do Sul
- *baixa* temperatura
- *alta* taxa de inflação
  
- *pequeno* erro de aproximação
- *rápida* resposta de um sistema dinâmico
- *mal condicionamento* de um sistema de equações lineares

## Problema da dicotomia

“One seed does not constitute a pile nor two nor three... from the other side everybody will agree that 100 million seeds constitute a pile. What therefore is the appropriate limit? Can we say that 325 647 seeds don't constitute a pile but 325 648 do?”

[Borel, 1950]





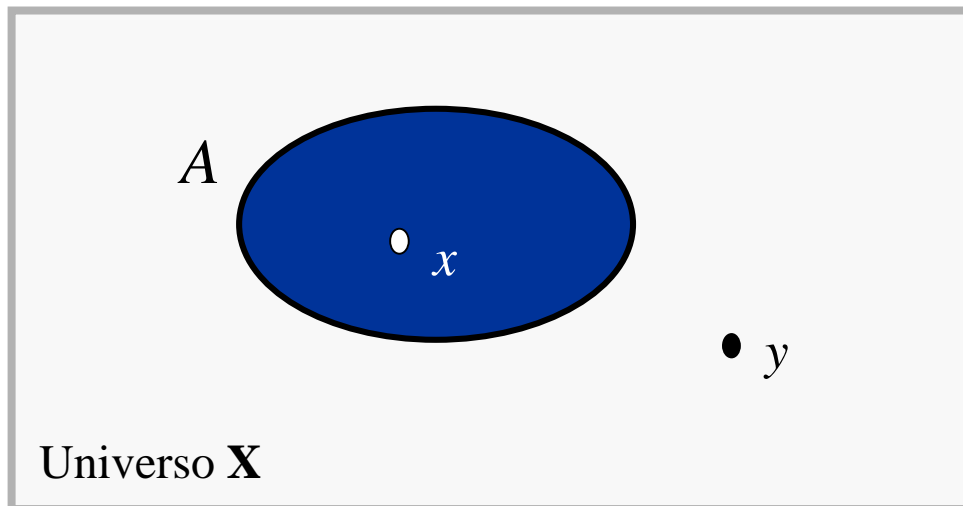
## Paradoxo do barbeiro (Russell)

“I shave all, and only, those man who don´t shave themselves”

$$T(S) = T(\neg S)$$

$$T(\neg S) = 1 - T(S)$$

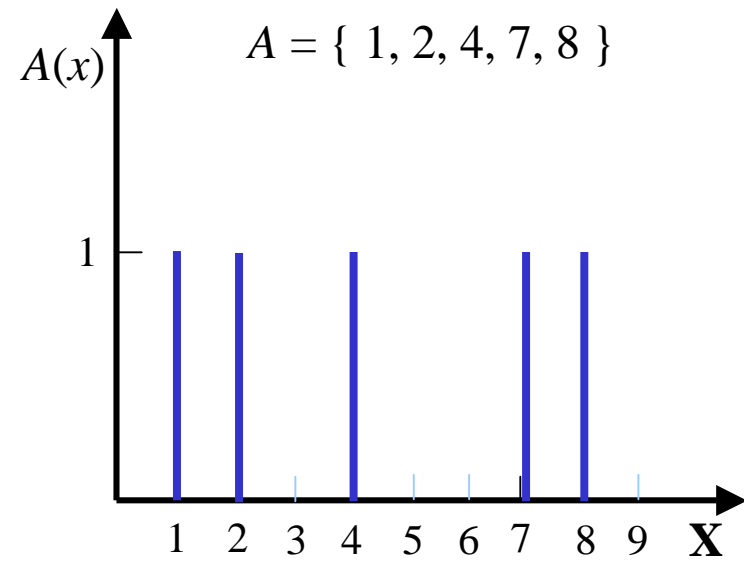
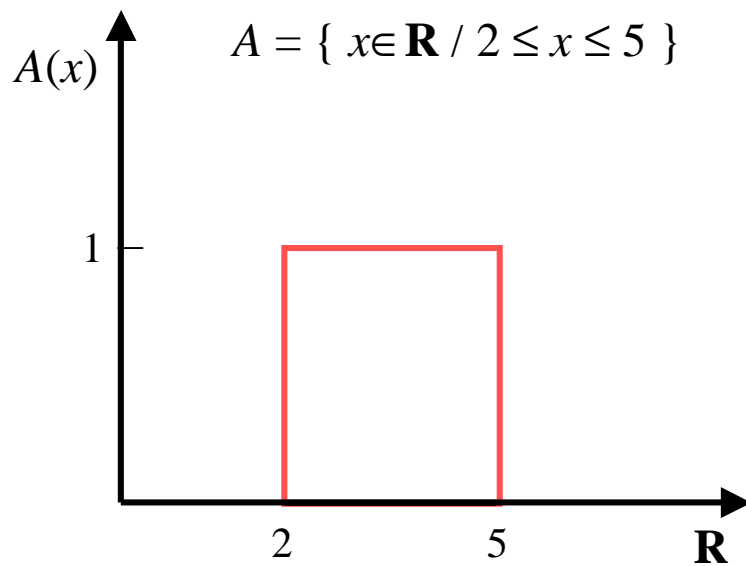
# Conjuntos



$$x \in A$$

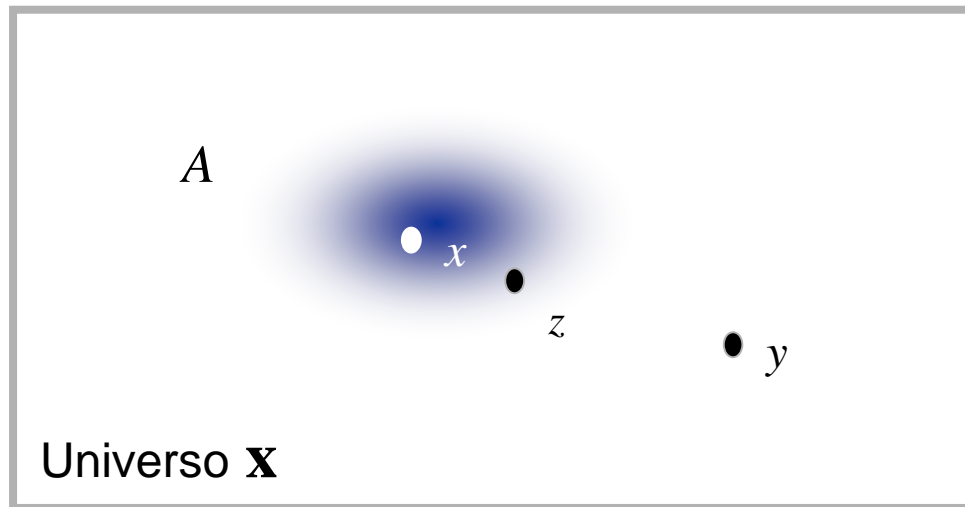
$$y \notin A$$

■ Função característica

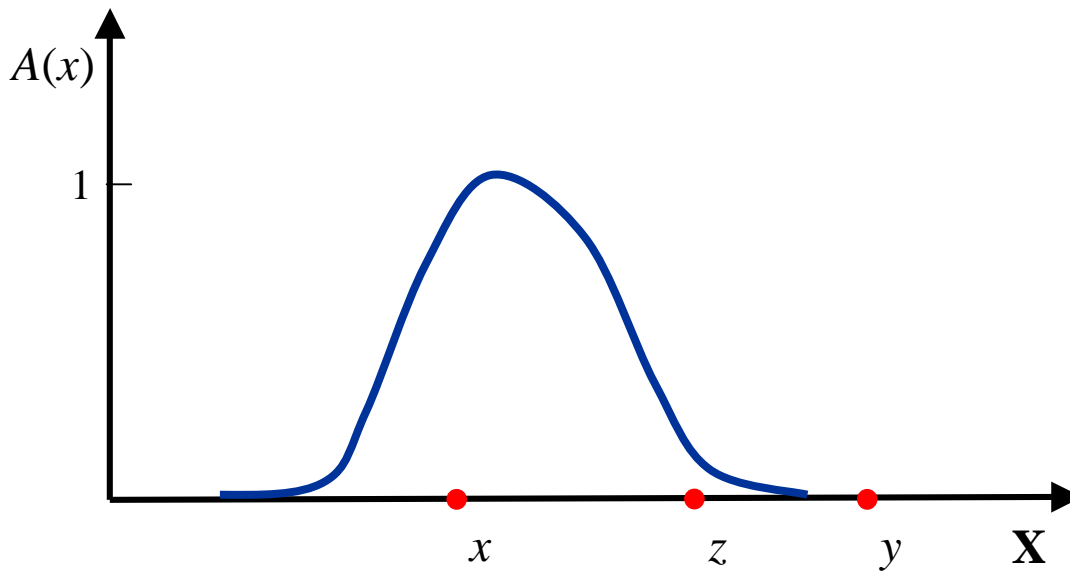


$$A: \mathbf{X} \rightarrow \{0, 1\} \quad A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

# Conjuntos fuzzy



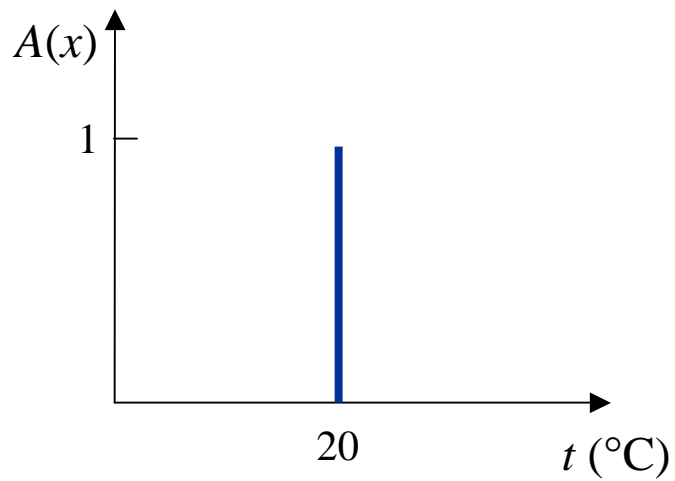
■ Função de pertinência



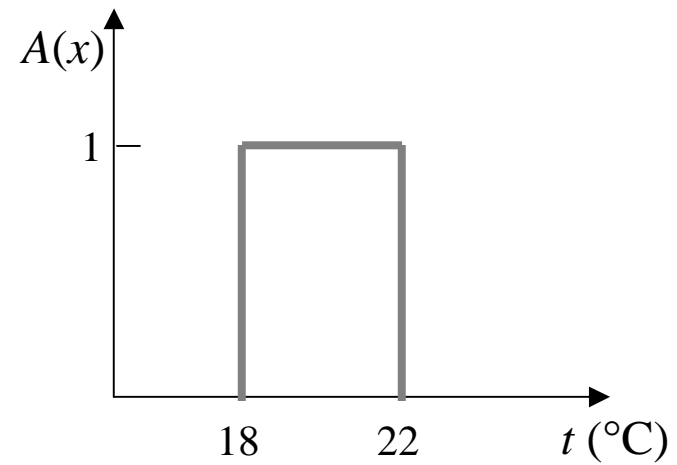
$$A: \mathbf{X} \rightarrow [0, 1]$$

■ Exemplo

temperatura  $t = \sim 20\text{ }^{\circ}\text{C}$

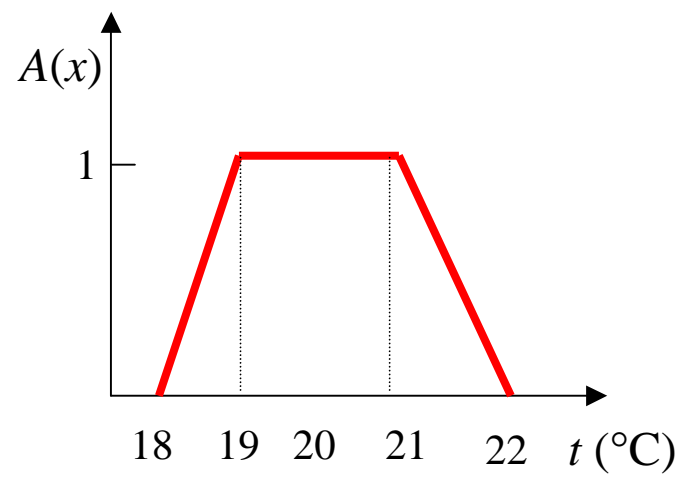


número  
(conjunto unitário)



intervalo  
(conjunto)

temperatura  $t = \sim 20\text{ }^{\circ}\text{C}$



conjunto fuzzy



■ Operações e operadores

União

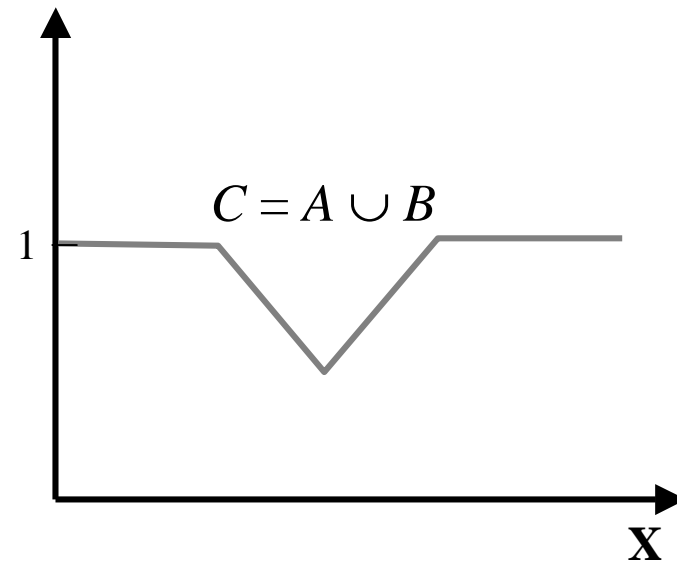
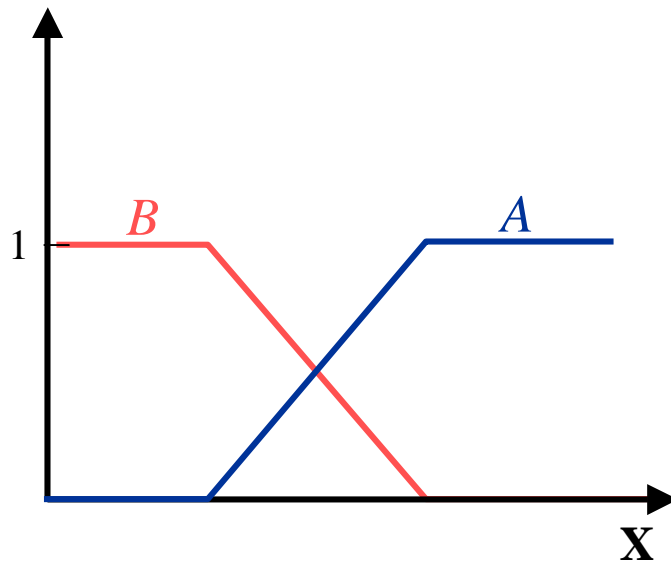
$$C = A \cup B$$

