



*Institut de Formació Contínua-IL3*  
UNIVERSITAT DE BARCELONA

---

# Técnicas avanzadas de Machine Learning

---

Alberto Julián Rigau

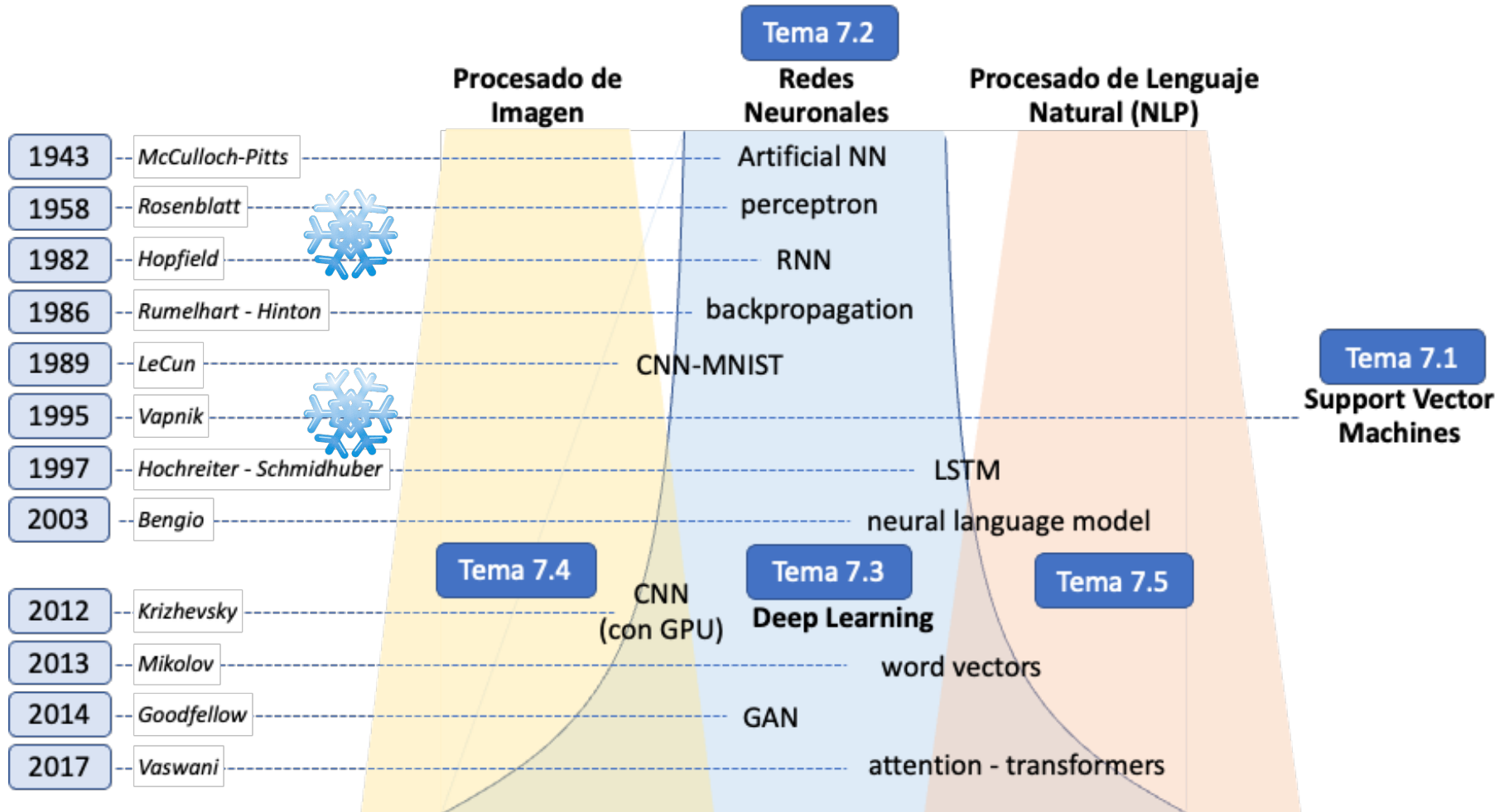
© de esta edición: Fundació IL3-UB 2021

---

## Agenda

1. Relación entre los temas
2. Plan de tutorías y actividades
3. Actividades
4. T7.1 Support Vector Machines

## Relación entre los temas



## Plan de tutorías y actividades – Noviembre

Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16 <b>7.1 SVM, AI1</b>	17	18	19 Tut 1 17:00 (16:00 UTC) <b>7.1 SVM, AI1</b>	20	21
22  <b>AI1</b>	23 <b>7.2 NN, 7.3 DL, AI2</b>	24	25	26 Tut 2 17:00 (16:00 UTC) <b>7.2 NN, 7.3 DL, AI2</b>	27	28
