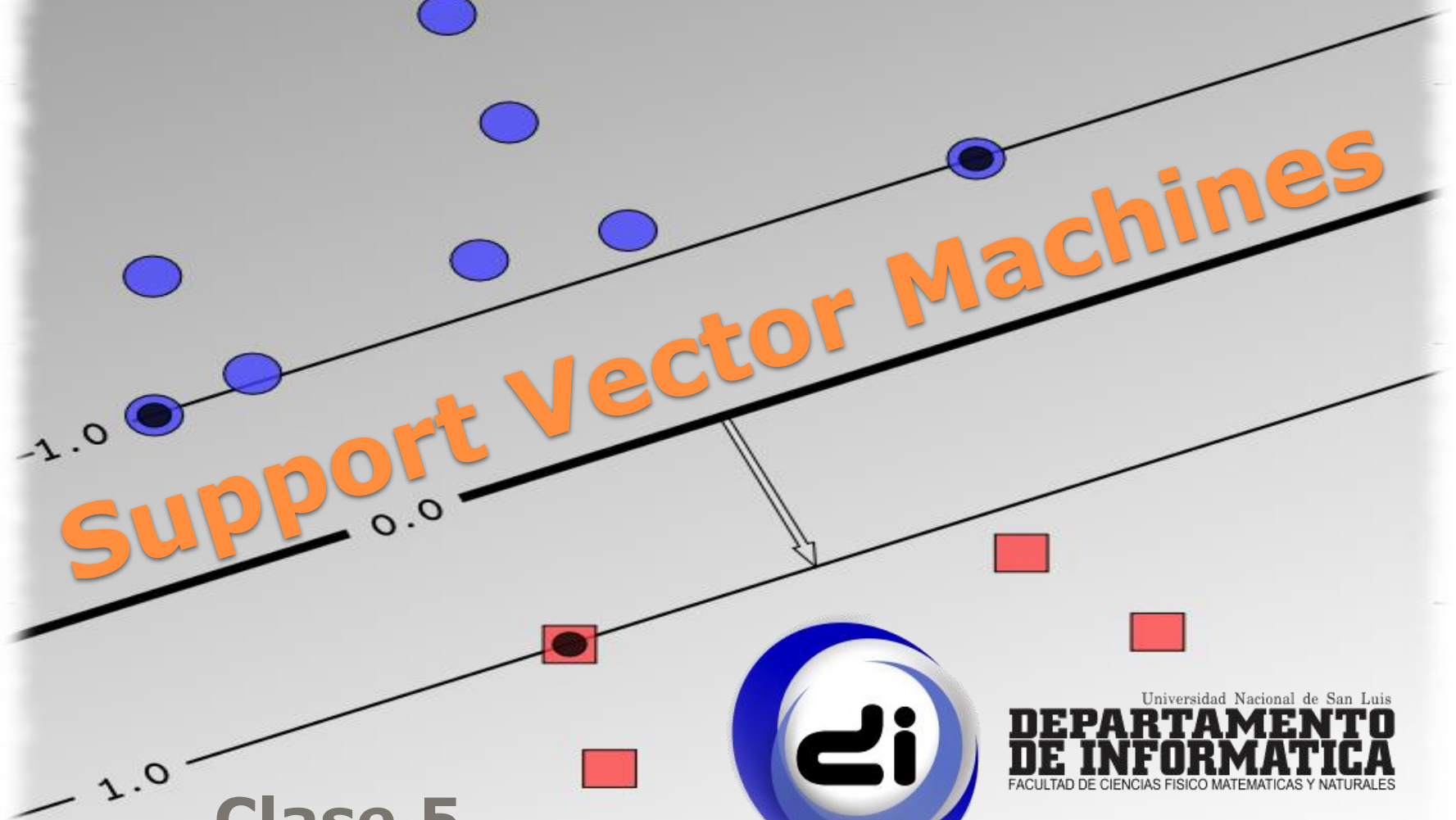


Curso: Minería de Datos (22 de noviembre de 2018)



Universidad Nacional de San Luis
**DEPARTAMENTO
DE INFORMATICA**
FACULTAD DE CIENCIAS FISICO MATEMATICAS Y NATURALES

Outline*

- A brief history of SVM
- Linear classifiers in a nutshell
 - Linear separable
 - Nonlinear separable
- Creating nonlinear classifiers: the kernel trick
- SVM kernel functions
- SVM extensions
- Software
- Conclusions

* Credits: These slides are a modified version of "A Simple Introduction to Support Vector Machines" by Martin Law. Lecture for CSE 802, Department of Computer Science and Engineering, Michigan State University.

History of SVM

- SVM was first introduced in 1992.^[1]
- SVM is related to statistical learning theory.^[2]
- SVM became popular because of its success in handwritten digit recognition.^[3]
- SVM is now regarded as an important example of “kernel methods”,^[4] a key area in machine learning.

[1] B.E. Boser et al. A Training Algorithm for Optimal Margin Classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory 5 144-152, Pittsburgh, 1992.

[2] V. Vapnik. The Nature of Statistical Learning Theory. 2nd edition, Springer, 1999.

[3] L. Bottou et al. Comparison of classifier methods: a case study in handwritten digit recognition. Proceedings of the 12th IAPR International Conference on Pattern Recognition, vol. 2, pp. 77-82, 1994.

[4] <http://www.kernel-machines.org/publications/pdfs/0701907.pdf>

Outline

- A brief history of SVM
- Linear classifiers in a nutshell
 - Linear separable
 - Nonlinear separable
- Creating nonlinear classifiers: the kernel trick
- SVM kernel functions
- SVM extensions
- Software
- Conclusions

Linear classifiers in a nutshell

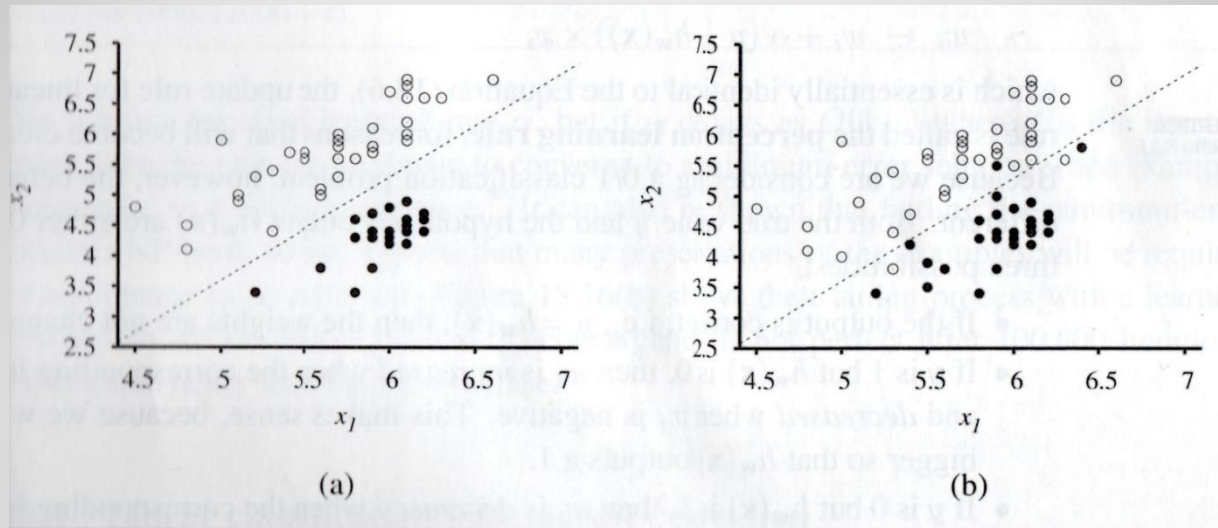
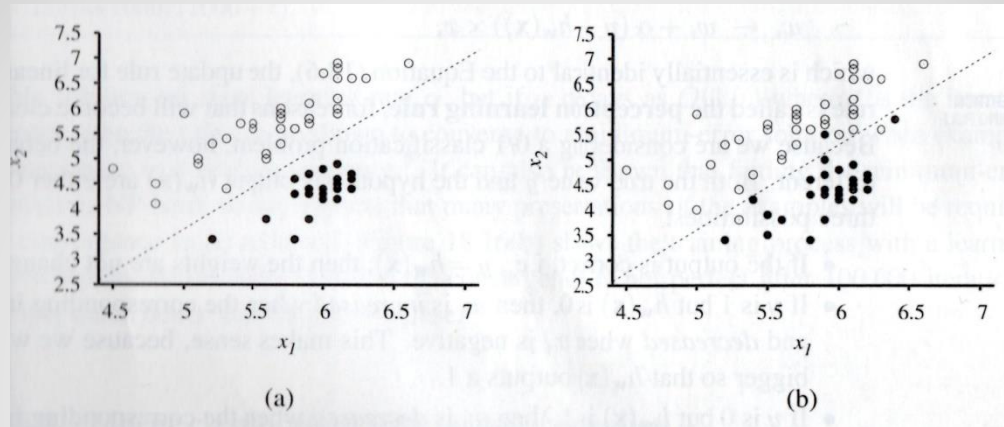


Figure 18.15 (a) Plot of two seismic data parameters, body wave magnitude x_1 and surface wave magnitude x_2 , for earthquakes (white circles) and nuclear explosions (black circles) occurring between 1982 and 1990 in Asia and the Middle East (Kebeasy *et al.*, 1998). Also shown is a decision boundary between the classes. (b) The same domain with more data points. The earthquakes and explosions are no longer linearly separable. [5]

- A **decision boundary** is a line (or a surface, in higher dimensions) that separates the two classes.

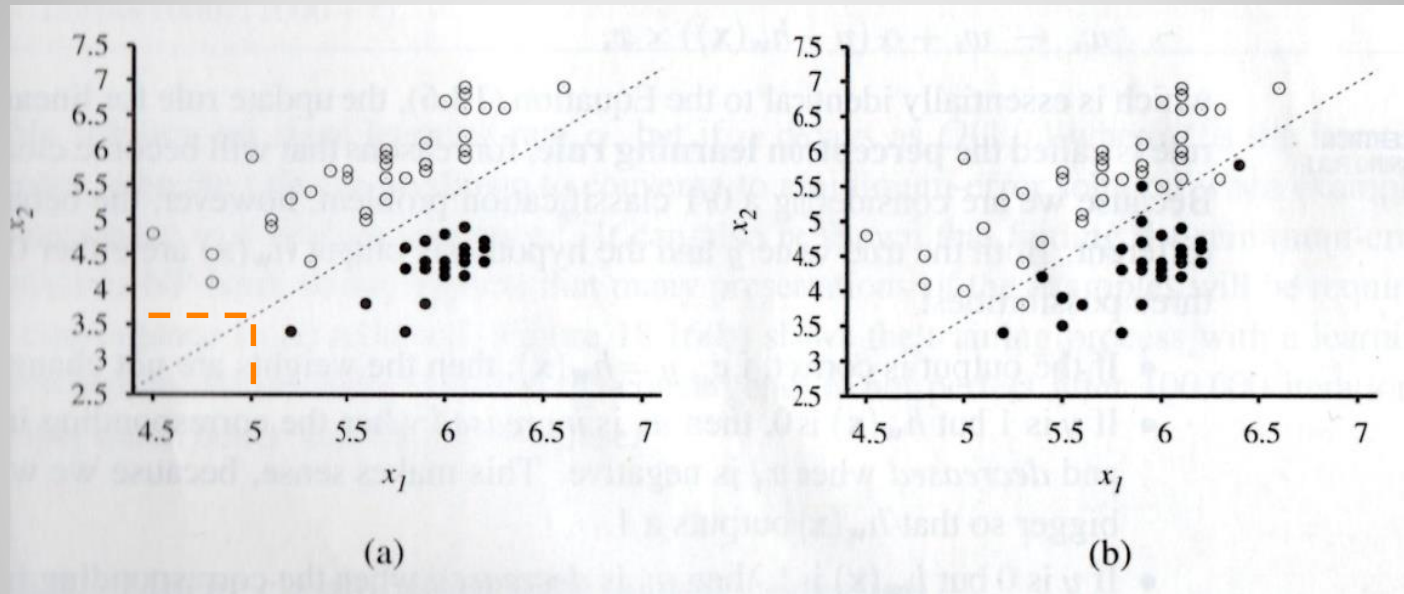
Linear classifiers in a nutshell



- A linear decision boundary is called a **linear separator** and data that admit such a separator are called **linearly separable**. The linear separator in this case is:

