Clasificación Supervisada

Máquinas de Vectores de Soporte

Jesús Ariel Carrasco Ochoa
Instituto Nacional de Astrofísica, Óptica y Electrónica

Máquinas de Vectores de Soporte

- Idea central:

  Trasformar los vectores de entrada X (n-dimensionales) en vectores de dimensión más alta Z (incluso de dimensión infinita) en los que el problema pueda solucionarse linealmente

$$
W \cdot Z + b \quad \begin{array}{lll} \geq 0 & \rightarrow & X \in C_0 \\ < 0 & \rightarrow & X \in C_1 \end{array}
$$

Máquinas de Vectores de Soporte Lineales

- Si se tiene una muestra $\boldsymbol{T}$ con $\boldsymbol{m}$ objetos $\boldsymbol{O_1, O_2, \ldots, O_m}$
- Descritos por $\boldsymbol{n}$ atributos reales $\boldsymbol{x_1, x_2, \ldots, x_n}$
- Los objetos están distribuidos en 2 clase $\boldsymbol{C_0}$, $\boldsymbol{C_1}$
- Cada objeto $\boldsymbol{O_j}$ tiene asociada la descripción $\boldsymbol{X_j}$ y el valor $\boldsymbol{y_j}$ tal que $\boldsymbol{y_j}=1$ si $\boldsymbol{O_j} \in \boldsymbol{C_0}$ y $\boldsymbol{y_j}=-1$ si $\boldsymbol{O_j} \in \boldsymbol{C_1}$

Máquinas de Vectores de Soporte Lineales

- Si las clases son linealmente separables podemos encontrar un hiperplano tal que:

