

# **Práctica 5: Clasificación usando método multiclase**

Asignatura: Introducción a la Minería de Datos, 4º Grado de Ingeniería  
Informática Escuela Politécnica Superior de Córdoba - Universidad de  
Córdoba 2020 - 2021

**Trabajo realizado por:**

-Antonio Gómez Giménez (32730338G)

[i72gogia@uco.es](mailto:i72gogia@uco.es)



---

## Índice:

Ejercicio 2	2
Ejercicio 3	2
Ejercicio 4 y 5	2



## Ejercicio 2

Para este ejercicio se han escogido las siguientes seis bases de datos: soybean\_mod, segment-challenge, iris, glass, segment-test\_mod y wine. El resto de bases de datos usadas en la práctica anterior han sido desechadas ya que no eran multiclase.

Respecto al clasificador utilizado para esta práctica, se ha utilizado el clasificador de árboles de decisión. Los resultados obtenidos al realizar árbol de decisión sobre estas bases de datos son los siguientes:

```
soybean_mod.arff 77.22074468085107
segment-challenge.arff 94.57142857142857
wine.arff 91.08974358974359
iris.arff 93.0
glass.arff 59.142857142857146
segment-test_mod.arff 91.19360902255639
```

## Ejercicio 3

Se han aplicado los métodos multiclase one-vs.-one (OVO), one-vs.all (OVA) y error correcting output codes (ECOC) a cada uno de los conjuntos de datos y los resultados obtenidos son los siguientes, la columna uno es la base de datos, la columna dos, los resultados usando OVO, la columna tres, los resultados usando OVA y la columna cuatro, los resultados usando ECOC:

```
soybean_mod.arff 78.48847517730496 73.46631205673759 79.14007092198581
segment-challenge.arff 95.04761904761902 94.76190476190474 95.23809523809521
wine.arff 90.1923076923077 91.98717948717947 74.35897435897435
iris.arff 90.27272727272728 93.0 93.0
glass.arff 60.42857142857143 52.33333333333333 59.857142857142854
segment-test_mod.arff 92.25250626566415 92.05513784461152 91.89849624060149
```

## Ejercicio 4 y 5

Tras obtener estos resultados, comparamos si hay diferencias significativas entre ellos usando el test de Iman-Davenport, esto no ocurre, pero aun así, he aplicado el procedimiento de Wilcoxon para comparar cada método multiclase con el clasificador base y los diferentes métodos entre ellos.

Tras realizar el test de Iman-Davenport, el resultado es el siguiente:

```
Statistics_Iman-Davenport= 1.9285714285714408 , 0.5873635060521845
```



Como podemos ver, no hay diferencias significativas entre estos resultados ya que no desechamos la hipótesis nula porque el p-value es mayor que la probabilidad crítica. Aun así se han realizado el procedimiento de Wilcoxon para comprobar que no hay diferencias significativas entre cada método multiclase con el clasificador multiclase:

```
Statistics_wilcoxon_tree_base_y_tree_ovo= 8.0 , 0.6875
```

Como podemos observar, para OVO no se rechaza la hipótesis nula, por tanto no hay diferencias significativas.

```
Statistics_wilcoxon_tree_base_y_tree_ova= 6.0 , 0.6858304344516057  
Statistics_wilcoxon_tree_base_y_tree_ecoc= 5.0 , 0.5001842570707944
```

Como podemos observar, para OVA como para ECOC, no se rechaza la hipótesis nula, por tanto no hay diferencias significativas.

Como hemos visto no hay diferencias significativas aunque estos clasificadores deberían de mejorar los resultados. Por ejemplo, en el caso de OVA, se separa la clase que queramos comprobar y se analiza respecto al resto, esto en algunos casos mejora como en la base de datos iris pero en el resto no se ha conseguido llegar a una mejora.

Esto se puede deber a que las bases de datos pueden ser muy pequeñas con pocos patrones de tal forma que al separar una base de datos donde una clase tenga pocos patrones, puede ser que de peores resultados el clasificador, lo más rentable sería que todas las clases tuvieran el mismo número de patrones para evitar así este tipo de problemas que se pueden ocasionar.