$$s : [0,1] \times [0,1] \rightarrow [0,1]$$

$$C(x) = A(x) s B(x)$$

$A, B$  e  $C$  em  $\mathbf{X}$

$\forall x \in \mathbf{X}$



## Intersecção

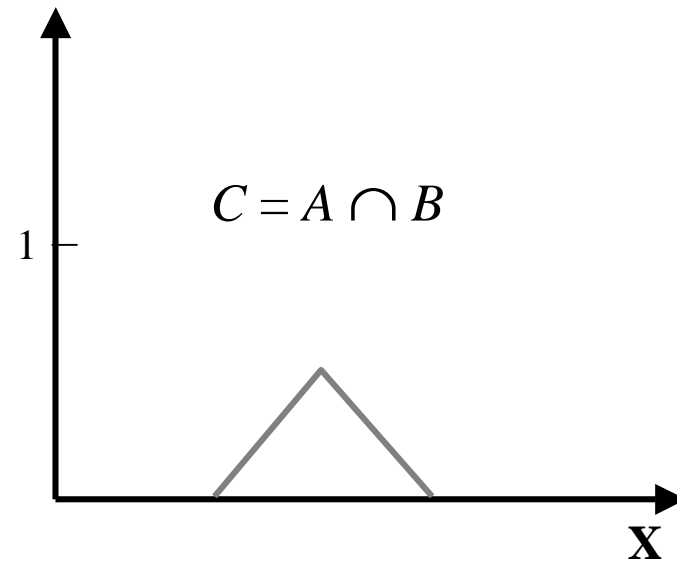
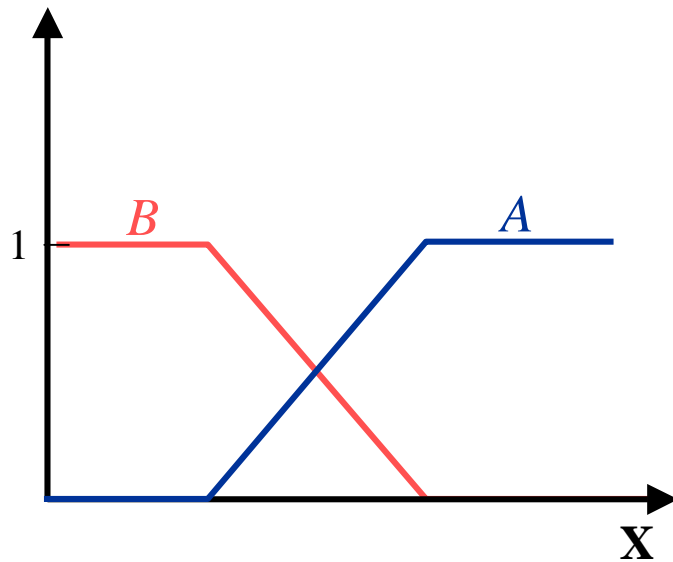
$$C = A \cap B$$

$$t : [0,1] \times [0,1] \rightarrow [0,1]$$

$$C(x) = A(x) \ t B(x)$$

$A, B$  e  $C$  em  $\mathbf{X}$

$\forall x \in \mathbf{X}$

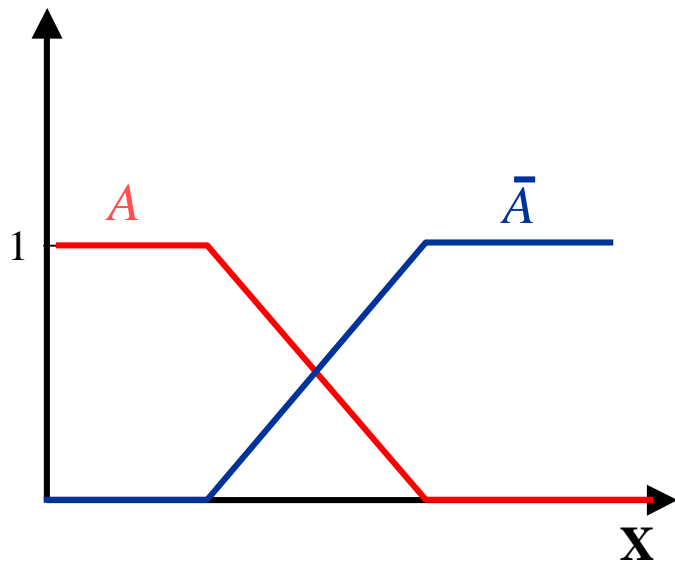


## Complemento

$A$  e  $\bar{A}$  em  $\mathbf{X}$

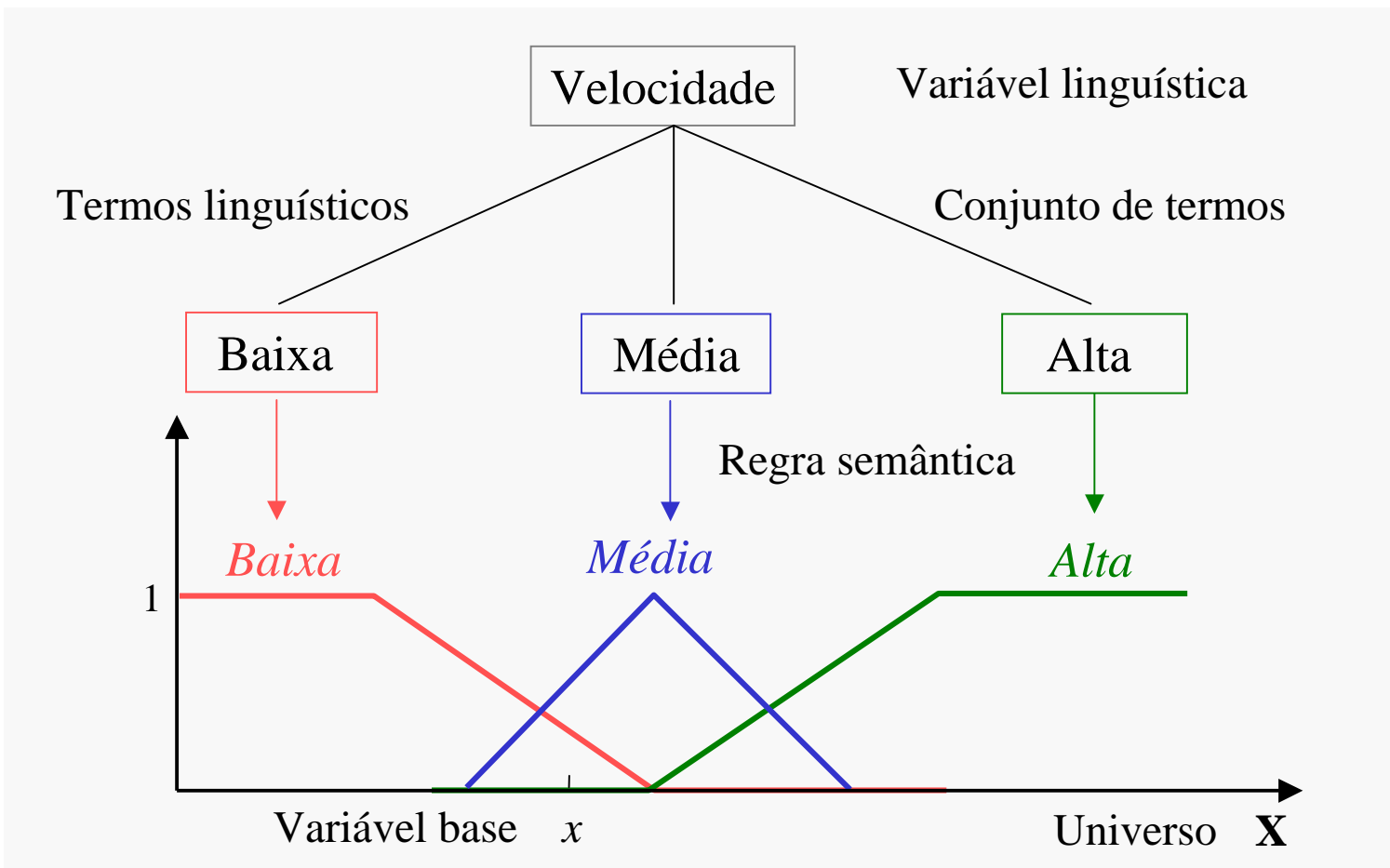
$C : [0,1] \rightarrow [0,1]$

$\bar{A}(x) = C(A(x)) \quad \forall x \in \mathbf{X}$



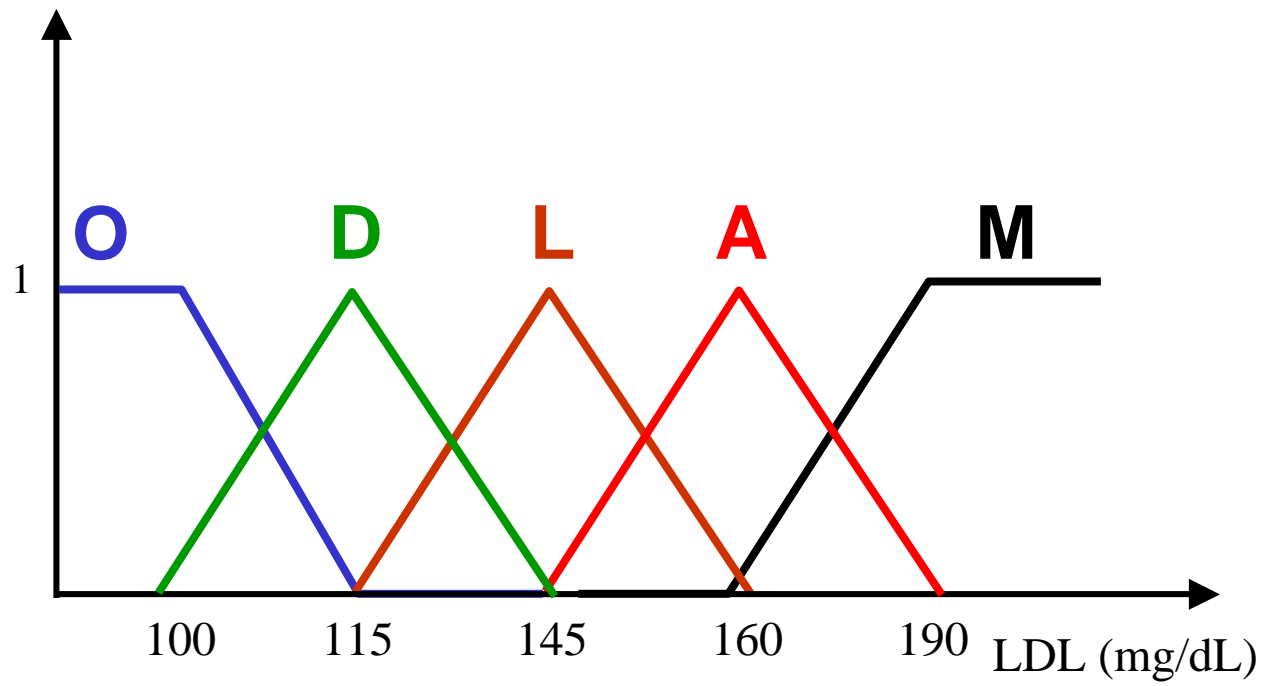
$$A(x) = 1 - \bar{A}(x)$$

■ Variáveis linguísticas



$$(X, T(X), \mathbf{X}, G, M)$$

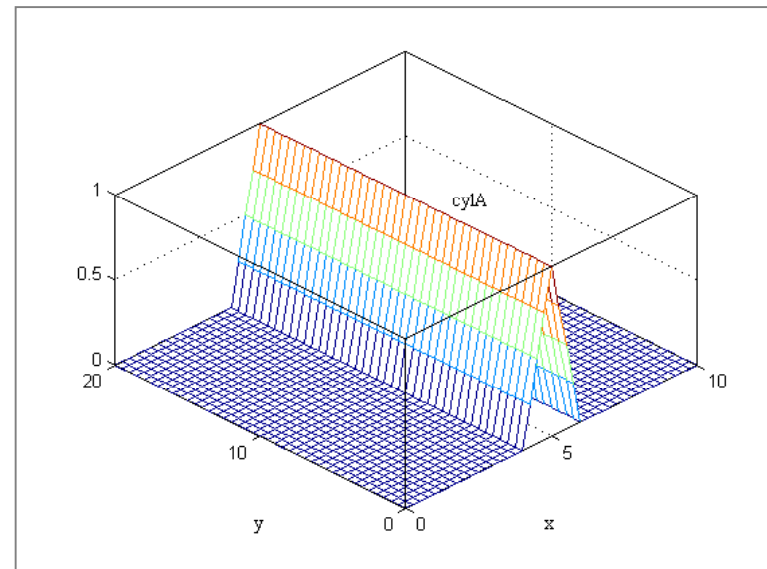
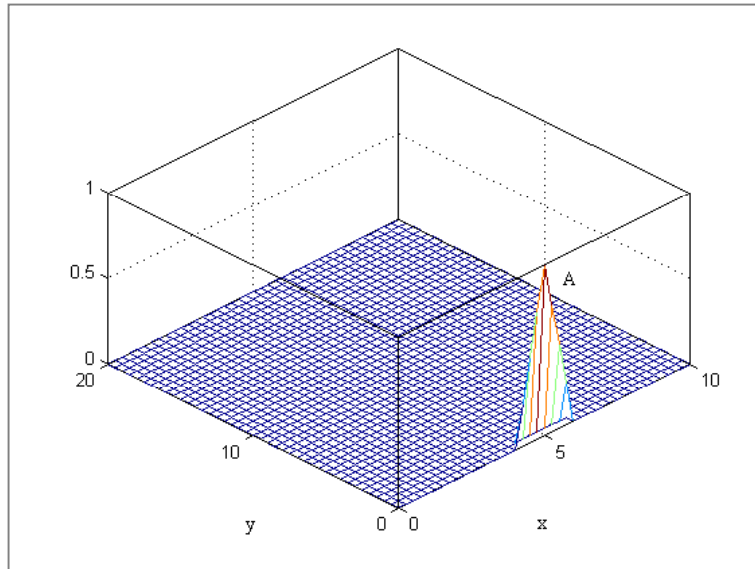
■ Exemplo: colesterol LDL



■ Extensão cilíndrica

$$\text{cyl } A(x,y) = A(x), \quad \forall x \in \mathbf{X}$$

$$\text{cyl } A = A_c$$



- Relação fuzzy

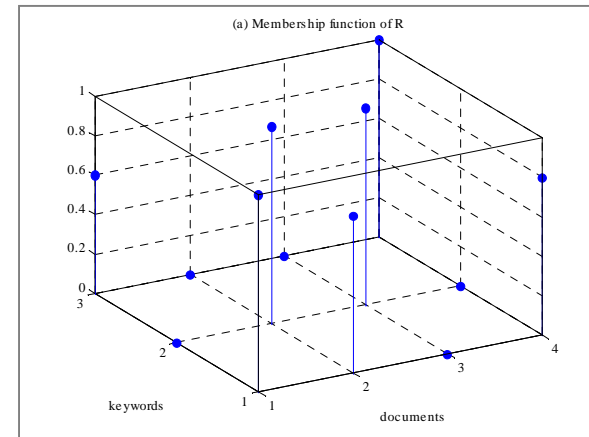
$$R : \mathbf{X} \times \mathbf{Y} \rightarrow [0,1]$$