$$-4.9 + 1.7x_1 - x_2 = 0$$

Linear classifiers in a nutshell

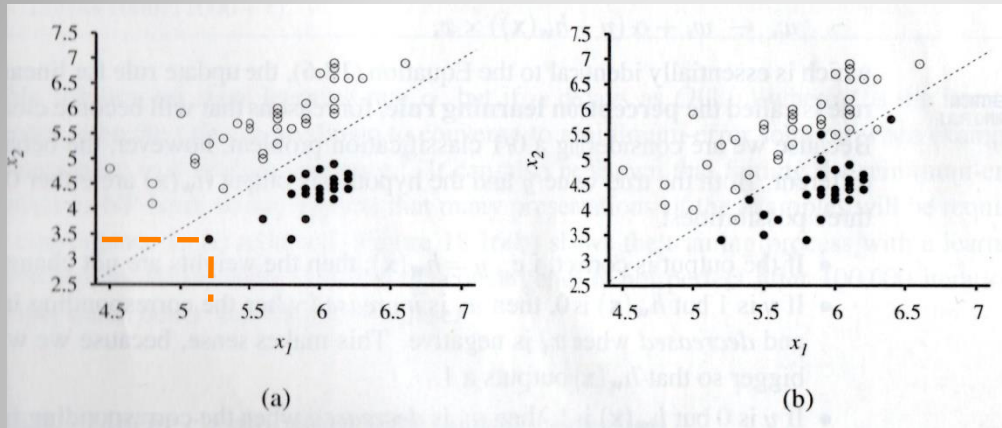


5

3.6

$$-4.9 + 1.7x_1 - x_2 = 0$$

Linear classifiers in a nutshell

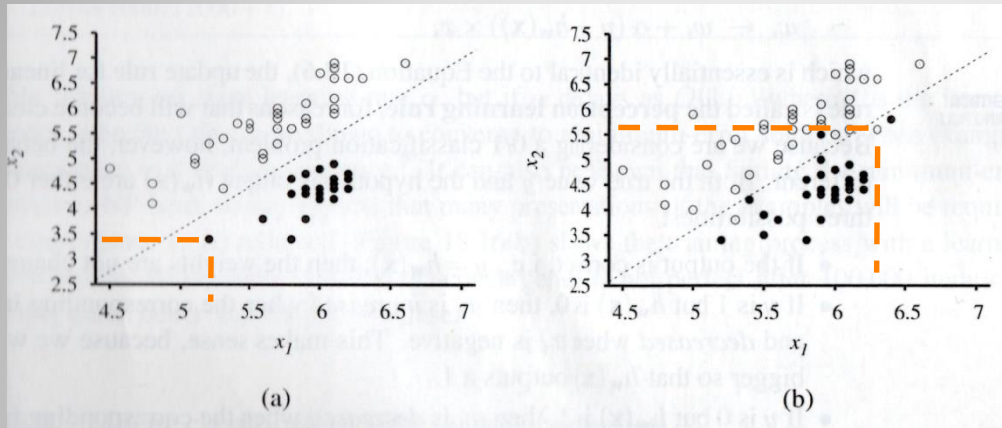


$$-4.9 + 1.7x_1 - x_2 = 0$$

- Nuclear explosions classification: $-4.9 + 1.7x_1 - x_2 > 0$

$$\underbrace{\begin{matrix} \uparrow & \uparrow \\ 5.25 & 3.4 \end{matrix}}_{0.625}$$

Linear classifiers in a nutshell



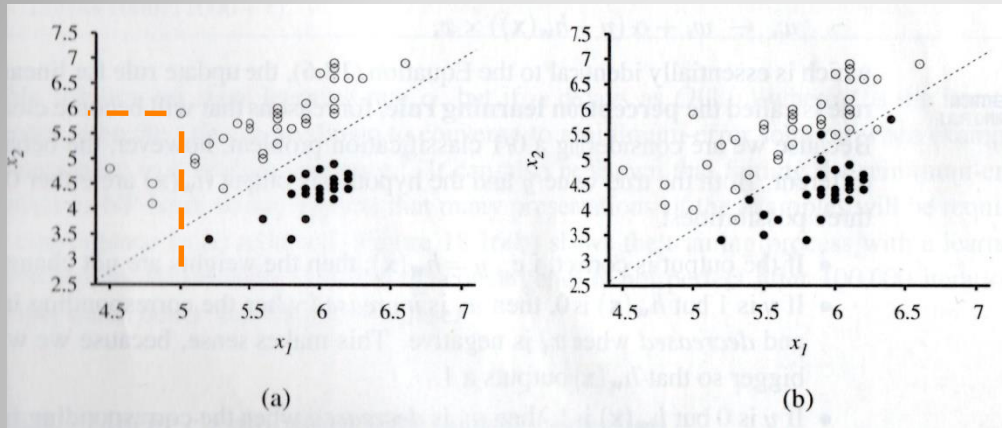
$$-4.9 + 1.7x_1 - x_2 = 0$$

- Nuclear explosions classification: $-4.9 + 1.7x_1 - x_2 > 0$

$$-4.9 + 1.7 \underbrace{(6.35)}_{x_1} - \underbrace{5.6}_{x_2} = 0.295$$

$$\underbrace{5.25 \quad 3.4}_{0.625}$$

Linear classifiers in a nutshell

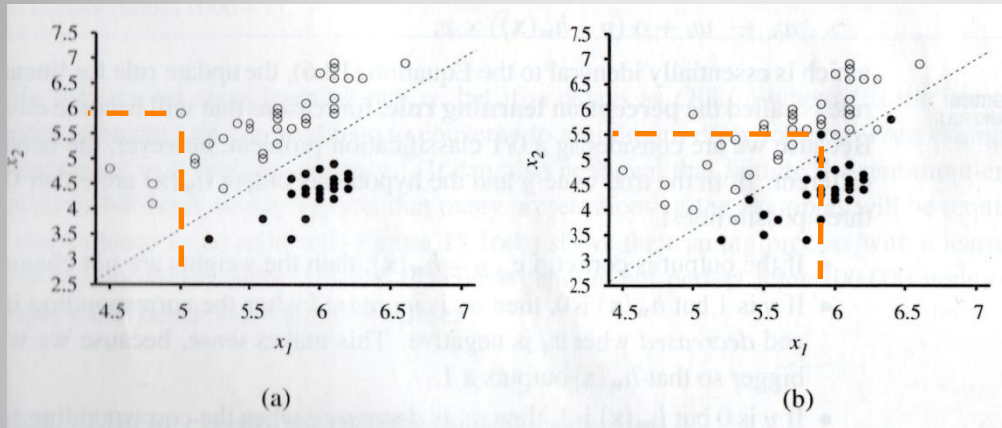


$$-4.9 + 1.7x_1 - x_2 = 0$$

- Nuclear explosions classification: $-4.9 + 1.7x_1 - x_2 > 0$
- Earthquakes classification: $-4.9 + 1.7x_1 - x_2 < 0$

$$\underbrace{\begin{matrix} \uparrow & \uparrow \\ 5 & 5.9 \end{matrix}}_{-2.3}$$

Linear classifiers in a nutshell



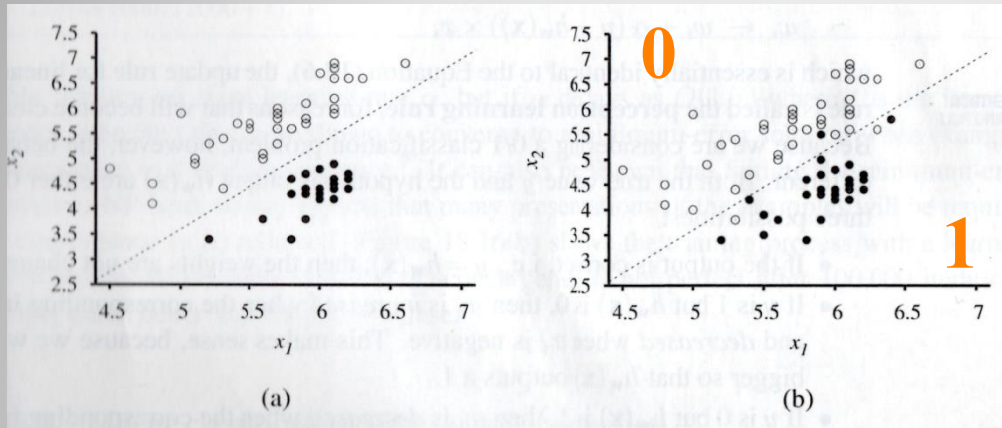
$$-4.9 + 1.7x_1 - x_2 = 0$$

- Nuclear explosions classification: $-4.9 + 1.7x_1 - x_2 > 0$
- Earthquakes classification: $-4.9 + 1.7x_1 - x_2 < 0$

$$-4.9 + 1.7 \underbrace{(5.85)}_{x_1} - \underbrace{5.5}_{x_2} = -0.455$$

$$\underbrace{\begin{matrix} \uparrow & \uparrow \\ 5 & 5.9 \end{matrix}}_{-2.3}$$

Linear classifiers in a nutshell



$$-4.9 + 1.7x_1 - x_2 = 0$$

- Nuclear explosions classification: $-4.9 + 1.7x_1 - x_2 > 0$
- Earthquakes classification: $-4.9 + 1.7x_1 - x_2 < 0$
- Given these training data, the classification task consists of learning a hypothesis that will take new (x_1, x_2) points and return either 0 for earthquakes or 1 for explosions.

Linear classifiers in a nutshell

- Using the convention of a dummy input $X_0=1$, we can write the classification hypothesis as:

$$h_{\mathbf{w}}(\mathbf{x}) = 1 \quad \text{if } \mathbf{w} \cdot \mathbf{x} \geq 0 \quad \text{and } 0 \text{ otherwise}$$

$$\mathbf{w} \cdot \mathbf{x} = (w_1, w_2, \dots, w_n) \cdot (x_1, x_2, \dots, x_n) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n$$

Linear classifiers in a nutshell

- Using the convention of a dummy input $X_0=1$, we can write the classification hypothesis as:

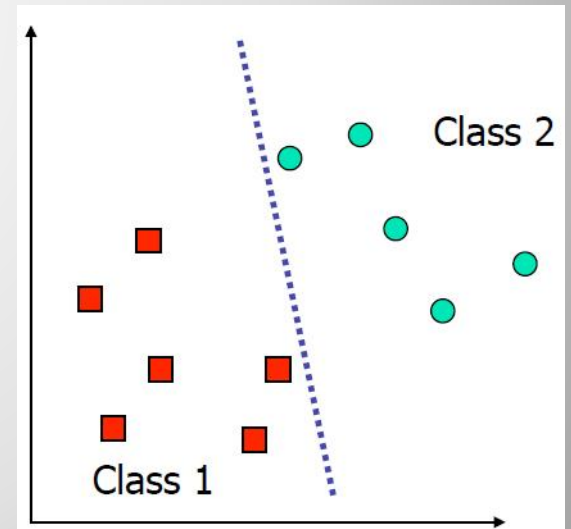
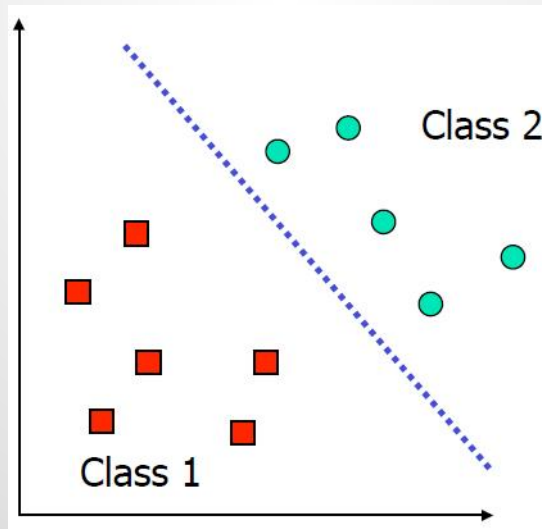
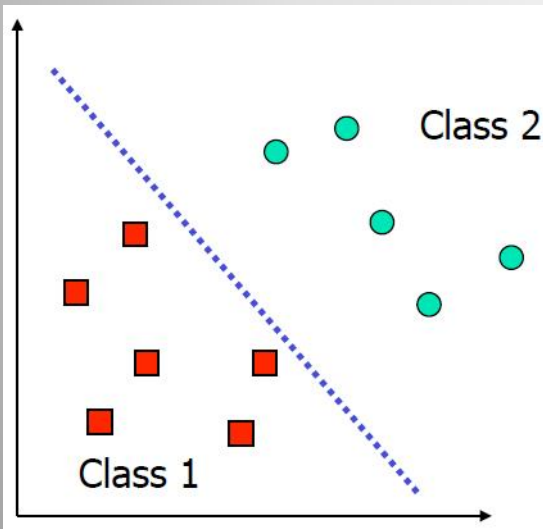
$$h_{\mathbf{w}}(\mathbf{x}) = 1 \text{ if } \mathbf{w} \cdot \mathbf{x} \geq 0 \text{ and } 0 \text{ otherwise}$$

