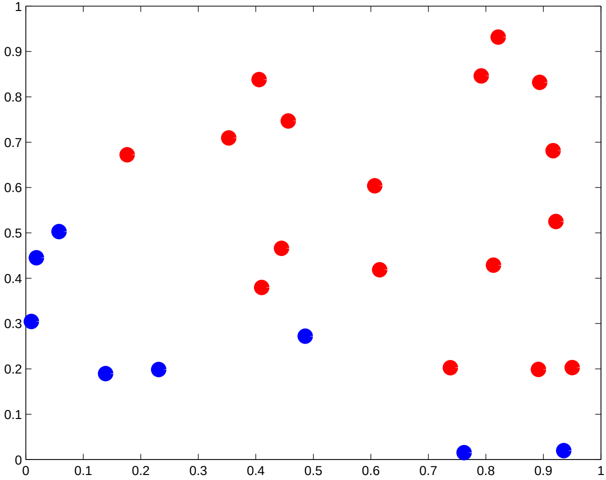
Support Vector Machines

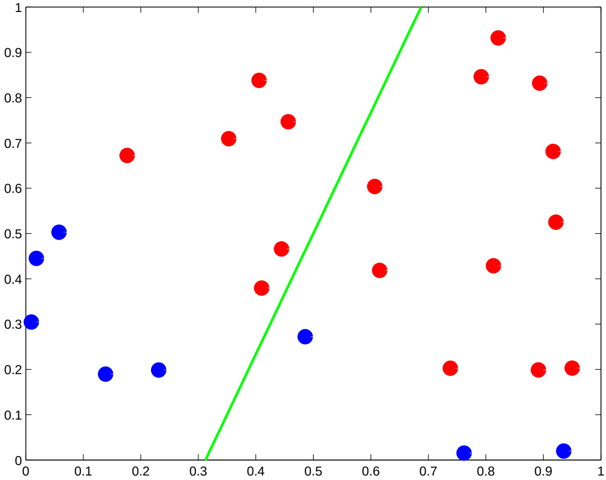
Fernando Lozano

Universidad de los Andes

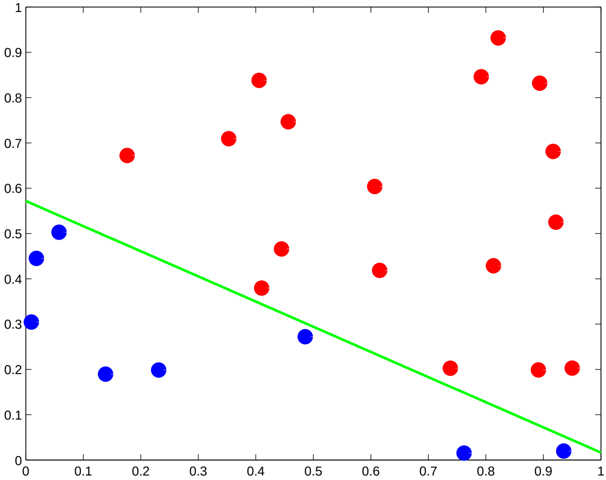
10 de octubre de 2022

El perceptrón





1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Convergencia del Perceptrón

Teorema

Suponga:

Convergencia del Perceptrón

Teorema

Suponga:

- $\|\mathbf{x}_i\| \leq K \in \mathbb{R}, \quad i = 1 \ldots, n.$

Convergencia del Perceptrón

Teorema

Suponga:

- $\|\mathbf{x}_i\| \leq K \in \mathbb{R}, \quad i = 1 \ldots, n.$
- $\exists \hat{\mathbf{w}} \in \mathbb{R}^{d+1}, \delta > 0 \quad \textit{tal que} \quad \hat{\mathbf{w}}^T \mathbf{x}_i \geq \delta \quad i = 1, \ldots, n.$

Convergencia del Perceptrón

Teorema

Suponga:

- $\|\mathbf{x}_i\| \leq K \in \mathbb{R}, \quad i = 1 \ldots, n.$
- $\exists \hat{\mathbf{w}} \in \mathbb{R}^{d+1}, \delta > 0 \quad$ *tal que* $\quad \hat{\mathbf{w}}^T \mathbf{x}_i \geq \delta \quad i = 1, \ldots, n.$

Entonces el algoritmo del perceptrón ejecuta el paso de actualización a lo sumo $\left(\frac{K\|\hat{\mathbf{w}}\|}{\delta}\right)^2$ *veces.*

Convergencia del Perceptrón

Teorema

Suponga:

- $\|\mathbf{x}_i\| \leq K \in \mathbb{R}, \quad i = 1\ldots, n.$
- $\exists \hat{\mathbf{w}} \in \mathbb{R}^{d+1}, \delta > 0 \quad \textit{tal que} \quad \hat{\mathbf{w}}^T \mathbf{x}_i \geq \delta \quad i = 1, \ldots, n.$

Entonces el algoritmo del perceptrón ejecuta el paso de actualización a lo sumo $\left(\frac{K\|\hat{\mathbf{w}}\|}{\delta}\right)^2$ *veces.*

- Deseable tener márgen δ grande.

Convergencia del Perceptrón

Teorema

Suponga:

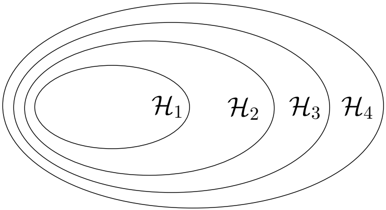
- $\|\mathbf{x}_i\| \leq K \in \mathbb{R}, \quad i = 1 \ldots, n.$
- $\exists \hat{\mathbf{w}} \in \mathbb{R}^{d+1}, \delta > 0 \quad \textit{tal que} \quad \hat{\mathbf{w}}^T \mathbf{x}_i \geq \delta \quad i = 1, \ldots, n.$

Entonces el algoritmo del perceptrón ejecuta el paso de actualización a lo sumo $\left(\frac{K\|\hat{\mathbf{w}}\|}{\delta}\right)^2$ *veces.*

- Deseable tener márgen δ grande.
- Algoritmo del perceptrón no tiene en cuenta el márgen.

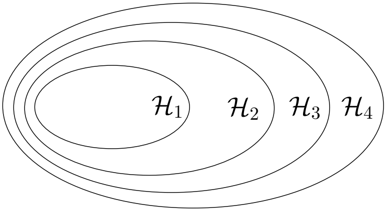
Motivación desde teoría de aprendizaje

- Structural Risk Minimization:



Motivación desde teoría de aprendizaje

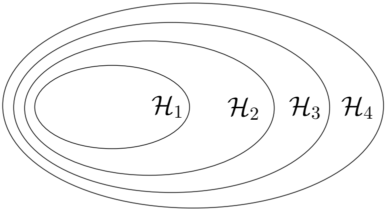
- Structural Risk Minimization:



- Algoritmo de aprendizaje determina la complejidad apropiada de la clase de hipótesis.

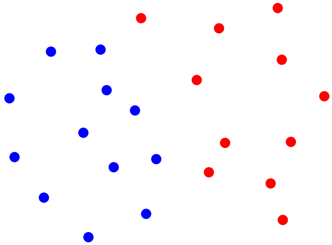
Motivación desde teoría de aprendizaje

- Structural Risk Minimization:

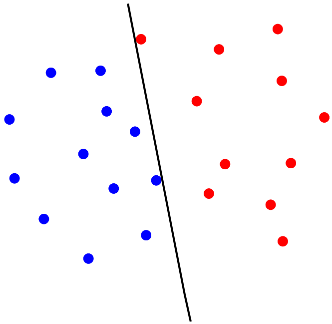


- Algoritmo de aprendizaje determina la complejidad apropiada de la clase de hipótesis.
- Cómo variar suavemente la complejidad?