- Exemplo

Documentos:  $\mathbf{D} = \{d_{fs}, d_{nf}, d_{ns}, d_{gf}\}$

Palavras chaves:  $\mathbf{W} = \{w_f, w_n, w_g\}$

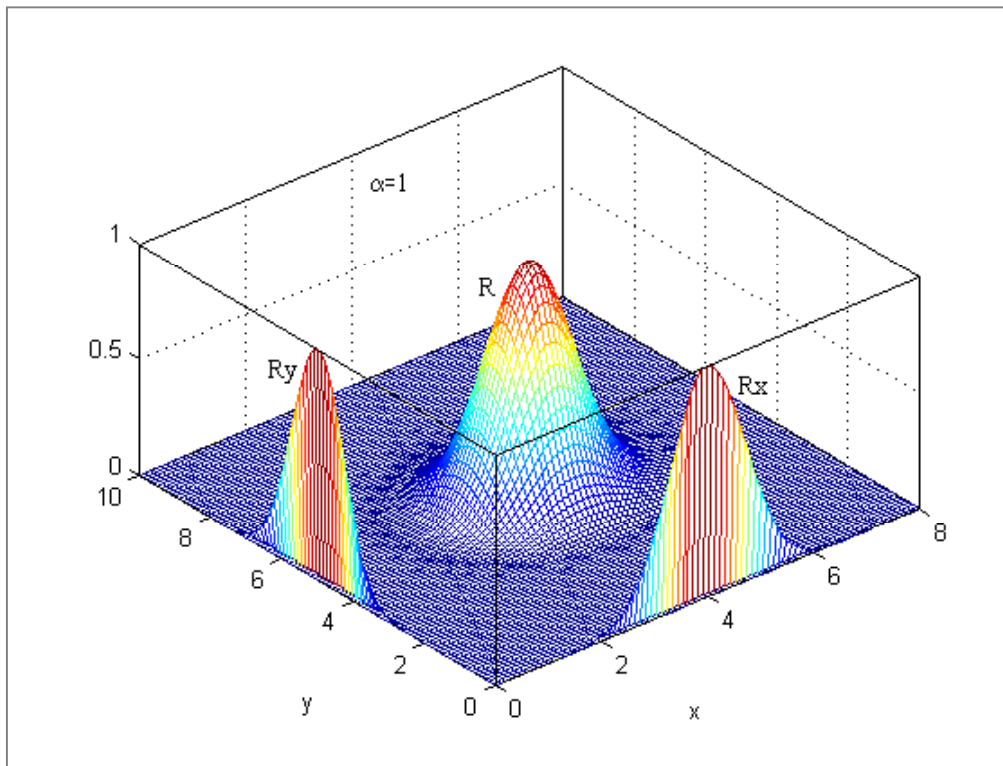
$$R : \mathbf{D} \times \mathbf{W} \rightarrow [0,1]$$



$$R = \begin{matrix} & \begin{matrix} w_f & w_n & w_g \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0.6 \\ 0.8 & 1 & 0 \\ 0 & 1 & 0 \\ 0.8 & 0 & 1 \end{bmatrix} & \begin{matrix} d_{fs} \\ d_{nf} \\ d_{ns} \\ d_{gf} \end{matrix} \end{matrix}$$

- **Projeção de uma relação fuzzy**

$$R(x, y) = \exp\{-\alpha[(x - 4)^2 + (y - 5)^2]\}, \quad \alpha = 1$$

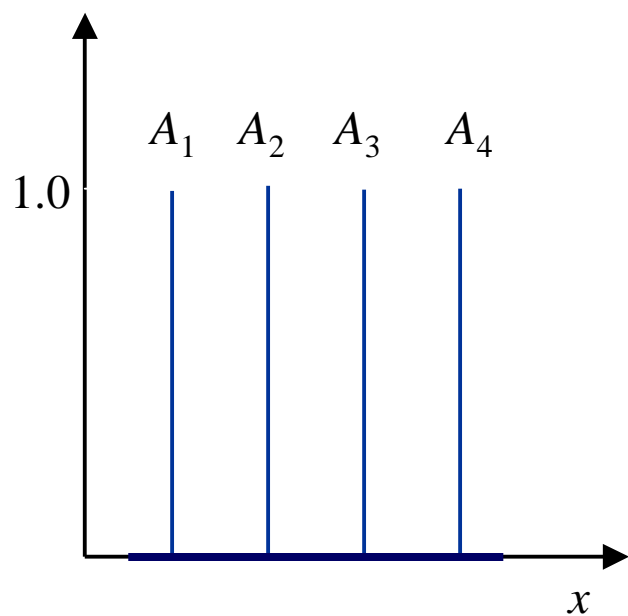


$$R_{\mathbf{X}}(x) = \text{Proj}_{\mathbf{X}} R(x, y) = \sup_y R(x, y)$$

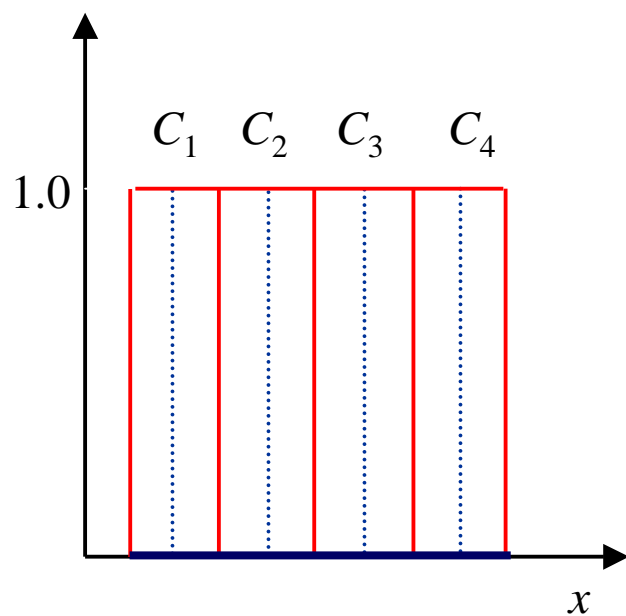
$$R_{\mathbf{Y}}(y) = \text{Proj}_{\mathbf{Y}} R(x, y) = \sup_x R(x, y)$$



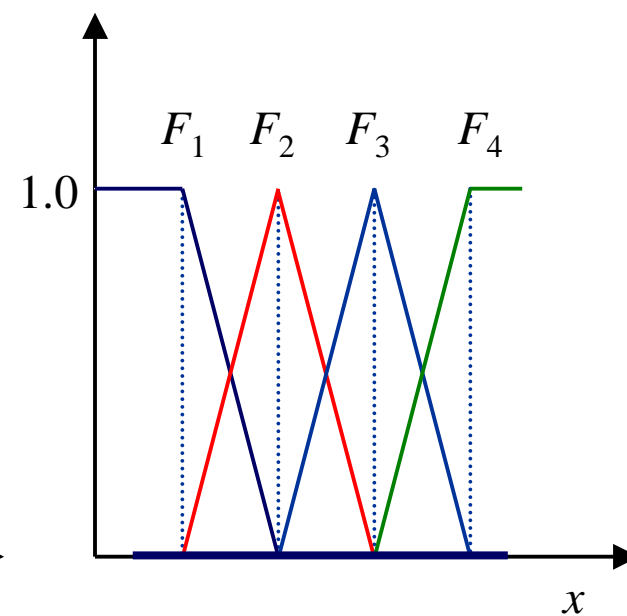
## ■ Granulação



Discretização



Quantização

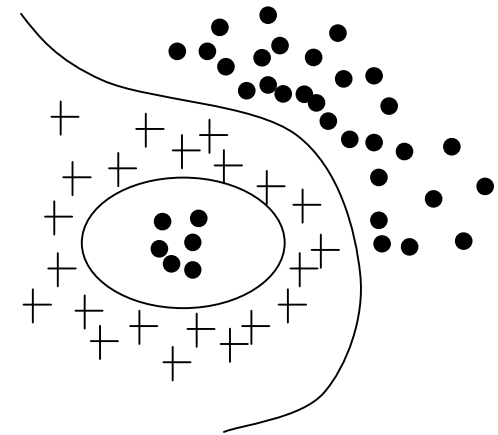
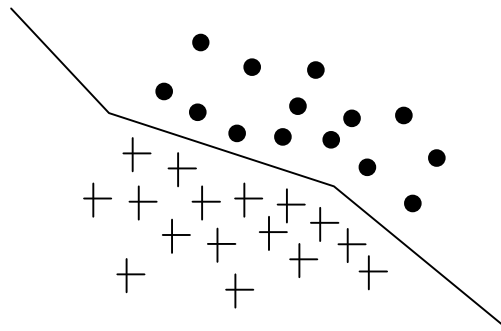
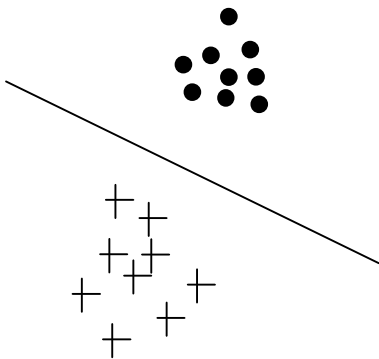
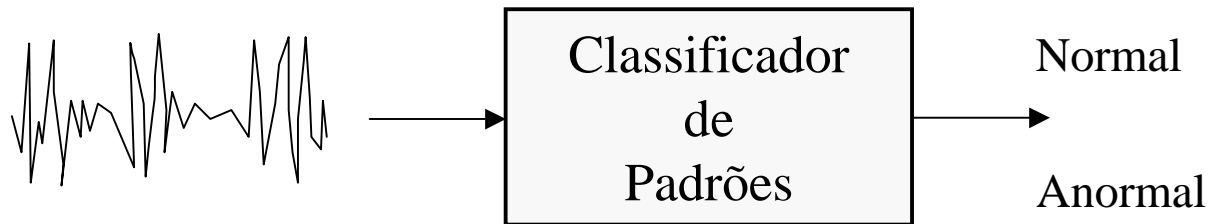


Granulação

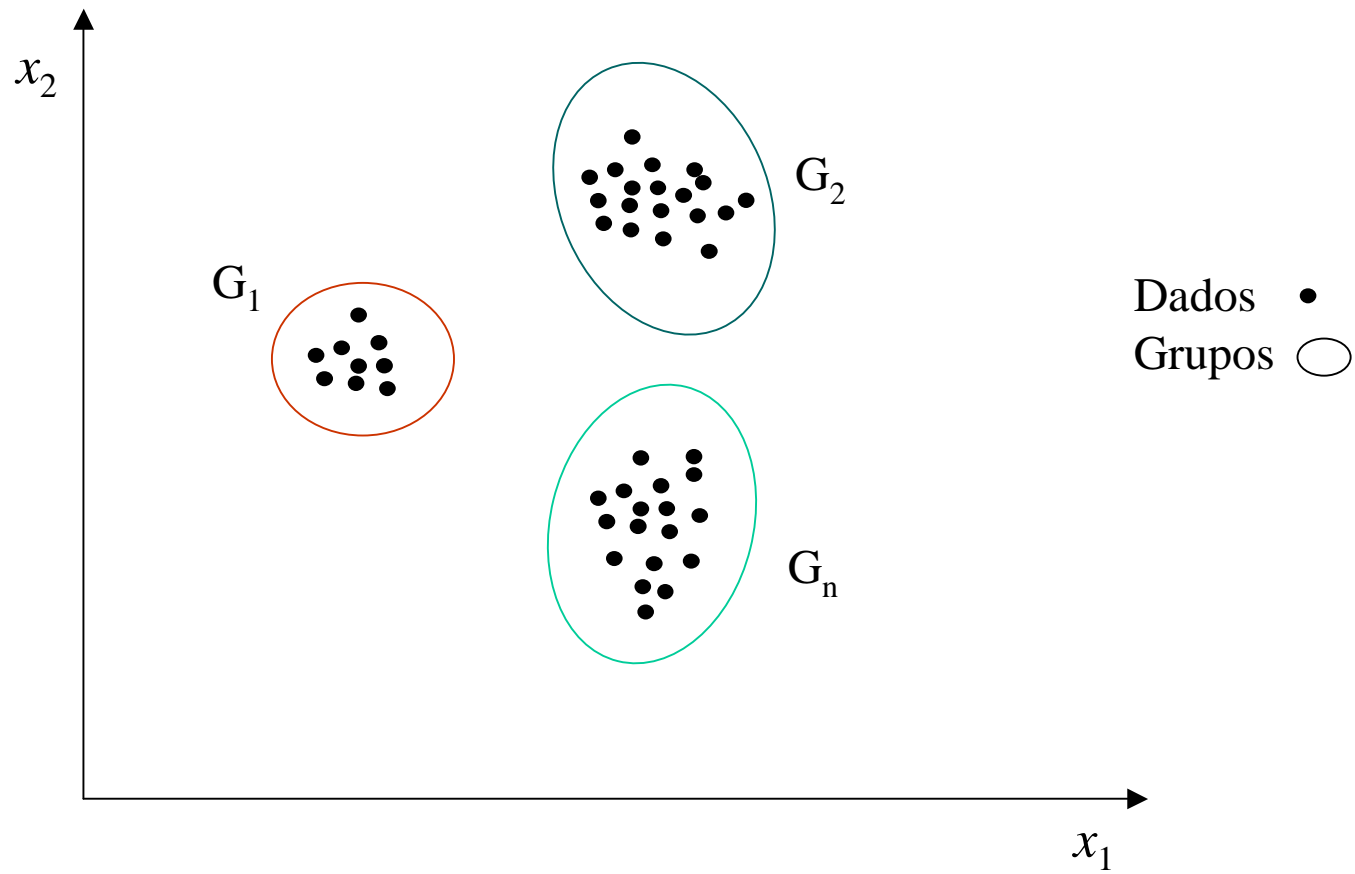
- Sistemas e modelos fuzzy



## ■ Classificação de Padrões



■ Agrupamento

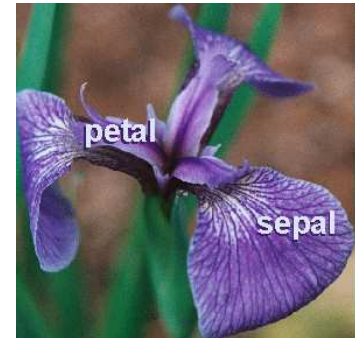


## ■ Exemplo: Iris

### Iris flower database

- dados 4-dimensional com 50 amostras de cada um dos três tipos de flores
- $p = 150$ ,  $n = 4$ ,  $c = 3$
- atributos
  - Sepal length, sepal width, petal length, petal width
- classes
  - Setosa, Versicolor, Virginica