- Para los objetos de la muestra en $\boldsymbol{C_0}$

$$W \cdot X_j + b \geq 1$$

- Y para los objetos de la muestra en $\boldsymbol{C_1}$

$$W \cdot X_j + b \leq -1$$

Máquinas de Vectores de Soporte Lineales

- Para todos los objetos de la muestra

$$y_j(W \cdot X_j + b) \geq 1$$

- La distancia de una descripción X_j al hiperplano está dada por

$$d_j = \frac{W \cdot X_j + b}{W \cdot W}$$

Máquinas de Vectores de Soporte Lineales

Entonces

$$d_j W \cdot W = W \cdot X_j + b$$

Multiplicando por y_j

$$y_j d_j W \cdot W = y_j (W \cdot X_j + b) \geq 1$$

Con lo cual

$$y_j d_j \geq \frac{1}{W \cdot W}$$

Máquinas de Vectores de Soporte Lineales

Para los objetos más cercanos al hiperplano

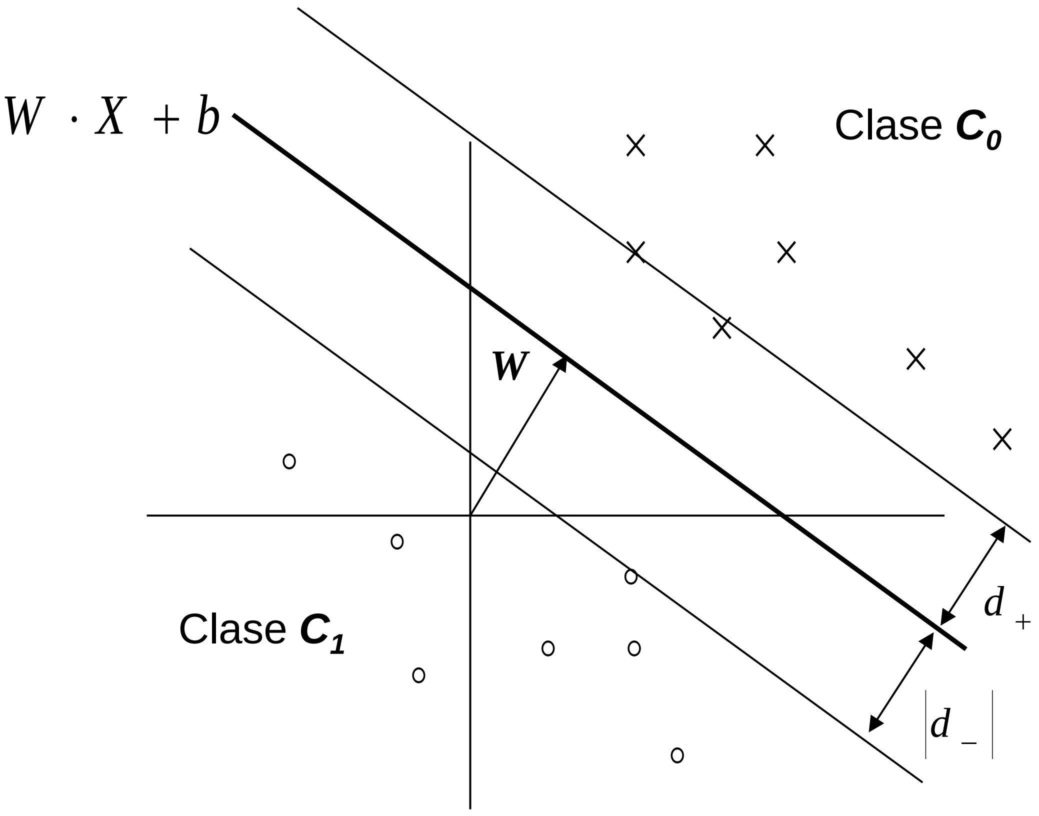
$$y_j d_j = \frac{1}{W \cdot W}$$

Máquinas de Vectores de Soporte Lineales

- Sean $\boldsymbol{d_+}$ y $\boldsymbol{d_-}$ las distancias mínimas al hiperplano, de los objetos de $\boldsymbol{C_0}$ y $\boldsymbol{C_1}$ respectivamente

$$d_+ - d_- = \frac{2}{W \cdot W}$$

Máquinas de Vectores de Soporte Lineales



Máquinas de Vectores de Soporte Lineales

- El objetivo es maximizar esta diferencia, es decir

$$\underset{W,b}{Minimizar} \quad \left\{ \frac{1}{2} W \cdot W \right\}$$

Con las restricciones

$$y_j (W \cdot X_j + b) \geq 1 \quad \forall j = 1, \ldots, m$$

Máquinas de Vectores de Soporte Lineales

Para esto consideramos

$$\frac{1}{2} W \cdot W - \sum_{i=1}^{m} \alpha_i \left[y_j (W \cdot X_j + b) - 1 \right]$$

Máquinas de Vectores de Soporte Lineales

- Para esto se resuelve

$$\underset{\alpha_j}{Maximizar} \left\{ \sum_{j=1}^{m} \alpha_j - \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j X_i \cdot X_j \right\}$$

- Con las restricciones

$$\alpha_j \geq 0 \quad \forall j = 1,\ldots, m$$

$$\sum_{j=1}^{m} \alpha_j y_j = 0$$

Máquinas de Vectores de Soporte Lineales

- Los $\boldsymbol{X_j}$ para los cuales $\boldsymbol{\alpha_j}$>0 se llaman vectores de soporte.

- La solución para $\boldsymbol{W}$ se obtiene como

$$W = \sum_{j=1}^{m} \alpha_j y_j X_j = \sum_{\text{Vectores de Soporte}} \alpha_j y_j X_j$$

- Para los vectores de soporte debe cumplirse

$$y_j (W \cdot X_j + b) = 1$$

Máquinas de Vectores de Soporte Lineales

- Por lo tanto ***b*** se calcula como

$$b = \frac{1}{\#\,\text{Vectores de Soporte}} \sum_{\text{Vectores de Soporte}} \left(y_j - W \cdot X_j \right)$$

