

# DESCOMPOSICIÓN DE PROBLEMAS MULTICLASE

---

Minería de Datos: Preprocesamiento y clasificación

# Descripción

- **Temario:**

- Modelos no lineales.
- Árboles de Decisión. Multiclasificadores.
- Descomposición de problemas multiclase.
- Aprendizaje de Reglas.
- Máquinas soporte vectorial (SVM).
- Preprocesamiento de Datos.

- **Bibliografía:**

- “An Introduction to Statistical Learning with Applications in R”, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Springer, 2013
- “Introduction to Data Mining”, Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Pearson, 2013 .
- “Foundations of Rule Learning”, Johannes Fürnkranz, Dragan Gamberger, Nada Lavrac, Springer, 2012.
- “Data Preprocessing in Data Mining”. Salvador García, Julián Luengo, Francisco Herrera, Springer, 2015.

- **Relacionado con:** Introducción a la Programación para Ciencia de Datos e Introducción a la Ciencia de Datos

# Descripción

- **Temario:**

- Modelos no lineales.
- Árboles de Decisión. Multiclasificadores.
- **Descomposición de problemas multiclase.**
- Aprendizaje de Reglas.
- Máquinas soporte vectorial (SVM).
- Preprocesamiento de Datos.

- **Bibliografía:**

- “An Introduction to Statistical Learning with Applications in R”, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Springer, 2013.
- “Introduction to Data Mining”, Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Pearson, 2013.
- “Foundations of Rule Learning”, Johannes Fürnkranz, Dragan Gamberger, Nada Lavrac, Springer, 2012.
- “Data Preprocessing in Data Mining”. Salvador García, Julián Luengo, Francisco Herrera, Springer, 2015.

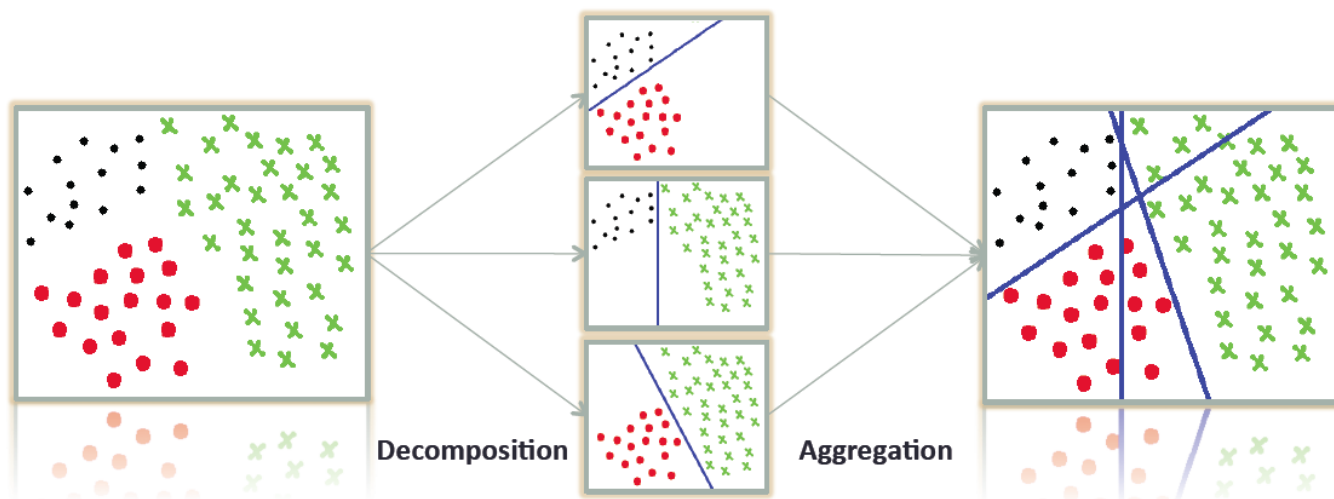
- **Relacionado con:** Introducción a la Programación para Ciencia de Datos e Introducción a la Ciencia de Datos

## Descomposición de problemas multiclase

- Diversos algoritmos de clasificación están limitados a resolver **problemas binarios**, es decir, con dos clases.
- Una manera de abordar el problema de clasificación multi-clase es **descomponerlo en problemas binarios** y construir un modelo de clasificación para cada uno.
- Dos estrategias comunes de descomposición de problemas multi-clase son:
  - **Uno contra uno** (OVO, one-versus-one).
  - **Uno contra todos** (OVA, one-versus-all).
- Otra estrategia menos común es la denominada “**todos y uno**” (A&O, all-and-one), la cual es una combinación de OVO y OVA.