Atributos



Setosa



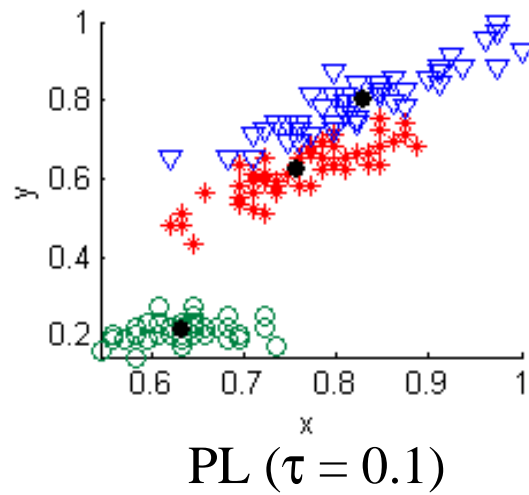
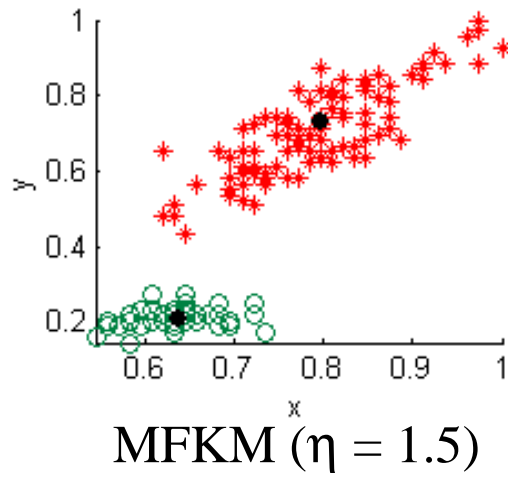
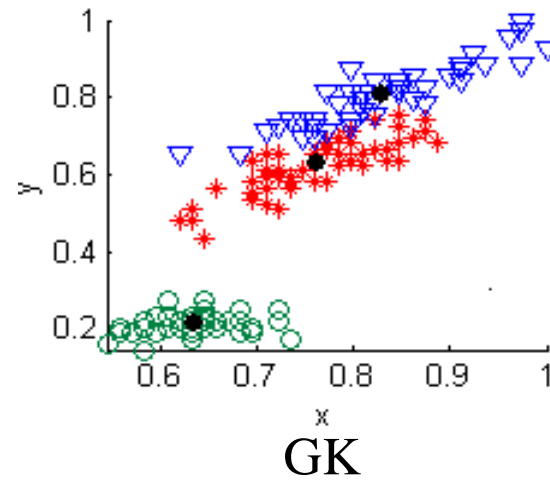
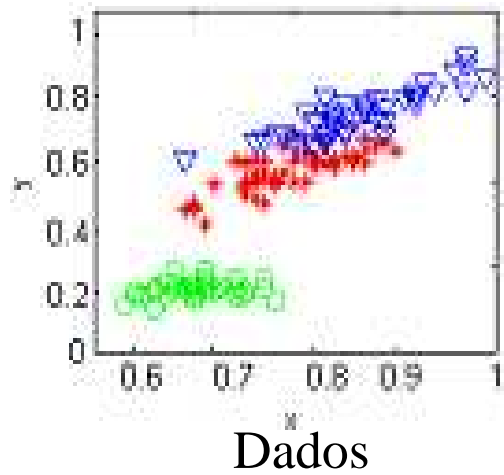
Versicolor



Virginica



■ Exemplo: Iris



## ■ Regras linguísticas

Forma canônica: Se  $\langle$  variável entrada é  $A_i$   $\rangle$  Então  $\langle$  variável saída é  $B_i$   $\rangle$

$$i = 1, 2, \dots, N$$

$A_i$  e  $B_i$  são conjuntos fuzzy em  $\mathbf{X}$  e  $\mathbf{Y}$

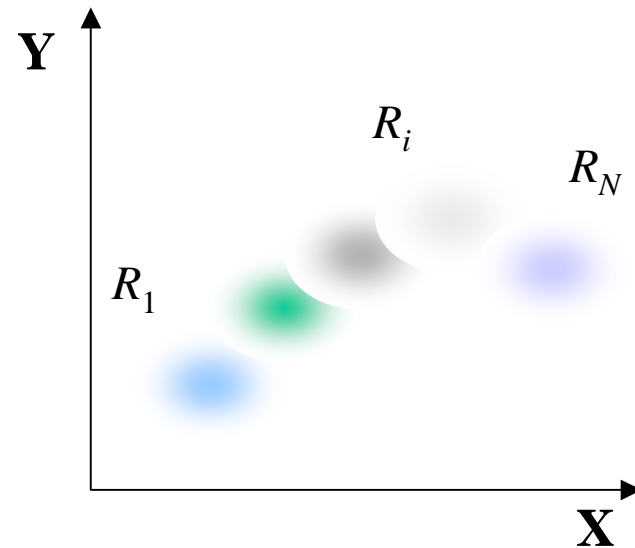
$R$  :

Se  $X$  é  $A_1$  Então  $Y$  é  $B_1$

Se  $X$  é  $A_2$  Então  $Y$  é  $B_2$

.....

Se  $X$  é  $A_N$  Então  $Y$  é  $B_N$



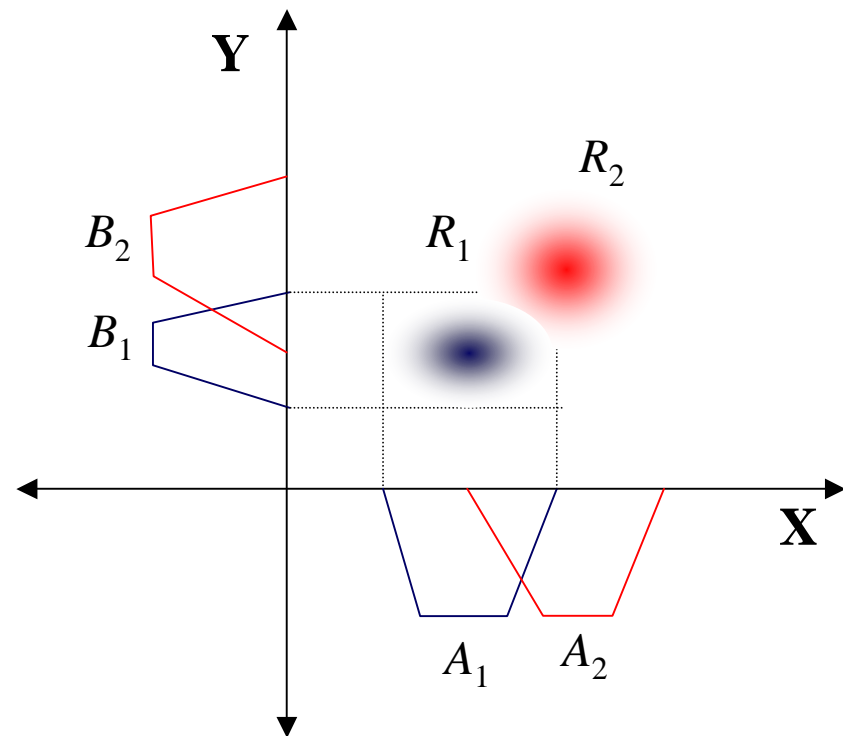
■ Exemplo

*Se temperatura é alta Então consumo de energia é alto*

*Se temperatura é média Então consumo de energia é médio*

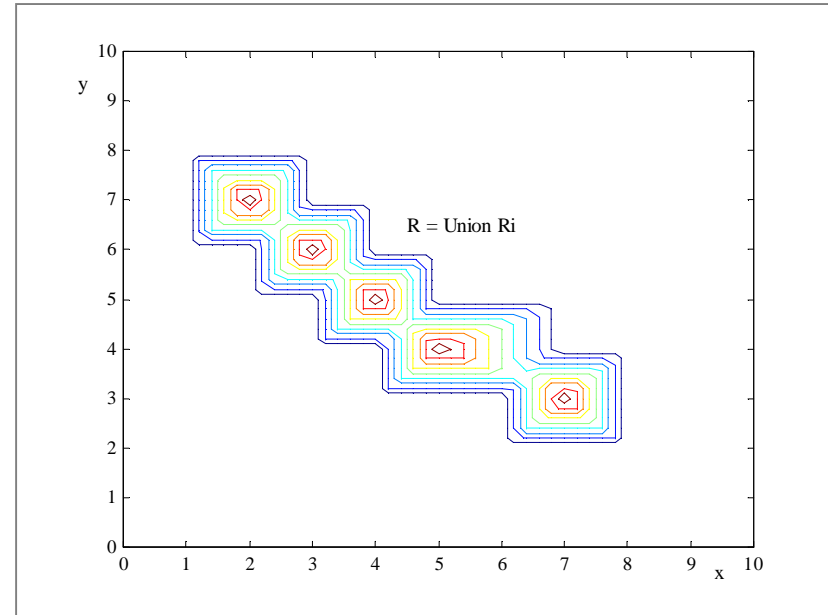
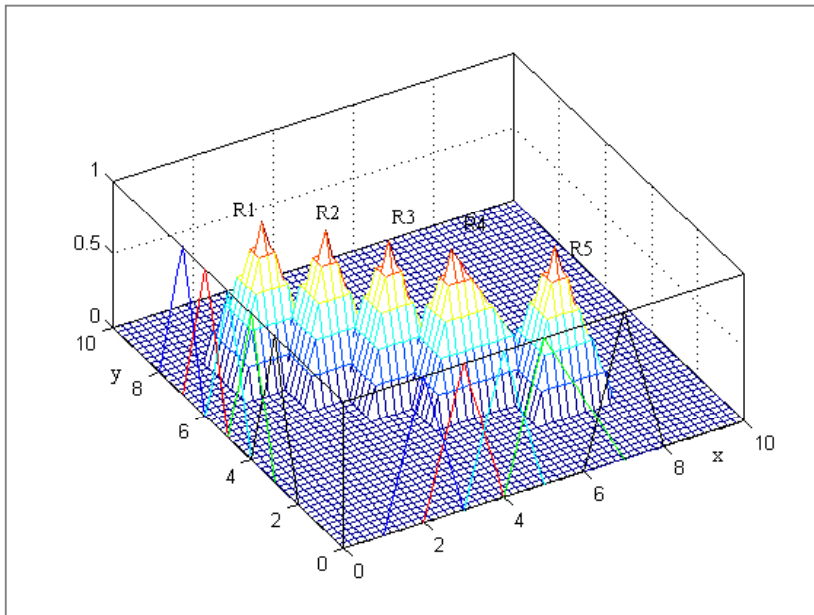
$R_1$  : Se  $X$  é  $A_1$  Então  $Y$  é  $B_1$

$R_2$  : Se  $X$  é  $A_2$  Então  $Y$  é  $B_2$





## ■ Regras linguísticas



## ■ Regras funcionais

Forma canônica: Se  $\langle \text{variável entrada é } A_i \rangle$  Então  $\langle \text{variável saída é } f_i(x) \rangle$

$$i = 1, 2, \dots, N$$

$A_i$  é um conjunto fuzzy em  $\mathbf{X}$

$f_i$  é uma função  $f_i : \mathbf{X} \rightarrow \mathbf{Y}$

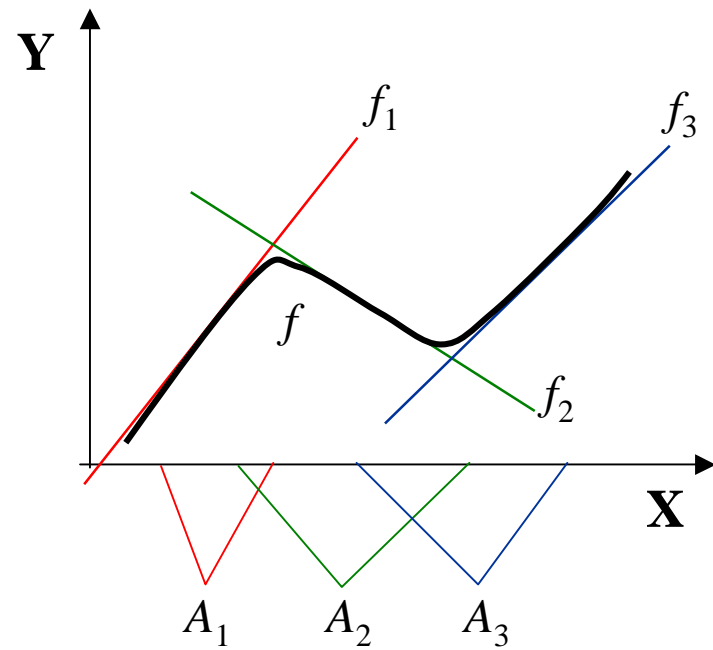
■ Exemplo

Se  $X$  é  $A_1$  Então  $y$  é  $f_1(x)$

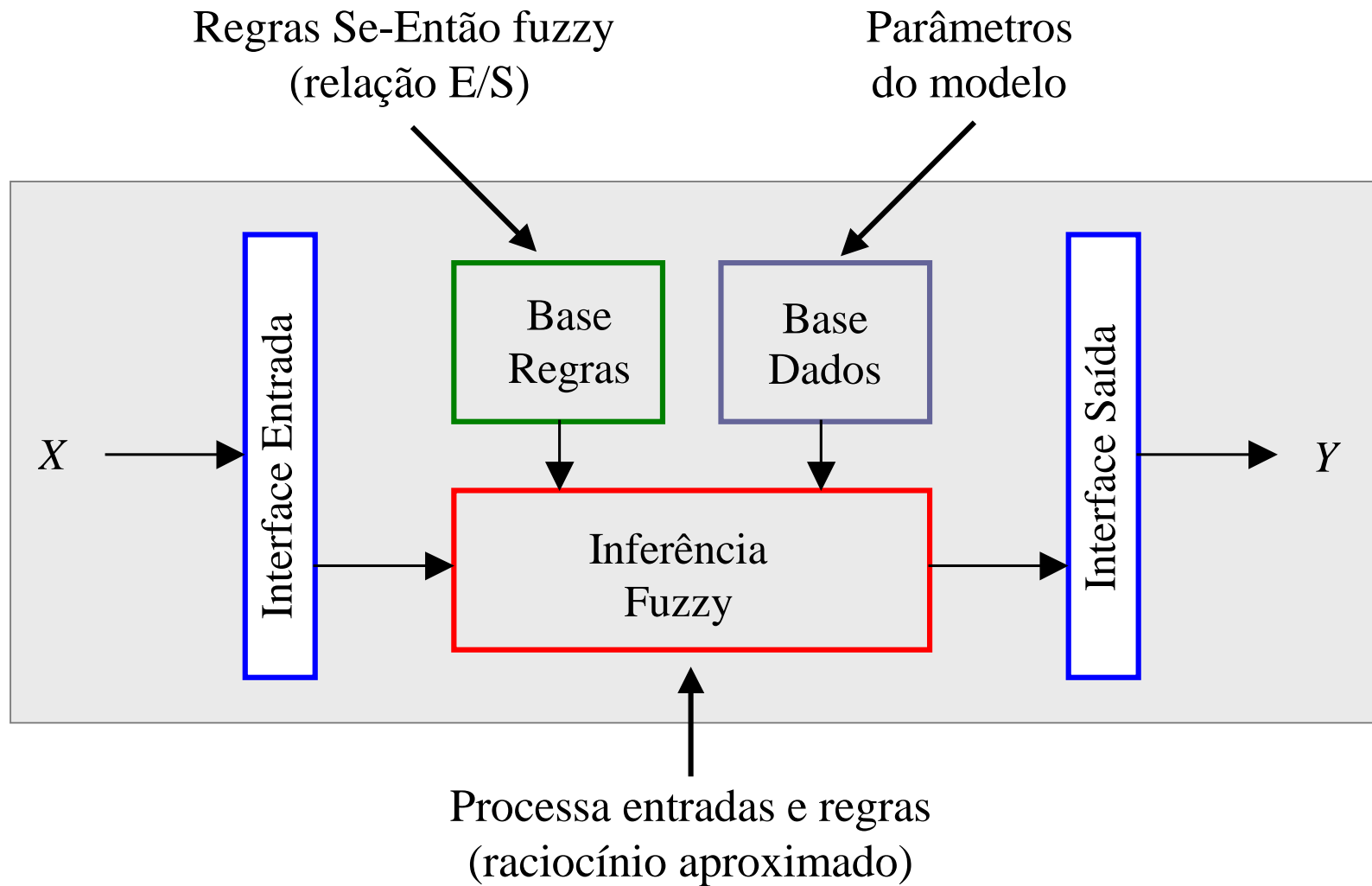
$R$  : Se  $X$  é  $A_2$  Então  $y$  é  $f_2(x)$

.....