- Traditionally SVMs use the convention that class labels are 1 and -1, instead of the 1 and 0 we have been using so far.
- The linear separator is thus defined as:

$$\{\mathbf{x} : \mathbf{w} \cdot \mathbf{x} + b = 0\}$$

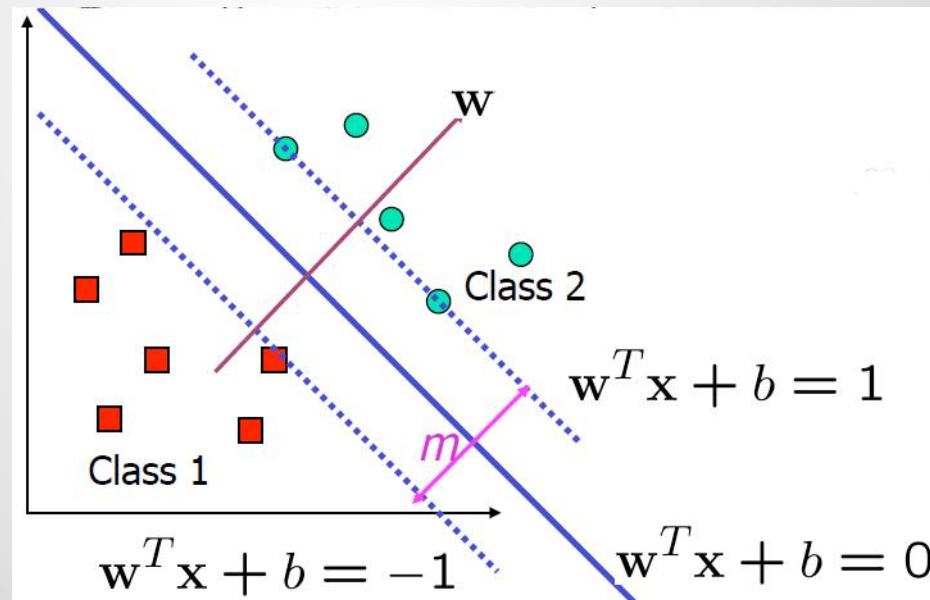
What is a good Decision Boundary?

- Consider a two-class, linearly separable classification problem
- Many decision boundaries!
- Are all decision boundaries equally good?



Large-margin Decision Boundary

- The decision boundary should be as far away as possible from the data of both classes.

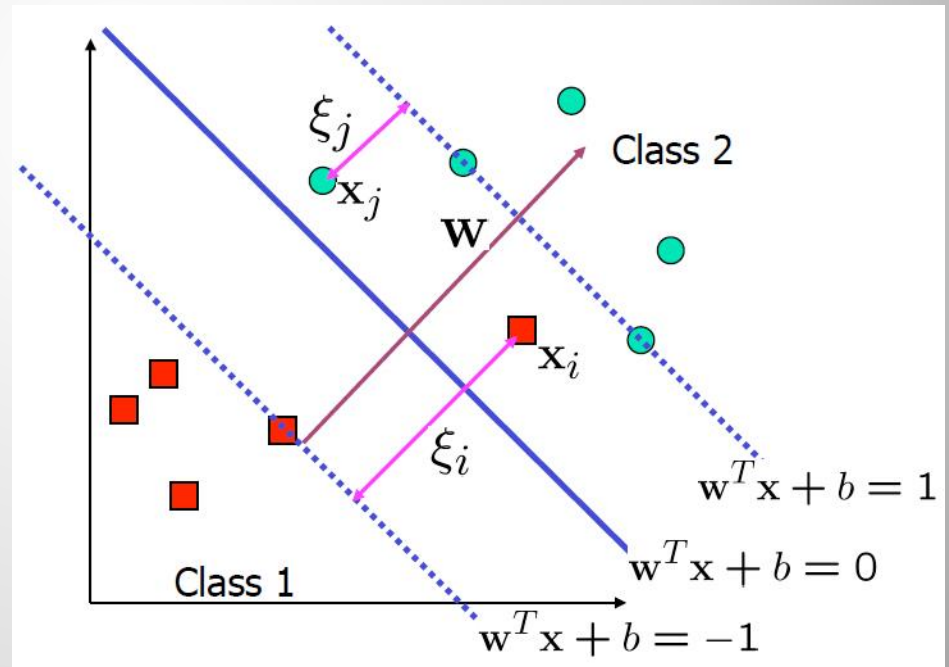


Outline

- A brief history of SVM
- Linear classifiers in a nutshell
 - Linear separable
 - Nonlinear separable
- Creating nonlinear classifiers: the kernel trick
- SVM kernel functions
- SVM extensions
- Software
- Conclusions

Non-linearly Separable Problems

- We allow “error” ξ_i in classification; it is based on the output of the discriminant function $\mathbf{w}^T \mathbf{x} + b$



How do we find the decision boundary?

The optimization problem becomes^[6]

$$\begin{aligned} &\text{Minimize } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \\ &\text{subject to } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \end{aligned}$$

- C : tradeoff parameter between error and margin.

[6] Cortes, C. and Vapnik, V.N. (1995) Support vector networks. Machine Learning 20, 273–297.

The Optimization Problem

The dual of this new constrained optimization problem is

$$\max. W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

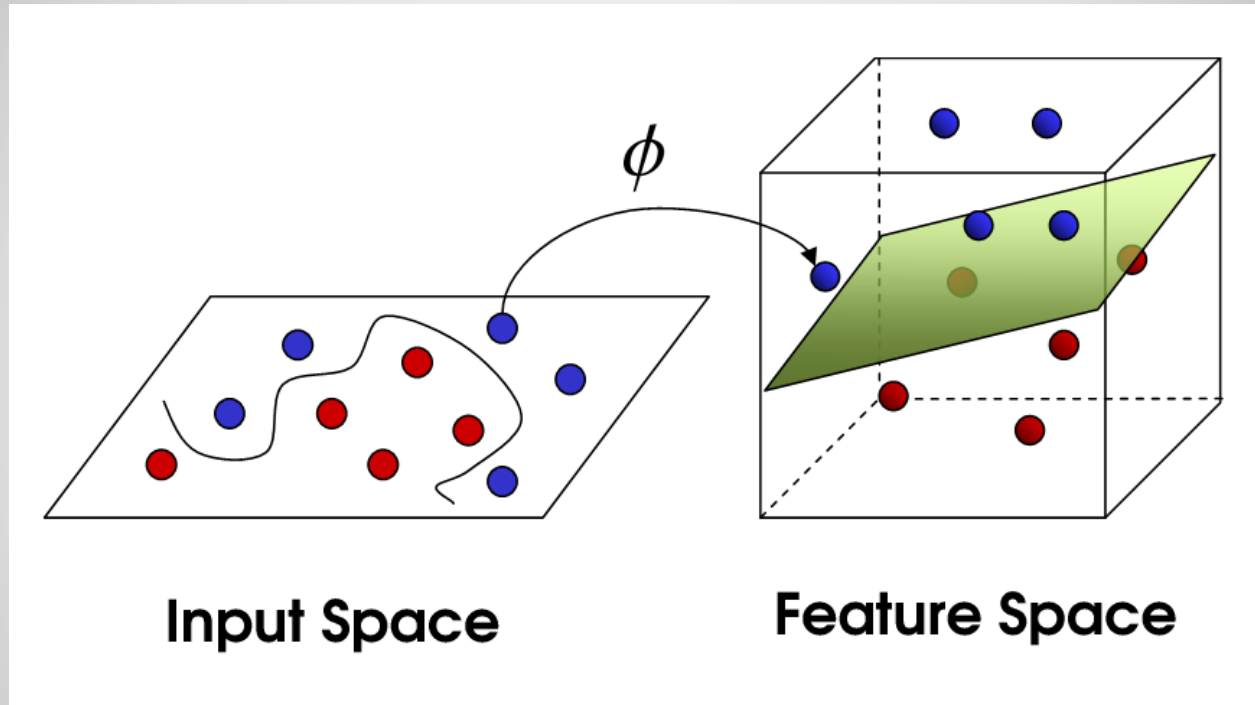
$$\text{subject to } C \geq \alpha_i \geq 0, \sum_{i=1}^n \alpha_i y_i = 0$$

w is recovered as $\mathbf{w} = \sum_{j=1}^n \alpha_{t_j} y_{t_j} \mathbf{x}_{t_j}$

Extension to Non-linear Decision Boundary

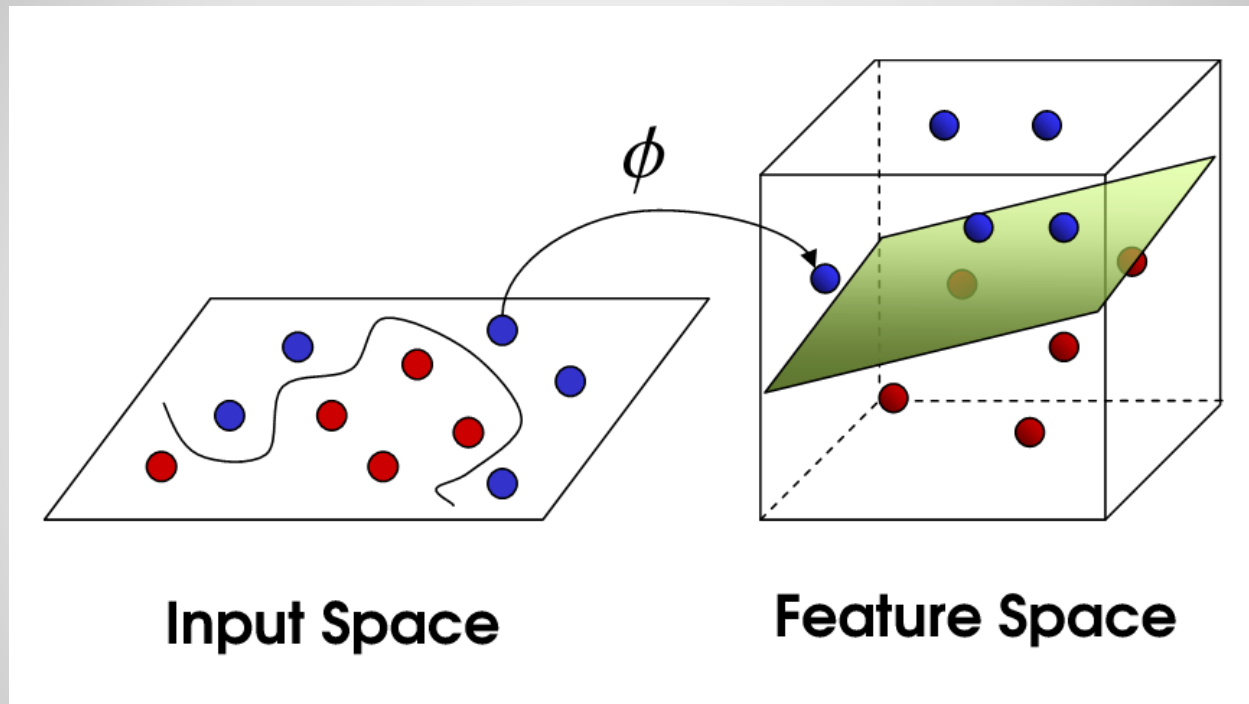
- So far, we have only considered large-margin classifier with a linear decision boundary.
- How to generalize it to become nonlinear?
- Key idea: transform \mathbf{x}_i to a higher dimensional space to “make life easier”.
 - Input space: the space the points \mathbf{x}_i are located
 - Feature space: the space of $\phi(\mathbf{x}_i)$ after transformation.

Transforming the Data



- Why transform?
 - Linear operation in the feature space is equivalent to nonlinear operation in input space.

Transforming the Data



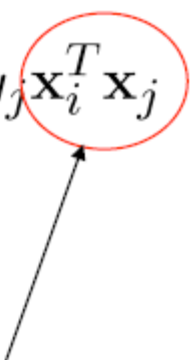
- Computation in the feature space can be costly because it is high dimensional.
 - The feature space is typically infinite-dimensional!
- The kernel trick comes to rescue.