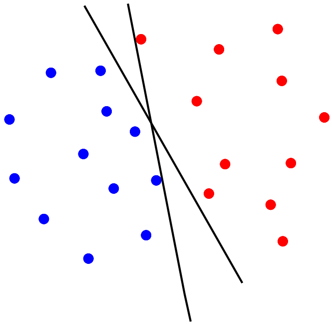
Clasificador con márgen (caso linealmente separable)



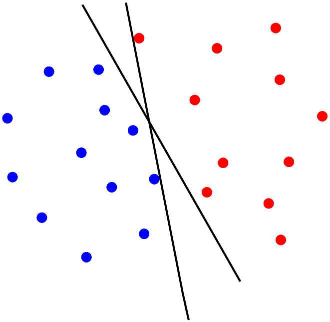
Clasificador con márgen (caso linealmente separable)



Clasificador con márgen (caso linealmente separable)

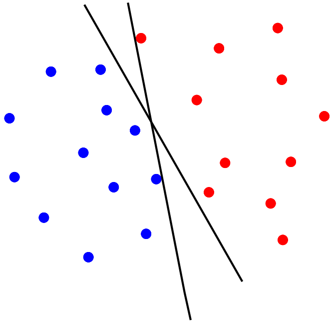


Clasificador con márgen (caso linealmente separable)



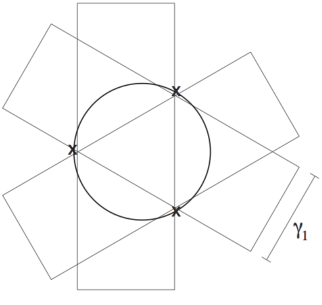
- Márgen: Distancia de un punto a la superficie de separación.

Clasificador con márgen (caso linealmente separable)



- Márgen: Distancia de un punto a la superficie de separación.
- Es deseable tener márgenes grandes.

Margen grande vs. Complejidad



Cómo encontrar un separador lineal con márgen grande?

Cómo encontrar un separador lineal con márgen grande?

- Hipótesis $h(\mathbf{x}) = \text{sign}(\langle \mathbf{w}, \mathbf{x} \rangle + b)$

Cómo encontrar un separador lineal con márgen grande?

- Hipótesis $h(\mathbf{x}) = \text{sign}(\langle \mathbf{w}, \mathbf{x} \rangle + b)$
- Separación:

Cómo encontrar un separador lineal con márgen grande?

- Hipótesis $h(\mathbf{x}) = \text{sign}(\langle \mathbf{w}, \mathbf{x} \rangle + b)$
- Separación:

$$\begin{array}{rl} \langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq 1 & \text{Si } y_i = 1 \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b \leq -1 & \text{Si } y_i = -1 \end{array}$$

Cómo encontrar un separador lineal con márgen grande?

- Hipótesis $h(\mathbf{x}) = \text{sign}(\langle \mathbf{w}, \mathbf{x} \rangle + b)$
- Separación:

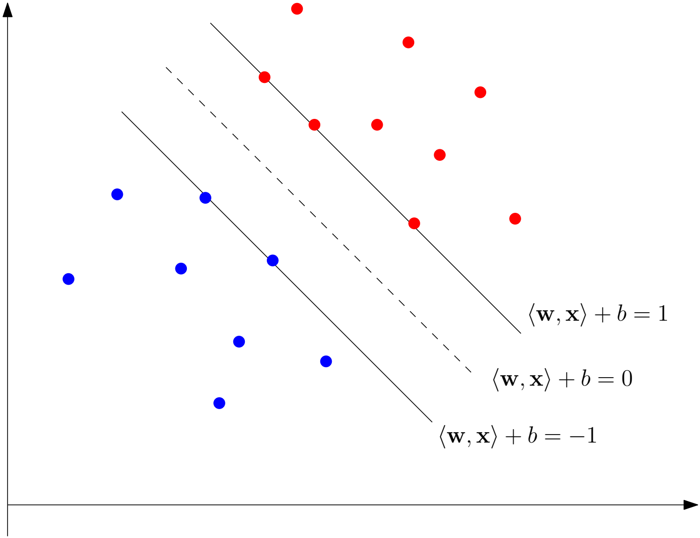
$$\left.\begin{array}{cc} \langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq 1 & \text{Si } y_i = 1 \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b \leq -1 & \text{Si } y_i = -1 \end{array}\right\} y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \geq 0 \quad \forall i$$

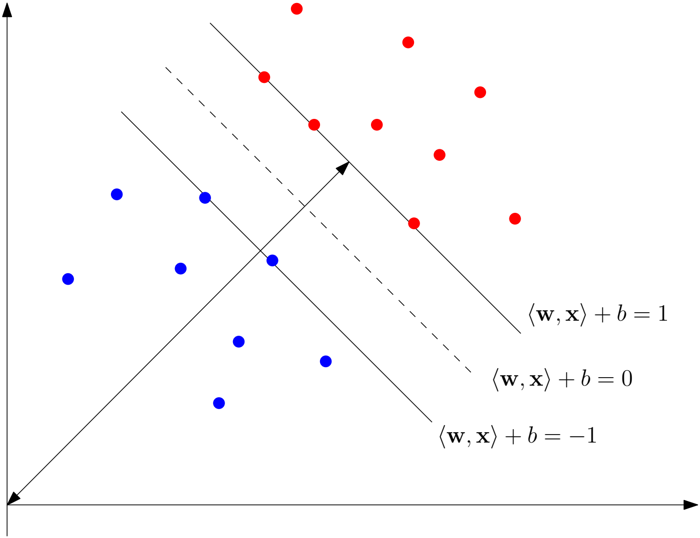
Cómo encontrar un separador lineal con márgen grande?

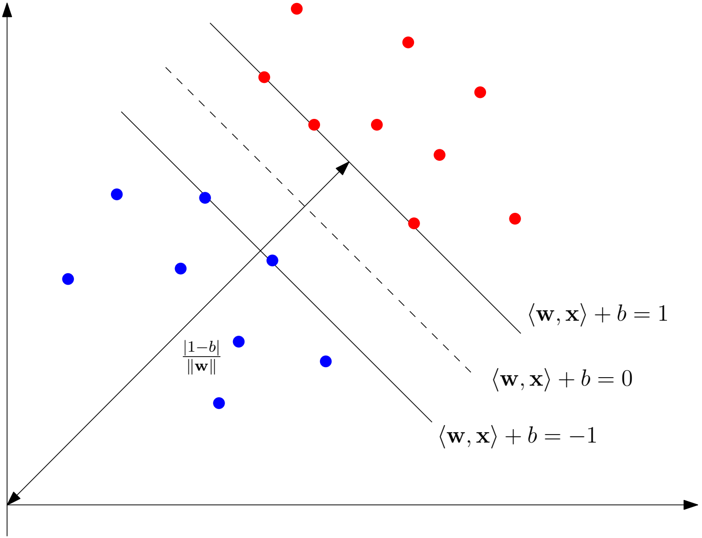
- Hipótesis $h(\mathbf{x}) = \text{sign}(\langle \mathbf{w}, \mathbf{x} \rangle + b)$
- Separación:

