

# Support Vector Machine

1

## Agenda

- ▶ Motivação;
- ▶ Conceito;
- ▶ SVM com Margem Rígida;
- ▶ SVM com Margem Suave;
- ▶ SVM Não Linear;
- ▶ Funções de Kernel;
- ▶ Classificação Multiclasse;
- ▶ SVM One-Class;
- ▶ Aplicações;
- ▶ Vantagens e Desvantagens;
- ▶ Bibliotecas disponíveis.

2

3

## Motivação

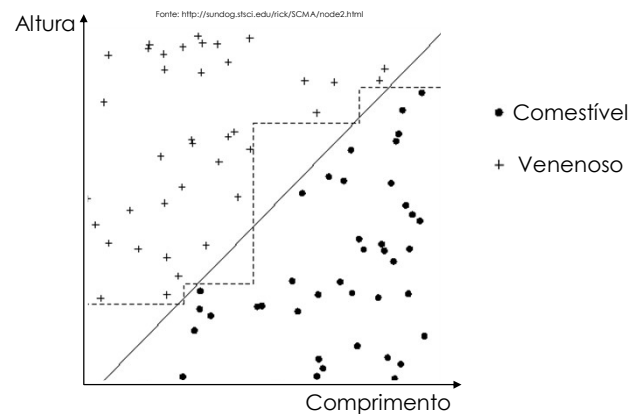
- Considere o seguinte exemplo:
  - Um pesquisador foi a campo e coletou diversos tipos de cogumelos. Ao chegar em seu laboratório, ele mediu a altura e o comprimento de cada cogumelo. Ele também classificou como venenoso e comestível.