Outline

- A brief history of SVM
- Linear classifiers in a nutshell
 - Linear separable
 - Nonlinear separable
- Creating nonlinear classifiers: the kernel trick
- SVM kernel functions
- SVM extensions
- Software
- Conclusions

The Kernel Trick

Recall the SVM optimization problem

$$\begin{aligned} \max. \quad W(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{subject to } C &\geq \alpha_i \geq 0, \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned}$$


The data points only appear as **inner product**

As long as we can calculate the inner product in the feature space, we do not need the mapping explicitly

Many common geometric operations (angles, distances) can be expressed by inner products

Define the kernel function K by

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$$

Modification Due to Kernel Function

Original

$$\begin{aligned} \max. \quad W(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{subject to } C &\geq \alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned}$$

With kernel function

$$\begin{aligned} \max. \quad W(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to } C &\geq \alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned}$$

Modification Due to Kernel Function

Original

$$\begin{aligned} \max. \quad W(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{subject to } C &\geq \alpha_i \geq 0, \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned}$$

With kernel function

$$\begin{aligned} \max. \quad W(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to } C &\geq \alpha_i \geq 0, \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned}$$

Characteristics of the solution

- Many of the α_i are zero
 - \mathbf{w} is a linear combination of a small number of data points
- \mathbf{x}_i with non-zero α_i are called support vectors (SV)
 - The decision boundary is determined only by the SV
 - Let t_j ($j=1, \dots, s$) be the indices of the s support vectors. We can write
$$\mathbf{w} = \sum_{j=1}^s \alpha_{t_j} y_{t_j} \mathbf{x}_{t_j}$$
- For testing with a new data \mathbf{z}
 - Compute $\mathbf{w}^T \mathbf{z} + b = \sum_{j=1}^s \alpha_{t_j} y_{t_j} (\mathbf{x}_{t_j}^T \mathbf{z}) + b$ and classify \mathbf{z} as class 1 if the sum is positive, and class 2 otherwise

Outline

- A brief history of SVM
- Linear classifiers in a nutshell
 - Linear separable
 - Nonlinear separable
- Creating nonlinear classifiers: the kernel trick
- SVM kernel functions
- SVM extensions
- Software
- Conclusions

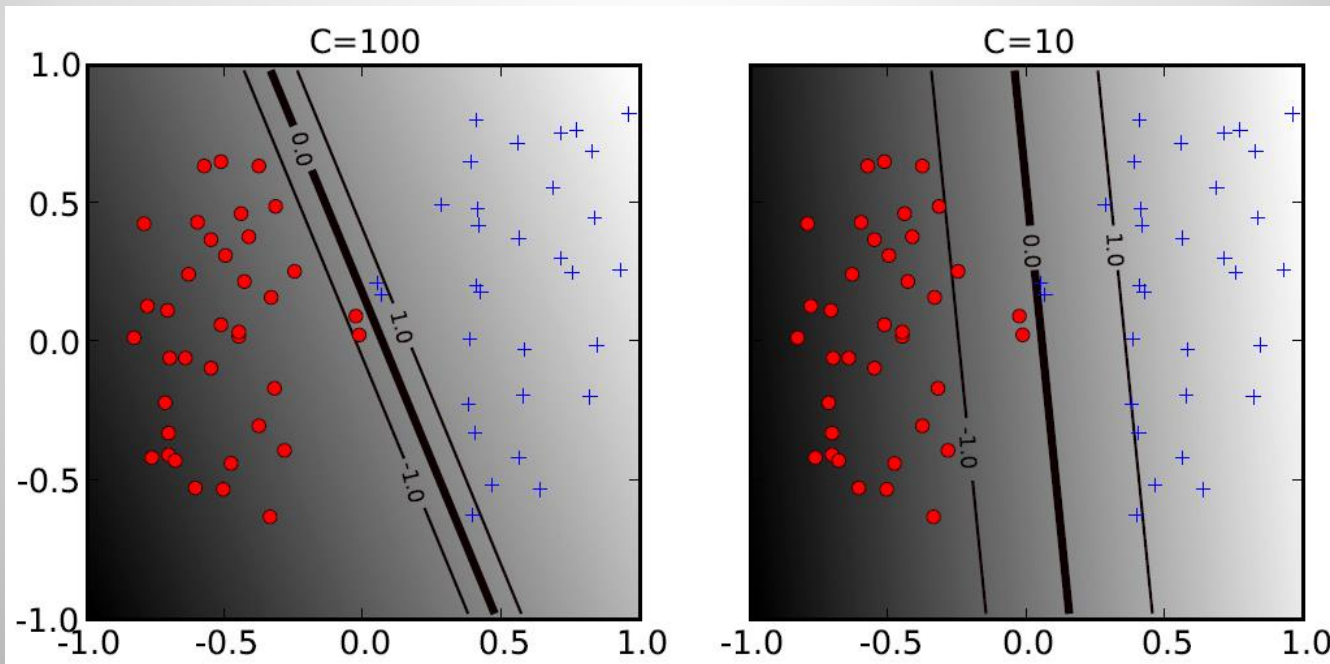
SVM four basic kernels

- linear: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$.
- polynomial: $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d, \gamma > 0$.
- radial basis function (RBF): $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \gamma > 0$.
- sigmoid: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$.

Here, γ , r , and d are kernel parameters. [7]

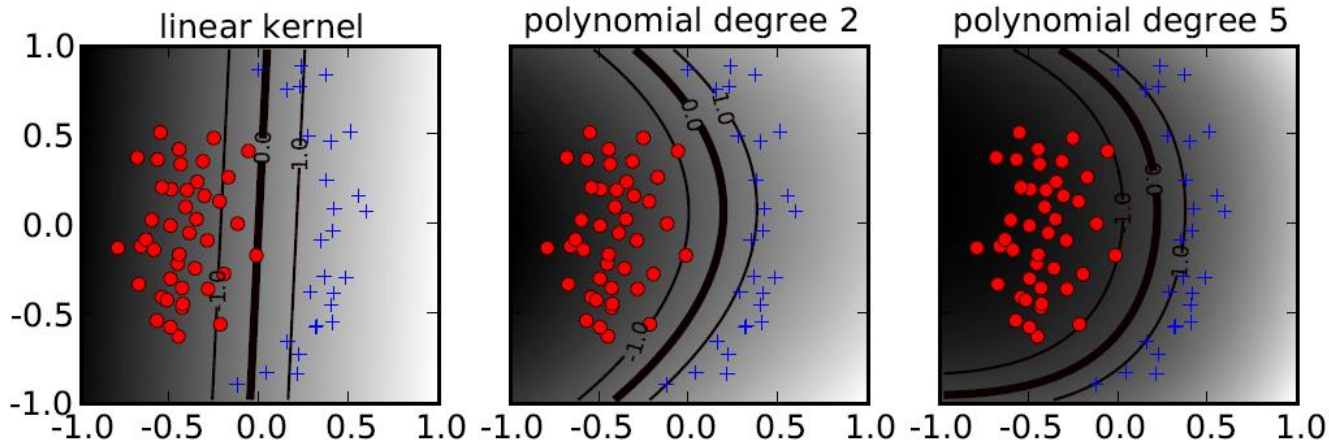
Linear kernel

- C is not a parameter of this kernel, but its values are usually tune up to improve the classifier performance.

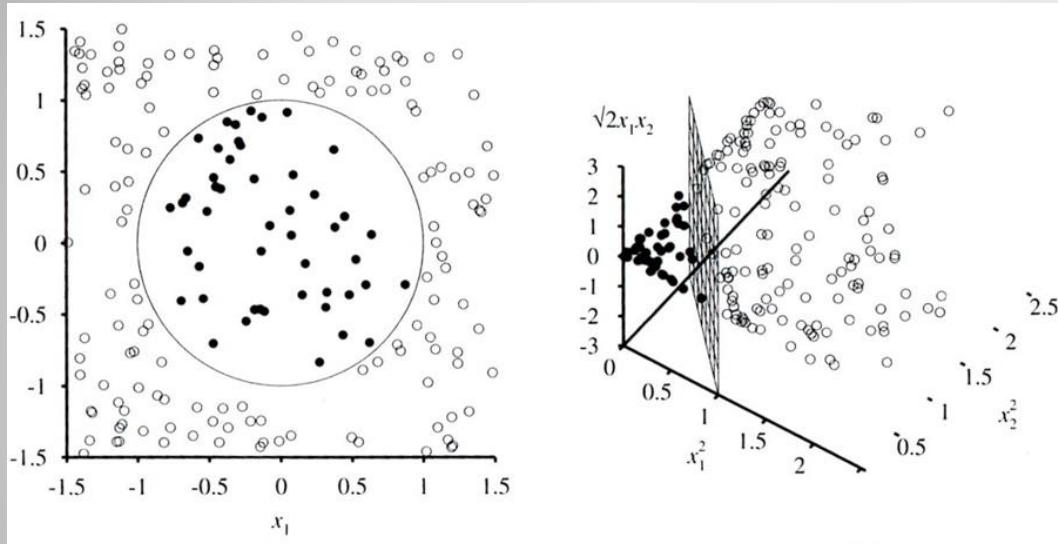


Polynomial kernel

- Parameter d increases the features space:
 - Number of monomials with sum of exponents lower or equal to d .
- Higher degree polynomial kernels allow a more flexible decision boundary, but may lead to overfitting.



Polynomial kernel



$$f_1 = x_1^2$$

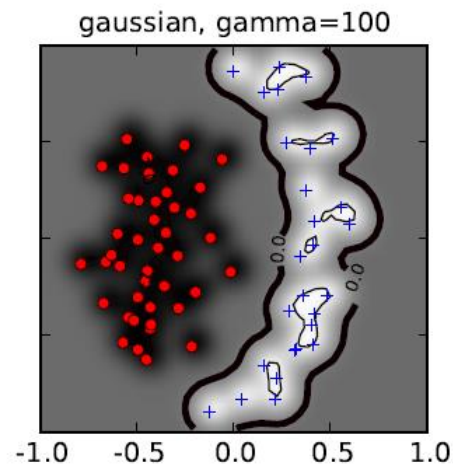
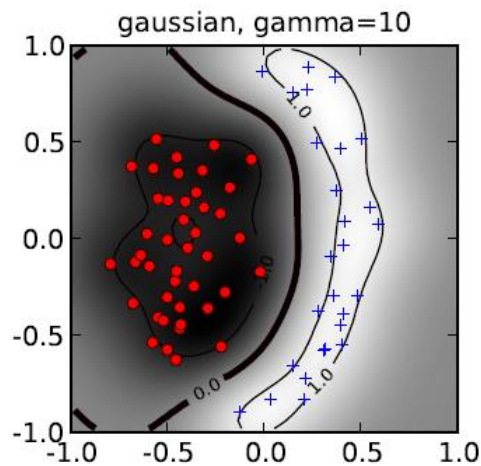
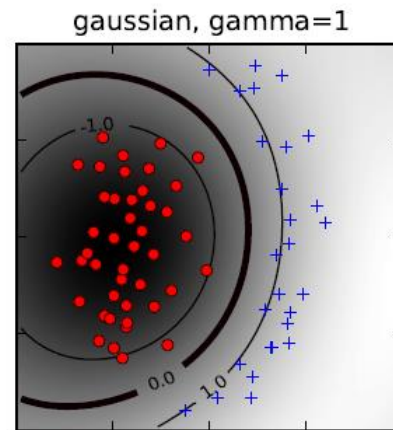
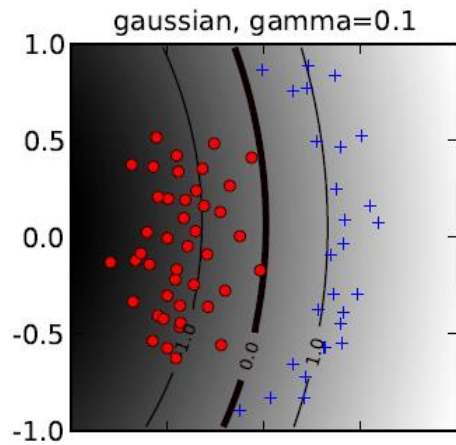
$$f_2 = x_2^2$$

$$f_3 = \sqrt{2}x_1x_2$$

- The numbers of monomials is calculated as:

$$\binom{d+m-1}{d} = \frac{(d+m-1)!}{d!(m-1)!} = \frac{(2+2-1)!}{2!(2-1)!} = 3$$

RBF kernel



Sigmoid Kernel

- This kernel was quite popular for support vector machines due to its origin from neural network theory.
- An SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network.
- The sigmoid kernel is not better than the RBF kernel in general.^[8]

[8] Hsuan-Tien Lin and Chih-Jen Lin. A Study on Sigmoid Kernels for SVM and the Training of non-PSD Kernels by SMO-type Methods. <http://home.caltech.edu/~htlin/publication/doc/tanh.pdf>