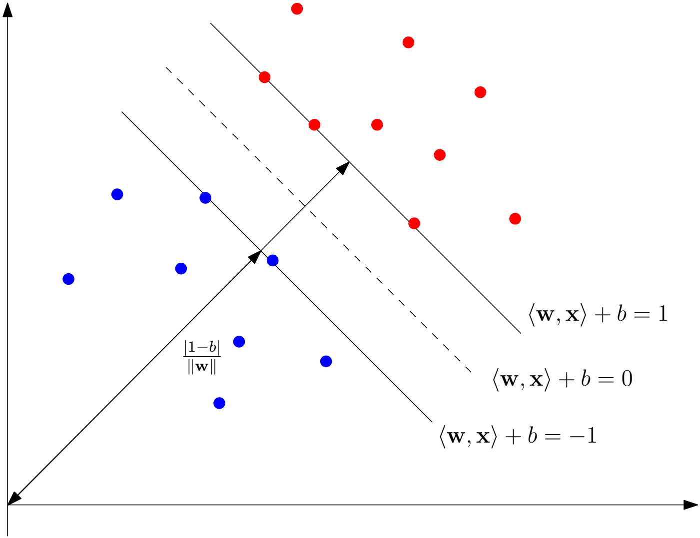
$$\left.\begin{array}{ll} \langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq 1 & \text{Si } y_i = 1 \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b \leq -1 & \text{Si } y_i = -1 \end{array}\right\} y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \geq 0 \quad \forall i$$

- Márgen?

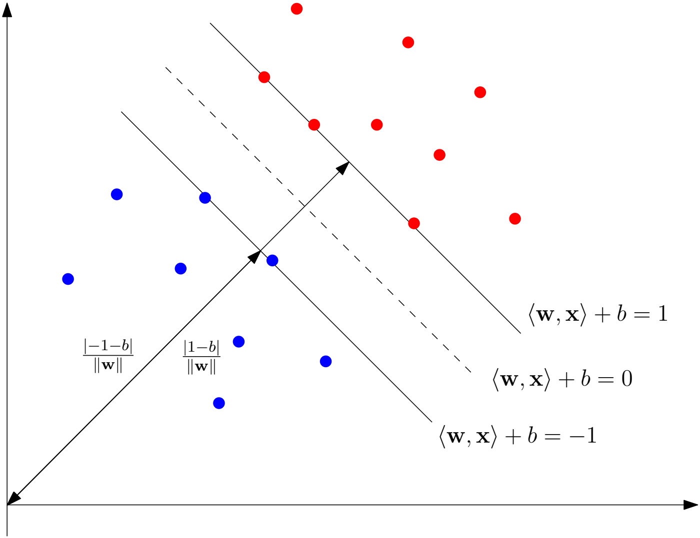

$\langle \mathbf{w}, \mathbf{x} \rangle + b = 1$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = -1$

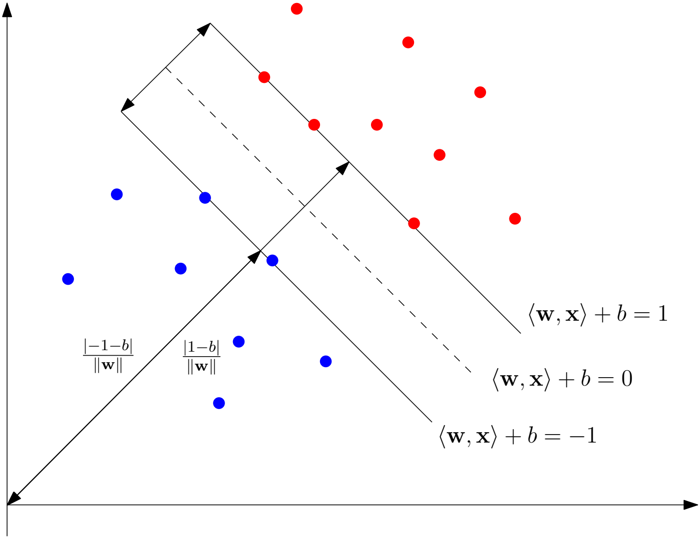

$\langle \mathbf{w}, \mathbf{x} \rangle + b = 1$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = -1$

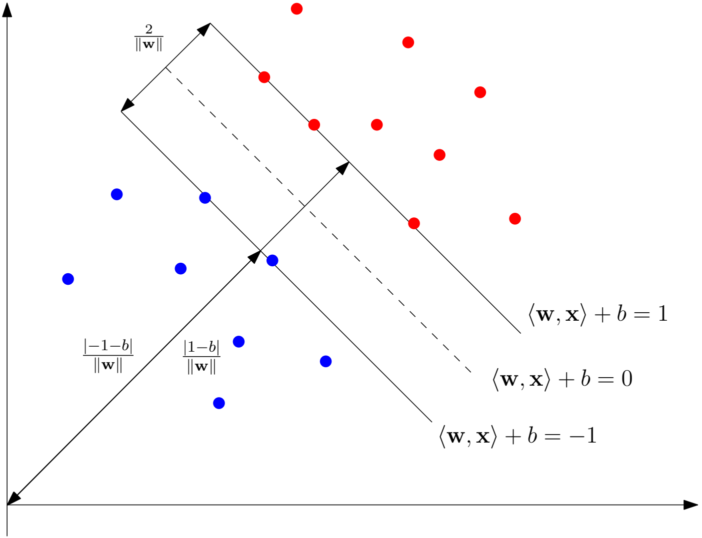

$\frac{|1-b|}{\|\mathbf{w}\|}$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 1$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = -1$



$\langle \mathbf{w}, \mathbf{x} \rangle + b = 1$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = -1$
$\frac{|1-b|}{\|\mathbf{w}\|}$


$\langle \mathbf{w}, \mathbf{x} \rangle + b = 1$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = -1$
$\frac{|-1-b|}{\|\mathbf{w}\|}$
$\frac{|1-b|}{\|\mathbf{w}\|}$


$\langle \mathbf{w}, \mathbf{x} \rangle + b = 1$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = -1$
$\frac{|-1-b|}{\|\mathbf{w}\|}$
$\frac{|1-b|}{\|\mathbf{w}\|}$


$\frac{2}{\|\mathbf{w}\|}$
$\frac{|-1-b|}{\|\mathbf{w}\|}$
$\frac{|1-b|}{\|\mathbf{w}\|}$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 1$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = -1$

- Problema de optimización:

$$
\begin{aligned}
\text{mín} \quad & \frac{1}{2}\|\mathbf{w}\|^2 \\
\text{sujeto a} \quad & y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1 \geq 0 \quad i = 1, \ldots, n
\end{aligned}
$$

- Problema de optimización:

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{sujeto a} \quad & y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1 \geq 0 \quad i = 1, \ldots, n \end{aligned}$$

- Problema de programación cuadrática.

- Problema de optimización:

$$
\begin{array}{rl}
\text{mín} & \dfrac{1}{2}\|\mathbf{w}\|^2 \\
\text{sujeto a} & y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 \geq 0 \quad i = 1, \ldots, n
\end{array}
$$

- Problema de programación cuadrática.
- Problema convexo:

- Problema de optimización:

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{sujeto a} \quad & y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1 \geq 0 \quad i = 1, \ldots, n \end{aligned}$$

- Problema de programación cuadrática.
- Problema convexo:
  - Unico mínimo global.