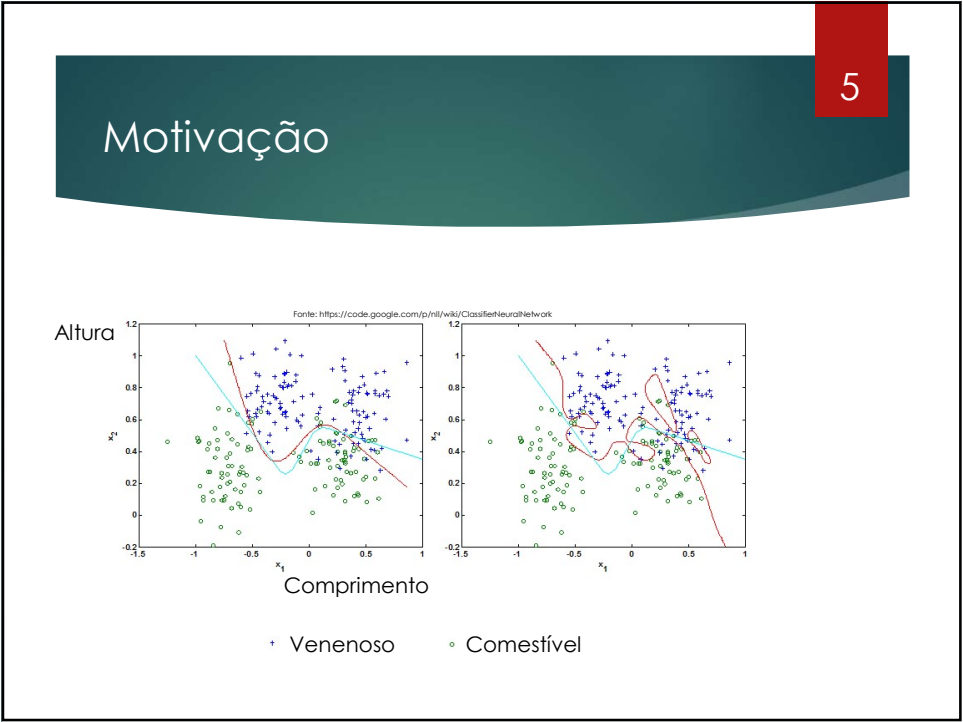
3

4

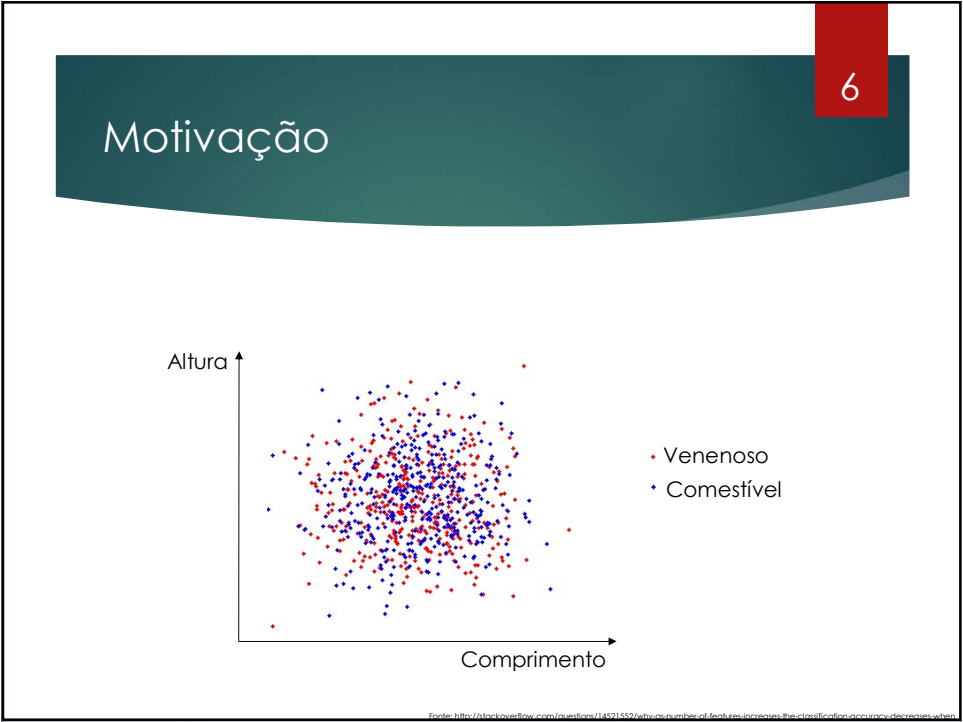
## Motivação



4



5



6

7

Conceito

► SVM é uma técnica de Aprendizagem de Máquina que procura por um hiperplano que divida duas classes com maior margem possível;

Fonte: [http://docs.opencv.org/doc/tutorials/ml/introduction\\_to\\_svm/introduction\\_to\\_svm.html](http://docs.opencv.org/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html)

7

8

Conceito – SVM com Margens Rígidas

► Determinar hiperplano de separação:

1. Dado  $n$  objetos  $m$ -dimensionais  $x = (x_1, x_2, \dots, x_m)$  nas classes  $y = \{+1, -1\}$ ;
2. Para cada conjunto de treinamento linearmente separável  $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$ ;
3. Seja  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$  solução do seguinte problema de otimização com restrições:

$$\text{Maximizar: } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j$$
$$\text{Sujeito a: } \begin{cases} \alpha_i \geq 0, i = 1, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases}$$

3. O par  $(w, b)$  representado a seguir define o hiperplano ótimo:

$$w = \sum_{i=1}^n \alpha_i y_i x_i$$
$$b = \frac{1}{2} \left[ \max_{\{i|y_i = -1\}} (w \cdot x_i) + \min_{\{i|y_i = +1\}} (w \cdot x_i) \right]$$

4. O classificador é expresso por:

$$g(x) = \text{sgn}(f(x)) = \text{sgn} \left( \sum_{i=1}^n \alpha_i y_i x_i \cdot x - b \right) = \begin{cases} +1 & \text{se } \sum_{i=1}^n \alpha_i y_i x_i \cdot x - b > 0 \\ -1 & \text{se } \sum_{i=1}^n \alpha_i y_i x_i \cdot x - b < 0 \end{cases}$$

8

## Conceito – SVM com Margens Rígidas

9

- ▶ Os dados devem ser *linearmente separáveis*;
- ▶ Determinar hiperplano de separação:

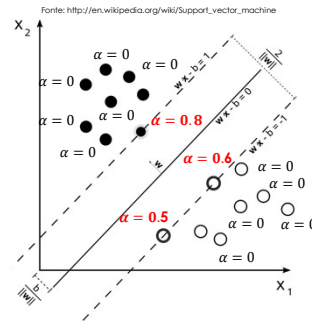
$$\text{Minimizar: } \frac{1}{2} \|w\|^2$$

$$\text{Sujeito a: } y_i(w \cdot x_i + b) \geq 1, \forall x_i$$

O classificador é expresso por:

$$g(x) = \begin{cases} +1 & \text{se } w \cdot x_i + b > 0 \\ -1 & \text{se } w \cdot x_i + b \leq 0 \end{cases}$$

- ▶ Para resolução do problema é utilizado a função Lagrangiana:
  - ▶ Os Multiplicadores de Lagrange não nulos são utilizados como vetores de suporte.