# Descomposición OVO

- Para un problema de clasificación con  $c$  clases se crean  $c(c-1)/2$  modelos de clasificación binarios, los cuales corresponden a todas las posibles combinaciones entre pares de clases.
- Cada clasificador binario aprende a distinguir entre las  $i$ -ésima y  $j$ -ésima clases, por lo que se entrena únicamente con instancias de dichas clases.
- Para clasificar un patrón, se combinan las respuestas de todos los  $c(c-1)/2$  clasificadores construidos.
- Clasificación con OVO para tres clases:



# Descomposición OVO

- Sea  $r_{i,j}$  la respuesta de un clasificador entrenado para discriminar entre las clases  $\omega_i$  y  $\omega_j$ , tal que  $r_{i,j} \in [-\infty, \infty]$  y  $r_{j,i} = -r_{i,j}$ .
- La regla de decisión es:

$$\text{Decidir} \begin{cases} \omega_i & \text{si } r_{i,j} > 0 \\ \omega_j & \text{otro caso} \end{cases}$$

- Sea  $R$  la matriz de respuestas de todos los clasificadores creados mediante la descomposición OVO para un patrón arbitrario:

$$R = \begin{pmatrix} 0 & r_{1,2} & \cdots & r_{1,j} & \cdots & r_{1,c} \\ r_{2,1} & 0 & \cdots & r_{2,j} & \cdots & r_{2,c} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ r_{i,1} & r_{i,2} & \cdots & 0 & \cdots & r_{i,c} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{c,1} & r_{c,2} & \cdots & r_{c,j} & \cdots & 0 \end{pmatrix}$$

- La predicción de clase se hace usando la información de la matriz  $R$  aplicando una estrategia de agregación.

# Descomposición OVO-MV

- En la estrategia de agregación “voto mayoritario” (MV) la clase ganadora es aquella que haya tenido un mayor número de respuestas positivas y se expresa como:

$$\omega_i = \arg \max_{i=1,\dots,c} \sum_{1 \leq j \leq c} v_{i,j}, \quad \text{para } i \neq j$$

- donde

$$v_{i,j} = \begin{cases} 1 & \text{si } r_{i,j} > 0 \\ 0 & \text{otro caso} \end{cases}$$

- Ventaja: es una estrategia simple con resultados competitivos.
- Desventaja: es posible que resulten zonas en el espacio de características que produzcan empates en el número de votos.

# Descomposición OVO-WV

- En la estrategia de agregación “voto ponderado” (WV) la clase ganadora es aquella cuya suma de respuestas sea la de mayor magnitud y se expresa como:

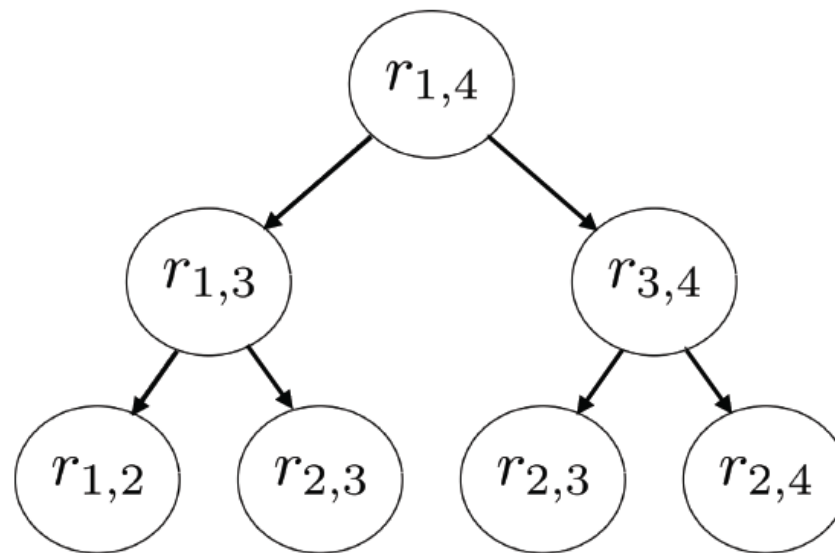
$$\omega_i = \arg \max_{i=1,\dots,c} \sum_{1 \leq j \leq c} r_{i,j}, \quad \text{para } i \neq j$$

- Ventaja: es poco probable que existan zonas en el espacio de características que produzcan empates con una misma suma de respuestas.
- Desventaja: es posible que la clase con máxima suma de respuestas no sea la clase verdadera.



