

## Curso 2 – CD, AM e DM

# SVM

## *Support Vector Machine*

Roseli Ap. Francelin Romero  
SCC – ICMC - USP



# Tópicos

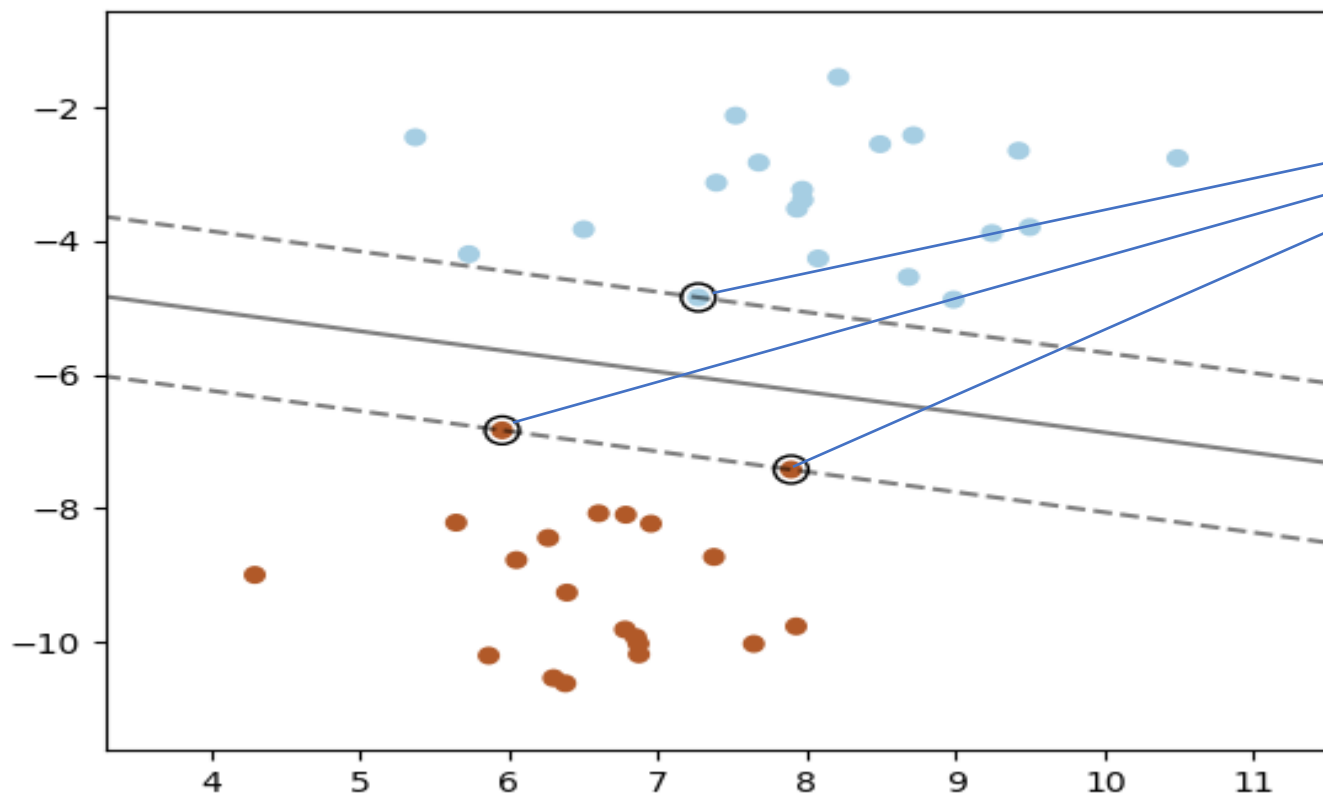
- Introdução
- SVMs lineares
- SVMs não lineares
- SVMs em problemas multiclases
- Conclusões