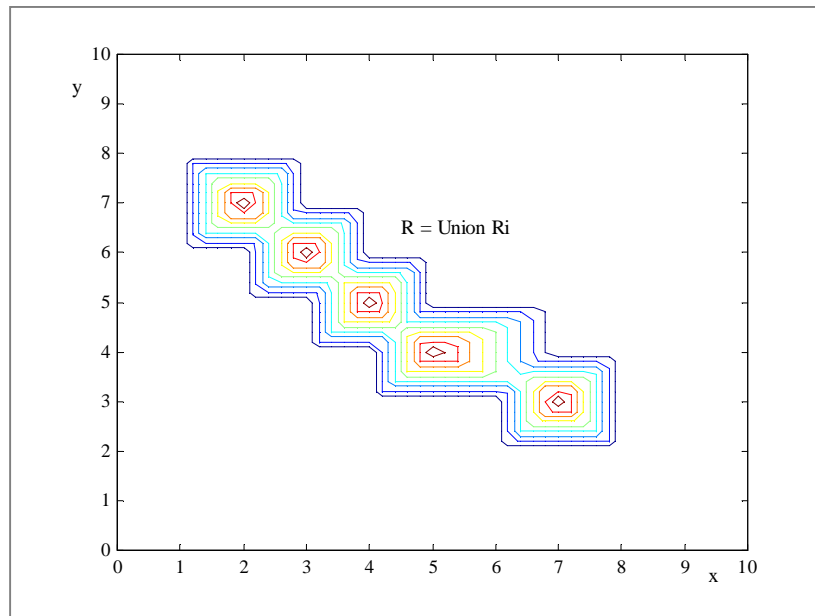
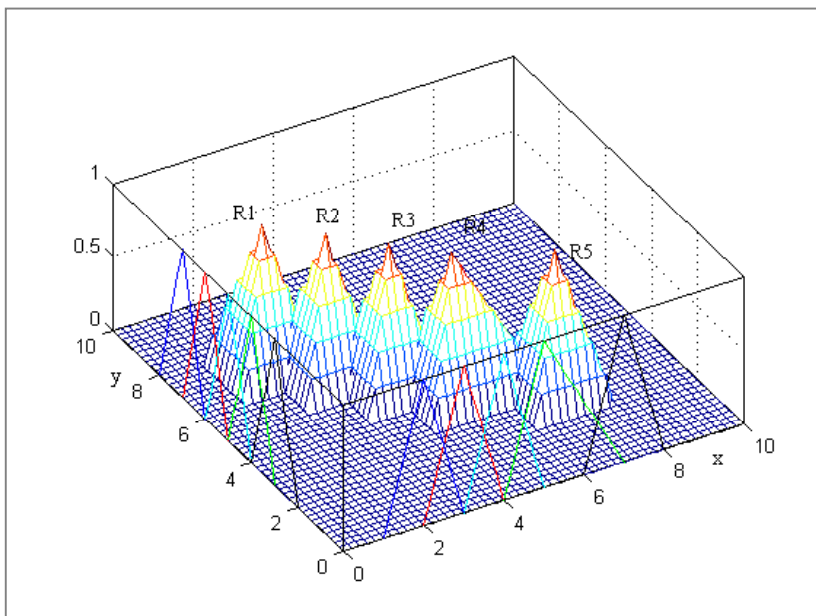
Se  $X$  é  $A_N$  Então  $y$  é  $f_N(x)$



■ Estrutura de modelos fuzzy



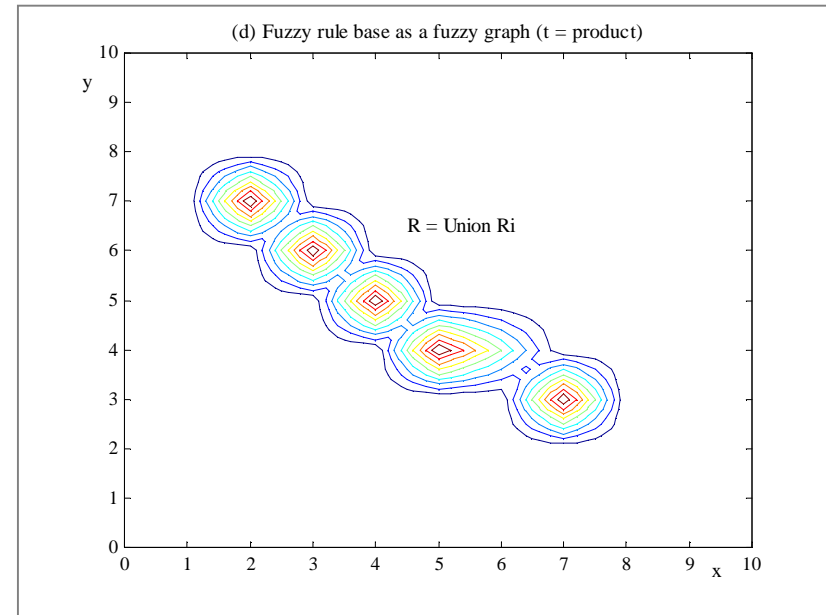
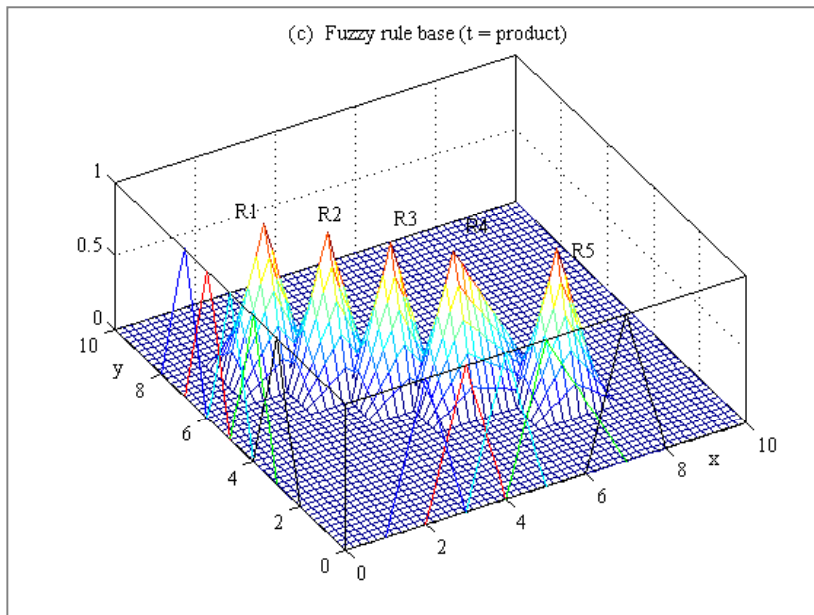
- Base reglas lingüísticas  $\equiv$  grafo nebuloso



$$R_i = A_i \times B_i \Rightarrow R_i(x,y) = \min [A_i(x), B_i(y)]$$

$$R = \cup R_i \Rightarrow R(x,y) = \max [R_i(x,y), i = 1, \dots, N]$$

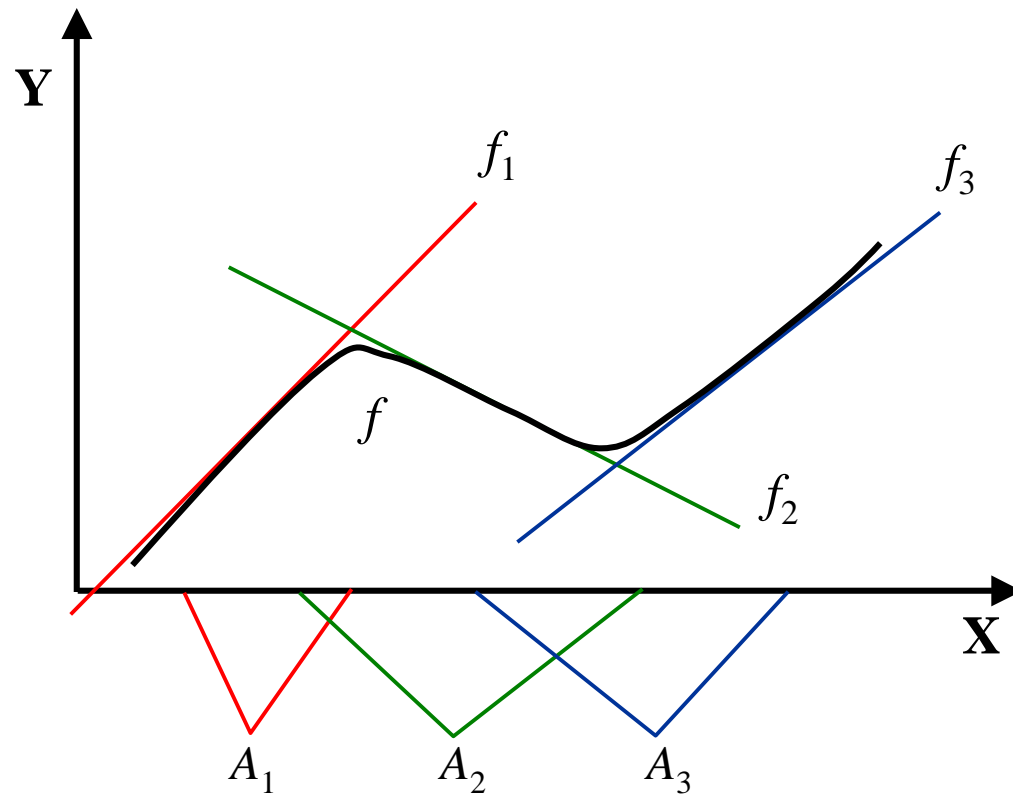
■ Base regras linguísticas



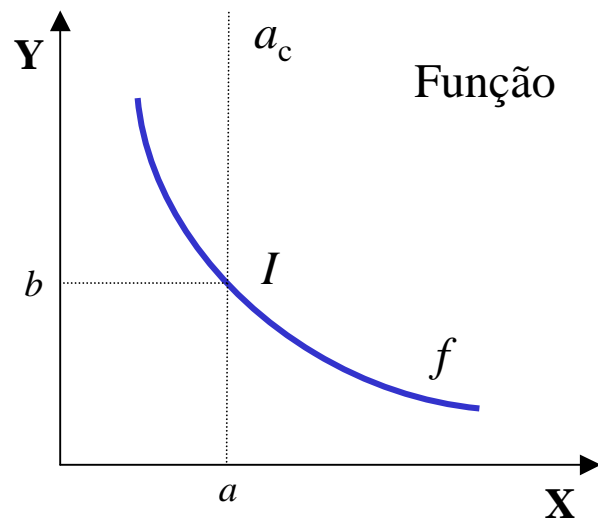
$$R_i = A_i \text{ t } B_i \Rightarrow R_i(x,y) = A_i(x) B_i(y)$$

$$R = \cup R_i \Rightarrow R(x,y) = \max [R_i(x,y), i = 1, \dots, N]$$

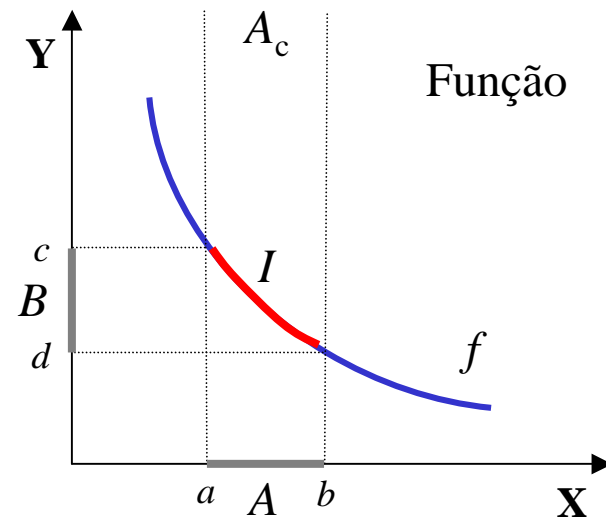
- Base de regras fuzzy funcionais



- Modelos linguísticos e inferência

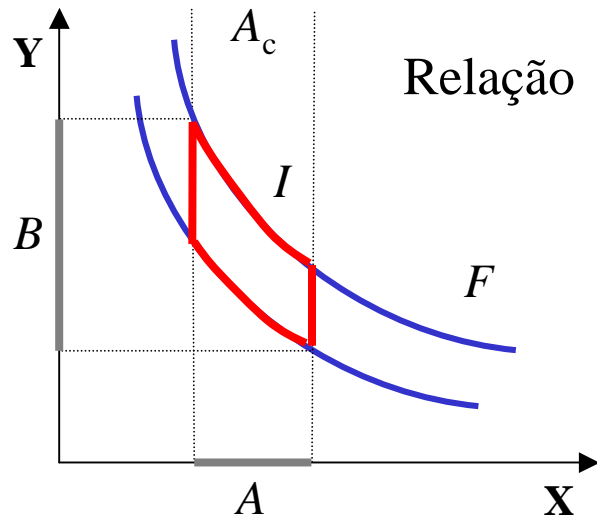


$$\frac{\begin{array}{l} x = a \\ y = f(x) \end{array}}{y = b}$$

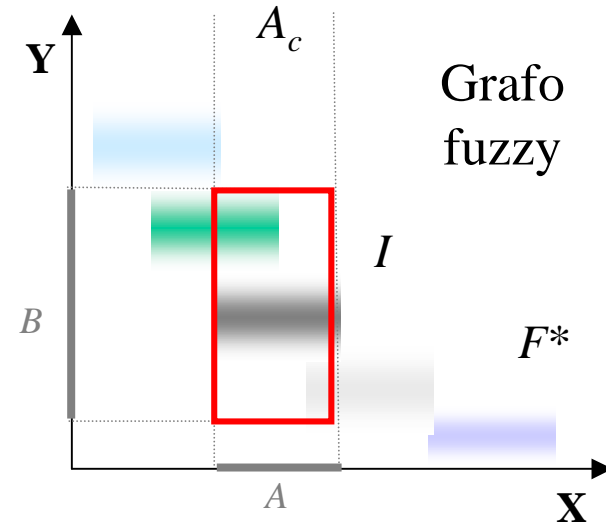


$$\frac{\begin{array}{l} x \text{ é } A \\ (x, y) \text{ é } f \end{array}}{y \text{ é } B}$$