- Problema de optimización:

$$
\begin{array}{rl}
\text{mín} & \dfrac{1}{2}\|\mathbf{w}\|^2 \\
\text{sujeto a} & y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 \geq 0 \quad i = 1, \ldots, n
\end{array}
$$

- Problema de programación cuadrática.
- Problema convexo:
  - Unico mínimo global.
  - Condiciones de Karush-Kuhn-Tucker (KKT) son suficientes y necesarias para mínimo global y máximo global del problema dual.

- Problema de optimización:

$$
\begin{aligned}
\text{mín} \quad & \frac{1}{2}\|\mathbf{w}\|^2 \\
\text{sujeto a} \quad & y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1 \geq 0 \quad i = 1, \ldots, n
\end{aligned}
$$

- Problema de programación cuadrática.
- Problema convexo:
    - Unico mínimo global.
    - Condiciones de Karush-Kuhn-Tucker (KKT) son suficientes y necesarias para mínimo global y máximo global del problema dual.
    - Solución eficiente.

- Problema de optimización:

$$\begin{aligned} \min \quad & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{sujeto a} \quad & y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1 \geq 0 \quad i = 1, \ldots, n \end{aligned}$$

- Problema de programación cuadrática.
- Problema convexo:
  - Unico mínimo global.
  - Condiciones de Karush-Kuhn-Tucker (KKT) son suficientes y necesarias para mínimo global y máximo global del problema dual.
  - Solución eficiente.
  - Tamaño puede ser grande.

Problema dual

Problema dual

- Introducimos multiplicadores de Lagrange $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n] \geq 0$.

Problema dual

- Introducimos multiplicadores de Lagrange $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n] \geq 0$. El Lagrangiano:

Problema dual

- Introducimos multiplicadores de Lagrange $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n] \geq 0$. El Lagrangiano:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1\right)$$

Problema dual

- Introducimos multiplicadores de Lagrange $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n] \geq 0$. El Lagrangiano:

$$
\begin{aligned}
L(\mathbf{w}, b, \boldsymbol{\alpha}) &= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \right) \\
&= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i (\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) + \sum_{i=1}^{n} \alpha_i
\end{aligned}
$$

Problema dual

- Introducimos multiplicadores de Lagrange $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n] \geq 0$. El Lagrangiano:

$$
\begin{aligned}
L(\mathbf{w}, b, \boldsymbol{\alpha}) &= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \right) \\
&= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i (\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) + \sum_{i=1}^{n} \alpha_i
\end{aligned}
$$

- Minimizamos $L(\mathbf{w}, b, \boldsymbol{\alpha})$ con respecto a $\mathbf{w}, b$ para obtener la función dual:

Problema dual

- Introducimos multiplicadores de Lagrange $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n] \geq 0$. El Lagrangiano:

$$
\begin{aligned}
L(\mathbf{w}, b, \boldsymbol{\alpha}) &= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \right) \\
&= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i (\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) + \sum_{i=1}^{n} \alpha_i
\end{aligned}
$$

- Minimizamos $L(\mathbf{w}, b, \boldsymbol{\alpha})$ con respecto a $\mathbf{w}, b$ para obtener la función dual:

$$
\frac{\partial L}{\partial b} = 0
$$

Problema dual

- Introducimos multiplicadores de Lagrange $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n] \geq 0$. El Lagrangiano:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \right)$$

$$= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i (\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) + \sum_{i=1}^{n} \alpha_i$$

- Minimizamos $L(\mathbf{w}, b, \boldsymbol{\alpha})$ con respecto a $\mathbf{w}, b$ para obtener la función dual:

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0$$

Problema dual

- Introducimos multiplicadores de Lagrange $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n] \geq 0$. El Lagrangiano:

$$
\begin{aligned}
L(\mathbf{w}, b, \boldsymbol{\alpha}) &= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \right) \\
&= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i (\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) + \sum_{i=1}^{n} \alpha_i
\end{aligned}
$$

- Minimizamos $L(\mathbf{w}, b, \boldsymbol{\alpha})$ con respecto a $\mathbf{w}, b$ para obtener la función dual:

$$
\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0
$$

$$
\frac{\partial L}{\partial \mathbf{w}} = 0
$$

Problema dual

- Introducimos multiplicadores de Lagrange $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_n] \geq 0$. El Lagrangiano:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1\right)$$

$$= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) + \sum_{i=1}^{n} \alpha_i$$

- Minimizamos $L(\mathbf{w}, b, \boldsymbol{\alpha})$ con respecto a $\mathbf{w}, b$ para obtener la función dual:

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \mathbf{x}_i$$

- Reemplazando en el Lagrangiano:

- Reemplazando en el Lagrangiano:

$$L = \frac{1}{2}\left\langle \sum_{i=1}^{n}(\alpha_i y_i)\mathbf{x}_i, \sum_{j=1}^{n}(\alpha_j y_j)\mathbf{x}_j \right\rangle - \sum_{i=1}^{n}\alpha_i y_i \left(\left\langle \sum_{j=1}^{n}(\alpha_j y_j)\mathbf{x}_j, \mathbf{x}_i \right\rangle + b \right) + \sum_{i=1}^{n}\alpha_i$$

- Reemplazando en el Lagrangiano:

$$L = \frac{1}{2}\left\langle \sum_{i=1}^{n}(\alpha_i y_i)\mathbf{x}_i, \sum_{j=1}^{n}(\alpha_j y_j)\mathbf{x}_j \right\rangle$$
$$- \sum_{i=1}^{n}\alpha_i y_i \left(\left\langle \sum_{j=1}^{n}(\alpha_j y_j)\mathbf{x}_j, \mathbf{x}_i \right\rangle + b \right) + \sum_{i=1}^{n}\alpha_i$$

- Obtenemos la función dual:

$$g(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

- Reemplazando en el Lagrangiano:

$$L = \frac{1}{2}\left\langle \sum_{i=1}^{n}(\alpha_i y_i)\mathbf{x}_i, \sum_{j=1}^{n}(\alpha_j y_j)\mathbf{x}_j \right\rangle$$
$$- \sum_{i=1}^{n}\alpha_i y_i \left(\left\langle \sum_{j=1}^{n}(\alpha_j y_j)\mathbf{x}_j, \mathbf{x}_i \right\rangle + b \right) + \sum_{i=1}^{n}\alpha_i$$

- Obtenemos la función dual:

$$g(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle = \mathbf{1}^T\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^T\bar{\mathbf{K}}\boldsymbol{\alpha}$$

donde $[\bar{\mathbf{K}}]_{ij} = y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle$

Problema Dual

$$\begin{aligned} \text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\ \text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\ & \boldsymbol{\alpha} \geq 0 \end{aligned}$$

Problema Dual

$$
\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle \\
\text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\
& \boldsymbol{\alpha} \geq 0
\end{aligned}
$$

- Problema de programación cuadrática cóncavo.

Problema Dual

$$
\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\
\text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\
& \boldsymbol{\alpha} \geq 0
\end{aligned}
$$

- Problema de programación cuadrática cóncavo.
- Los datos $\mathbf{x}_i$ sólo aparecen en productos punto.