- Para clasificar un nuevo objeto ***X***

$$\text{Si} \quad W \cdot X + b \quad \begin{cases} \geq 0 & \rightarrow \quad X \in C_0 \\ < 0 & \rightarrow \quad X \in C_1 \end{cases}$$

Máquinas de Vectores de Soporte Lineales

- Si las clases no son linealmente separables no habrá solución para el problema de optimización
- Para resolver esto se introducen variables para relajar las restricciones
- Así el objetivo es

$$\underset{W,b}{Minimizar} \left\{ \frac{1}{2} W \cdot W + C \sum_{j=1}^{m} \beta_j \right\}$$

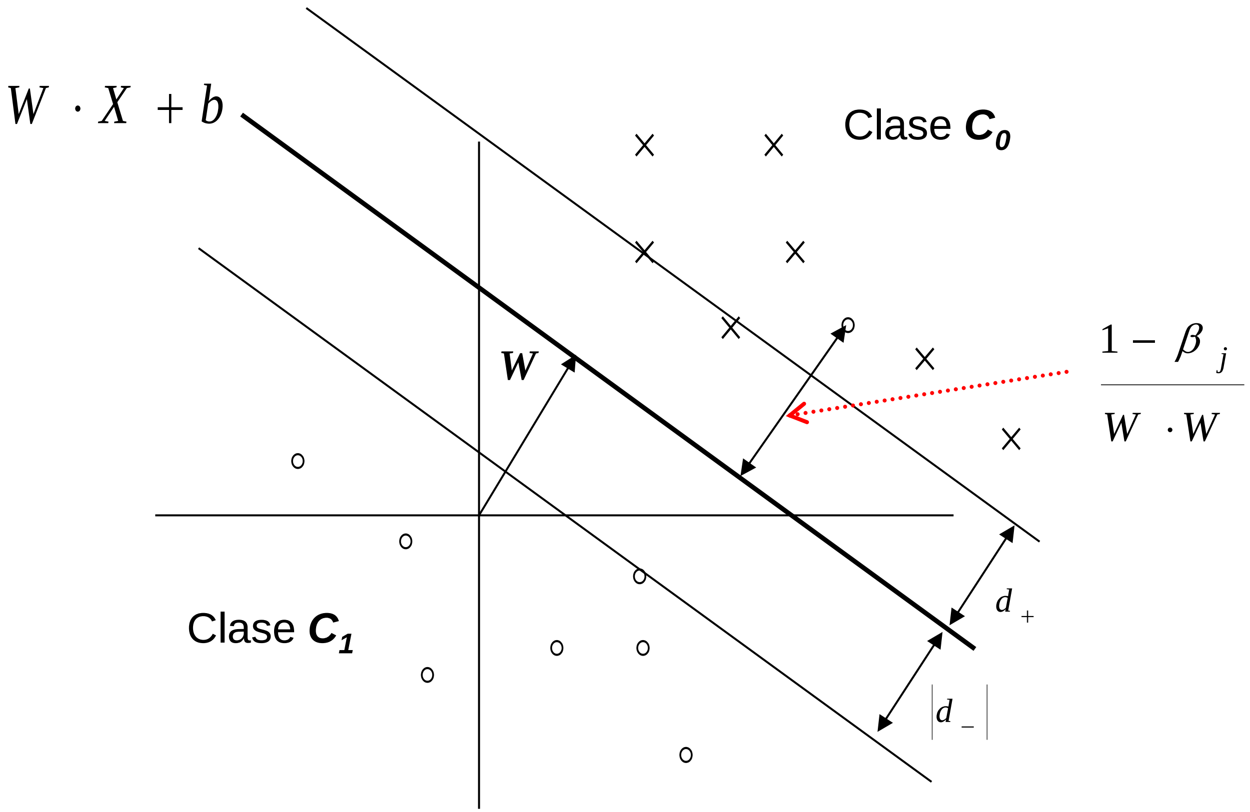
- Con las restricciones

$$y_j (W \cdot X_j + b) \geq 1 - \beta_j \quad \beta_j \geq 0, \quad \forall j = 1,\ldots, m$$