29  <b>AI2</b>	30 <b>7.4 Imagen, AI3</b>	1	2	3 Tut 3 17:00 (16:00 UTC) <b>7.4 Imagen, AI3</b>	4	5

## Plan de tutorías y actividades – Diciembre

Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
29  AI2	30 7.4 Imagen, AI3	1	2	3 Tut 3 17:00 (16:00 UTC) 7.4 Imagen, AI3	4	5
6  AI3	7 7.5 Texto, AI4, AG1	8	9	10 Tut 4 17:00 (16:00 UTC) 7.5 Texto, AI4 y AG1	11	12
13  AI4	14	15	16	17 Tut 5 17:00 (16:00 UTC) 7.5 Texto, AG1	18	19
20  AG1	21	22  Test	23	24	25	26
27	28	29	30	31	1	2

## Actividades

AI1 SVM (T7.1)

**AI2 Frameworks de Deep Learning (T.7.3)**

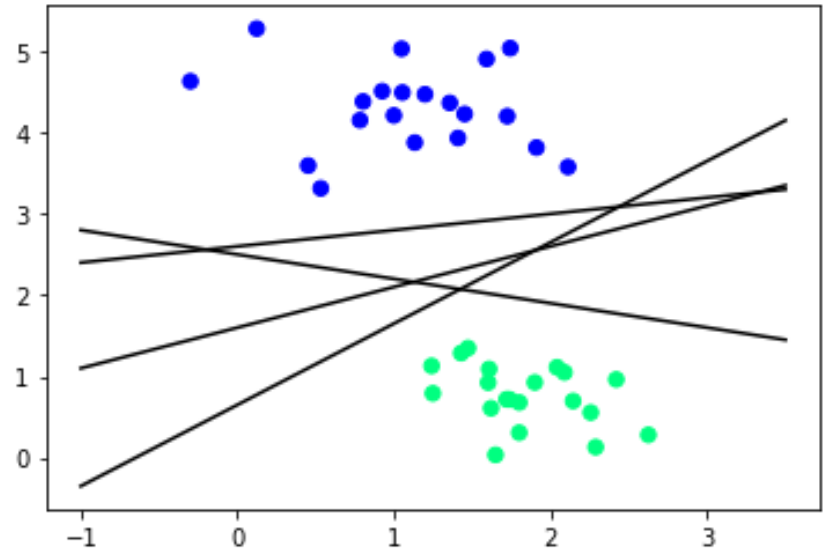
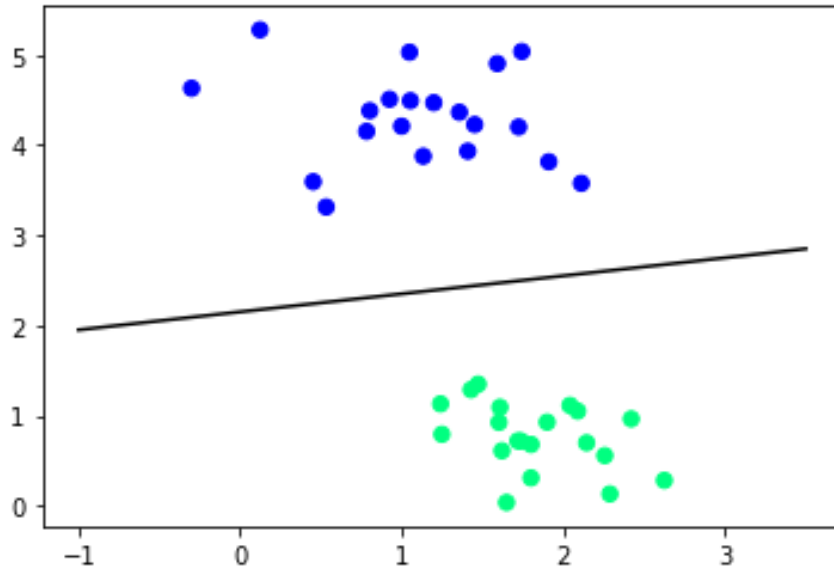
AI3 Style Transfer (T7.4)