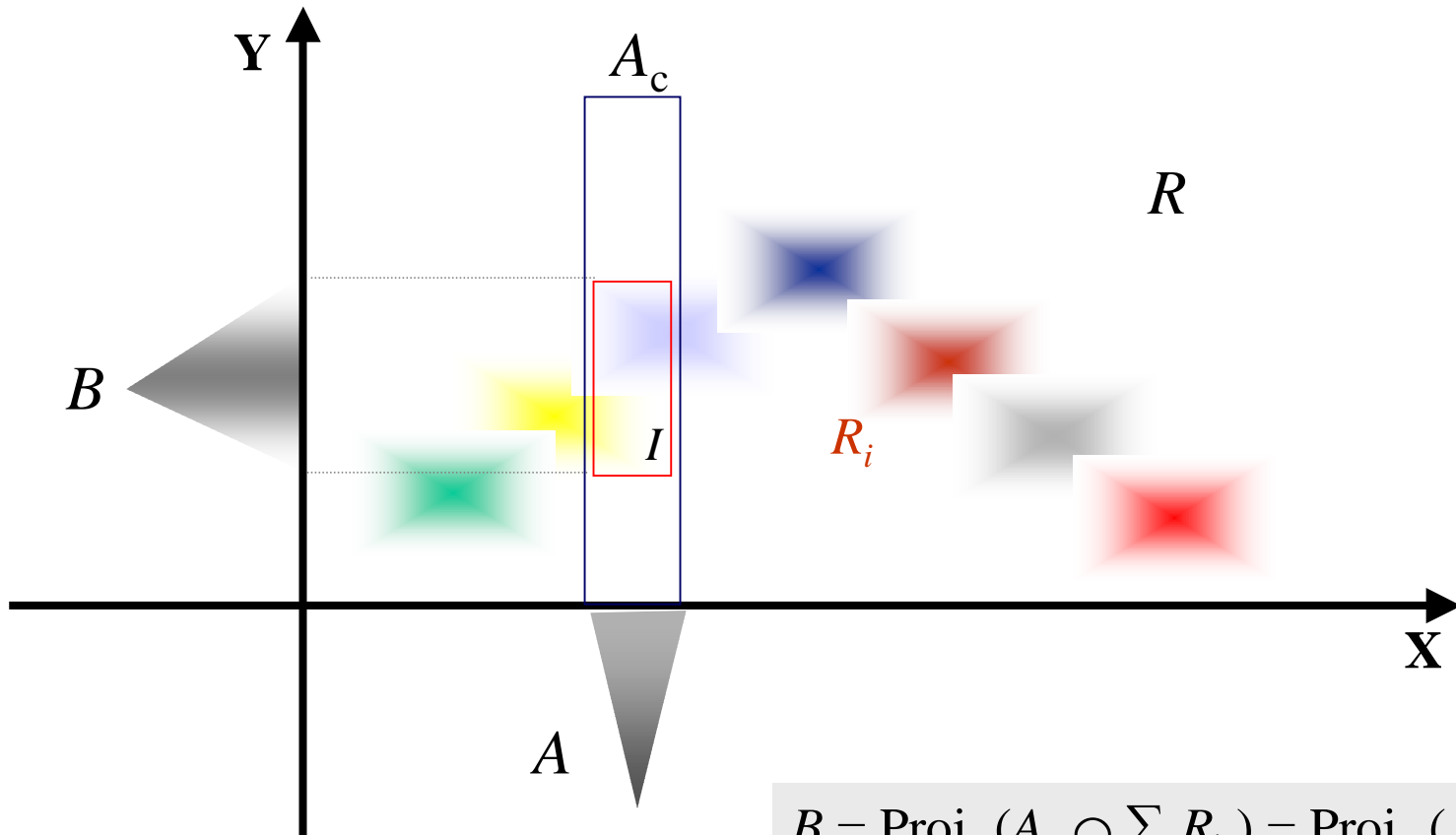


$$\frac{x \in A \quad (x, y) \in F}{y \in B}$$



$$\frac{X \in A \quad (X, Y) \in F^*}{Y \in B}$$

# Inferência fuzzy



$X \text{ é } A$   
 $(X, Y) \text{ é } R$   


---

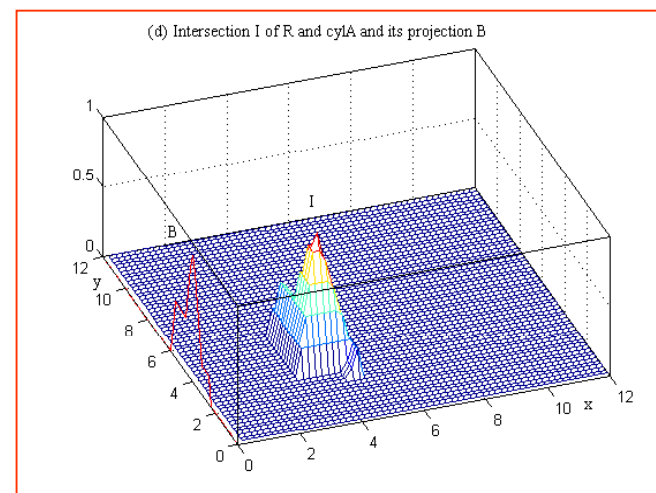
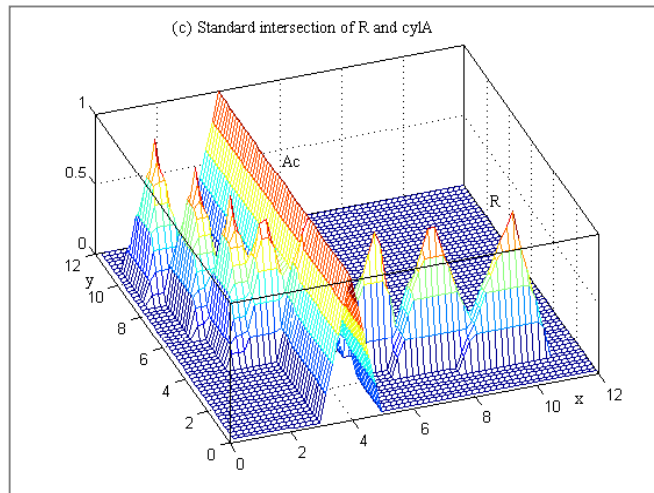
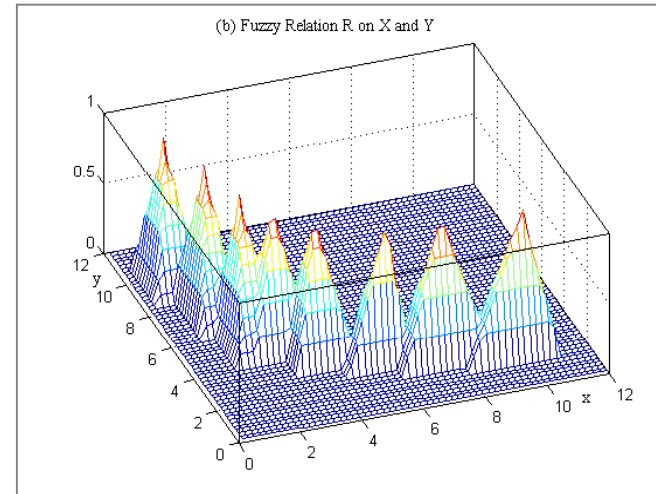
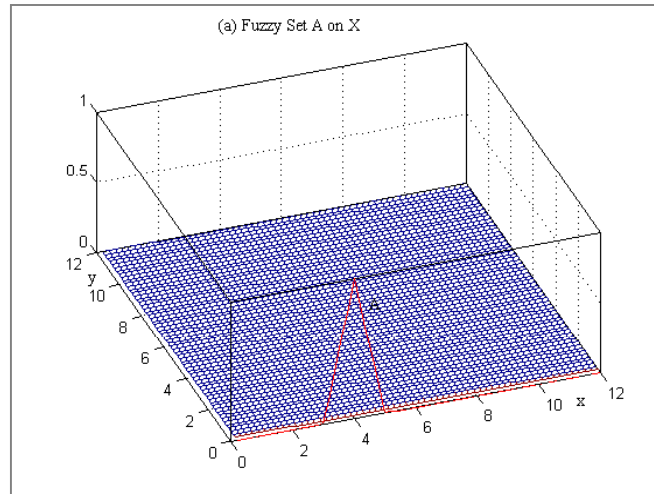
 $Y \text{ é } A \circ R$

$$B = \text{Proj}_Y(A_c \cap \sum_i R_i) = \text{Proj}_Y(A_c \cap R)$$

$$B = A \circ R \quad \text{regra da composição}$$

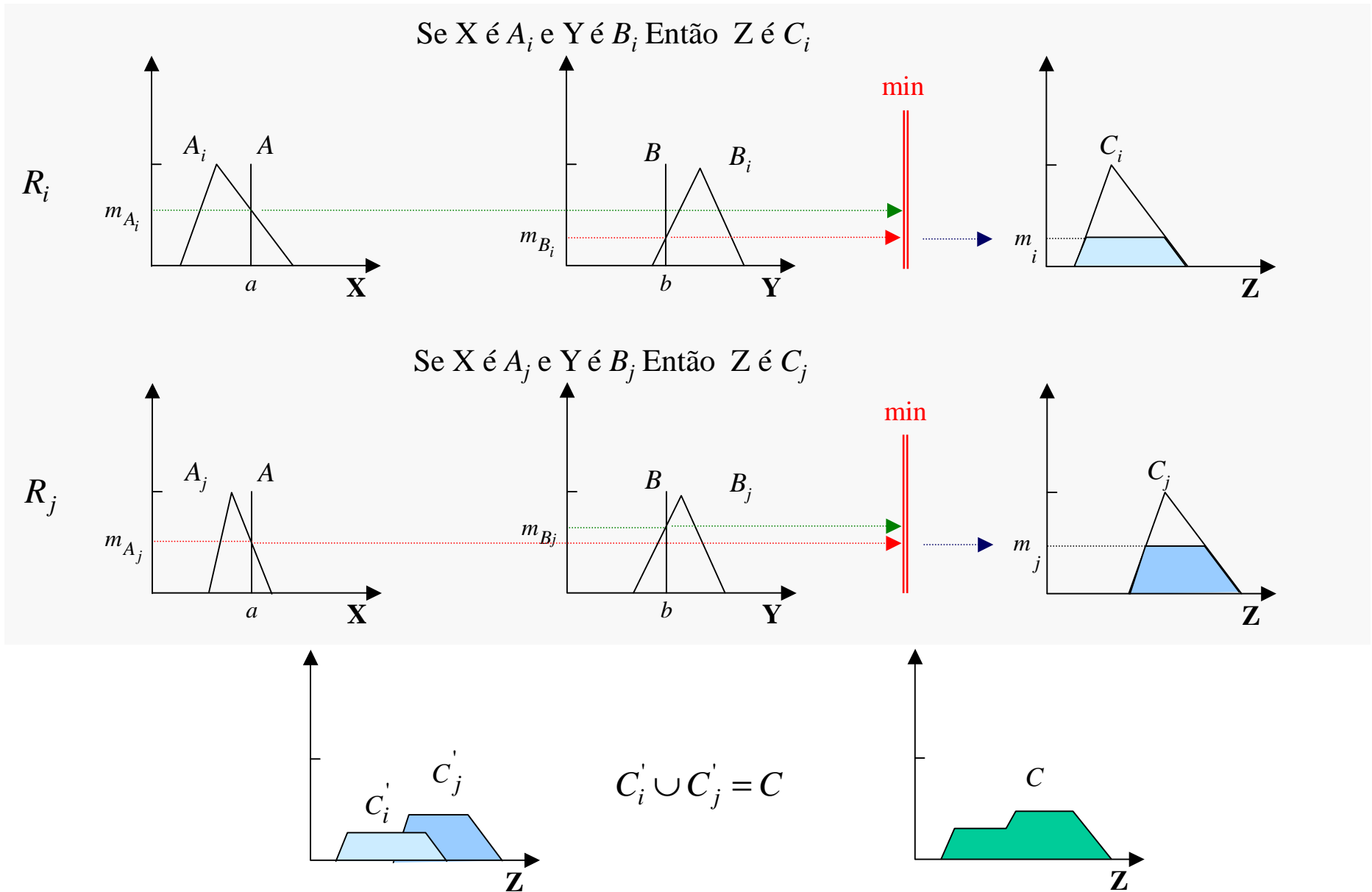
$$B(y) = \sup_x [ A(x) \text{ t } R(x, y) ]$$

# ■ Exemplo



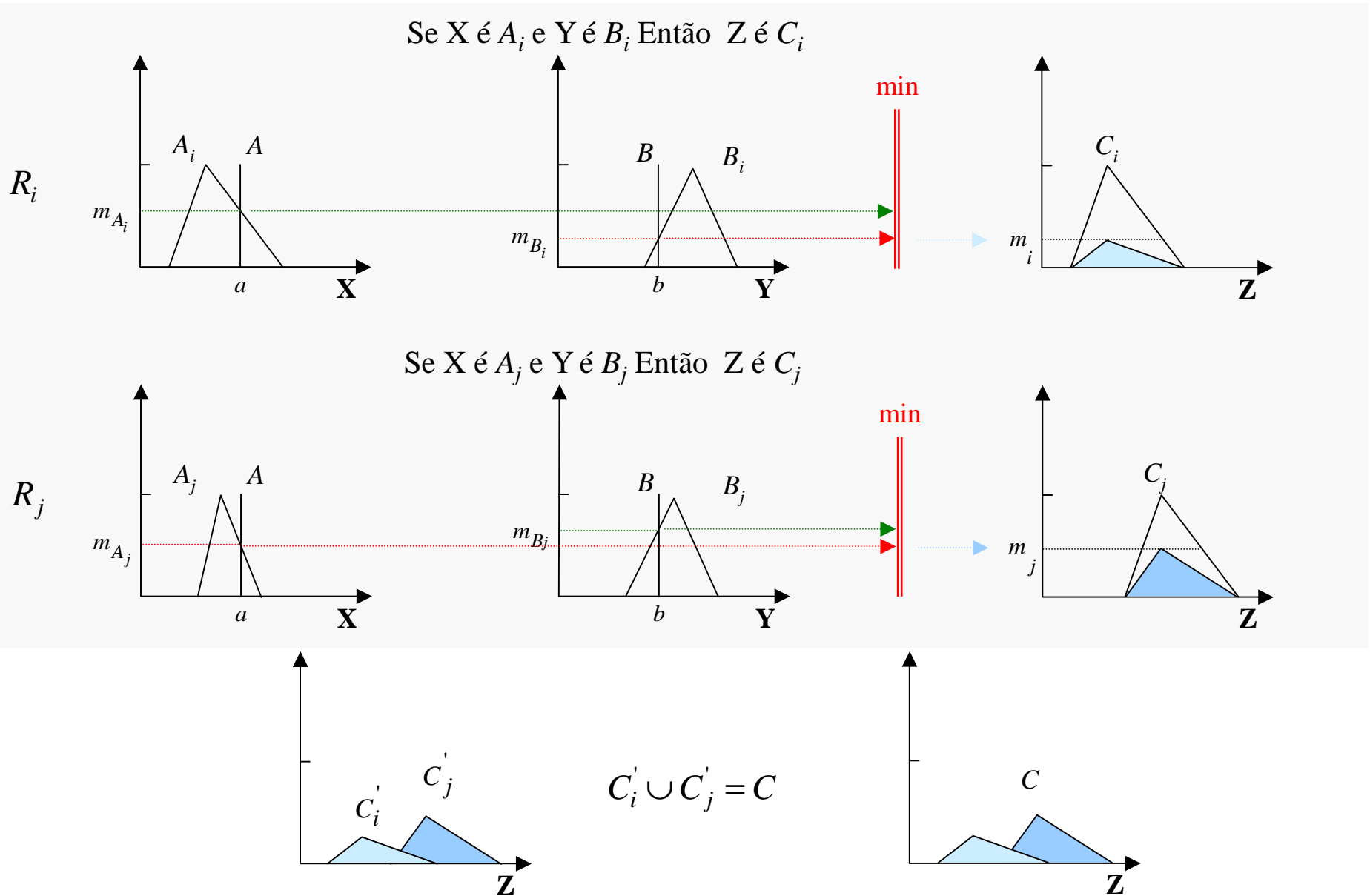
■ Exemplo modelo linguístico (Mamdani)