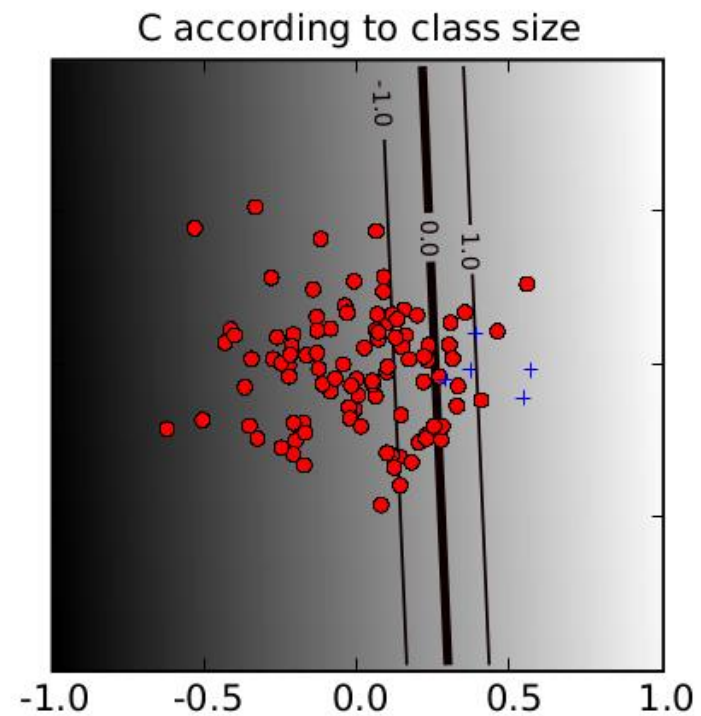
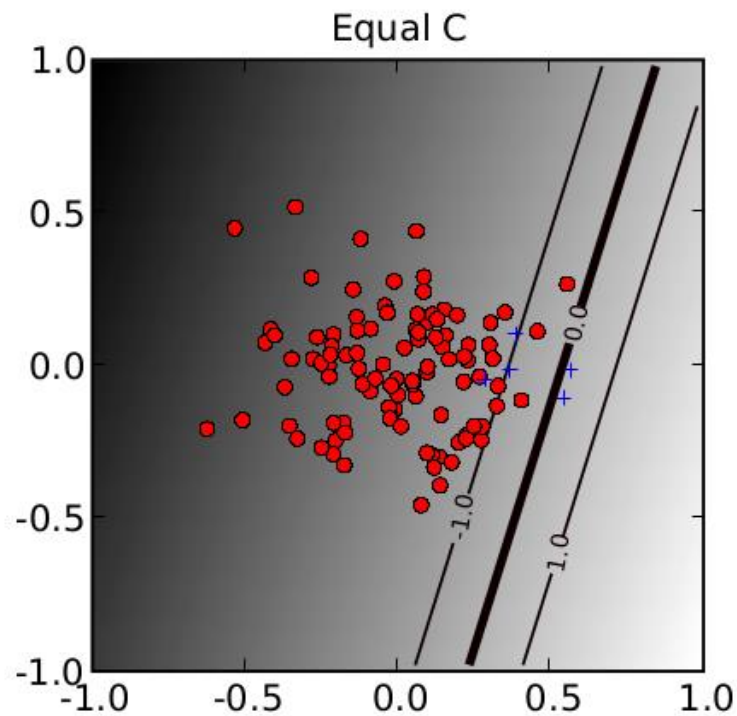
Outline

- A brief history of SVM
- Linear classifiers in a nutshell
 - Linear separable
 - Nonlinear separable
- Creating nonlinear classifiers: the kernel trick
- SVM kernel functions
- SVM extensions
- Software
- Conclusions

SVMs for Unbalanced Data

$$C \sum_{i=1}^n \xi_i \longrightarrow C_+ \sum_{i \in I_+} \xi_i + C_- \sum_{i \in I_-} \xi_i$$

$$C_+ n_+ = C_- n_-$$



SVM for multi-class classification?

- One can change the QP formulation to become multi-class.^[9,10]
- More often, building binary classifiers which distinguish between:^[11]
 - one of the labels and the rest (*one-versus-all*).
 - between every pair of classes (*one-versus-one*).

[9] Ma, Y; Guo G. (Eds.) Support Vector Machines Applications. Chapter 2: "Multi-Class Support Vector Machine" Springer-Verlag (2014)

[10] Gerrit J.J. van den Burg, Patrick J.F. Groenen, GenSVM: A Generalized Multiclass Support Vector Machine, Journal of Machine Learning Research 17 (2016) 1-42

[11] M. Galar, A. Fernández, E. Barrenechea, H. Bustince and F. Herrera, An Overview of Ensemble Methods for Binary Classifiers in Multi-class Problems: Experimental Study on One-vs-One and One-vs-All Schemes. Pattern Recognition 44:8 (2011) 1761-1776. <http://sci2s.ugr.es/ovo-ova>

SVM for multi-class classification?

- One of the labels and the rest (*one-versus-all*)
- Between every pair of classes (*one-versus-one*)

WIKIPEDIA La enciclopedia libre

Artículo **Departamento de San José**

Coordenadas: 34°20'37"S 56°42'37"O [mapa]

San José
Departamento de Uruguay

San José es uno de los 19 departamentos que conforman la República Oriental del Uruguay. Está situado en la región meridional del país. Limita al norte con Flores, al este con Florida, Canelones y Montevideo, al oeste con Colonia y Soriano (mediante una pequeña intersección en la Villa Sancti Spiritus), y al sur con el Río de la Plata. Su superficie es de 4882 km² y tiene una población estimada de 108 399 habitantes según el censo de 2011, lo que lo convierte en uno de los departamentos más poblados del Uruguay.

Su franja meridional es parte del Área Metropolitana de Montevideo, debido a que esta zona está ocupada por la expansión de la mancha urbana del Gran Montevideo. La capital departamental es la ciudad de San José de Mayo.

Indice [ver]

- 1 Toponimia
- 2 Historia
 - 2.1 San José en la historia de Uruguay
 - 2.2 1831: Año de fundación del
 - 2.3 34: Fundación
 - 2.4 1.º de Abril: Día del departamento
 - 2.5 1881: La batalla de Caganchal
 - 2.6 Monarquía y Regencia
- 3 Geografía
 - 3.1 Orografía
 - 3.2 Hidrografía
- 4 Clima
- 5 Flora y fauna
- 6 Economía
- 7 Comunicaciones
- 8 Gobierno
 - 8.1 Ejecutivo
 - 8.2 Legislativo
 - 8.3 Municipios
- 9 Educación y cultura
- 10 Población
- 10.1 Centros urbanos
- 11 Atracciones turísticas
- 11.1 Bos del Cofre
- 11.2 Ríos

San José

Coordenadas: 34°20'37"S 56°42'37"O

Capital: San José de Mayo

Entidad: Departamento

País: Uruguay

Intendente: José Luis Fariña

Partido Nacional

Superficie: Párrafo 16.º

- Total: 4882 km² (2.63 %)

Población (2011): Párrafo 5.º

SVM for multi-class classification?

- One of the labels and the rest (*one-versus-all*)
- Between every pair of classes (*one-versus-one*)

WIKIPEDIA La enciclopedia libre

Artículo | Discusión

Departamento de San José

San José es uno de los 19 departamentos que conforman la República Oriental del Uruguay. Está situado en la región meridional del país. Limita al norte con Flores, al este con Florida, Canelones y Montevideo, al oeste con Colonia y Soriano (mediante una pequeña intersección en la Villa Tanzi del Corral), y al sur con el Río de la Plata. Su superficie es de 4882 km² y tiene una población estimada de 108 399 habitantes según el censo de 2011, lo que lo convierte en uno de los departamentos más poblados del Uruguay.

Se trata de una zona meridional en parte del Área Metropolitana de Montevideo, debido a que esta zona está ocupada por la expansión de la mancha urbana del Gran Montevideo. La capital del departamento es San José de Mayo.

Coordenadas: 34°25'37"S 56°14'23"W (mapa) ✱

San José
Departamento de Uruguay

Brasão

Ubicación de San José

Coordenadas: 34°25'37"S 56°14'23"W

Capital: San José de Mayo

Entidad: Departamento

País: Uruguay

Intendente: José Luis Falero

Partido Nacional

Superficie: Puesto 16.º

- Total: 4882 km² (2.63 %)

Población (2011): Puesto 5.º

1 Toponimia

2 Historia

3 Demografía

4 Geografía

5 Flora y fauna

6 Economía

7 Comunicaciones

8 Gobierno

9 Educación y cultura

10 Población

11 Atractivos turísticos

12 Río

FA

GA

C

Stub

SVM

SVM for multi-class classification?

- One of the labels and the rest (*one-versus-all*)

The screenshot shows the Wikipedia page for 'Departamento de San José'. A diagram is overlaid on the page, illustrating the SVM classification process. The diagram consists of several colored squares representing different classes or features:

- A blue square with a star and the letters 'FA'.
- A green square with a plus sign and the letters 'GA'.
- A yellow square with the letter 'C'.
- A red square with the word 'Stub'.

An orange line connects these squares to a central icon representing a Support Vector Machine (SVM) model, which is a grid with a downward arrow and several colored dots. Below this icon, the text 'C₁ SVM' is written. The diagram also includes a large orange 'P' on the left and a large orange 'N' on the right, likely representing positive and negative classes in the SVM context.

SVM for multi-class classification?

- One of the labels and the rest (*one-versus-all*)

WIKIPEDIA La enciclopedia libre

Artículo | Discusión

Departamento de San José

San José es uno de los 19 departamentos que conforman la República Oriental del Uruguay. Está situado en la región meridional del país. Limita al norte con Flores, al este con Florida, Canelones y Montevideo, al oeste con Colonia y Soriano (mediante una pequeña intersección en la Villa Sarandí Corina), y al sur con el Río de la Plata. Su superficie es de 4882 km² y tiene una población estimada de 108 399 habitantes según el censo de 2011, lo que lo convierte en uno de los departamentos más poblados del Uruguay.

Se trata de una zona meridional en parte del Área Metropolitana de Montevideo, debido a que esta zona está ocupada por la expansión de la mancha urbana del Gran Montevideo. La capital

1 Toponimia
2 Historia
3 Geografía
4 Clima
5 Flora y fauna
6 Economía
7 Comunicaciones
8 Gobierno
9 Educación y cultura
10 Población
11 Atractivos turísticos
12 Río

11.1 Bosque del Cofre
11.2 Río

San José

San José de Uruguay

San José

Coordenadas: 34°25'37S 56°14'23W (mapa)

Capital: San José de Mayo
Entidad: Departamento
País: Uruguay
Intendente: José Luis Falero
Superficie: Puesto 16.º
• Total: 4882 km² (2.63 %)
Población (2011): Puesto 5.º

Diagram illustrating the SVM classification process for multi-class classification:

The input data is divided into two groups, **P** and **N**, based on the SVM classification results.

Group **P** contains four classes: **FA** (blue square), **GA** (green square), **C** (yellow square), and **Stub** (red square).

Group **N** contains the remaining classes, represented by a grid icon.

The SVM classifier is shown at the bottom, labeled **C₂ SVM**.

SVM for multi-class classification?

- One of the labels and the rest (*one-versus-all*)

WIKIPEDIA La enciclopedia libre

Artículo | Discusión

Departamento de San José

San José es uno de los 19 departamentos que conforman la República Oriental del Uruguay. Está situado en la región meridional del país. Limita al norte con Flores, al este con Florida, Carreteres y Montevideo, al oeste con Colonia y Soriano (mediante una pequeña intersección en la Villa Tanzi Corina), y al sur con el Río de la Plata. Su superficie es de 4882 km² y tiene una población estimada de 108 399 habitantes según el censo de 2011, lo que la convierte en uno de los departamentos más poblados del Uruguay.

5. Frontera meridional ex parte del Área Metropolitana de Montevideo, debido a que esta zona está ocupada por la expansión de la mancha urbana del Gran Montevideo.

Toponimia

1.1 Etimología

1.2 Historia

4. Clima

5. Flora y fauna

6. Economía

7. Comunicaciones

8. Gobierno

8.1 Ejecutivo

8.2 Legislativo

8.3 Municipios

9. Educación y cultura

10. Población

10.1 Centros urbanos

11. Atracciones turísticas

11.1 Bosca del Cofre

11.2 Ríos

San José

Departamento de Uruguay

Capital

Bravío

Ubicación de San José

Coordenadas: 34°25'37.5" S 64°23'10" O

Capital: San José de Mayo

Entidad: Departamento

País: Uruguay

Intendente: José Luis Falero

Partido Nacional

Superficie: Puesto 16.º

- Total: 4882 km² (2.63 %)

Población (2011): Puesto 5.º

Diagram illustrating the SVM classification process for multi-class classification:

- Four input classes are shown: FA (blue square with a star), GA (green square with a plus sign), C (yellow square), and Stub (red square).
- These classes are fed into an SVM (Support Vector Machine) model, represented by a box with a grid and arrows.
- The SVM output is labeled with 'C₃' and 'SVM'.
- The output is then mapped to a set of labels: 'P' (orange) and 'N' (orange).