Condiciones de Karush-Kuhn-Tucker (KKT)

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle\mathbf{w},\mathbf{x}_i\rangle + b\right) - 1 \geq 0 \quad i = 1,\ldots,n$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \geq 0 \quad i = 1, \ldots, n$
2. $\alpha_i \geq 0 \quad i = 1, \ldots, n$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 \geq 0 \quad i = 1, \ldots, n$
2. $\alpha_i \geq 0 \quad i = 1, \ldots, n$
3. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i)\mathbf{x}_i$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \geq 0 \quad i = 1, \ldots, n$
2. $\alpha_i \geq 0 \quad i = 1, \ldots, n$
3. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \mathbf{x}_i$
4. $\sum_{i=1}^{n} \alpha_i y_i = 0$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \geq 0 \quad i = 1, \ldots, n$
2. $\alpha_i \geq 0 \quad i = 1, \ldots, n$
3. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \mathbf{x}_i$
4. $\sum_{i=1}^{n} \alpha_i y_i = 0$
5. $\alpha_i \left(y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \right) = 0$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \geq 0 \quad i = 1, \ldots, n$
2. $\alpha_i \geq 0 \quad i = 1, \ldots, n$
3. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \mathbf{x}_i$
4. $\sum_{i=1}^{n} \alpha_i y_i = 0$
5. $\alpha_i \left(y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 \right) = 0$

- Las condiciones 3 y 5 implican que la suma $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \mathbf{x}_i$ sólo involucra vectores $\mathbf{x}_i$ para los cuales la restricción es activa.

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1 \geq 0 \quad i = 1, \ldots, n$
2. $\alpha_i \geq 0 \quad i = 1, \ldots, n$
3. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \mathbf{x}_i$
4. $\sum_{i=1}^{n} \alpha_i y_i = 0$
5. $\alpha_i \left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1\right) = 0$

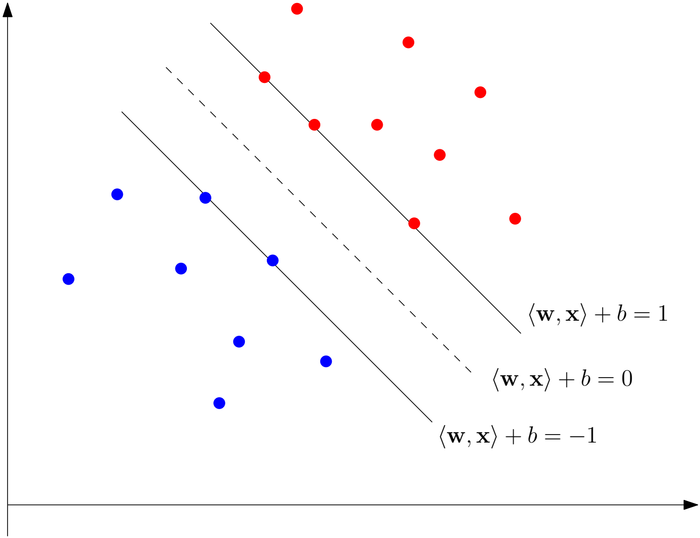
- Las condiciones 3 y 5 implican que la suma $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \mathbf{x}_i$ sólo involucra vectores $\mathbf{x}_i$ para los cuales la restricción es activa.
- Estos vectores se llaman vectores de soporte.

Condiciones de Karush-Kuhn-Tucker (KKT)

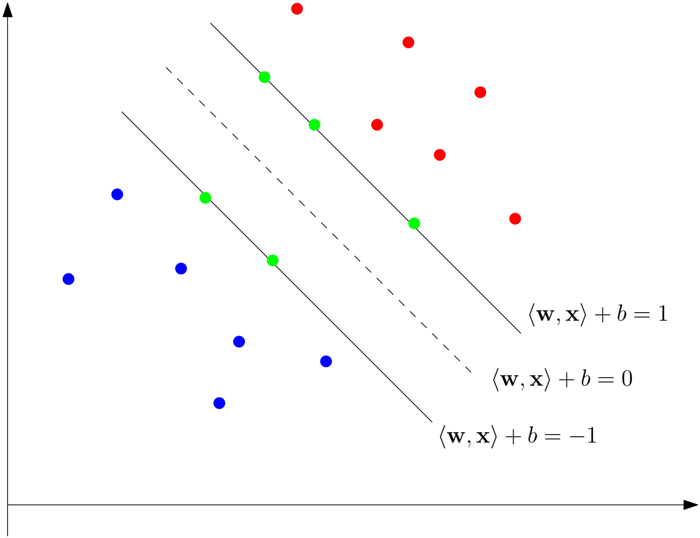
1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1 \geq 0 \quad i = 1, \ldots, n$
2. $\alpha_i \geq 0 \quad i = 1, \ldots, n$
3. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \mathbf{x}_i$
4. $\sum_{i=1}^{n} \alpha_i y_i = 0$
5. $\alpha_i \left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1\right) = 0$

- Las condiciones 3 y 5 implican que la suma $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \mathbf{x}_i$ sólo involucra vectores $\mathbf{x}_i$ para los cuales la restricción es activa.
- Estos vectores se llaman vectores de soporte.
- Si S es el conjunto de vectores de soporte tenemos:

$$\mathbf{w} = \sum_{i\,:\,\mathbf{x}_i \in S} (\alpha_i y_i) \mathbf{x}_i$$



$\langle \mathbf{w}, \mathbf{x} \rangle + b = 1$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = -1$


$\langle \mathbf{w}, \mathbf{x} \rangle + b = 1$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = -1$

Caso no separable

- En general, puede no existir una solución con error cero en los datos.

Caso no separable

- En general, puede no existir una solución con error cero en los datos.
- Se introducen variables de holgura $\zeta_1, \zeta_2, \ldots, \zeta_n \geq 0$:

Caso no separable

- En general, puede no existir una solución con error cero en los datos.
- Se introducen variables de holgura $\zeta_1, \zeta_2, \ldots, \zeta_n \geq 0$:

$$\begin{array}{rl} \langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq 1 - \zeta_i & \text{Si } y_i = 1 \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b \leq -1 + \zeta_i & \text{Si } y_i = -1 \end{array}$$

Caso no separable

- En general, puede no existir una solución con error cero en los datos.
- Se introducen variables de holgura $\zeta_1, \zeta_2, \ldots, \zeta_n \geq 0$:

$$\left.\begin{array}{cc} \langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq 1 - \zeta_i & \text{Si } y_i = 1 \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b \leq -1 + \zeta_i & \text{Si } y_i = -1 \end{array}\right\} y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 + \zeta_i \geq 0 \quad \forall i$$

Caso no separable

- En general, puede no existir una solución con error cero en los datos.
- Se introducen variables de holgura $\zeta_1, \zeta_2, \ldots, \zeta_n \geq 0$:

$$\left.\begin{array}{cc} \langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq 1 - \zeta_i & \text{Si } y_i = 1 \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b \leq -1 + \zeta_i & \text{Si } y_i = -1 \end{array}\right\} y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 + \zeta_i \geq 0 \quad \forall i$$

  - Si hay error en $\mathbf{x}_i$

Caso no separable

- En general, puede no existir una solución con error cero en los datos.
- Se introducen variables de holgura $\zeta_1, \zeta_2, \ldots, \zeta_n \geq 0$:

$$\left.\begin{array}{cc} \langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq 1 - \zeta_i & \text{Si } y_i = 1 \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b \leq -1 + \zeta_i & \text{Si } y_i = -1 \end{array}\right\} y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 + \zeta_i \geq 0 \quad \forall i$$

  - Si hay error en $\mathbf{x}_i \Rightarrow \zeta_i > 1$.