$X$  é  $A$  e  $Y$  é  $B$



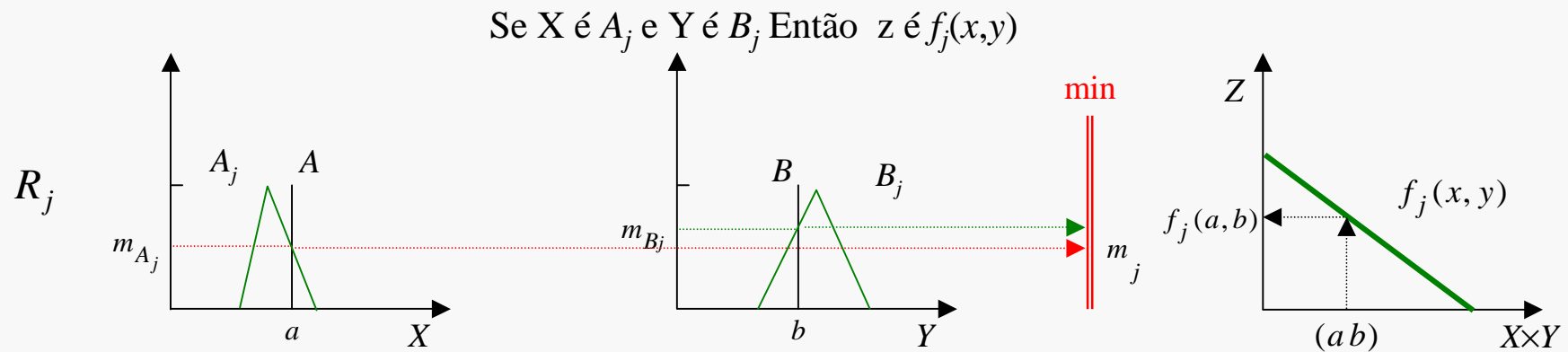
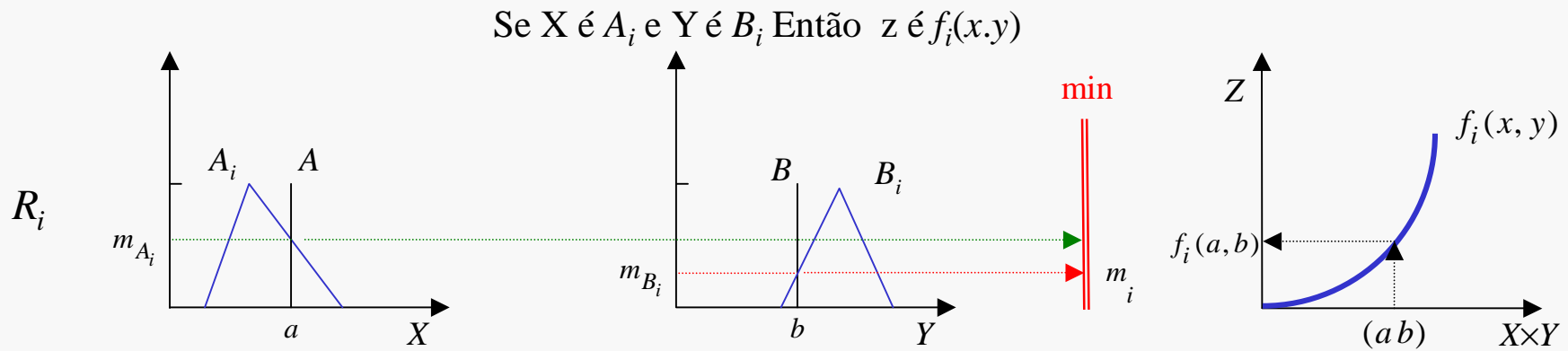
# Exemplo modelo linguístico (Larsen)

$X$  é  $A$  e  $Y$  é  $B$



■ Exemplo modelo funcional (TS)

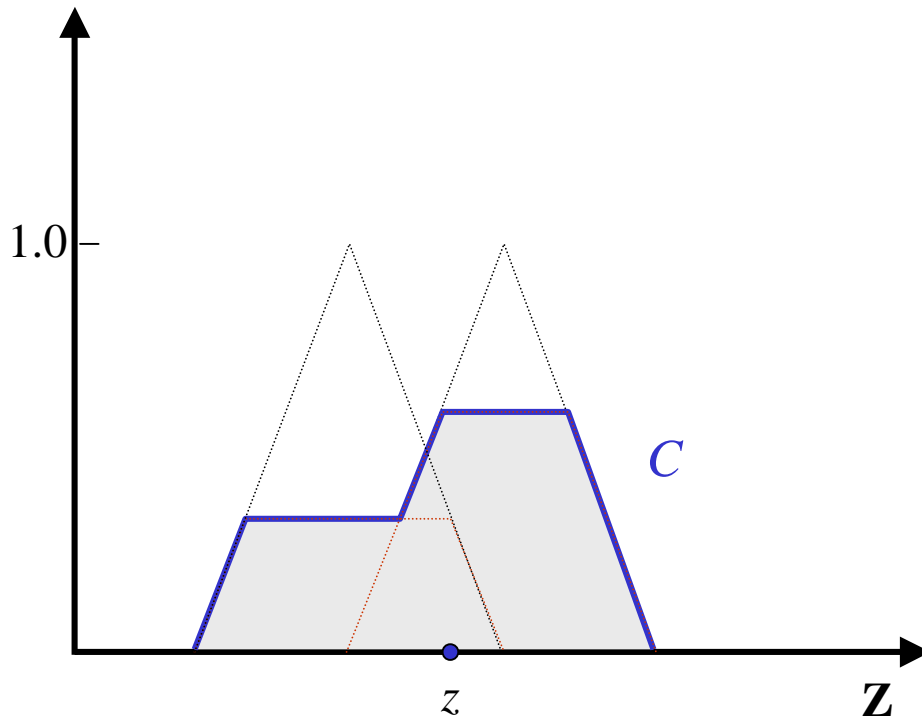
$X \text{ é } A \text{ e } Y \text{ é } B$



$$z = \frac{m_i f_i(a,b) + m_j f_j(a,b)}{m_i + m_j}$$

- Interface de saída

Centro de gravidade

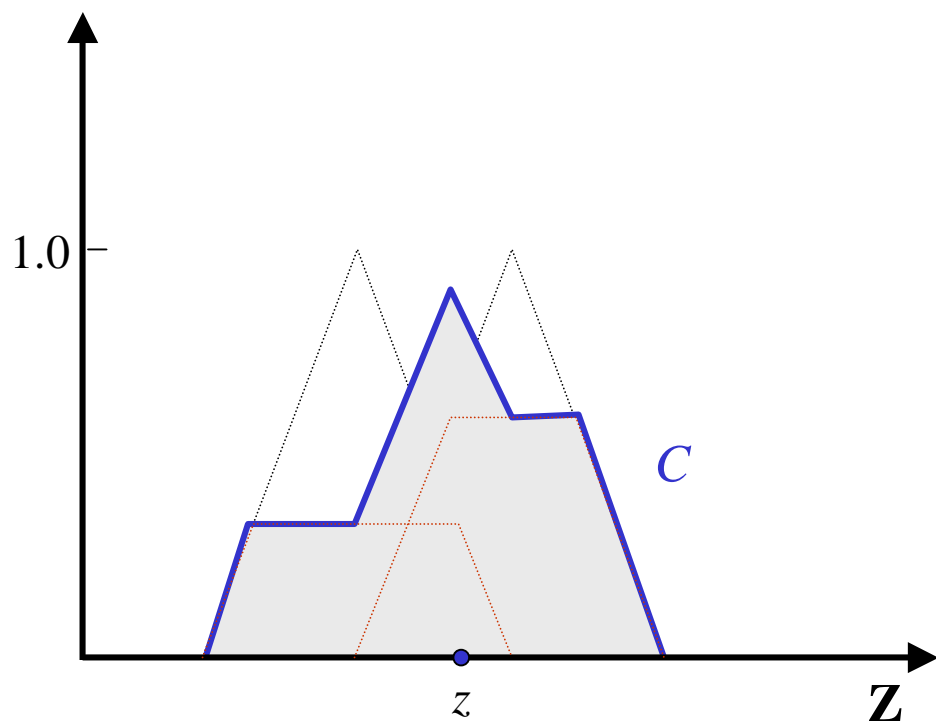


$$z = \frac{\sum_{i=1}^n z_i C(z_i)}{\sum_{i=1}^n C(z_i)}$$

$$Z = [z_1, \dots, z_n]$$

$$C = \bigcup_{k=1}^N C'_k$$

■ Centro da soma



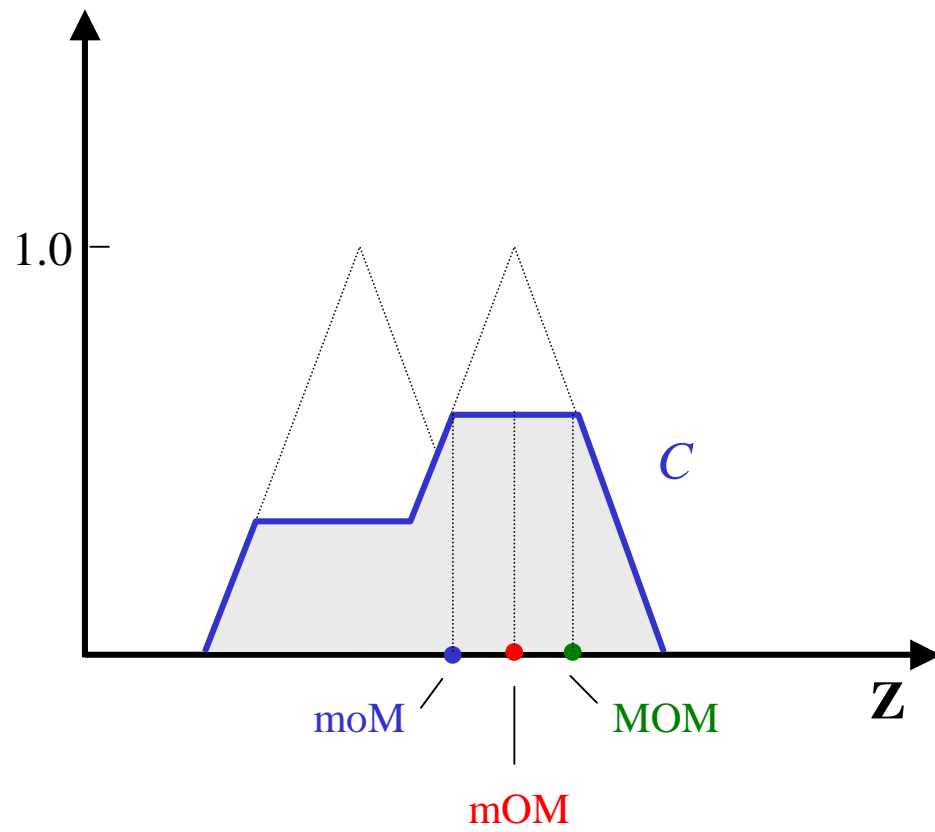
$$z = \frac{\sum_{i=1}^n z_i \sum_{k=1}^{N'} C'_k(z_i)}{\sum_{i=1}^n \sum_{k=1}^{N'} C'_k(z_i)}$$