IA  
BIG  
DATA



# Hiperplano de Margens separadoras máximas



SVC – Encontra  
Vetores suporte

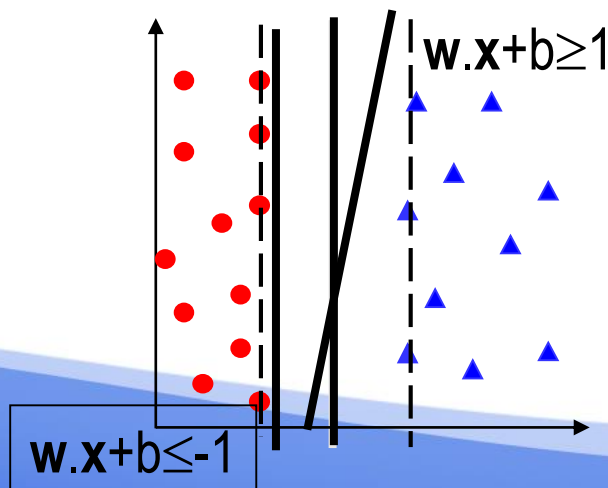


# SVMs lineares

- Funções do tipo

$$g(\mathbf{x}) = \text{sgn}(f(\mathbf{x})) = \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b) = \begin{cases} +1 & \text{se } \mathbf{w} \cdot \mathbf{x} + b > 0 \\ -1 & \text{se } \mathbf{w} \cdot \mathbf{x} + b < 0 \end{cases}$$

- Conjunto de dados linearmente separável

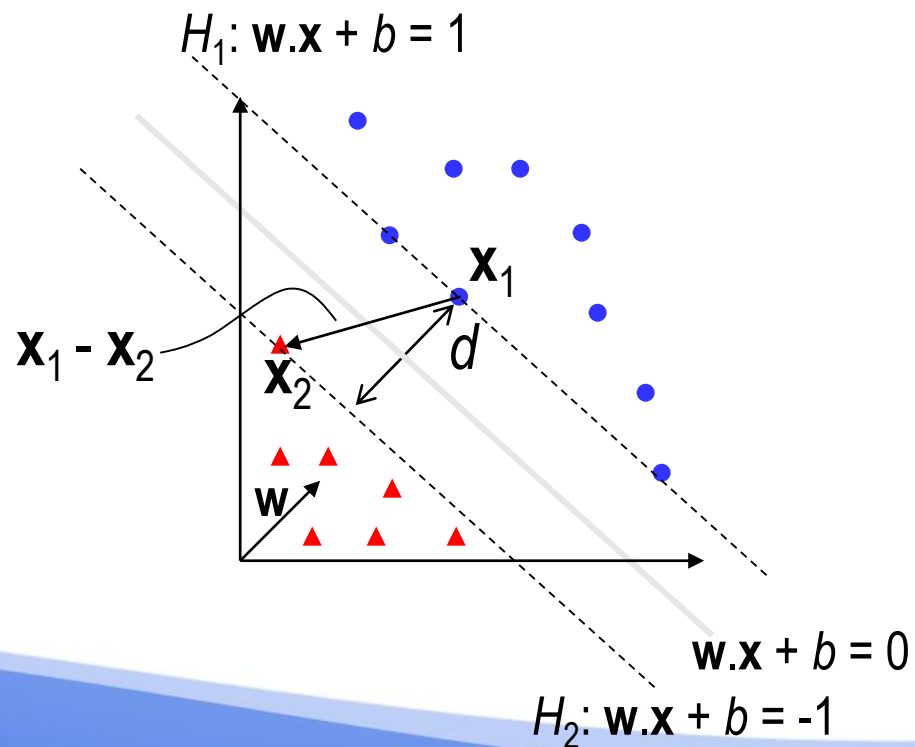


Hiperplano canônico em relação a T

$$\begin{cases} \mathbf{w} \cdot \mathbf{x}_i + b \geq +1 & \text{se } y_i = +1 \\ \mathbf{w} \cdot \mathbf{x}_i + b \leq -1 & \text{se } y_i = -1 \end{cases}$$



# SVMs lineares



$$\begin{cases} w \cdot x_1 + b = +1 \\ w \cdot x_2 + b = -1 \end{cases}$$
$$\frac{w \cdot (x_1 - x_2)}{\|w\| \cdot \|x_1 - x_2\|} = 2$$

Projeção  $x_1 - x_2$  na direção de  $w$

$$(x_1 - x_2) \left( \frac{w}{\|w\|} \cdot \frac{(x_1 - x_2)}{\|x_1 - x_2\|} \right)$$

$d$  = norma da projeção

$$d = \frac{2}{\|w\|}$$

$$\text{margem} = \frac{1}{\|w\|}$$



# Teorema de Cover

O teorema diz que se temos um vetor  $\mathbf{x}$  com dimensionalidade alta e aplicamos um conjunto de funções não lineares  $\varphi_i$ ,  $i=1, \dots, m_0$ , neste padrão gerando um vetor de dimensão mais alta, temos que a probabilidade deste vetor ser linearmente separável é maior.