Máquinas de Vectores de Soporte Lineales

- ***C***>0 es un parámetro que determina un compromiso entre el margen de separación y el número de puntos erróneamente clasificados
- Si ***C*** es grande se considera una gran penalización en el número de errores
- Si ***C*** es pequeño se considera poca penalización

Máquinas de Vectores de Soporte Lineales



Máquinas de Vectores de Soporte Lineales

- Para esto se resuelve

$$\underset{\alpha_j}{Maximizar} \left\{ \sum_{j=1}^{m} \alpha_j - \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j X_i \cdot X_j \right\}$$

- Con las restricciones

$$0 \leq \alpha_j \leq C \qquad \forall j = 1, \ldots, m$$

$$\sum_{j=1}^{m} \alpha_j y_j = 0$$

Máquinas de Vectores de Soporte Lineales

- Los $\boldsymbol{X_j}$ para los cuales $\boldsymbol{\alpha_j}$>0 se llaman vectores de soporte.

- La solución para $\boldsymbol{W}$ se obtiene como

$$W = \sum_{j=1}^{m} \alpha_j y_j X_j = \sum_{\substack{\text{Vectores de} \\ \text{Soporte}}} \alpha_j y_j X_j$$

Máquinas de Vectores de Soporte Lineales

- Para los vectores de soporte debe cumplirse
  - Si $0<\alpha_j<C$

$$y_j(W \cdot X_j + b) = 1 \quad \text{con} \quad \beta_j = 0$$

  - Si $\alpha_j=C$

$$y_j(W \cdot X_j + b) < 1 \quad \text{con} \quad \beta_j > 0$$

Máquinas de Vectores de Soporte Lineales

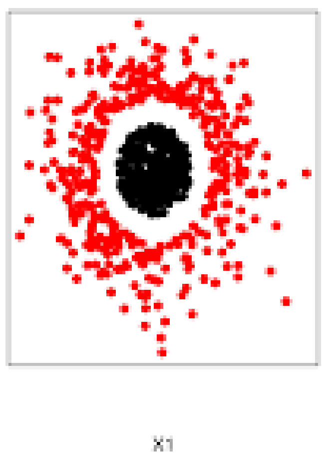
- Por lo tanto ***b*** se calcula como

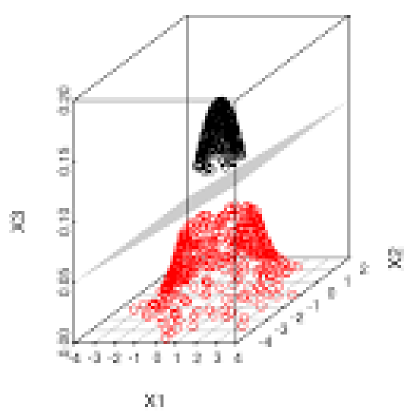
$$b = \frac{1}{\#\,\text{Vectores de Soporte con } \alpha_{\mathrm{j}} < C} \sum_{\substack{\text{Vectores de} \\ \text{Soporte con } \alpha_{\mathrm{j}} < C}} \left(y_j - W \cdot X_j \right)$$

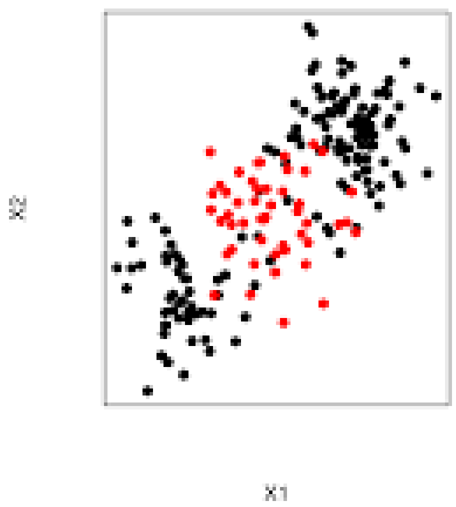
- Para clasificar un nuevo objeto ***X***