9

## Conceito – SVM com Margens Suaves

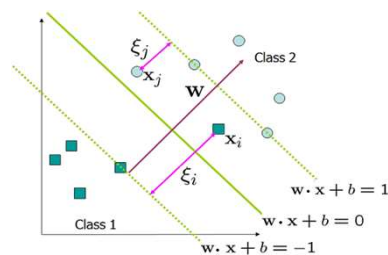
10

- ▶ Quando os dados não são linearmente separáveis, são utilizadas variáveis de relaxamento  $\xi$  que suavizam as restrições impostas na determinação do hiperplano ótimo;
- ▶ O problema de otimização para determinar o hiperplano é:

$$\text{Minimizar: } \frac{1}{2} \|w\|^2 + C \sum \xi_i$$

$$\text{Sujeito a: } \begin{cases} y_i(w \cdot x_i + b) \geq 1 - \xi_i, \forall x_i \\ \xi_i \geq 0 \end{cases}$$

- ▶  $C$  é o único parâmetro de ajuste.



10

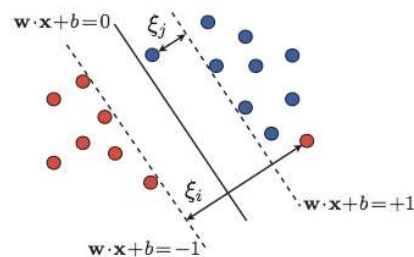
## Conceito – SVM com Margens Suaves

11

- ▶ Quando os dados não são linearmente separáveis, são utilizadas variáveis de relaxamento  $\xi$  que suavizam as restrições impostas na determinação do hiperplano ótimo.
- ▶ O problema de otimização tem sua restrição modificada em relação a SVM com Margens Rígidas:

$$\begin{aligned} \text{Maximizar: } & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j \\ \text{Sujeito a: } & \begin{cases} 0 \leq \alpha_i \leq C, i = 1, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \end{aligned}$$

- ▶  $C$  é o único parâmetro de ajuste.

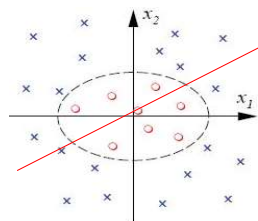


11

## SVM Não Linear

12

- ▶ E quando os dados não podem ser separados linearmente?



Fonte: <http://stackoverflow.com/questions/3997454/help-me-understand-linear-separability-in-a-binary-svm>

- ▶ São utilizadas funções de mapeamento  $\Phi$  para gerar um novo espaço de atributos, denominado *espaço de características*.

12

13

## SVM Não Linear

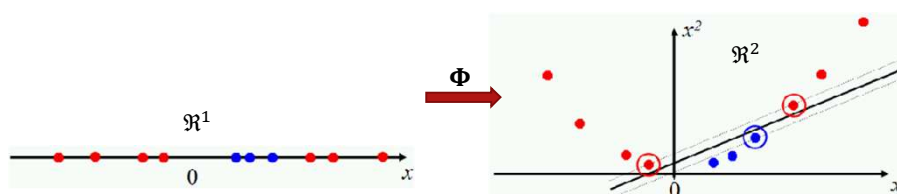
- ▶ A dimensão do novo espaço de características deve ser maior que o original:
- ▶ O Teorema de Cover diz que se aplicarmos uma transformação não-linear  $\Phi$  no espaço de características  $\mathcal{R}^n$  para um espaço de características  $\mathcal{R}^m$ , em que  $m > n$ , nós aumentamos a probabilidade das amostras serem linearmente separáveis em  $\mathcal{R}^m$ .