Seja:  $\mathbf{x} = (x_1, x_2, \dots, x_N)$

Separar linearmente  $\mathbf{x}$  em duas dicotomias, significa encontrar  $\mathbf{w}$  tal que:

$$\mathbf{w}^T \mathbf{x} > 0, \mathbf{x} \in H_1$$

$$\mathbf{w}^T \mathbf{x} < 0, \mathbf{x} \in H_2$$





# Teorema de Cover

Se aplicarmos  $\varphi$  em  $\mathbf{x}$ , geramos um novo vetor  $\varphi(\mathbf{x})$  com  $m_1$  componentes:

$$\varphi(\mathbf{x}) = [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \dots, \varphi_{m_1}(\mathbf{x})]^T$$

Separar linearmente este vetor, significa, da mesma forma, encontrar um  $\mathbf{w}$ , tal que:

$$\mathbf{w}^T \varphi(\mathbf{x}) > 0, \mathbf{x} \in H_1$$

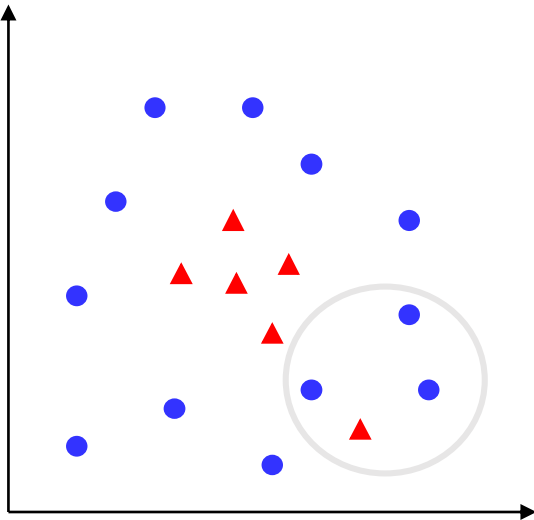
$$\mathbf{w}^T \varphi(\mathbf{x}) < 0, \mathbf{x} \in H_2$$

O teorema de Cover diz (em outras palavras) que a probabilidade de separação linear aumenta, a medida que  $m_1$  aumenta.



# SVMs não lineares

- Muitos conjuntos de dados são não lineares



- Mapeia dados para espaço de maior dimensão
- Teorema de Cover

Escolha apropriada de função de mapeamento



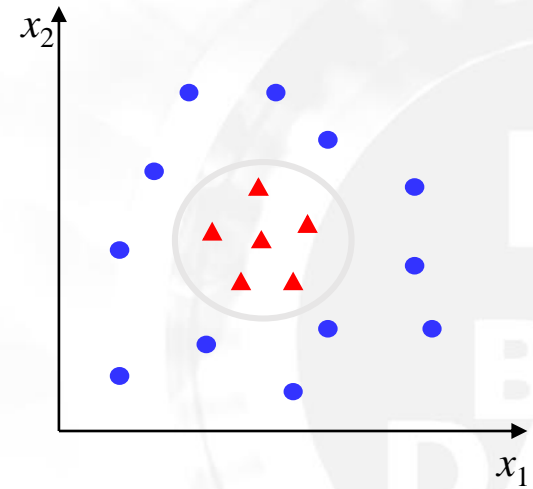
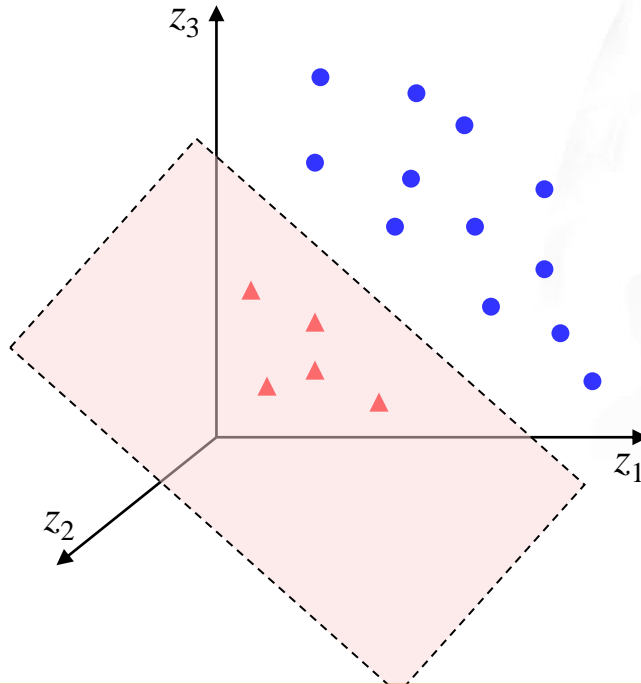
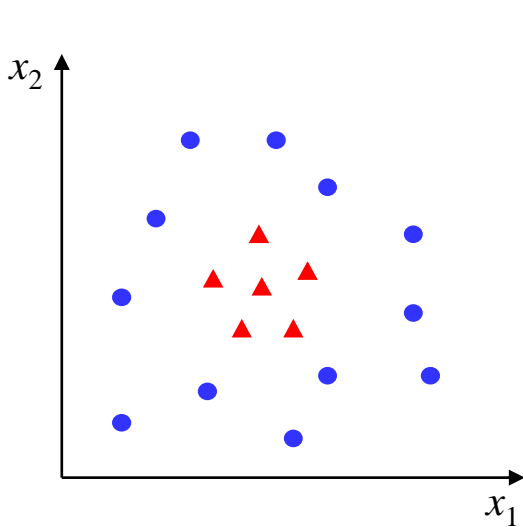
Dados podem ser separados por SVM linear