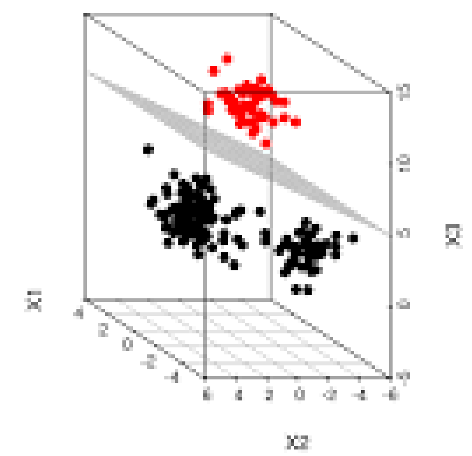
$$\text{Si} \quad W \cdot X + b \quad \begin{cases} \geq 0 & \rightarrow \quad X \in C_0 \\ < 0 & \rightarrow \quad X \in C_1 \end{cases}$$

Máquinas de Vectores de Soporte No Lineales









Máquinas de Vectores de Soporte No Lineales

- Si las clases no son linealmente separables la idea es transformar el espacio en uno de dimensión mayor (posiblemente infinita) donde sí sean linealmente separables

- Si tenemos una función Φ***(X)*** que transforma los vectores $\boldsymbol{X} \in \boldsymbol{R^n}$ en vectores en un espacio de mayor dimensión $\boldsymbol{F} > \boldsymbol{n}$

$$X \in R^{n} \rightarrow \Phi(X) \in R^{F}$$

Máquinas de Vectores de Soporte No Lineales

- Así el objetivo es

$$\underset{W,b}{Minimizar} \quad \left\{ \tfrac{1}{2} W \cdot W \right\}$$

- Con las restricciones

$$y_j (W \cdot \Phi(X_j) + b) \geq 1 \quad \forall j = 1,\ldots, m$$

Máquinas de Vectores de Soporte No Lineales

- Para esto se resuelve

$$\underset{\alpha_j}{Maximizar} \left\{ \sum_{j=1}^{m} \alpha_j - \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \Phi(X_i) \cdot \Phi(X_j) \right\}$$