AI4 Generación de texto con LSTM (T7.5)

**AG1 Análisis de sentimiento con LSTM (T7.5)**

Test – 16 preguntas

## T7.1 SVM: Clasificación Binaria, Hiperplanos y Separabilidad lineal



Ecuación del hiperplano:  $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p = 0$

Si se cumple, la distancia euclídea de un punto cualquiera  $x^*$  al hiperplano se obtiene sustituyendo las coordenadas del punto en la ecuación del hiperplano:

$$x^* = (x_1^*, x_2^*, \dots, x_p^*)$$

$$\beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \dots + \beta_p x_p^* = d$$

## T7.1 SVM: Clasificador de Margen Máximo

**Idea: Cuanto mejor separe el conjunto de entrenamiento, mejor separará el conjunto de prueba**

Clase  $y_i$  asignada a cada observación  $x_i$ :

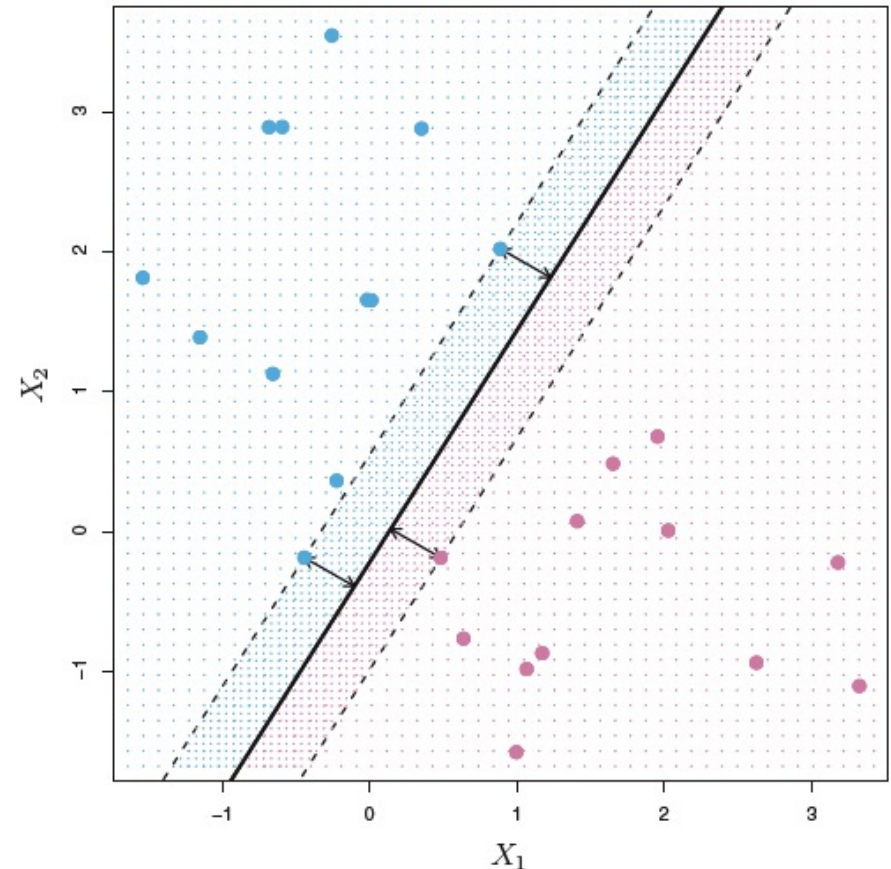
$$y_1, y_2, \dots, y_n \in \{1, -1\}$$

Maximizar  $M$  para  $\beta_0, \beta_1, \dots, \beta_p$

Sujeto a las siguientes restricciones:

$$\sum_{j=1}^p \beta_j^2 = 1$$

$$y_i(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}) \geq M \text{ para cada } i = 1, \dots, n$$