SVM for multi-class classification?

- One of the labels and the rest (*one-versus-all*)

WIKIPEDIA La enciclopedia libre

Artículo | Discusión

Departamento de San José

San José es uno de los 19 departamentos que conforman la República Oriental del Uruguay. Está situado en la región meridional del país. Limita al norte con Flores, al este con Florida, Carreteres y Montevideo, al oeste con Colonia y Soriano (mediante una pequeña intersección en la Villa Sarandí Corina), y al sur con el Río de la Plata. Su superficie es de 4882 km² y tiene una población estimada de 108 399 habitantes según el censo de 2011, lo que lo convierte en uno de los departamentos más poblados del Uruguay.

Se encuentra meridional en parte del Área Metropolitana de Montevideo, debido a que esta zona está ocupada por la expansión de la mancha urbana del Gran Montevideo. La capital del departamento es San José.

1 Topografía

2 Historia

3 Población

4 Clima

5 Flora y fauna

6 Economía

7 Comunicaciones

8 Gobierno

8.1 Ejecutivo

8.2 Legislativo

8.3 Municipios

9 Educación y cultura

10 Población

10.1 Centros urbanos

11 Atracciones turísticas

11.1 Bosque del Cofre

11.2 Ríos

12 Geografía

12.1 Hidrografía

12.2 Clima

12.3 Flora y fauna

12.4 Economía

12.5 Comunicaciones

12.6 Gobierno

12.7 Educación y cultura

12.8 Población

12.9 Atracciones turísticas

12.10 Bosque del Cofre

12.11 Ríos

12.12 Geografía

12.13 Hidrografía

12.14 Clima

12.15 Flora y fauna

12.16 Economía

12.17 Comunicaciones

12.18 Gobierno

12.19 Educación y cultura

12.20 Población

12.21 Atracciones turísticas

12.22 Bosque del Cofre

12.23 Ríos

12.24 Geografía

12.25 Hidrografía

12.26 Clima

12.27 Flora y fauna

12.28 Economía

12.29 Comunicaciones

12.30 Gobierno

12.31 Educación y cultura

12.32 Población

12.33 Atracciones turísticas

12.34 Bosque del Cofre

12.35 Ríos

12.36 Geografía

12.37 Hidrografía

12.38 Clima

12.39 Flora y fauna

12.40 Economía

12.41 Comunicaciones

12.42 Gobierno

12.43 Educación y cultura

12.44 Población

12.45 Atracciones turísticas

12.46 Bosque del Cofre

12.47 Ríos

12.48 Geografía

12.49 Hidrografía

12.50 Clima

12.51 Flora y fauna

12.52 Economía

12.53 Comunicaciones

12.54 Gobierno

12.55 Educación y cultura

12.56 Población

12.57 Atracciones turísticas

12.58 Bosque del Cofre

12.59 Ríos

12.60 Geografía

12.61 Hidrografía

12.62 Clima

12.63 Flora y fauna

12.64 Economía

12.65 Comunicaciones

12.66 Gobierno

12.67 Educación y cultura

12.68 Población

12.69 Atracciones turísticas

12.70 Bosque del Cofre

12.71 Ríos

12.72 Geografía

12.73 Hidrografía

12.74 Clima

12.75 Flora y fauna

12.76 Economía

12.77 Comunicaciones

12.78 Gobierno

12.79 Educación y cultura

12.80 Población

12.81 Atracciones turísticas

12.82 Bosque del Cofre

12.83 Ríos

12.84 Geografía

12.85 Hidrografía

12.86 Clima

12.87 Flora y fauna

12.88 Economía

12.89 Comunicaciones

12.90 Gobierno

12.91 Educación y cultura

12.92 Población

12.93 Atracciones turísticas

12.94 Bosque del Cofre

12.95 Ríos

12.96 Geografía

12.97 Hidrografía

12.98 Clima

12.99 Flora y fauna

13 Geografía

13.1 Hidrografía

13.2 Clima

13.3 Flora y fauna

13.4 Economía

13.5 Comunicaciones

13.6 Gobierno

13.7 Educación y cultura

13.8 Población

13.9 Atracciones turísticas

13.10 Bosque del Cofre

13.11 Ríos

13.12 Geografía

13.13 Hidrografía

13.14 Clima

13.15 Flora y fauna

13.16 Economía

13.17 Comunicaciones

13.18 Gobierno

13.19 Educación y cultura

13.20 Población

13.21 Atracciones turísticas

13.22 Bosque del Cofre

13.23 Ríos

13.24 Geografía

13.25 Hidrografía

13.26 Clima

13.27 Flora y fauna

13.28 Economía

13.29 Comunicaciones

13.30 Gobierno

13.31 Educación y cultura

13.32 Población

13.33 Atracciones turísticas

13.34 Bosque del Cofre

13.35 Ríos

13.36 Geografía

13.37 Hidrografía

13.38 Clima

13.39 Flora y fauna

13.40 Economía

13.41 Comunicaciones

13.42 Gobierno

13.43 Educación y cultura

13.44 Población

13.45 Atracciones turísticas

13.46 Bosque del Cofre

13.47 Ríos

13.48 Geografía

13.49 Hidrografía

13.50 Clima

13.51 Flora y fauna

13.52 Economía

13.53 Comunicaciones

13.54 Gobierno

13.55 Educación y cultura

13.56 Población

13.57 Atracciones turísticas

13.58 Bosque del Cofre

13.59 Ríos

13.60 Geografía

13.61 Hidrografía

13.62 Clima

13.63 Flora y fauna

13.64 Economía

13.65 Comunicaciones

13.66 Gobierno

13.67 Educación y cultura

13.68 Población

13.69 Atracciones turísticas

13.70 Bosque del Cofre

13.71 Ríos

13.72 Geografía

13.73 Hidrografía

13.74 Clima

13.75 Flora y fauna

13.76 Economía

13.77 Comunicaciones

13.78 Gobierno

13.79 Educación y cultura

13.80 Población

13.81 Atracciones turísticas

13.82 Bosque del Cofre

13.83 Ríos

13.84 Geografía

13.85 Hidrografía

13.86 Clima

13.87 Flora y fauna

13.88 Economía

13.89 Comunicaciones

13.90 Gobierno

13.91 Educación y cultura

13.92 Población

13.93 Atracciones turísticas

13.94 Bosque del Cofre

13.95 Ríos

13.96 Geografía

13.97 Hidrografía

13.98 Clima

13.99 Flora y fauna

14 Clima

14.1 Hidrografía

14.2 Clima

14.3 Flora y fauna

14.4 Economía

14.5 Comunicaciones

14.6 Gobierno

14.7 Educación y cultura

14.8 Población

14.9 Atracciones turísticas

14.10 Bosque del Cofre

14.11 Ríos

14.12 Geografía

14.13 Hidrografía

14.14 Clima

14.15 Flora y fauna

14.16 Economía

14.17 Comunicaciones

14.18 Gobierno

14.19 Educación y cultura

14.20 Población

14.21 Atracciones turísticas

14.22 Bosque del Cofre

14.23 Ríos

14.24 Geografía

14.25 Hidrografía

14.26 Clima

14.27 Flora y fauna

14.28 Economía

14.29 Comunicaciones

14.30 Gobierno

14.31 Educación y cultura

14.32 Población

14.33 Atracciones turísticas

14.34 Bosque del Cofre

14.35 Ríos

14.36 Geografía

14.37 Hidrografía

14.38 Clima

14.39 Flora y fauna

14.40 Economía

14.41 Comunicaciones

14.42 Gobierno

14.43 Educación y cultura

14.44 Población

14.45 Atracciones turísticas

14.46 Bosque del Cofre

14.47 Ríos

14.48 Geografía

14.49 Hidrografía

14.50 Clima

14.51 Flora y fauna

14.52 Economía

14.53 Comunicaciones

14.54 Gobierno

14.55 Educación y cultura

14.56 Población

14.57 Atracciones turísticas

14.58 Bosque del Cofre

14.59 Ríos

14.60 Geografía

14.61 Hidrografía

14.62 Clima

14.63 Flora y fauna

14.64 Economía

14.65 Comunicaciones

14.66 Gobierno

14.67 Educación y cultura

14.68 Población

14.69 Atracciones turísticas

14.70 Bosque del Cofre

14.71 Ríos

14.72 Geografía

14.73 Hidrografía

14.74 Clima

14.75 Flora y fauna

14.76 Economía

14.77 Comunicaciones

14.78 Gobierno

14.79 Educación y cultura

14.80 Población

14.81 Atracciones turísticas

14.82 Bosque del Cofre

14.83 Ríos

14.84 Geografía

14.85 Hidrografía

14.86 Clima

14.87 Flora y fauna

14.88 Economía

14.89 Comunicaciones

14.90 Gobierno

14.91 Educación y cultura

14.92 Población

14.93 Atracciones turísticas

14.94 Bosque del Cofre

14.95 Ríos

14.96 Geografía

14.97 Hidrografía

14.98 Clima

14.99 Flora y fauna

15 Flora y fauna

15.1 Hidrografía

15.2 Clima

15.3 Flora y fauna

15.4 Economía

15.5 Comunicaciones

15.6 Gobierno

15.7 Educación y cultura

15.8 Población

15.9 Atracciones turísticas

15.10 Bosque del Cofre

15.11 Ríos

15.12 Geografía

15.13 Hidrografía

15.14 Clima

15.15 Flora y fauna

15.16 Economía

15.17 Comunicaciones

15.18 Gobierno

15.19 Educación y cultura

15.20 Población

15.21 Atracciones turísticas

15.22 Bosque del Cofre

15.23 Ríos

15.24 Geografía

15.25 Hidrografía

15.26 Clima

15.27 Flora y fauna

15.28 Economía

15.29 Comunicaciones

15.30 Gobierno

15.31 Educación y cultura

15.32 Población

15.33 Atracciones turísticas

15.34 Bosque del Cofre

15.35 Ríos

15.36 Geografía

15.37 Hidrografía

15.38 Clima

15.39 Flora y fauna

15.40 Economía

15.41 Comunicaciones

15.42 Gobierno

15.43 Educación y cultura

15.44 Población

15.45 Atracciones turísticas

15.46 Bosque del Cofre

15.47 Ríos

15.48 Geografía

15.49 Hidrografía

15.50 Clima

15.51 Flora y fauna

15.52 Economía

15.53 Comunicaciones