13

14

## SVM Não Linear

- ▶ Exemplo 1 de transformação não linear:

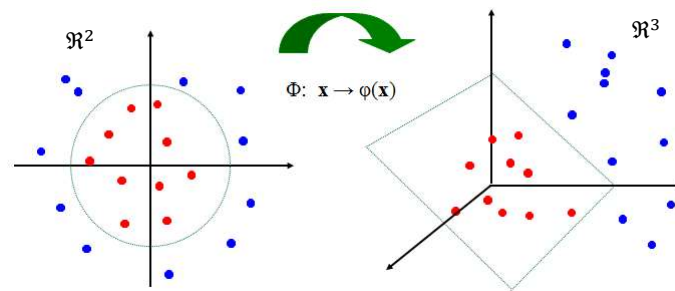
Fonte: <http://mlp.stanford.edu/IR-book/html/htmledition/nonlinear-svms-1.html>

14

15

## SVM Não Linear

► Exemplo 2 de transformação não linear:



Fonte: <http://cg.2010studio.wordpress.com/2012/05/20/sep9d99e75b759a7e6594eaf5e678c75b17e590e917e9b675b675e65a959f7e5999a8-non-linear-svm/>

15

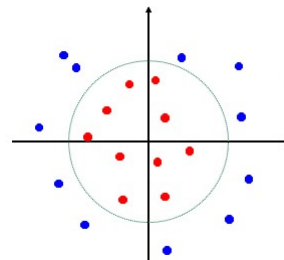
16

## SVM Não Linear

► Hiperplano de separação:

$$\text{Minimizar: } \frac{1}{2} \|w\|^2 + C \sum \xi_i$$

$$\text{Sujeito a: } \begin{cases} y_i(w \cdot \Phi(x_i) + b) \geq 1 - \xi_i, \forall x_i \\ \xi_i \geq 0 \end{cases}$$



Fonte: <http://cg.2010studio.wordpress.com/2012/05/20/sep9d99e75b759a7e6594eaf5e678c75b17e590e917e9b675b675e65a959f7e5999a8-non-linear-svm/>

16



17

## SVM Não Linear

### ► Determinar hiperplano de separação:

1. Dado  $n$  objetos  $m$ -dimensionais  $x = (x_1, x_2, \dots, x_m)$  nas classes  $y_i \in \{+1, -1\}$ ;
2. Para cada conjunto de treinamento linearmente separável  $\Phi(S) = \{(\Phi(x_1), y_1), \dots, (\Phi(x_n), y_n)\}$ ;
3. Seja  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$  solução do seguinte problema de otimização com restrições:

$$\text{Maximizar: } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \Phi(x_i) \cdot \Phi(x_j)$$

$$\text{Sujeito a: } \begin{cases} \alpha_i \geq 0, i = 1, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases}$$

3. O par  $(w, b)$  representado a seguir define o hiperplano ótimo:

$$w = \sum_{i=1}^n \alpha_i y_i \Phi(x_i)$$

$$b = \frac{1}{2} \left[ \max_{\{i|y_i = -1\}} (w \cdot \Phi(x_i)) + \min_{\{i|y_i = +1\}} (w \cdot \Phi(x_i)) \right]$$

4. O classificador é expresso por:

$$g(x) = \text{sgn}(f(x)) = \text{sgn} \left( \sum_{i=1}^n \alpha_i y_i \Phi(x_i) \cdot \Phi(x) - b \right) = \begin{cases} +1 & \text{se } \sum_{i=1}^n \alpha_i y_i \Phi(x_i) \cdot \Phi(x) - b > 0 \\ -1 & \text{se } \sum_{i=1}^n \alpha_i y_i \Phi(x_i) \cdot \Phi(x) - b < 0 \end{cases}$$

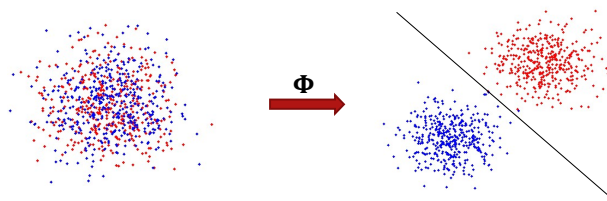
17

18

## SVM Não Linear

- Após mapeando para um novo espaço de características espera-se que os dados possam ser separados com uma SVM com Margem Suave.