- Con las restricciones

$$\alpha_j \geq 0 \quad \forall j = 1,\ldots, m$$

$$\sum_{j=1}^{m} \alpha_j y_j = 0$$

Máquinas de Vectores de Soporte No Lineales

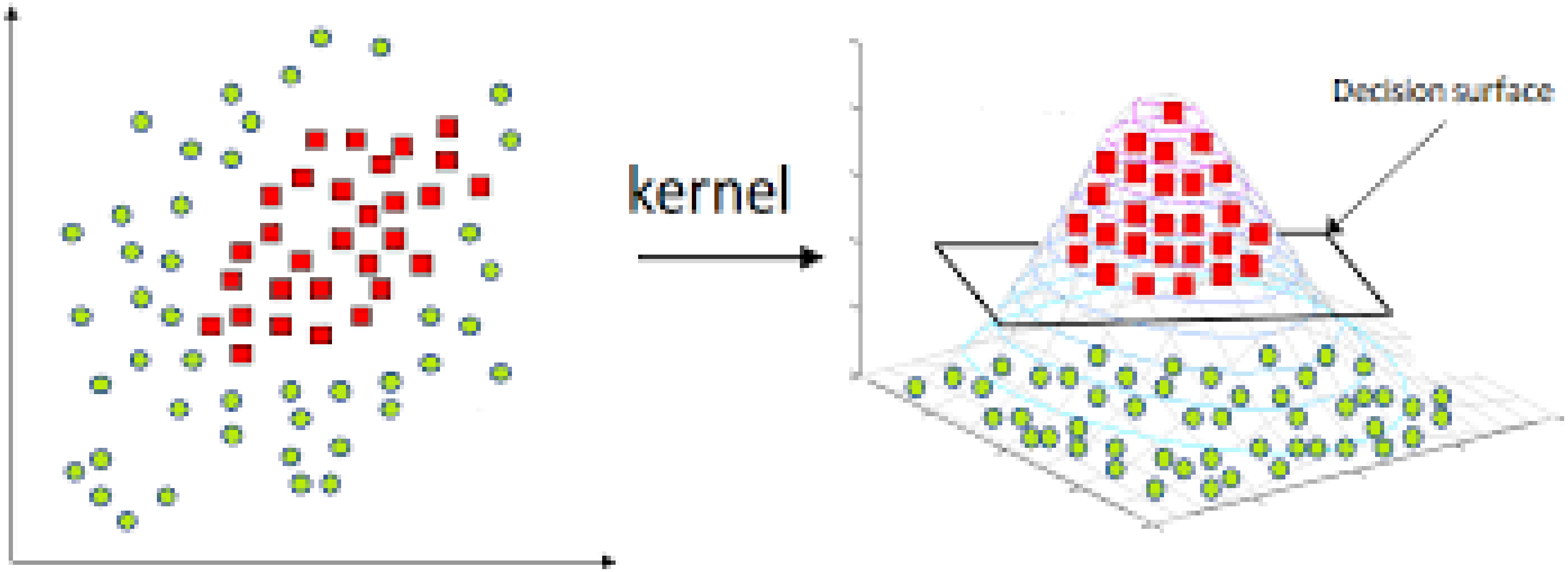
- El producto punto de vectores en el espacio de dimensión $\boldsymbol{F}$ puede ser muy costoso

- Además no conocemos la función Φ***(X)***

- Por lo cual usamos una función *kernel* ***K*** tal que:

$$K(X_i, X_j) = \Phi(X_i) \cdot \Phi(X_j)$$

Máquinas de Vectores de Soporte No Lineales



Máquinas de Vectores de Soporte No Lineales

- Entonces el problema de optimización se transforma de

$$\underset{\alpha_j}{Maximizar} \left\{ \sum_{j=1}^{m} \alpha_j - \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \Phi(X_i) \cdot \Phi(X_j) \right\}$$

- Con las restricciones

$$\alpha_j \geq 0 \quad \forall j = 1,\ldots, m$$

$$\sum_{j=1}^{m} \alpha_j y_j = 0$$

Máquinas de Vectores de Soporte No Lineales

- En

$$\underset{\alpha_j}{Maximizar} \left\{ \sum_{j=1}^{m} \alpha_j - \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j K(X_i, X_j) \right\}$$

- Con las restricciones

$$\alpha_j \geq 0 \quad \forall j = 1, \ldots, m$$

$$\sum_{j=1}^{m} \alpha_j y_j = 0$$

Máquinas de Vectores de Soporte No Lineales

- Algunas funciones kernel comunes

Polinomial $$K(X_i, X_j) = \left[(X_i \cdot X_j) + 1\right]^d$$

Base Radial $$K(X_i, X_j) = e^{\frac{\left\| X_i - X_j \right\|^2}{2\sigma^2}}$$

Sigmoidal $$K(X_i, X_j) = \frac{1}{1 + e^{-c(X_i \cdot X_j)}}$$

Inversión Multicuadrática $$K(X_i, X_j) = \left(\sqrt{\left\| X_i - X_j \right\|^2 + c} \right)^{-1}$$

Máquinas de Vectores de Soporte No Lineales

- Los $\boldsymbol{X_j}$ para los cuales $\boldsymbol{\alpha_j}>0$ se llaman vectores de soporte.