15.54 Gobierno

15.55 Educación y cultura

15.56 Población

15.57 Atracciones turísticas

15.58 Bosque del Cofre

15.59 Ríos

15.60 Geografía

15.61 Hidrografía

15.62 Clima

15.63 Flora y fauna

15.64 Economía

15.65 Comunicaciones

15.66 Gobierno

15.67 Educación y cultura

15.68 Población

15.69 Atracciones turísticas

15.70 Bosque del Cofre

15.71 Ríos

15.72 Geografía

15.73 Hidrografía

15.74 Clima

15.75 Flora y fauna

15.76 Economía

15.77 Comunicaciones

15.78 Gobierno

15.79 Educación y cultura

15.80 Población

15.81 Atracciones turísticas

15.82 Bosque del Cofre

15.83 Ríos

15.84 Geografía

15.85 Hidrografía

15.86 Clima

15.87 Flora y fauna

15.88 Economía

15.89 Comunicaciones

15.90 Gobierno

15.91 Educación y cultura

15.92 Población

15.93 Atracciones turísticas

15.94 Bosque del Cofre

15.95 Ríos

15.96 Geografía

15.97 Hidrografía

15.98 Clima

15.99 Flora y fauna

16 Economía

16.1 Comunicaciones

16.2 Gobierno

16.3 Educación y cultura

16.4 Población

16.5 Atracciones turísticas

16.6 Bosque del Cofre

16.7 Ríos

16.8 Geografía

16.9 Hidrografía

16.10 Clima

16.11 Flora y fauna

16.12 Economía

16.13 Comunicaciones

16.14 Gobierno

16.15 Educación y cultura

16.16 Población

16.17 Atracciones turísticas

16.18 Bosque del Cofre

16.19 Ríos

16.20 Geografía

16.21 Hidrografía

16.22 Clima

16.23 Flora y fauna

16.24 Economía

16.25 Comunicaciones

16.26 Gobierno

16.27 Educación y cultura

16.28 Población

16.29 Atracciones turísticas

16.30 Bosque del Cofre

16.31 Ríos

16.32 Geografía

16.33 Hidrografía

16.34 Clima

16.35 Flora y fauna

16.36 Economía

16.37 Comunicaciones

16.38 Gobierno

16.39 Educación y cultura

16.40 Población

16.41 Atracciones turísticas

16.42 Bosque del Cofre

16.43 Ríos

16.44 Geografía

16.45 Hidrografía

16.46 Clima

16.47 Flora y fauna

16.48 Economía

16.49 Comunicaciones

16.50 Gobierno

16.51 Educación y cultura

16.52 Población

16.53 Atracciones turísticas

16.54 Bosque del Cofre

16.55 Ríos

16.56 Geografía

16.57 Hidrografía

16.58 Clima

16.59 Flora y fauna

16.60 Economía

16.61 Comunicaciones

16.62 Gobierno

16.63 Educación y cultura

16.64 Población

16.65 Atracciones turísticas

16.66 Bosque del Cofre

16.67 Ríos

16.68 Geografía

16.69 Hidrografía

16.70 Clima

16.71 Flora y fauna

16.72 Economía

16.73 Comunicaciones

16.74 Gobierno

16.75 Educación y cultura

16.76 Población

16.77 Atracciones turísticas

16.78 Bosque del Cofre

16.79 Ríos

16.80 Geografía

16.81 Hidrografía

16.82 Clima

16.83 Flora y fauna

16.84 Economía

16.85 Comunicaciones

16.86 Gobierno

16.87 Educación y cultura

16.88 Población

16.89 Atracciones turísticas

16.90 Bosque del Cofre

16.91 Ríos

16.92 Geografía

16.93 Hidrografía

16.94 Clima

16.95 Flora y fauna

16.96 Economía

16.97 Comunicaciones

16.98 Gobierno

16.99 Educación y cultura

17 Comunicaciones

17.1 Gobierno

17.2 Educación y cultura

17.3 Población

17.4 Atracciones turísticas

17.5 Bosque del Cofre

17.6 Ríos

17.7 Geografía

17.8 Hidrografía

17.9 Clima

17.10 Flora y fauna

17.11 Economía

17.12 Comunicaciones

17.13 Gobierno

17.14 Educación y cultura

17.15 Población

17.16 Atracciones turísticas

17.17 Bosque del Cofre

17.18 Ríos

17.19 Geografía

17.20 Hidrografía

17.21 Clima

17.22 Flora y fauna

17.23 Economía

17.24 Comunicaciones

17.25 Gobierno

17.26 Educación y cultura

17.27 Población

17.28 Atracciones turísticas

17.29 Bosque del Cofre

17.30 Ríos

17.31 Geografía

17.32 Hidrografía

17.33 Clima

17.34 Flora y fauna

17.35 Economía

17.36 Comunicaciones

17.37 Gobierno

17.38 Educación y cultura

17.39 Población

17.40 Atracciones turísticas

17.41 Bosque del Cofre

17.42 Ríos

17.43 Geografía

17.44 Hidrografía

17.45 Clima

17.46 Flora y fauna

17.47 Economía

17.48 Comunicaciones

17.49 Gobierno

17.50 Educación y cultura

18 Gobierno

18.1 Educación y cultura

18.2 Población

18.3 Atracciones turísticas

18.4 Bosque del Cofre

18.5 Ríos

18.6 Geografía

18.7 Hidrografía

18.8 Clima

18.9 Flora y fauna

18.10 Economía

18.11 Comunicaciones

18.12 Gobierno

18.13 Educación y cultura

18.14 Población

18.15 Atracciones turísticas

18.16 Bosque del Cofre

18.17 Ríos

18.18 Geografía

18.19 Hidrografía

18.20 Clima

18.21 Flora y fauna

18.22 Economía

18.23 Comunicaciones

18.24 Gobierno

18.25 Educación y cultura

18.26 Población

18.27 Atracciones turísticas

18.28 Bosque del Cofre

18.29 Ríos

18.30 Geografía

18.31 Hidrografía

18.32 Clima

18.33 Flora y fauna

18.34 Economía

18.35 Comunicaciones

18.36 Gobierno

18.37 Educación y cultura

18.38 Población

18.39 Atracciones turísticas

18.40 Bosque del Cofre

18.41 Ríos

18.42 Geografía

18.43 Hidrografía

18.44 Clima

18.45 Flora y fauna

18.46 Economía

18.47 Comunicaciones

18.48 Gobierno

18.49 Educación y cultura

18.50 Población

18.51 Atracciones turísticas

18.52 Bosque del Cofre

18.53 Ríos

18.54 Geografía

18.55 Hidrografía

18.56 Clima

18.57 Flora y fauna

18.58 Economía

18.59 Comunicaciones

18.60 Gobierno

18.61 Educación y cultura

18.62 Población

18.63 Atracciones turísticas

18.64 Bosque del Cofre

18.65 Ríos

18.66 Geografía

18.67 Hidrografía

18.68 Clima

18.69 Flora y fauna

18.70 Economía

18.71 Comunicaciones

18.72 Gobierno

18.73 Educación y cultura

18.74 Población

18.75 Atracciones turísticas

18.76 Bosque del Cofre

18.77 Ríos

18.78 Geografía

18.79 Hidrografía

18.80 Clima

18.81 Flora y fauna

18.82 Economía

18.83 Comunicaciones

18.84 Gobierno

18.85 Educación y cultura

18.86 Población

18.87 Atracciones turísticas

18.88 Bosque del Cofre

18.89 Ríos

18.90 Geografía

18.91 Hidrografía

18.92 Clima

18.93 Flora y fauna

18.94 Economía

18.95 Comunicaciones

18.96 Gobierno

18.97 Educación y cultura

18.98 Población

18.99 Atracciones turísticas

19 Educación y cultura

19.1 Población

19.2 Atracciones turísticas

19.3 Bosque del Cofre

19.4 Ríos

19.5 Geografía

19.6 Hidrografía

19.7 Clima

19.8 Flora y fauna

19.9 Economía

19.10 Comunicaciones

19.11 Gobierno

19.12 Educación y cultura

19.13 Población

19.14 Atracciones turísticas

19.15 Bosque del Cofre

19.16 Ríos

19.17 Geografía

19.18 Hidrografía

19.19 Clima

19.20 Flora y fauna

19.21 Economía

19.22 Comunicaciones

19.23 Gobierno

19.24 Educación y cultura

19.25 Población

19.26 Atracciones turísticas

19.27 Bosque del Cofre

19.28 Ríos

19.29 Geografía

19.30 Hidrografía

19.31 Clima

19.32 Flora y fauna

19.33 Economía

19.34 Comunicaciones

19.35 Gobierno

19.36 Educación y cultura

19.37 Población

19.38 Atracciones turísticas

19.39 Bosque del Cofre

19.40 Ríos

19.41 Geografía

19.42 Hidrografía

19.43 Clima

19.44 Flora y fauna

19.45 Economía

19.46 Comunicaciones

19.47 Gobierno

19.48 Educación y cultura

19.49 Población

19.50 Atracciones turísticas

20 Población

20.1 Atracciones turísticas

20.2 Bosque del Cofre

20.3 Ríos

20.4 Geografía

20.5 Hidrografía

20.6 Clima

20.7 Flora y fauna

20.8 Economía

20.9 Comunicaciones

20.10 Gobierno

20.11 Educación y cultura

20.12 Población

20.13 Atracciones turísticas

20.14 Bosque del Cofre

20.15 Ríos

20.16 Geografía

20.17 Hidrografía

20.18 Clima

20.19 Flora y fauna

20.20 Economía

20.21 Comunicaciones

20.22 Gobierno

20.23 Educación y cultura

20.24 Población

20.25 Atracciones turísticas

20.26 Bosque del Cofre

20.27 Ríos

20.28 Geografía

20.29 Hidrografía

20.30 Clima

20.31 Flora y fauna

20.32 Economía

20.33 Comunicaciones

20.34 Gobierno

20.35 Educación y cultura

20.36 Población

20.37 Atracciones turísticas

20.38 Bosque del Cofre

20.39 Ríos

20.40 Geografía

20.41 Hidrografía

20.42 Clima

20.43 Flora y fauna

20.44 Economía

20.45 Comunicaciones

20.46 Gobierno

20.47 Educación y cultura

20.48 Población

20.49 Atracciones turísticas

20.50 Bosque del Cofre

20.51 Ríos

20.52 Geografía

20.53 Hidrografía

20.54 Clima

20.55 Flora y fauna

20.56 Economía

20.57 Comunicaciones

20.58 Gobierno

20.59 Educación y cultura

20.60 Población

20.61 Atracciones turísticas

20.62 Bosque del Cofre