- Problema dos cogumelos:



Fonte: <http://stackoverflow.com/questions/14591552/why-as-number-of-features-increases-the-classification-accuracy-decreases-when>

18

19

Funções de Kernels

►

 Funções de Kernels são utilizadas para gerar o novo espaço de características;

►

 A utilidade dos Kernels está na simplicidade de cálculo e na capacidade de representar espaços abstratos;

►

 Kernels mais utilizados na literatura:

Kernel	Função	Parâmetros de ajuste
Polinomial	$(x_i^T \cdot x_j + 1)^\rho$	$\rho$
Gaussiano (RBF)	$e^{-\frac{1}{2\sigma^2}\ x_i - x_j\ ^2}$	$\sigma^2$
Sigmoidal	$\tanh(\beta_0 x_i \cdot x_j + \beta_1)$	$\beta_0$ e $\beta_1$

20

SVM Multiclasse

►

 Diversos problemas envolvem mais de duas classes, como classificar esses problemas?

The scatter plot displays three distinct clusters of data points representing different species of Iris flowers. The x-axis is labeled 'Petal Length (cm)' and ranges from 1 to 7. The y-axis is labeled 'Petal Width (cm)' and ranges from 0 to 2.5. A legend in the top-left corner identifies the clusters: 'setosa' (red dots), 'versicolor' (green dots), and 'virginica' (blue dots). The 'setosa' cluster is located at the bottom-left, with low petal length and width. The 'versicolor' cluster is in the middle, showing a wider range of petal lengths and widths. The 'virginica' cluster is at the top-right, characterized by higher petal lengths and widths. The clusters are well-separated, indicating good linear separability in this 2D space.

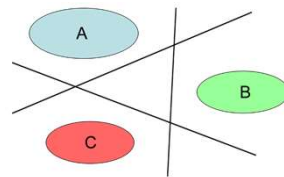
Fonte: [http://www.mathworks.com/help/stats/classificationsvm\\_figposterfor.html](http://www.mathworks.com/help/stats/classificationsvm_figposterfor.html)

21

## SVM Multiclasse

### ► Um-Contra-Todos:

- Gerar  $K$  SVMs, onde  $K$  é o número de classes. Para cada SVM gerada, considerar uma classe como positiva e demais negativas;



Fonte: <http://courses.media.mit.edu/2006fall/mas622j/Project/aiien-project/>

- A saída é dada para o classificador que apresente saída máxima.

21

22

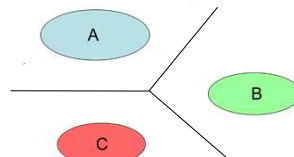
## SVM Multiclasse

### ► Todos-Contra-Todos:

- Consiste em comparar as classes duas a duas, sendo necessário  $\frac{K(K-1)}{2}$  SVMs, onde  $K$  é o número de classes;

Para  $K = 3$ :

$$\frac{3(3-1)}{2} = 3 * \frac{2}{2} = 3 \text{ SVMs}$$



Fonte: <http://courses.media.mit.edu/2006fall/mas622j/Project/aiien-project/>

Classificações
A+B
B+C
A+C

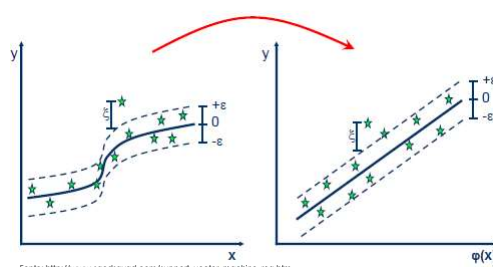
- A saída é dada para a classe que recebeu mais indicações;
- Deve-se considerar condições de empate.

22

23

## SVM Regressão

- Consiste na determinação de valores contínuos;
- O parâmetro de ajuste  $\epsilon$  define a margem de tolerância para realizar o mapeamento da função;

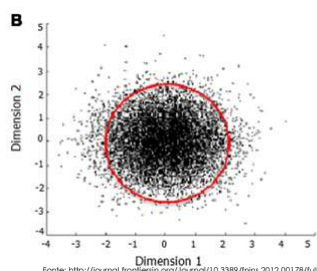


23

24

## SVM One-Class

- Consiste em encontrar uma hiperesfera que concentre grande parte dos exemplos de treinamento;
- O treinamento é utilizado apenas com exemplos positivos.