Caso no separable

- En general, puede no existir una solución con error cero en los datos.
- Se introducen variables de holgura $\zeta_1, \zeta_2, \ldots, \zeta_n \geq 0$:

$$\left.\begin{array}{cc} \langle \mathbf{w}, \mathbf{x}_i \rangle + b \geq 1 - \zeta_i & \text{Si } y_i = 1 \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b \leq -1 + \zeta_i & \text{Si } y_i = -1 \end{array}\right\} y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 + \zeta_i \geq 0 \quad \forall i$$

  - Si hay error en $\mathbf{x}_i \Rightarrow \zeta_i > 1$.
  - Luego $\sum_{i=1}^{n} \zeta_i$ es una cota superior del número de errores.

Problema de optimización

$$
\begin{aligned}
\text{mín} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + {\color{red}C}\sum_{i=1}^{n}\zeta_i \\
\text{sujeto a} \quad & y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n \\
& \zeta_i \geq 0
\end{aligned}
$$

Problema de optimización

$$\begin{aligned} \text{mín} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\zeta_i \\ \text{sujeto a} \quad & y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n \\ & \zeta_i \geq 0 \end{aligned}$$

- C es un párametro que indica el balance deseado entre márgen y error.

Problema de optimización

$$
\begin{aligned}
\text{mín} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} \zeta_i \\
\text{sujeto a} \quad & y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n \\
& \zeta_i \geq 0
\end{aligned}
$$

- C es un párametro que indica el balance deseado entre márgen y error.

Problema Dual

$$
\begin{aligned}
L(\mathbf{w}, b, \boldsymbol{\zeta}, \boldsymbol{\alpha}, \boldsymbol{\mu}) = \frac{1}{2}\|\mathbf{w}\|^2 + {\color{red}C} \sum_{i=1}^{n} \zeta_i \\
- \sum_{i=1}^{n} \alpha_i \left(y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 + \zeta_i \right) - \sum_{i=1}^{n} \mu_i \zeta_i
\end{aligned}
$$

Problema Dual

$$L(\mathbf{w}, b, \boldsymbol{\zeta}, \boldsymbol{\alpha}, \boldsymbol{\mu}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\zeta_i$$
$$- \sum_{i=1}^{n}\alpha_i\left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) - \sum_{i=1}^{n}\mu_i\zeta_i$$

Derivando con respecto a las variables primales:

Problema Dual

$$L(\mathbf{w}, b, \boldsymbol{\zeta}, \boldsymbol{\alpha}, \boldsymbol{\mu}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\zeta_i$$

$$- \sum_{i=1}^{n}\alpha_i\left(y_i\left(\langle\mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) - \sum_{i=1}^{n}\mu_i\zeta_i$$

Derivando con respecto a las variables primales:

$$\frac{\partial L}{\partial b} = 0$$

Problema Dual

$$L(\mathbf{w}, b, \boldsymbol{\zeta}, \boldsymbol{\alpha}, \boldsymbol{\mu}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\zeta_i - \sum_{i=1}^{n}\alpha_i\left(y_i\left(\langle\mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) - \sum_{i=1}^{n}\mu_i\zeta_i$$

Derivando con respecto a las variables primales:

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n}\alpha_i y_i = 0$$

Problema Dual

$$L(\mathbf{w}, b, \boldsymbol{\zeta}, \boldsymbol{\alpha}, \boldsymbol{\mu}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\zeta_i$$
$$- \sum_{i=1}^{n}\alpha_i\left(y_i\left(\langle\mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) - \sum_{i=1}^{n}\mu_i\zeta_i$$

Derivando con respecto a las variables primales:

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n}\alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0$$

Problema Dual

$$L(\mathbf{w}, b, \boldsymbol{\zeta}, \boldsymbol{\alpha}, \boldsymbol{\mu}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\zeta_i - \sum_{i=1}^{n}\alpha_i\left(y_i\left(\langle\mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) - \sum_{i=1}^{n}\mu_i\zeta_i$$

Derivando con respecto a las variables primales:

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n}\alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{n}(\alpha_i y_i)\mathbf{x}_i$$

Problema Dual

$$L(\mathbf{w}, b, \boldsymbol{\zeta}, \boldsymbol{\alpha}, \boldsymbol{\mu}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\zeta_i$$

$$- \sum_{i=1}^{n}\alpha_i\left(y_i\left(\langle\mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) - \sum_{i=1}^{n}\mu_i\zeta_i$$

Derivando con respecto a las variables primales:

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n}\alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{n}(\alpha_i y_i)\mathbf{x}_i$$

$$\frac{\partial L}{\partial \zeta_i} = 0$$

Problema Dual

$$L(\mathbf{w}, b, \boldsymbol{\zeta}, \boldsymbol{\alpha}, \boldsymbol{\mu}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\zeta_i - \sum_{i=1}^{n}\alpha_i\left(y_i\left(\langle\mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) - \sum_{i=1}^{n}\mu_i\zeta_i$$

Derivando con respecto a las variables primales:

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n}\alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{n}(\alpha_i y_i)\mathbf{x}_i$$

$$\frac{\partial L}{\partial \zeta_i} = 0 \Rightarrow \alpha_i + \mu_i = C$$

Problema Dual

$$L(\mathbf{w}, b, \boldsymbol{\zeta}, \boldsymbol{\alpha}, \boldsymbol{\mu}) = \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^{n} \zeta_i$$

$$- \sum_{i=1}^{n} \alpha_i \left(y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 + \zeta_i \right) - \sum_{i=1}^{n} \mu_i \zeta_i$$

Derivando con respecto a las variables primales:

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \mathbf{x}_i$$

$$\frac{\partial L}{\partial \zeta_i} = 0 \Rightarrow \alpha_i + \mu_i = C \Rightarrow \alpha_i \leq C$$

$$
\begin{aligned}
L = \frac{1}{2}\left\langle \sum_{i=1}^{n}(\alpha_i y_i)\mathbf{x}_i, \sum_{j=1}^{n}(\alpha_j y_j)\mathbf{x}_j \right\rangle + {\color{red}C}\sum_{i=1}^{n}\zeta_i \\
- \sum_{i=1}^{n}\alpha_i\left(y_i\left(\left\langle \sum_{j=1}^{n}(\alpha_j y_j)\mathbf{x}_j, \mathbf{x}_i \right\rangle + b\right) - 1 + \zeta_i\right) - \sum_{i=1}^{n}\mu_i\zeta_i
\end{aligned}
$$

$$L = \frac{1}{2}\left\langle \sum_{i=1}^{n}(\alpha_i y_i)\mathbf{x}_i, \sum_{j=1}^{n}(\alpha_j y_j)\mathbf{x}_j \right\rangle + C\sum_{i=1}^{n}\zeta_i$$

$$- \sum_{i=1}^{n}\alpha_i \left(y_i \left(\left\langle \sum_{j=1}^{n}(\alpha_j y_j)\mathbf{x}_j, \mathbf{x}_i \right\rangle + b \right) - 1 + \zeta_i \right) - \sum_{i=1}^{n}\mu_i\zeta_i$$

Obtenemos la misma función dual:

$$g(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

$$L = \frac{1}{2}\left\langle \sum_{i=1}^{n}(\alpha_i y_i)\mathbf{x}_i, \sum_{j=1}^{n}(\alpha_j y_j)\mathbf{x}_j \right\rangle + C\sum_{i=1}^{n}\zeta_i$$
$$-\sum_{i=1}^{n}\alpha_i \left(y_i \left(\left\langle \sum_{j=1}^{n}(\alpha_j y_j)\mathbf{x}_j, \mathbf{x}_i \right\rangle + b \right) - 1 + \zeta_i \right) - \sum_{i=1}^{n}\mu_i\zeta_i$$

Obtenemos la misma función dual:

$$g(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle = \mathbf{1}^T\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^T\bar{\mathbf{K}}\boldsymbol{\alpha}$$

donde $[\bar{\mathbf{K}}]_{ij} = y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle$

Problema dual

$$
\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle \\
\text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq C
\end{aligned}
$$

Problema dual

$$
\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle \\
\text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq C
\end{aligned}
$$

- Unico cambio en el dual es cota superior en los multiplicadores α_i.