## T7.1 SVM: Clasificador de Vectores Soporte. Soft Margin y vulneraciones

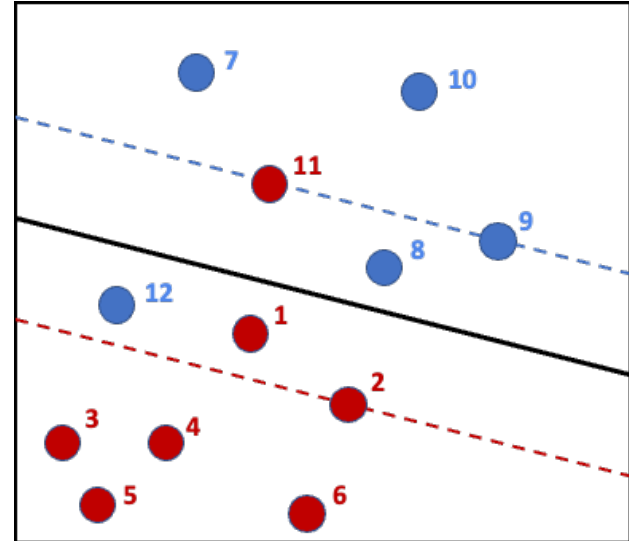
Maximizar  $M$  para  $\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n$

Restricciones:

$$y_i(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}) \geq M(1 - \epsilon_i)$$

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

$$\sum_{j=1}^p \beta_j^2 = 1 \quad \epsilon_i \geq 0, \sum_{i=1}^n \epsilon_i \leq C$$



- Observaciones **rojas**:
  - 3, 4, 5 y 6 están en el lado correcto del margen:  $\{\epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6\} = 0$
  - 2 está sobre el margen:  $\epsilon_2 = 0$
  - 1 está en el lado incorrecto del margen:  $0 < \epsilon_1 < 1$
  - 11 está en el lado incorrecto del hiperplano:  $\epsilon_{11} > 1$
  - **2, 1 y 11 son vectores soporte**

- Observaciones **azules**:
  - 7 y 10 están en el lado correcto del margen:  $\{\epsilon_7, \epsilon_{10}\} = 0$
  - 9 está sobre el margen:  $\epsilon_9 = 0$
  - 8 está en el lado incorrecto del margen:  $0 < \epsilon_8 < 1$
  - 12 está en el lado incorrecto del hiperplano:  $\epsilon_{12} > 1$
  - **9, 8 y 12 son vectores soporte**