margens suaves





# SVMs não lineares



$$\Phi(\mathbf{x}) = \left( x_1^2, \sqrt{2}x_1x_2, x_2^2 \right)$$

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b = w_1x_1^2 + w_2\sqrt{2}x_1x_2 + w_3x_2^2 + b$$



# SVMs não lineares

- Principais tipos de Kernel

| Tipo            | $K(\mathbf{x}_i, \mathbf{x}_j)$                      | Parâmetros     |
|-----------------|--|----------------|
| Polinomial      | $(\delta(\mathbf{x}_i \cdot \mathbf{x}_j) + c)^d$    | $\delta, c, d$ |
| Gaussiano (RBF) | $\exp(-\gamma \ \mathbf{x}_i - \mathbf{x}_j\ ^2)$    | $\gamma$       |
| Sigmoidal       | $\tanh(\delta(\mathbf{x}_i \cdot \mathbf{x}_j) + c)$ | $\delta, c$    |

Pol. Grau 2  
Pol. Grau 3

seleção de modelo

tipo de Kernel

parâmetros do Kernel



# SVMs para problemas multiclass

- SVMs: originalmente para problemas binários
  - Classes +1 e -1

Problemas multiclass

Combinação de classificadores binários

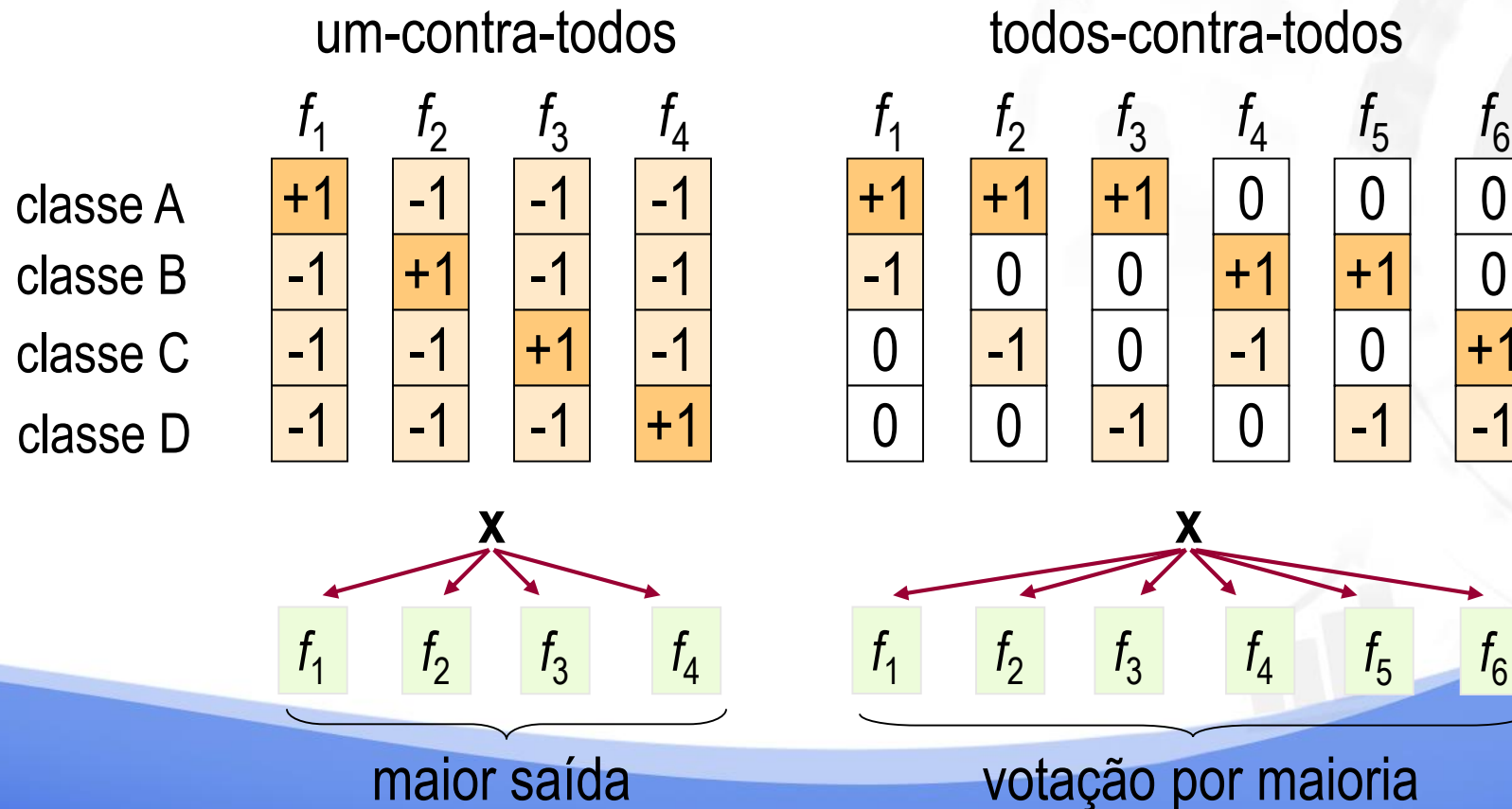
Modificação no algoritmo de otimização

Algoritmos computacionalmente custosos



# SVMs para problemas multiclases

- Decomposições comuns:



# Exemplos de aplicações

- Benchmarks:*

| Conj dados | SVM                               | RBF                               |
|------------|-----------------------------------|-----------------------------------|