20.63 Ríos

20.64 Geografía

20.65 Hidrografía

20.66 Clima

20.67 Flora y fauna

20.68 Economía

20.69 Comunicaciones

20.70 Gobierno

20.71 Educación y cultura

20.72 Población

20.73 Atracciones turísticas

20.74 Bosque del Cofre

20.75 Ríos

20.76 Geografía

20.77 Hidrografía

20.78 Clima

20.79 Flora y fauna

20.80 Economía

20.81 Comunicaciones

20.82 Gobierno

20.83 Educación y cultura

20.84 Población

20.85 Atracciones turísticas

20.86 Bosque del Cofre

20.87 Ríos

20.88 Geografía

20.89 Hidrografía

20.90 Clima

20.91 Flora y fauna

20.92 Economía

20.93 Comunicaciones

20.94 Gobierno

20.95 Educación y cultura

20.96 Población

20.97 Atracciones turísticas

20.98 Bosque del Cofre

20.99 Ríos

20.100 Geografía

21 Atracciones turísticas

21.1 Bosque del Cofre

21.2 Ríos

21.3 Geografía

21.4 Hidrografía

21.5 Clima

21.6 Flora y fauna

21.7 Economía

21.8 Comunicaciones

21.9 Gobierno

21.10 Educación y cultura

21.11 Población

21.12 Atracciones turísticas

21.13 Bosque del Cofre

21.14 Ríos

21.15 Geografía

21.16 Hidrografía

21.17 Clima

21.18 Flora y fauna

21.19 Economía

21.20 Comunicaciones

21.21 Gobierno

21.22 Educación y cultura

21.23 Población

21.24 Atracciones turísticas

21.25 Bosque del Cofre

21.26 Ríos

21.27 Geografía

21.28 Hidrografía

21.29 Clima

21.30 Flora y fauna

21.31 Economía

21.32 Comunicaciones

21.33 Gobierno

21.34 Educación y cultura

21.35 Población

21.36 Atracciones turísticas

21.37 Bosque del Cofre

21.38 Ríos

21.39 Geografía

21.40 Hidrografía

21.41 Clima

21.42 Flora y fauna

21.43 Economía

21.44 Comunicaciones

21.45 Gobierno

21.46 Educación y cultura

21.47 Población

21.48 Atracciones turísticas

21.49 Bosque del Cofre

21.50 Ríos

21.51 Geografía

21.52 Hidrografía

21.53 Clima

21.54 Flora y fauna

21.55 Economía

21.56 Comunicaciones

21.57 Gobierno

21.58 Educación y cultura

21.59 Población

21.60 Atracciones turísticas

21.61 Bosque del Cofre

21.62 Ríos

21.63 Geografía

21.64 Hidrografía

21.65 Clima

21.66 Flora y fauna

21.67 Economía

21.68 Comunicaciones

21.69 Gobierno

21.70 Educación y cultura

21.71 Población

21.72 Atracciones turísticas

21.73 Bosque del Cofre

21.74 Ríos

21.75 Geografía

21.76 Hidrografía

21.77 Clima

21.78 Flora y fauna

21.79 Economía

21.80 Comunicaciones

21.81 Gobierno

21.82 Educación y cultura

21.83 Población

21.84 Atracciones turísticas

21.85 Bosque del Cofre

21.86 Ríos

21.87 Geografía

21.88 Hidrografía

21.89 Clima

21.90 Flora y fauna

21.91 Economía

21.92 Comunicaciones

21.93 Gobierno

21.94 Educación y cultura

21.95 Población

21.96 Atracciones turísticas

21.97 Bosque del Cofre

21.98 Ríos

21.99 Geografía

22 Bosque del Cofre

22.1 Ríos

22.2 Geografía

22.3 Hidrografía

22.4 Clima

22.5 Flora y fauna

22.6 Economía

22.7 Comunicaciones

22.8 Gobierno

22.9 Educación y cultura

22.10 Población

22.11 Atracciones turísticas

22.12 Bosque del Cofre

22.13 Ríos

22.14 Geografía

22.15 Hidrografía

22.16 Clima

22.17 Flora y fauna

22.18 Economía

22.19 Comunicaciones

22.20 Gobierno

22.21 Educación y cultura

22.22 Población

22.23 Atracciones turísticas

22.24 Bosque del Cofre

22.25 Ríos

22.26 Geografía

22.27 Hidrografía

22.28 Clima

22.29 Flora y fauna

22.30 Economía

22.31 Comunicaciones

22.32 Gobierno

22.33 Educación y cultura

22.34 Población

22.35 Atracciones turísticas

22.36 Bosque del Cofre

22.37 Ríos

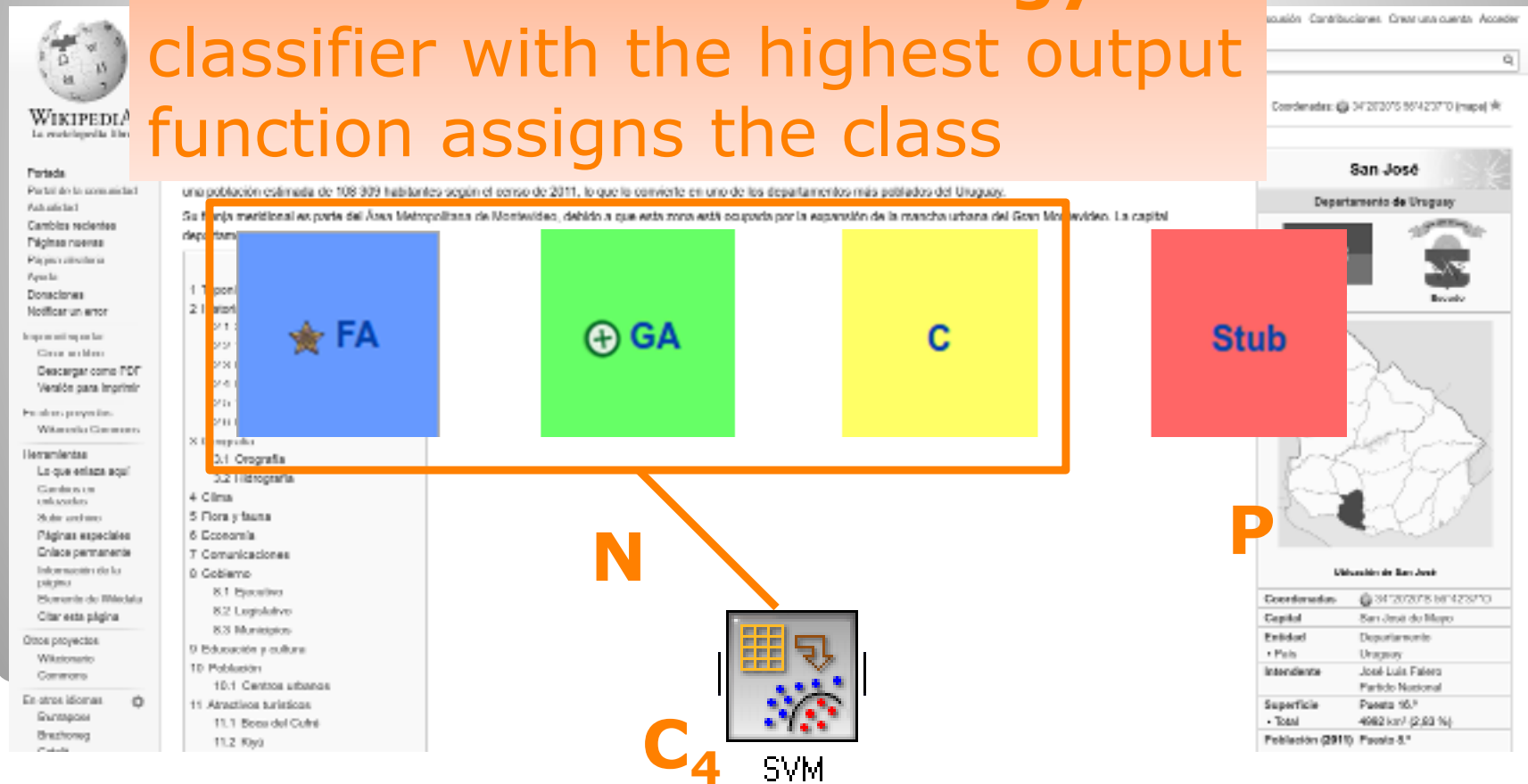
22.38

SVM for multi-class classification?

- One of the labels and the rest (*one-versus-all*)

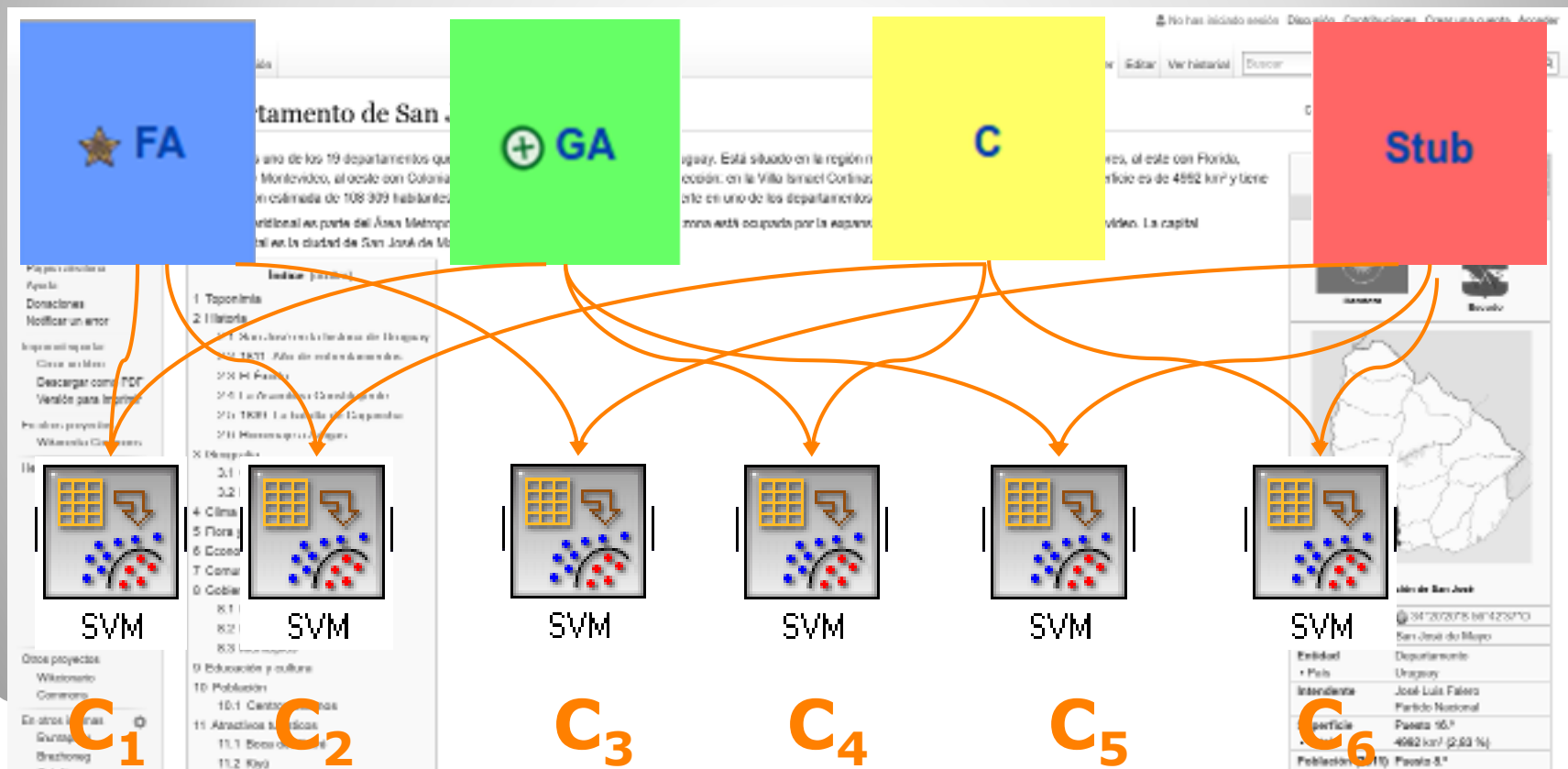
winner-takes-all strategy:

classifier with the highest output function assigns the class



SVM for multi-class classification?

- Between every pair of classes (*one-versus-one*)



SVM for multi-class classification?

- Between every pair of classes (*one-versus-one*)

max-wins voting strategy: every classifier assigns the instance to one of two classes. Class assigned the vote is increased by one. The class with more votes determines the instance classification

The diagram illustrates the max-wins voting strategy for multi-class classification using SVMs. It shows six individual SVM classifiers, each represented by a small icon with a blue and red dot pattern. Below each icon is a label: C₁, C₂, C₃, C₄, C₅, and C₆. The classifiers C₁ and C₂ are shown with additional details: C₁ has a sidebar with 'Otros proyectos' (Wikcionario, Commons, etc.) and 'En otros idiomas' (Gunsberg, Brezhnev, etc.); C₂ has a sidebar with 'T Comar' (0 Gobie, 8.1, 8.2, 8.3), 'D Educación y cultura', '10 Población', '10.1 Centro', '11 Atractivos turísticos', '11.1 Bosc', and '11.2 Roga'. The classifiers C₃, C₄, C₅, and C₆ are shown with a sidebar that includes 'Entidad' (País, Departamento, Intendencia, Municipio, Partido), 'Población' (2010, Población 5*), and 'Población' (2010, Población 5*). The classifiers C₁ and C₂ are shown with a sidebar that includes 'Entidad' (País, Departamento, Intendencia, Municipio, Partido), 'Población' (2010, Población 5*), and 'Población' (2010, Población 5*). The classifiers C₃, C₄, C₅, and C₆ are shown with a sidebar that includes 'Entidad' (País, Departamento, Intendencia, Municipio, Partido), 'Población' (2010, Población 5*), and 'Población' (2010, Población 5*).

Outline

- A brief history of SVM
- Linear classifiers in a nutshell
 - Linear separable
 - Nonlinear separable
- Creating nonlinear classifiers: the kernel trick
- SVM kernel functions
- SVM extensions
- **Software**
- **Conclusions**

Software

- A list of SVM implementation can be found at <http://www.kernel-machines.org/software.html>
- Some implementation (such as LIBSVM) can handle multi-class classification.
- SVMLight is among one of the earliest implementation of SVM.
- Several Matlab toolboxes for SVM are also available.
- PyML - Machine Learning in Python
http://pyml.sourceforge.net/svm_howto.html
- <https://scikit-learn.org/stable/modules/svm.html>

Outline

- A brief history of SVM
- Linear classifiers in a nutshell
 - Linear separable
 - Nonlinear separable
- Creating nonlinear classifiers: the kernel trick
- SVM kernel functions
- SVM extensions
- Software
- **Conclusions**

Conclusions

- Kernel methods owe their name to the use of **kernel functions**, which enable them to operate in a high-dimensional, **implicit** feature space without ever computing the coordinates of the data in that space, but rather by simply computing the **inner products** between the images of all pairs of data in the feature space.
- This operation is often computationally cheaper than the explicit computation of the coordinates. This approach is called the **kernel trick**.
- SVM is one of the most popular kernel-based methods.

Conclusions

- Choosing the kernel function: probably the most tricky part of using SVM.
 - In practice, a low degree polynomial kernel or RBF kernel with a reasonable width is a good initial try. Grid search to perform parameters tuning. See e.g. [7].
- How to use SVM for multi-class classification?
 - One can change the QP formulation to become multi-class.
 - More often, building binary classifiers which distinguish between:
 - one of the labels and the rest (*one-versus-all*).
 - between every pair of classes (*one-versus-one*).

References

- A Simple Introduction to Support Vector Machines by Martin Law. Lecture for CSE 802, Department of Computer Science and Engineering, Michigan State University.
- Artificial Intelligence: A Modern Approach. Stuart Russell and Peter Norvig. 3rd Edition. Section 18.9, pp.744-748
- Data Mining: Practical Machine Learning Tools and Techniques. Ian H. Witten and Eibe Frank. 2nd Edition. Section 6.3, pp. 214-222
- A Practical Guide to Support Vector Classification. Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin.
<http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- A User's Guide to Support Vector Machines. Asa Ben-Hur and Jason Weston. In Data Mining Techniques for the Life Sciences, 2010, Springer.