## T7.1 SVM: Definición alternativa del parámetro C

Definición clásica:  $\epsilon_i \geq 0, \sum_{i=1}^n \epsilon_i \leq C$

Al aumentar C, aumenta el margen

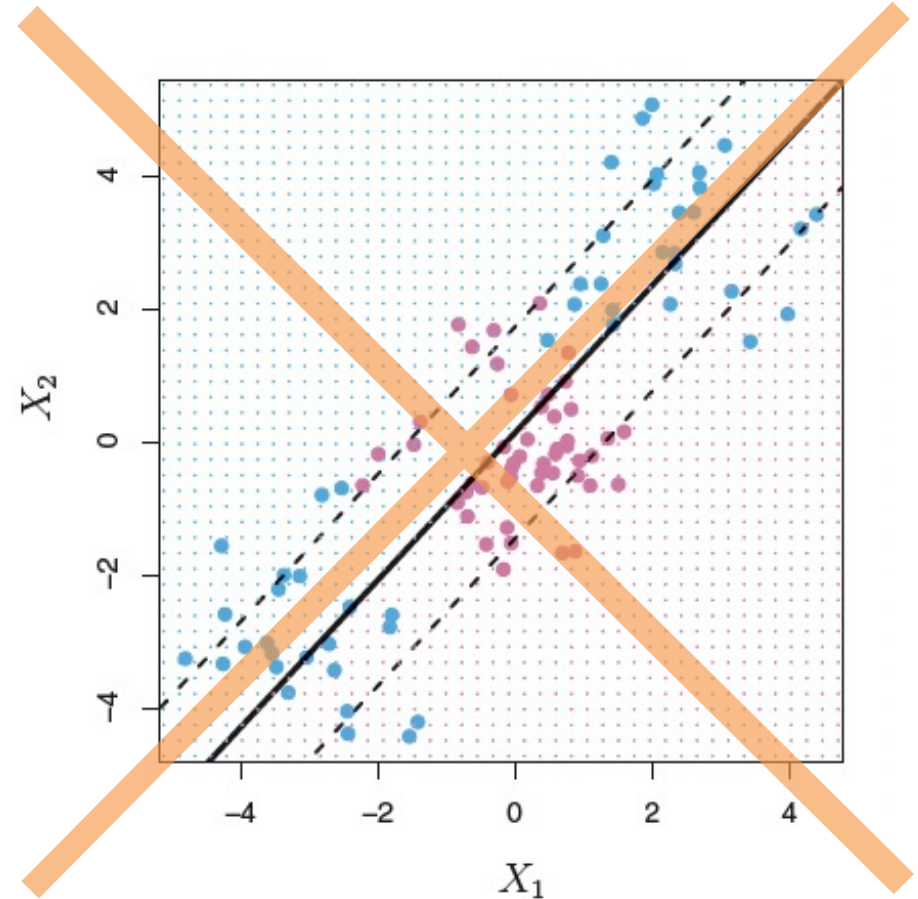
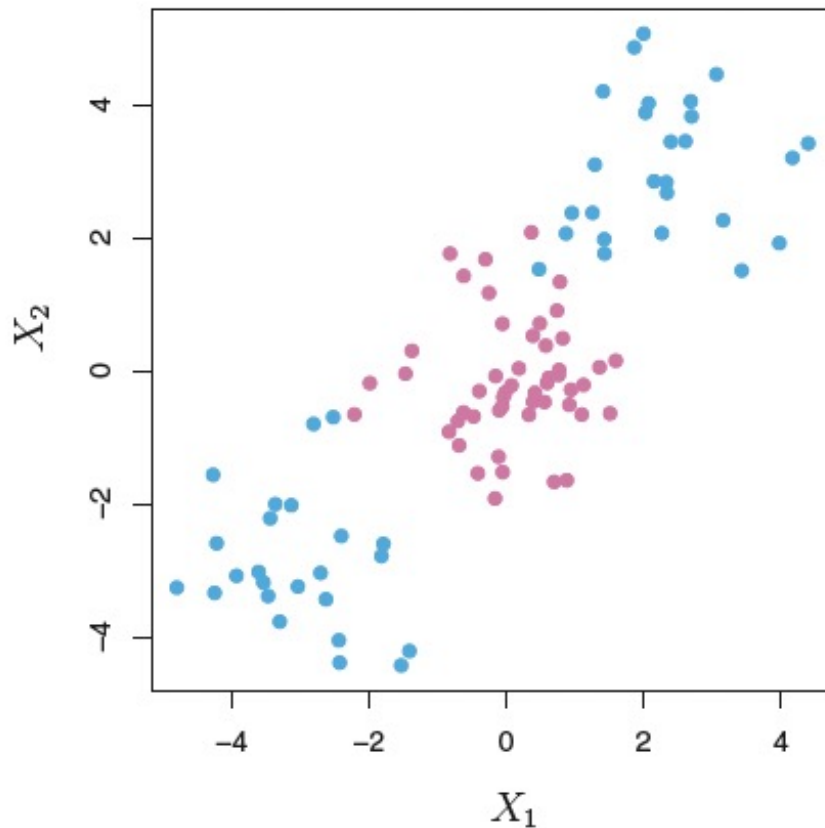
Definición nueva (sklearn):

*Regularization parameter. The strength of the regularization is inversely proportional to C*

Al aumentar C, disminuye el margen

**Tenemos dos efectos opuestos dependiendo de qué definición del parámetro C estemos considerando, por lo que debemos verificar qué definición se implementa en el modelo o paquete con el que estemos trabajando.**

## T7.1 SVM: Conjuntos de datos no tratables con Clasificadores de Vector Soporte



## T7.1 SVM: Máquina de Vectores Soporte. Kernel Lineal y Polinómico

$$f(x) = \beta_0 + \sum_{i \in S} \alpha_i K(x, x_i)$$

$K$  es el **kernel** que transforma el espacio de entrada de manera no lineal al nuevo espacio de características.