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$
2. $\zeta_i \geq 0$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$
2. $\zeta_i \geq 0$
3. $\mu_i \geq 0$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$
2. $\zeta_i \geq 0$
3. $\mu_i \geq 0$
4. $\alpha_i + \mu_i = C$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$
2. $\zeta_i \geq 0$
3. $\mu_i \geq 0$
4. $\alpha_i + \mu_i = C$
5. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \mathbf{x}_i$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$
2. $\zeta_i \geq 0$
3. $\mu_i \geq 0$
4. $\alpha_i + \mu_i = C$
5. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i)\mathbf{x}_i$
6. $\sum_{i=1}^{n} \alpha_i y_i = 0$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$
2. $\zeta_i \geq 0$
3. $\mu_i \geq 0$
4. $\alpha_i + \mu_i = C$
5. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i)\mathbf{x}_i$
6. $\sum_{i=1}^{n} \alpha_i y_i = 0$
7. $\alpha_i\left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) = 0$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$
2. $\zeta_i \geq 0$
3. $\mu_i \geq 0$
4. $\alpha_i + \mu_i = C$
5. $\mathbf{w} = \sum_{i=1}^{n}(\alpha_i y_i)\mathbf{x}_i$
6. $\sum_{i=1}^{n} \alpha_i y_i = 0$
7. $\alpha_i\left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) = 0$
8. $\zeta_i \mu_i = 0$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$
2. $\zeta_i \geq 0$
3. $\mu_i \geq 0$
4. $\alpha_i + \mu_i = C$
5. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i)\mathbf{x}_i$
6. $\sum_{i=1}^{n} \alpha_i y_i = 0$
7. $\alpha_i\left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) = 0$
8. $\zeta_i \mu_i = 0$

$$\alpha_i = C$$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$
2. $\zeta_i \geq 0$
3. $\mu_i \geq 0$
4. $\alpha_i + \mu_i = C$
5. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i)\mathbf{x}_i$
6. $\sum_{i=1}^{n} \alpha_i y_i = 0$
7. $\alpha_i\left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) = 0$
8. $\zeta_i \mu_i = 0$

$$\alpha_i = C \Rightarrow \mu_i = 0$$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$
2. $\zeta_i \geq 0$
3. $\mu_i \geq 0$
4. $\alpha_i + \mu_i = C$
5. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i)\mathbf{x}_i$
6. $\sum_{i=1}^{n} \alpha_i y_i = 0$
7. $\alpha_i\left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) = 0$
8. $\zeta_i \mu_i = 0$

$$\alpha_i = C \Rightarrow \mu_i = 0 \Rightarrow \zeta_i \geq 0,$$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$
2. $\zeta_i \geq 0$
3. $\mu_i \geq 0$
4. $\alpha_i + \mu_i = C$
5. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i)\mathbf{x}_i$
6. $\sum_{i=1}^{n} \alpha_i y_i = 0$
7. $\alpha_i\left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) = 0$
8. $\zeta_i \mu_i = 0$

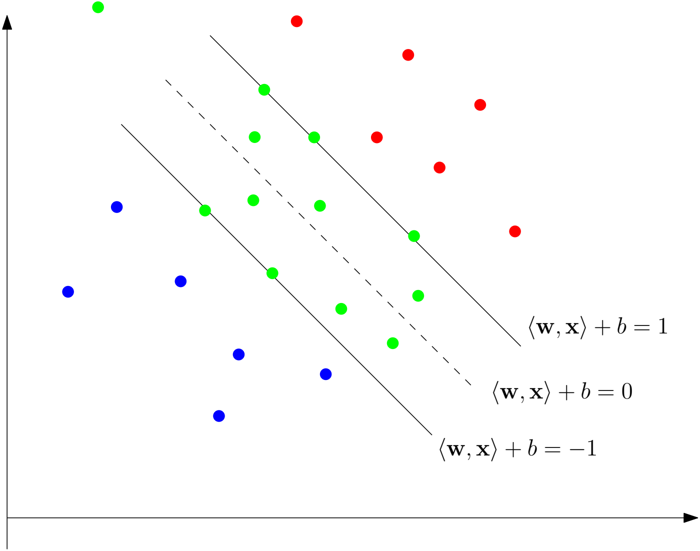
$$\alpha_i = C \Rightarrow \mu_i = 0 \Rightarrow \zeta_i \geq 0, \quad y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) = 1 - \zeta_i$$

Condiciones de Karush-Kuhn-Tucker (KKT)

1. $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i \geq 0 \quad i = 1, \ldots, n$
2. $\zeta_i \geq 0$
3. $\mu_i \geq 0$
4. $\alpha_i + \mu_i = C$
5. $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i)\mathbf{x}_i$
6. $\sum_{i=1}^{n} \alpha_i y_i = 0$
7. $\alpha_i\left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \zeta_i\right) = 0$
8. $\zeta_i \mu_i = 0$

$$\alpha_i = C \Rightarrow \mu_i = 0 \Rightarrow \zeta_i \geq 0, \quad y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) = 1 - \zeta_i$$

- Ahora los vectores de soporte incluyen vectores en los que hay error de márgen o hay error de clasificación.


$\langle \mathbf{w}, \mathbf{x} \rangle + b = 1$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$
$\langle \mathbf{w}, \mathbf{x} \rangle + b = -1$

Formulación equivalente

Formulación equivalente

- Sea $f(\mathbf{x}_i) = \langle \mathbf{w}, \mathbf{x}_i \rangle + b$,

Formulación equivalente

- Sea $f(\mathbf{x}_i) = \langle \mathbf{w}, \mathbf{x}_i \rangle + b$, $m_i = y_i f(\mathbf{x}_i)$ es el margen de clasificación.