$$Z = [z_1, \dots, z_n]$$

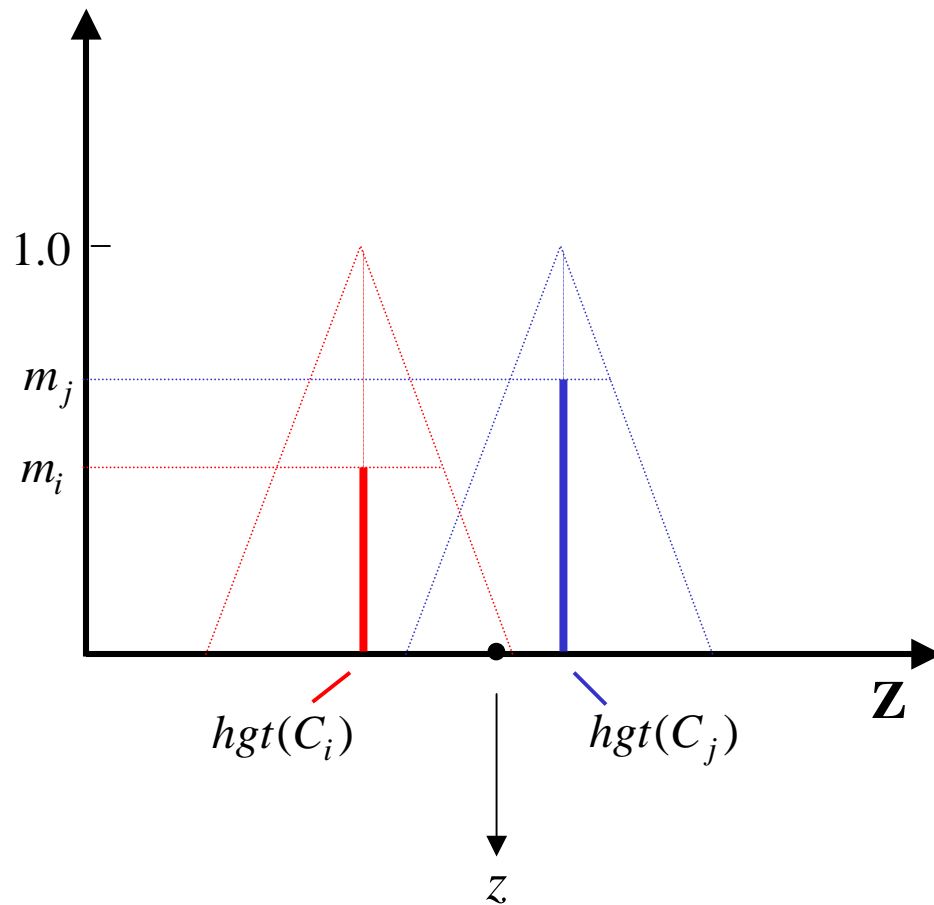
$N' = \#$  regras ativas



■ Método dos máximos



■ Método das alturas

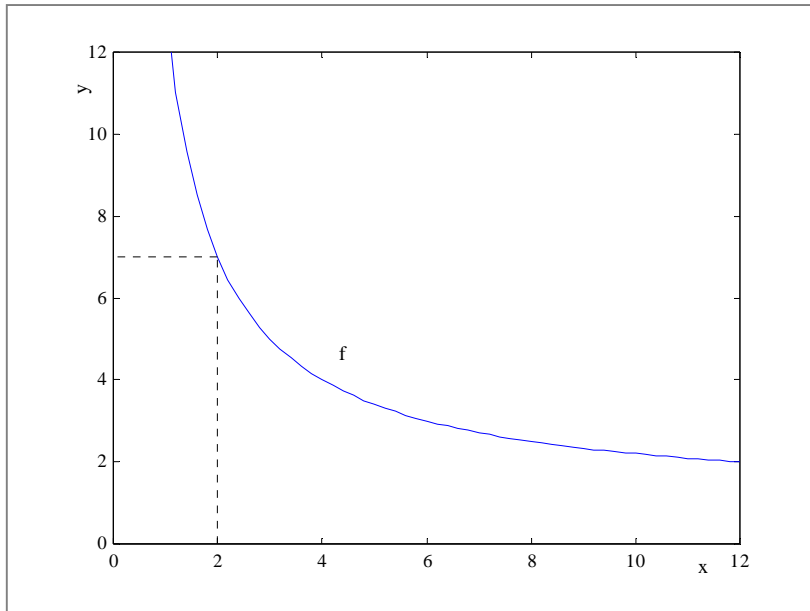


$$z = \frac{\sum_{k=1}^{N'} m_k hgt(C_k)}{\sum_{k=1}^{N'} hgt(C_k)}$$

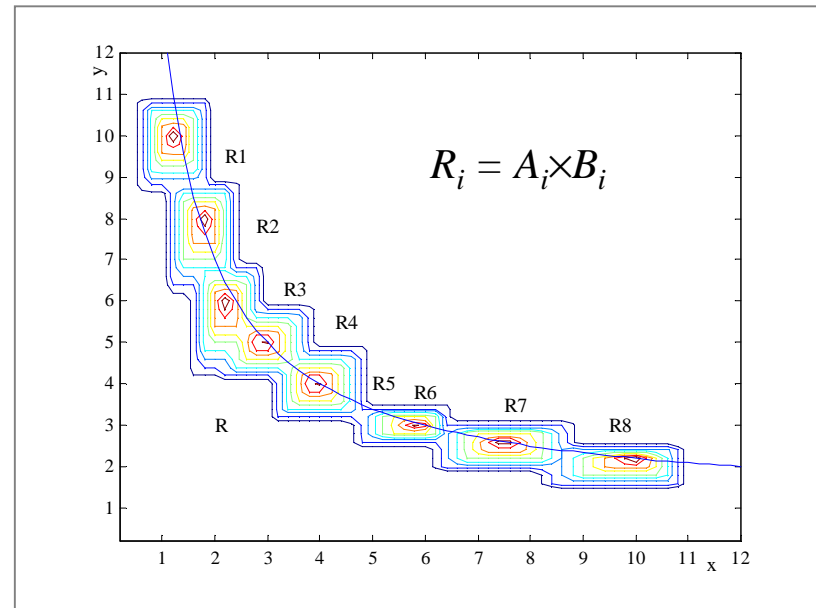
$N' = \#$  regras ativas

## ■ Aproximação universal

Grafo de  $f$  e sua aproximação granular  $R$

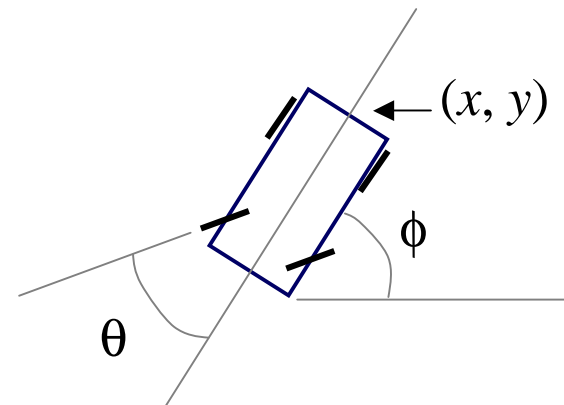
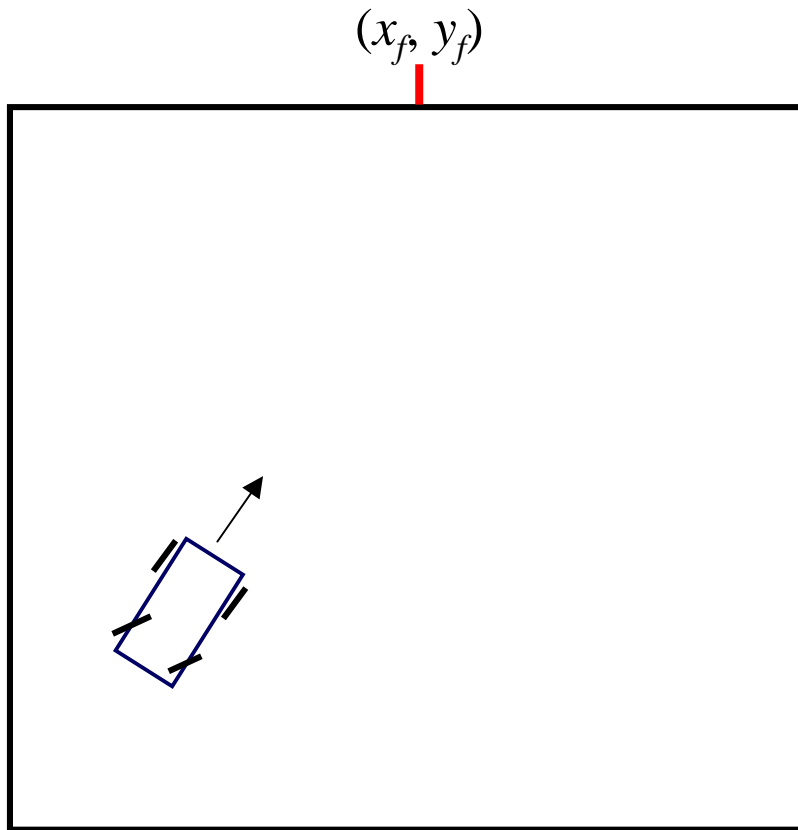


$f$

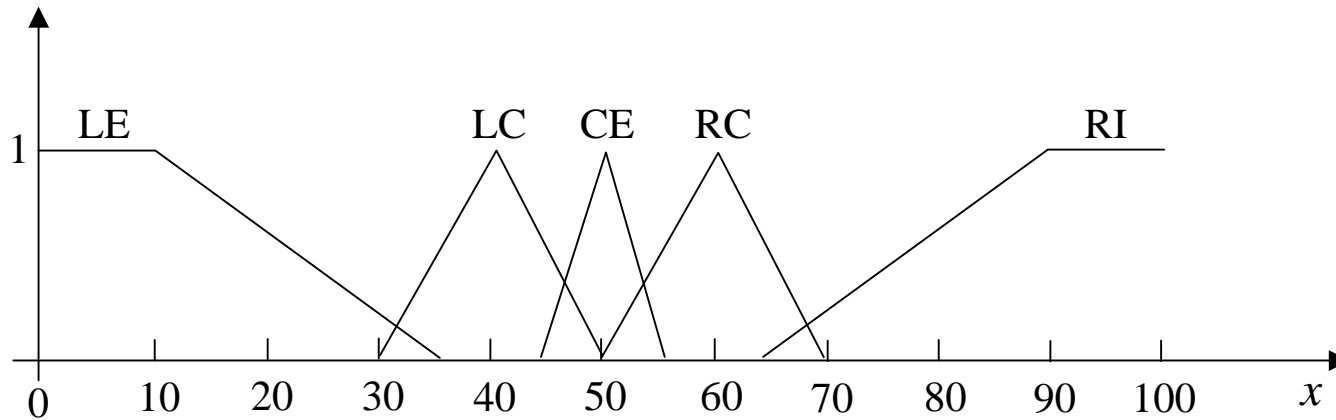


$R$

- Exemplo: guiagem autônoma

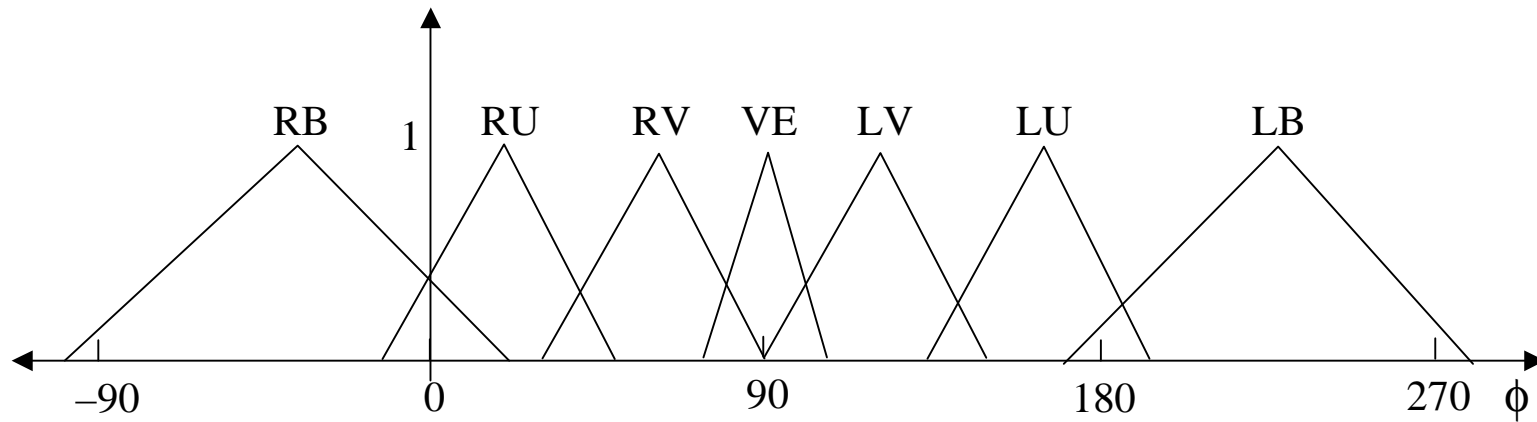


■ Granularização da posição



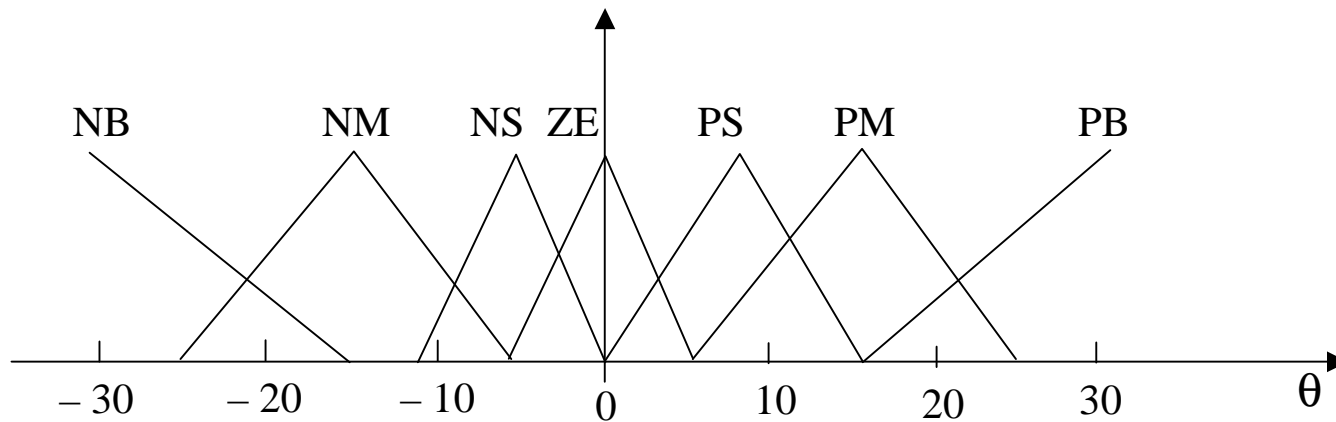
$$\mathbf{X} = \{x \mid 0 \leq x \leq 100\}$$

- Granularização da orientação



$$\Phi = \{\phi \mid -90 \leq \phi \leq 270\}$$

- Granularização do ângulo do volante



$$\Theta = \{\theta \mid -30 \leq \theta \leq 30\}$$

■ Controlador





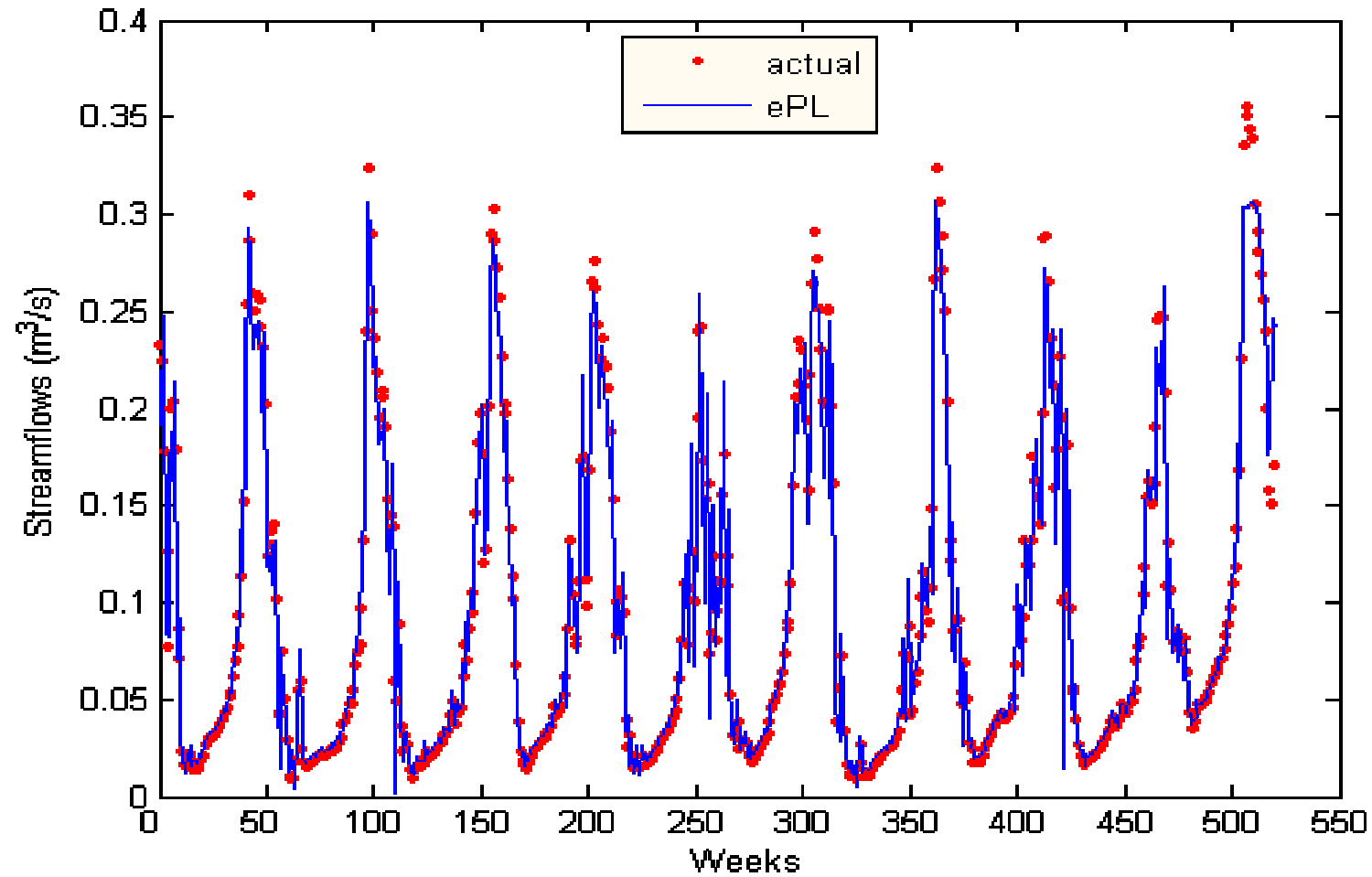
■ Base de regras

		$x$						
		LE	LC	CE	RC	RI		
	RB	PS	PM	PM	PB	PB		
	RU	NS	PS	PM	PB	PB		
$\phi$	RV	NM	NS	PS	PM	PB		
	VE	NM	NM	ZE	PM	PM		
	LV	NB	NM	NS	PS	PM		
	LU	NB	NB	NM	NS	PS		
	LB	NB	NB	NM	NM	NS		

Se  $x$  é LE e  $\phi$  é RB Então  $\theta$  é PS

Se  $x$  é CE e  $\phi$  é VE Então  $\theta$  é ZE

■ Previsão de vazão



## 6-Resumo

- Métodos para estimar densidades genéricas
- Duas abordagens importantes
  - estimação das densidades  $p(\mathbf{x} | \omega_i)$  e utilizar Bayes (Parzen)
  - estimação direta das probabilidades *a posteriori*  $P(\omega_i | \mathbf{x})$  ( $k_n$  vizinhos)
- Teoria de conjuntos fuzzy
- Classificadores fuzzy

## Observação

Este material refere-se às notas de aula do curso CT 720 Tópicos Especiais em Aprendizagem de Máquina e Classificação de Padrões da Faculdade de Engenharia Elétrica e de Computação da Unicamp e do Centro Federal de Educação Tecnológica do Estado de Minas Gerais. Não substitui o livro texto, as referências recomendadas e nem as aulas expositivas. Este material não pode ser reproduzido sem autorização prévia dos autores. Quando autorizado, seu uso é exclusivo para atividades de ensino e pesquisa em instituições sem fins lucrativos.