**Kernel Lineal** 
$$f(x) = \beta_0 + \sum_{i \in S} \alpha_i \langle x, x_i \rangle$$

**Kernel Polinómico** de grado “d” 
$$K(x_i, x_{i'}) = (1 + \sum_{j=1}^p x_{ij} x_{i'j})^d$$

## T7.1 SVM: Máquina de Vectores Soporte. Kernel Radial o Gaussiano

### Kernel Radial o Gaussiano

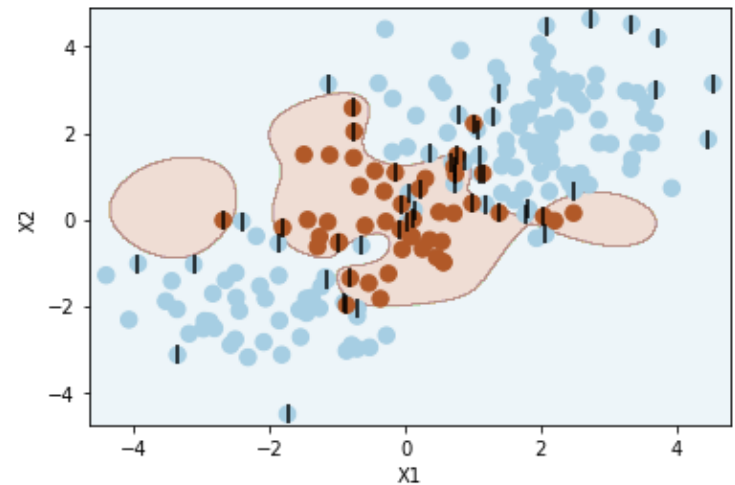
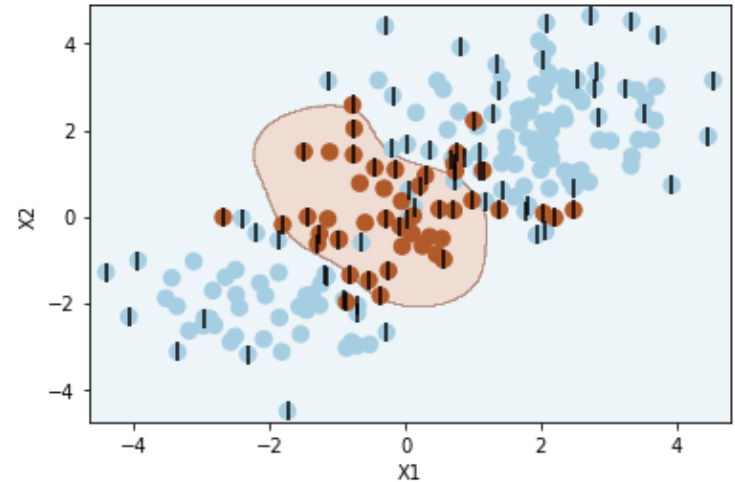
$$K(x_i, x_{i'}) = \exp(-\gamma \sum_{j=1}^p (x_{ij} - x_{i'j})^2)$$

C=1  
90 SV

$$\gamma = \frac{1}{2\sigma^2} > 0$$

Este kernel suele utilizarse cuando se dispone de **pocas características** y **muchos datos de entrenamiento**

C=100  
56 SV



## T7.1 SVM: Actividades

### Actividad guiada

Detección de Cáncer de Mama. Dataset "Breast Cancer Wisconsin"

Se compara la accuracy con varios tipos de Kernel y varios parámetros

### Actividad individual

Clasificación de imágenes.

Se compara la accuracy con:

- dos dataset: MNIST y Fashion-MNIST
- dos tipos de Kernel: Lineal y Gaussiano

Nota: si los modelos tardan mucho en entrenar, se puede reducir el tamaño de los dataset a la mitad (10.000 imágenes)



*Institut de Formació Contínua-IL3*  
UNIVERSITAT DE BARCELONA