Formulación equivalente

- Sea $f(\mathbf{x}_i) = \langle \mathbf{w}, \mathbf{x}_i \rangle + b$, $m_i = y_i f(\mathbf{x}_i)$ es el margen de clasificación.
- Función de costo del márgen:

$$l(m) = \begin{cases} 1 - m & m \leq 1 \\ 0 & m > 1 \end{cases}$$

Formulación equivalente

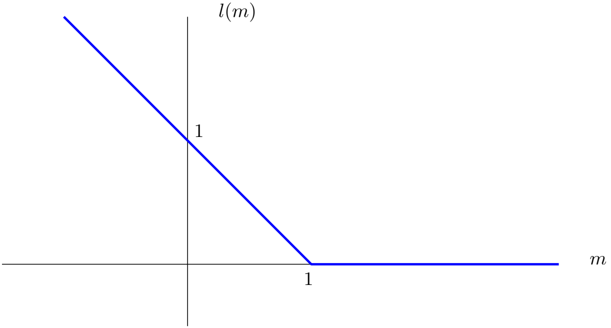
- Sea $f(\mathbf{x}_i) = \langle \mathbf{w}, \mathbf{x}_i \rangle + b$, $m_i = y_i f(\mathbf{x}_i)$ es el margen de clasificación.
- Función de costo del márgen:

$$l(m) = \begin{cases} 1 - m & m \leq 1 \\ 0 & m > 1 \end{cases} = \max \{1 - m, 0\}$$

Formulación equivalente

- Sea $f(\mathbf{x}_i) = \langle \mathbf{w}, \mathbf{x}_i \rangle + b$, $m_i = y_i f(\mathbf{x}_i)$ es el margen de clasificación.
- Función de costo del márgen:

$$l(m) = \begin{cases} 1 - m & m \leq 1 \\ 0 & m > 1 \end{cases} = \max\{1 - m, 0\}$$



$$\min_{\mathbf{w}} \quad l(\mathbf{w}) + \frac{1}{2C}\|\mathbf{w}\|^2$$

$$\min_{\mathbf{w}} \quad l(\mathbf{w}) + \frac{1}{2C}\|\mathbf{w}\|^2$$

- $l(\mathbf{w})$ función de error convexa.

$$\min_{\mathbf{w}} \quad l(\mathbf{w}) + \frac{1}{2C}\|\mathbf{w}\|^2$$

- $l(\mathbf{w})$ función de error convexa.
- $\|\mathbf{w}\|^2$: regularización.

$$\min_{\mathbf{w}} \quad l(\mathbf{w}) + \frac{1}{2C}\|\mathbf{w}\|^2$$

- $l(\mathbf{w})$ función de error convexa.
- $\|\mathbf{w}\|^2$: regularización.
- $\frac{1}{2C}$: constante de regularización.

Caso no lineal

Caso no lineal

- Idea: Proyectar datos a un espacio donde sean más fácilmente separables.

Caso no lineal

- Idea: Proyectar datos a un espacio donde sean más fácilmente separables.

$$
\begin{aligned}
\mathcal{X} &\to \mathcal{H} \\
\mathbf{x} &\mapsto \phi(\mathbf{x})
\end{aligned}
$$

Caso no lineal

- Idea: Proyectar datos a un espacio donde sean más fácilmente separables.

$$
\begin{aligned}
\mathcal{X} &\to \mathcal{H} \\
\mathbf{x} &\mapsto \phi(\mathbf{x})
\end{aligned}
$$

Donde $\mathcal{H}$ es un espacio de Hilbert:

Caso no lineal

- Idea: Proyectar datos a un espacio donde sean más fácilmente separables.

$$\begin{aligned} \mathcal{X} &\to \mathcal{H} \\ \mathbf{x} &\mapsto \phi(\mathbf{x}) \end{aligned}$$

Donde $\mathcal{H}$ es un espacio de Hilbert:

  - Producto punto.

Caso no lineal

- Idea: Proyectar datos a un espacio donde sean más fácilmente separables.

$$\mathcal{X} \to \mathcal{H}$$
$$\mathbf{x} \mapsto \phi(\mathbf{x})$$

Donde $\mathcal{H}$ es un espacio de Hilbert:
  - Producto punto.
  - Completo.

Caso no lineal

- Idea: Proyectar datos a un espacio donde sean más fácilmente separables.

$$
\begin{aligned}
\mathcal{X} &\to \mathcal{H} \\
\mathbf{x} &\mapsto \boldsymbol{\phi}(\mathbf{x})
\end{aligned}
$$

  Donde $\mathcal{H}$ es un espacio de Hilbert:
  - Producto punto.
  - Completo.

- Operamos en el espacio de características $\mathcal{H}$:

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle \longrightarrow \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) \rangle_{\mathcal{H}}$$

Caso no lineal

- Idea: Proyectar datos a un espacio donde sean más fácilmente separables.

$$\mathcal{X} \to \mathcal{H}$$
$$\mathbf{x} \mapsto \boldsymbol{\phi}(\mathbf{x})$$

  Donde $\mathcal{H}$ es un espacio de Hilbert:
  - Producto punto.
  - Completo.
- Operamos en el espacio de características $\mathcal{H}$:

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle \longrightarrow \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) \rangle_{\mathcal{H}}$$

- $\mathcal{X}$ puede ser un conjunto arbitrario.

Problema de optimización

$$
\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \, \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\
\text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq C
\end{aligned}
$$

Problema de optimización

$$
\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \, \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) \rangle_{\mathcal{H}} \\
\text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq C
\end{aligned}
$$

Problema de optimización

$$
\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \left\langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) \right\rangle_{\mathcal{H}} \\
\text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq C
\end{aligned}
$$

- Solución: $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \boldsymbol{\phi}(\mathbf{x}_i)$

Problema de optimización

$$\begin{aligned} \text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \, \color{red}{\langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) \rangle_{\mathcal{H}}} \\ \text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \end{aligned}$$

- Solución: $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \boldsymbol{\phi}(\mathbf{x}_i)$
- Evaluación:

Problema de optimización

$$
\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \left\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \right\rangle_{\mathcal{H}} \\
\text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq C
\end{aligned}
$$

- Solución: $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \phi(\mathbf{x}_i)$
- Evaluación:

$$
\text{sign}\left(\langle \mathbf{w}, \phi(\mathbf{x}) \rangle_{\mathcal{H}} + b \right)
$$

Problema de optimización

$$
\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \, {\color{red} \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) \rangle_{\mathcal{H}}} \\
\text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq C
\end{aligned}
$$

- Solución: $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \boldsymbol{\phi}(\mathbf{x}_i)$
- Evaluación:

$$
\text{sign}\left(\langle \mathbf{w}, \boldsymbol{\phi}(\mathbf{x}) \rangle_{\mathcal{H}} + b \right) = \text{sign}\left(\sum_{i=1}^{n} (\alpha_i y_i) \, {\color{red} \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}) \rangle_{\mathcal{H}}} + b \right)
$$

Problema de optimización

$$
\begin{aligned}
\text{máx} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \left\langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) \right\rangle_{\mathcal{H}} \\
\text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq C
\end{aligned}
$$

- Solución: $\mathbf{w} = \sum_{i=1}^{n} (\alpha_i y_i) \boldsymbol{\phi}(\mathbf{x}_i)$
- Evaluación:

$$
\text{sign}\left(\langle \mathbf{w}, \boldsymbol{\phi}(\mathbf{x}) \rangle_{\mathcal{H}} + b\right) = \text{sign}\left(\sum_{i=1}^{n} (\alpha_i y_i) \left\langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}) \right\rangle_{\mathcal{H}} + b \right)
$$

- Datos sólo aparecen en términos de productos internos.

Ejemplo

$$\phi \, : \, \mathbb{R}^2 \to \mathbb{R}^3$$
$$(x_1, x_2) \mapsto (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1 x_2, x_2^2)$$

Ejemplo

$$\phi \ : \ \mathbb{R}^2 \to \mathbb{R}^3$$

$$(x_1, x_2) \mapsto (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$