- La solución para $\boldsymbol{W}$ se obtiene como

$$W = \sum_{j=1}^{m} \alpha_j y_j \Phi(X_j) = \sum_{\substack{\text{Vectores de} \\ \text{Soporte}}} \alpha_j y_j \Phi(X_j)$$

Máquinas de Vectores de Soporte No Lineales

- Para los vectores de soporte debe cumplirse

$$y_j(W \cdot \Phi(X_j) + b) = 1$$

Máquinas de Vectores de Soporte No Lineales

- Por lo tanto ***b*** se calcula como

$$b = \frac{1}{\#\,\text{Vectores de Soporte}} \sum_{\text{Vectores de Soporte}} \left(y_j - W \cdot \Phi(X_j) \right)$$

- Sustituyendo ***W***

$$b = \frac{1}{\#\,\text{Vectores de Soporte}} \sum_{\substack{\text{Vectores de} \\ \text{Soporte}\ (j)}} \left(y_j - \left(\sum_{\substack{\text{Vectores de} \\ \text{Soporte}\ (i)}} \alpha_i y_i \Phi(X_i) \right) \cdot \Phi(X_j) \right)$$

Máquinas de Vectores de Soporte No Lineales

- Por lo tanto ***b*** se calcula como

$$b = \frac{1}{\#\,\text{Vectores de Soporte}} \sum_{\substack{\text{Vectores de} \\ \text{Soporte } (j)}} \left(y_j - \sum_{\substack{\text{Vectores de} \\ \text{Soporte } (i)}} \alpha_i y_i \Phi(X_i) \cdot \Phi(X_j) \right)$$

- Usando la función *kernel*

$$b = \frac{1}{\#\,\text{Vectores de Soporte}} \sum_{\substack{\text{Vectores de} \\ \text{Soporte } (j)}} \left(y_j - \sum_{\substack{\text{Vectores de} \\ \text{Soporte } (i)}} \alpha_i y_i K(X_i, X_j) \right)$$

Máquinas de Vectores de Soporte No Lineales

- Para clasificar un nuevo objeto ***X***

$$\text{Si} \quad W \cdot \Phi(X) + b \quad \begin{array}{l} \geq 0 \quad \rightarrow \quad X \in C_0 \\ < 0 \quad \rightarrow \quad X \in C_1 \end{array}$$

Máquinas de Vectores de Soporte No Lineales

- Sustituyendo ***W***

$$\text{Si} \quad \left(\sum_{\substack{\text{Vectores de} \\ \text{Soporte}}} \alpha_j y_j \Phi(X_j) \right) \cdot \Phi(X) + b \quad \begin{array}{l} \geq 0 \quad \rightarrow \quad X \in C_0 \\ < 0 \quad \rightarrow \quad X \in C_1 \end{array}$$

Máquinas de Vectores de Soporte No Lineales

- Nos queda

$$\text{Si} \quad \left(\sum_{\substack{\text{Vectores de} \\ \text{Soporte}}} \alpha_j y_j \Phi(X_j) \cdot \Phi(X) \right) + b \quad \begin{array}{lll} \geq 0 & \rightarrow & X \in C_0 \\ < 0 & \rightarrow & X \in C_1 \end{array}$$

Máquinas de Vectores de Soporte No Lineales

- Usando la función *kernel, tenemos que* para clasificar un nuevo objeto ***X***

$$\text{Si} \quad \left(\sum_{\substack{\text{Vectores de} \\ \text{Soporte}}} \alpha_j y_j K(X_j, X) \right) + b \quad \begin{cases} \geq 0 & \rightarrow \quad X \in C_0 \\ < 0 & \rightarrow \quad X \in C_1 \end{cases}$$

Clasificación Supervisada

Máquinas de Vectores de Soporte

Jesús Ariel Carrasco Ochoa
Instituto Nacional de Astrofísica, Óptica y Electrónica