# Descomposición OVO-DDAG

- En la estrategia de agregación “grafo acíclico dirigido de decisión” (DDAG) se crea un árbol binario de decisión donde cada nodo representa la respuesta de un clasificador.
- De esta manera, cuando una se predice una clase, otra clase es descartada, por lo que es necesario descartar la pertenencia a  $c-1$  clases.
- La clase ganadora es aquella con respuesta positiva en el último nodo.

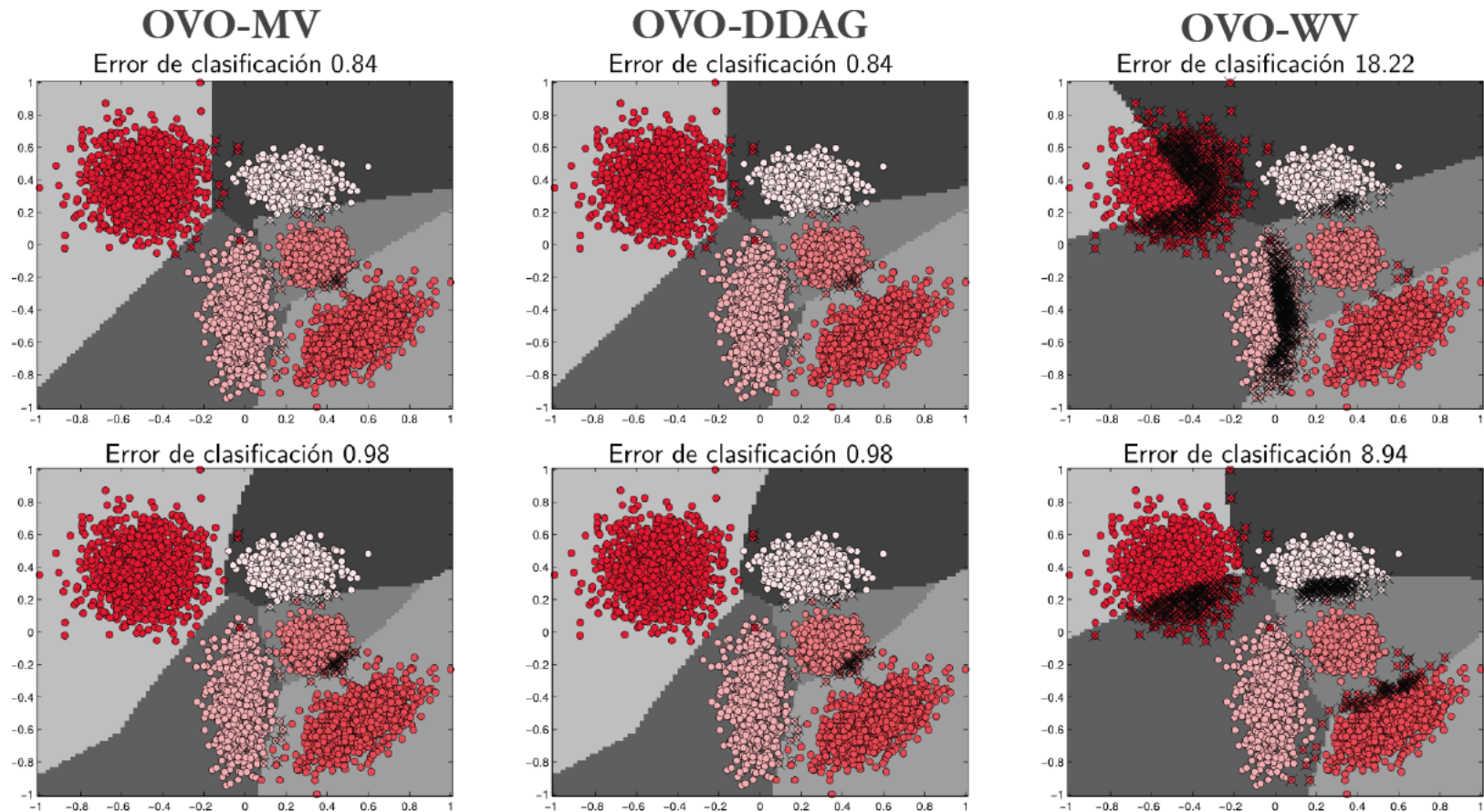


Árbol binario de decisión para cuatro clases.