| B. Cancer  | <b><math>26.0 \pm 0.47</math></b> | $27.6 \pm 0.47$                   |
| Diabetes   | <b><math>23.5 \pm 0.17</math></b> | $23.2 \pm 0.16$                   |
| German     | <b><math>23.6 \pm 0.21</math></b> | $24.7 \pm 0.24$                   |
| Heart      | <b><math>16.0 \pm 0.33</math></b> | $17.6 \pm 0.33$                   |
| Image      | <b><math>3.0 \pm 0.06</math></b>  | $3.3 \pm 0.06$                    |
| Ringnorm   | <b><math>1.7 \pm 0.01</math></b>  | <b><math>1.7 \pm 0.02</math></b>  |
| F. Sonar   | <b><math>32.4 \pm 0.18</math></b> | $34.4 \pm 0.20$                   |
| Splice     | $10.9 \pm 0.07$                   | <b><math>10.0 \pm 0.10</math></b> |
| Titanic    | <b><math>22.4 \pm 0.10</math></b> | $23.3 \pm 0.13$                   |
| Waveform   | <b><math>9.9 \pm 0.04</math></b>  | $10.8 \pm 0.06$                   |





# Exemplos de aplicações

- Reconhecimento de dígitos manuscritos:

6 5 4  
6 5

| Técnica          | Taxa erro |
|------------------|-----------|
| K-NN             | 5.7%      |
| RBF              | 4.2%      |
| SVM              | 4.0%      |
| Virtual SVM      | 3.0%      |
| Boosting         | 2.6%      |
| Tangent Distance | 2.5%      |
| Humano           | 2.5%      |

