

Entregable I - Multilabel Datasets

Isabella Hernández García,
z22hegai@uco.es
31 de enero de 2023

ÍNDICE

I. Introducción	1
II. Respuesta a los ejercicios	1
II-A. Ejercicio 2	1
II-A1. Información representada en los datasets	1
II-B. Ejercicio 3	1
II-B1. Script <i>car.py</i>	1
II-B2. Resumen de las medidas de caracterización	1
II-C. Ejercicio 4	1
II-D. Ejercicio 5	1
II-E. Ejercicio 6	1
III. Conclusiones	1
Referencias	1

I. INTRODUCCIÓN

El objetivo de la práctica es familiarizarse con el tema de clasificación multietiqueta utilizando la biblioteca *Scikit-multilearn*[1]

En el *Link a Github* se encuentran implementadas las soluciones a cada uno de los ejercicios.

II. RESPUESTA A LOS EJERCICIOS

En esta sección se da respuesta a cada una de las preguntas que se encontraban explícitas en la orden.

II-A. Ejercicio 2

Familiarízate con los datasets presentes en *Scikit-multilearn* y con los que trabajaremos durante las próximas sesiones. ¿Qué información representan?. Concretamente, dispone de los siguientes 17 datasets:

```
{'scene', 'Corel5k', 'bibtex',  
'enron', 'rcv1subset5', 'tmc2007_500',  
'rcv1subset3', 'rcv1subset1', 'delicious',  
'rcv1subset4', 'genbase', 'birds',  
'emotions', 'rcv1subset2', 'mediamill',  
'medical', 'yeast'}
```

II-A1. Información representada en los datasets:

- **Scene:** Un conjunto de datos de imágenes con 2407 imágenes etiquetadas hasta en 6 clases: playa, puesta de sol, hojas de otoño, campo, montaña y urbana. Cada imagen se describe con 294 características numéricas visuales correspondientes a momentos de color espacial en el espacio LUV.
- **Corel5k:** Es una prueba popular para métodos de clasificación y anotación de imágenes. Se basa en 5000 imágenes Corel.

II-B. Ejercicio 3

II-B1. Script *car.py*: TODO

II-B2. Resumen de las medidas de caracterización: TODO

II-C. Ejercicio 4

Busca en Internet dos datasets extra. Descárgalos localmente y realiza un script en Python para cargarlos y recalcular todas las medidas del paso anterior sobre los mismos. Debes escoger datasets variados, con diferente número de etiquetas, variables e instancias.

La orden no tenía preguntas explícitas para este ejercicio pero se pueden observar los resultados en *Github* TODO

II-D. Ejercicio 5

Familiarízate con la documentación *scikit-multilearn* en: <http://scikit.ml/> Prueba los métodos ML disponibles pertenecientes a las dos categorías (transformación y adaptación) que hemos visto en teoría.

La orden no tenía preguntas explícitas para este ejercicio pero se pueden observar los resultados en *Github*

II-E. Ejercicio 6

III. CONCLUSIONES

Reportar en tercera persona las diferentes conclusiones producto de la práctica de laboratorio desarrollada.

REFERENCIAS

- [1] Pński@ and T. Kajdanowicz. A scikit-based Python environment for performing multi-label classification. *ArXiv e-prints*, February 2017.