# Descomposición OVO-DDAG

- En los nodos superiores se colocan los clasificadores que produzcan una mayor separación entre clases, aunque no hay evidencia sustancial de que el ordenamiento de los nodos con alguna preferencia mejore la clasificación.
- Ventajas: tiene el mismo desempeño que OVO-MV para los casos en que el número de respuestas positivas para una clase es  $c-1$ . Además, no hay ambigüedad en la predicción de clase.
- Desventaja: el proceso de construcción y evaluación del grafo puede resultar más costoso en comparación de la estrategia OVO-MV.

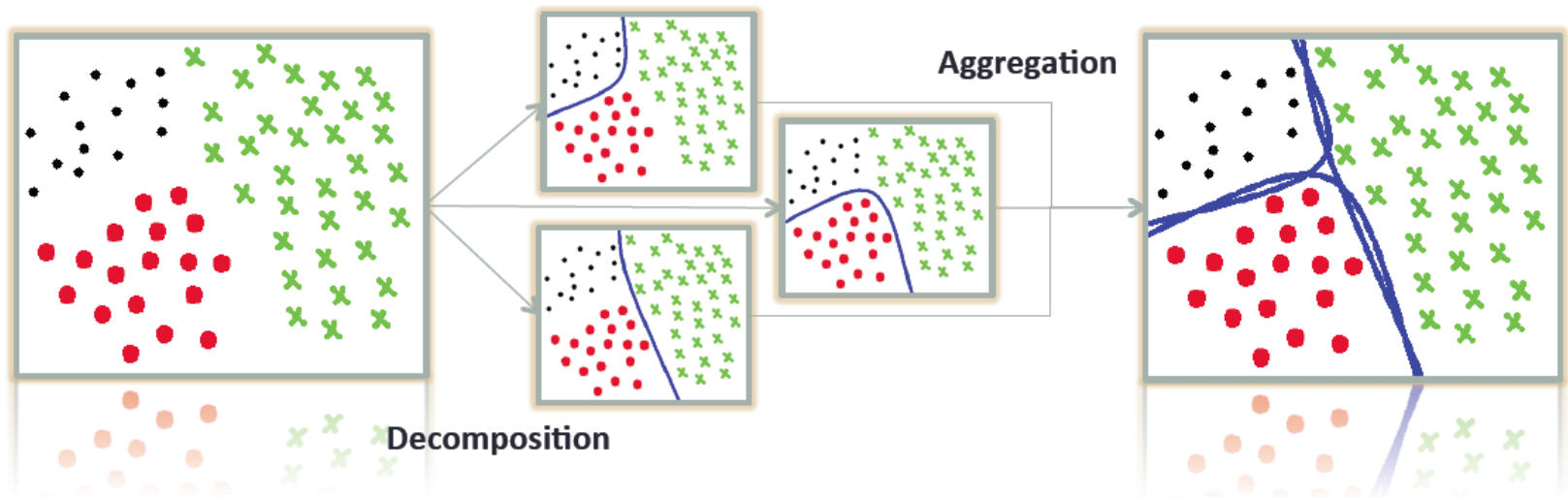
# Comparativo OVO



Regiones de decisión creadas por la estrategia de descomposición OVO para dos tipos de clasificadores binarios: LDA (fila superior) y SVM-RBF (fila inferior).

# Descomposición OVA

- Para un problema de clasificación con  $c$  clases se crean  $c$  modelos de clasificación binarios, donde cada uno de ellos aprende a distinguir una clase determinada del resto.
- El vector de respuestas es  $\mathbf{r}=[r_1, \dots, r_c]$ , tal que  $r_i \in [-\infty, \infty]$  y  $r_i > 0$  para muestras de la clase  $\omega_i$ .
- La principal desventaja de esta descomposición es el alto desbalance en el número de muestras de entrenamiento para cada clasificador base.



# Descomposición OVA-MAX

- La estrategia de agregación más común se denomina de “máxima confianza” (MAX), en la cual la clase ganadora es aquella con la máxima respuesta positiva:

$$\omega_i = \arg \max_{i=1, \dots, c} r_i$$

- Ventaja: es poco probable que haya empates en la máxima respuesta positiva.
- Desventaja: tomar el valor máximo no garantiza una clasificación correcta. Puede suceder que no existan respuestas positivas, aunque se toma la respuesta negativa más cercana a cero.

