

# Máquinas de Vectores de Soporte

## Caso Multiclase

Luis Norberto Zúñiga Morales

29 de agosto de 2022

# Contenido

- 1 Clasificación Multiclase
- 2 Modelo Uno Contra Todos
- 3 Modelo Uno Contra Uno

# Clasificación Multiclase

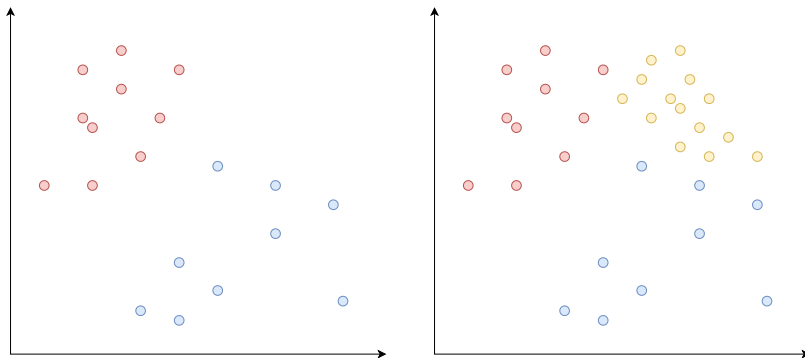


Figura: Clasificación binaria vs clasificación multiclase.

# Clasificación Multiclase

## Three Type of Classification Tasks

YAHOO!  
JAPAN

### Binary Classification



- Spam
- Not spam

### Multiclass Classification



- Dog
- Cat
- Horse
- Fish
- Bird
- ...

### Multi-label Classification



- Dog
- Cat
- Horse
- Fish
- Bird
- ...

**Figura:** Clasificación binaria vs Clasificación multiclase vs Clasificación Multietiqueta.

# Clasificación Multiclase

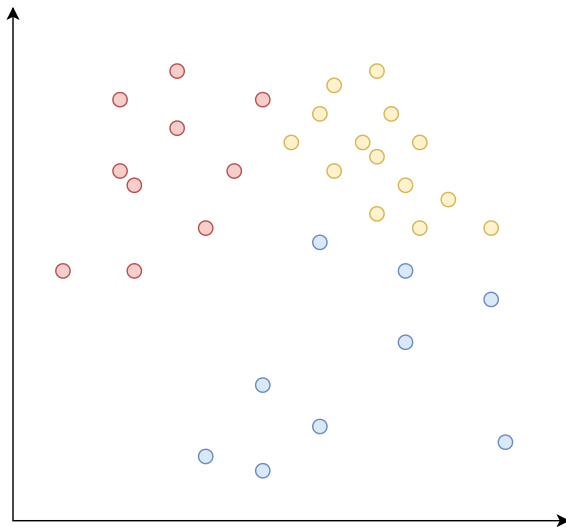
- Originalmente las máquinas de vectores de soporte se plantearon considerando un modelo de **clasificación binaria**, es decir, solo separan dos clases.
- Múltiples enfoques se han considerado para extender el algoritmo para el caso de clasificación multiclase.
- Se consideran dos formas de atacar el problema:
  - **Uno contra todos**
  - **Uno contra uno**

# Modelo Uno Contra Todos

## Modelo Uno Contra Todos

- El enfoque **uno contra todos** construye  $n$  modelos de MVS, uno para cada clase considerada.
- El  $i$ -ésimo modelo considera la  $j$ -ésima clase y sus elementos como la clase positiva y las  $n - 1$  clases restantes se consideran como las negativas.

# Modelo Uno Contra Todos



# Modelo Uno Contra Todos

Dado el conjunto de datos  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_L, y_L)$ , la  $j$ -ésima MVS resuelve el siguiente problema:

$$\begin{aligned} \text{mín} \quad & \frac{1}{2} \|\mathbf{w}^j\|^2 + C \sum_{i=1}^L \xi_i^j \\ \text{sujeto a} \quad & (\mathbf{w}^j)^T \cdot \phi(\mathbf{x}_i) + b^j \geq 1 - \xi_i^j \quad \text{si } y_j = i \\ & (\mathbf{w}^j)^T \cdot \phi(\mathbf{x}_i) + b^j \leq -1 + \xi_i^j \quad \text{si } y_j \neq i \\ & \xi_i^j \geq 0 \end{aligned} \tag{1}$$



# Modelo Uno Contra Todos

Al resolver la ecuación (1) se tienen  $n$  funciones de decisión:

$$(\mathbf{w}^1)^T \phi(\mathbf{x}) + b^1$$

$$\vdots$$

$$(\mathbf{w}^n)^T \phi(\mathbf{x}) + b^n$$

$\mathbf{x}_i$  pertenece a la clase con el mayor valor en la función de decisión:

$$y = \operatorname{argmax}_{j \in \{1, \dots, n\}} (\mathbf{w}^j)^T \phi(\mathbf{x}_i) + b^j. \quad (2)$$

# Modelo Uno Contra Uno

## Modelo Uno Contra Uno

- Este enfoque construye  $n(n - 1)/2$  clasificadores donde cada uno se entrena con información de dos clases.
- Considerando los datos de la clase  $j$  y la clase  $k$  se resuelve el siguiente problema:

$$\begin{aligned} \text{mín} \quad & \frac{1}{2} \|\mathbf{w}^{jk}\|^2 + C \sum_{i=1}^L \xi_i^{jk} \\ \text{sujeto a} \quad & (\mathbf{w}^{jk})^T \cdot \phi(\mathbf{x}_i) + b^{jk} \geq 1 - \xi_i^{jk} \quad \text{si } y_i = j \\ & (\mathbf{w}^{jk})^T \cdot \phi(\mathbf{x}_i) + b^{jk} \leq -1 + \xi_i^{jk} \quad \text{si } y_i = k \\ & \xi_i^{jk} \geq 0 \end{aligned} \quad (3)$$

# Modelo Uno Contra Uno

Hsu y Lin [1] deciden usar la siguiente estrategia basada en votos:

- Si  $\text{sign}((\mathbf{w}^{jk})^T \phi(\mathbf{x}) + b^{jk})$  dice que  $\mathbf{x}$  pertenece a la  $j$ -ésima clase, se suma un voto a esa clase.
- Si dice que pertenece a la clase  $k$ , entonces se da el voto a la clase  $k$ .
- La clase que se elige para  $\mathbf{x}$  es aquella que tenga más votos.

- [1] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2):415–425, 2002.