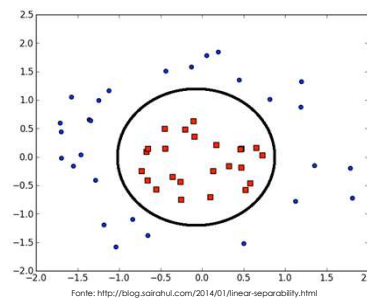


24

25

## SVM One-Class

- ▶ Quando é utilizada SVM One-Class?
  - ▶ Remover outliers do conjunto de dados;
  - ▶ O problemas contém um número muito reduzido de uma classe (Ex: cogumelos venenosos em extinção).

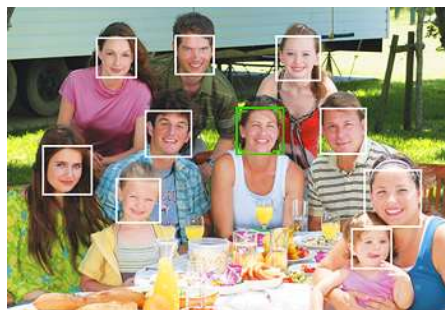


25

26

## Aplicações

- ▶ Detecção de Faces;



26

27

Aplicações

► Reconhecimento de Expressões Faciais:

			
DISGUST	FEAR	ANGER	CONTEMPT
			
SADNESS	SURPRISE	HAPPINESS	?

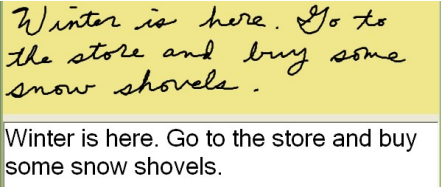
Fonte: <http://www.humansensing.cs.cmu.edu/projects/autofacial.html>

27

28

Aplicações

► Identificação de manuscritos:


Winter is here. Go to the store and buy some snow shovels.

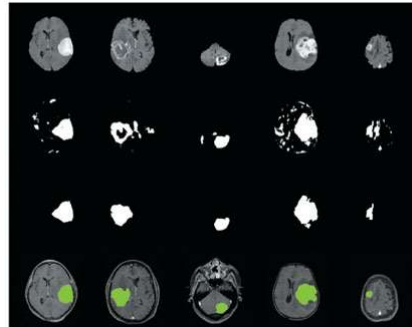
Fonte: [http://www.jonpythor.info/index.php?file=Basic\\_Handwriting\\_Recognition](http://www.jonpythor.info/index.php?file=Basic_Handwriting_Recognition)

28

29

## Aplicações

### ► Detecção de tumores:



Fonte: [http://www.scielo.br/scielo.php?pid=S1679-4508201200200008&script=sci\\_arttext&lng=eng](http://www.scielo.br/scielo.php?pid=S1679-4508201200200008&script=sci_arttext&lng=eng)

29

30

## Vantagens e Desvantagens

### ► Vantagens:

- Consegue lidar bem com grandes conjuntos de exemplos;
- Atinge bom performance com dados de alta dimensão;
- A classificação é rápida.

### ► Desvantagens:

- Definir um bom Kernel;
- Treinamento pode ser longo dependendo do número de exemplos e dimensão dos dados;
- Difícil interpretação do modelo.

30

31

## Bibliotecas disponíveis

- ▶ SMO (C++) – contém Kernel Linear e RBF;
- ▶ SVMLight (Multiplataforma) – contém vários Kernels e pode ser utilizada em problemas de classificação, binários;
- ▶ \*LibSVM (C++, JAVA e MATLAB) – possui código fonte disponível. Pode ser utilizada em problemas de classificação Multiclasse e Regressão.

31

32

## Referências

- ▶ A. C. Loreira e A. C. P. L. F. de Carvalho. ***Uma Introdução às Support Vector Machines***. Revista de Informática Teórica e Aplicada. Volume 14, número 2, 2007.
- ▶ Eduardo Akira Kinto. ***Otimização e análise das máquinas de vetores de suporte aplicadas à classificação de documentos***. USP, 2011.
- ▶ Alessandro L. Koerich. ***Aprendizagem de Máquina: Máquinas de Vetor de Suporte***. UFPR.
- ▶ Edirlei Soares de Lima. ***Inteligência Artificial: Suport Vector Machine***. PUCRIO, 2012.

32