# Descomposición A&O

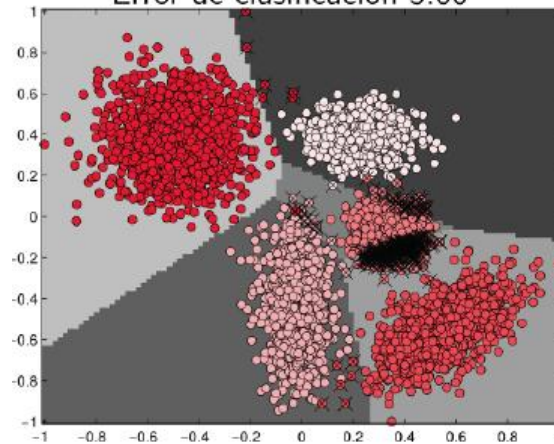
- En la estrategia A&O es necesario entrenar los clasificadores base de las estrategias OVO y OVA.
- La clasificación de una instancia se realiza como:
  - Obtener las dos clases,  $\omega_a$  y  $\omega_b$ , cuyos correspondientes clasificadores en la estrategia OVA,  $c_a$  y  $c_b$ , generan las respuestas más altas.
  - Clasificar la instancia con el clasificador  $c_{a,b}$  en la estrategia OVO.
- Desventaja: es necesario entrenar  $c(c+1)/2$  clasificadores, por lo que aumenta el costo espacial y temporal.
- Ventajas: en algunos casos se obtienen mejores resultados que usar de manera independiente OVO ó OVA.



# Comparativa OVA

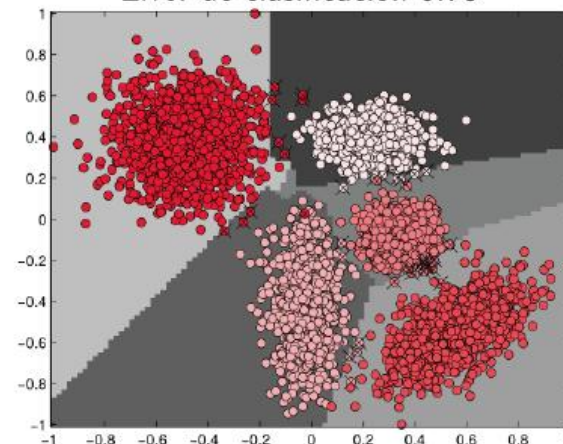
## OVA-MAX

Error de clasificación 5.66

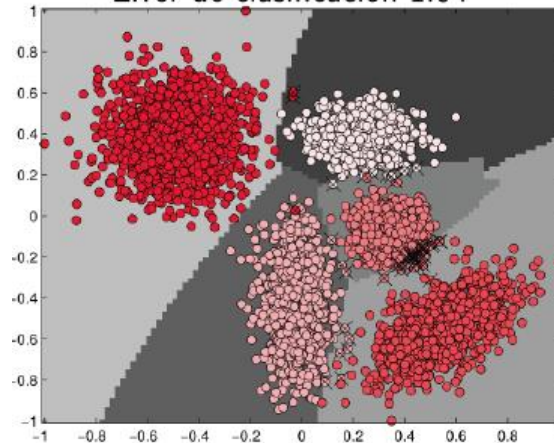


## A&O

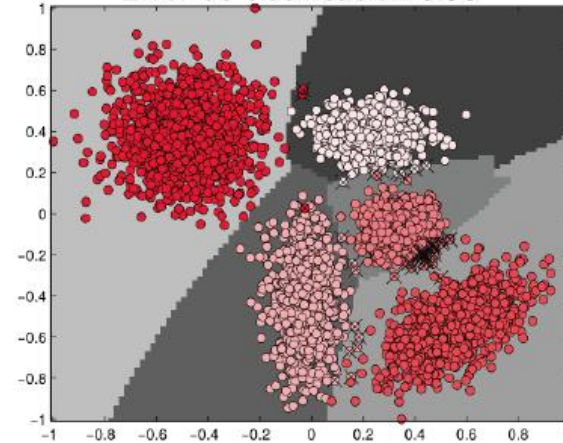
Error de clasificación 0.78



Error de clasificación 1.04



Error de clasificación 0.98



# Estudios sobre la descomposición binaria

- M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, 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, [doi: 10.1016/j.patcog.2011.01.017](#)
- Anderson Rocha and Siome Goldenstein. [Multiclass from Binary: Expanding One-vs-All, One-vs-One and ECOC-based Approaches](#). *IEEE Transactions on Neural Networks and Learning Systems* 25:2 (2014) 289–302 [doi:10.1109/TNNLS.2013.2274735](#)