



## Práctica 3 Clasificación

### Objetivo

*Desarrollar un Jupyter-Notebook por entregable (4 en total) que permitan implementar algoritmos para la obtención de reglas de asociación y patrones secuenciales.*

### Herramientas

- Lenguaje de programación: Python
- Librerías: Numpy, pandas, scikit-learn, matplotlib
- Entorno de gestión de librerías: Anaconda
- Editor: Jupyter

### Información de la entrega

La entrega se realizará a través de la tarea LAB3 disponible en la página Canvas de la asignatura.

Consistirá en un fichero comprimido (.zip, .tar.gz) con nombre LAB03-GRUPOxx.zip que contendrá:

- 1- Un Jupyter-Notebook por cada entregable (archivos con extensión .ipynb)
- 2- La memoria del laboratorio se entregará integrada en el Notebook de manera que explique y complemente el código entregado
- 3- El código entregado tiene que ser funcional, correcto y completo.

**Las entregas que no se ajusten exactamente a este formato NO SE EVALUARÁN.**

### Rúbrica

#### Código

El valor de cada entregable para la nota final de la práctica se indica en el enunciado, así como el valor de cada uno de los apartados.

Todos los aspectos de programación se dan por supuestos.

El código debe ser:

- Funcional: debe ejecutar sin errores y el resultado debe ser el esperable
- Original: el código no puede ser una copia de trabajos publicados en Internet o de otros compañeros. Grupos con código igual serán suspendidos.
- No redundante: se penalizará el código que no sea útil o redundante
- Comentado: es obligatorio incluir comentarios en el código, en su justa medida
- Gráficas: deben incluir todos los datos que sean necesarios

#### Memoria

La memoria estará incluida en los Jupyter-Notebook que se entreguen de manera que complementen el código entregado. La redacción debe ser clara y correcta ortográfica y gramaticalmente. Debe incluir la justificación de cada paso que se realice para la resolución de los problemas planteados.



## Entregable 1 – Clasificador de Flores

*Este entregable vale 3 puntos de la nota final de la práctica 3.*

El jardín botánico quiere agilizar el proceso de clasificación de flores. En particular, quieren poder identificar los tipos de Iris según las características de las flores. Para crear un clasificador que les ayude en esta tarea, tendremos en cuenta el lugar que ocupa cada tipo de iris en un espacio  $n$ -dimensional, siendo  $n$  el número de características de cada flor de iris.

Utilizaremos el dataset iris.csv que recoge la información de 50 muestras de cada una de las 3 especies de Iris que trabajan. En el archivo iris.info encontraréis más información sobre el dataset.

Elige el clasificador más adecuado de entre los vistos en clase y usa scikit-learn junto con las librerías que necesites para resolver las siguientes cuestiones. Muestra todos los resultados del algoritmo paso a paso.

- 1) Haz todo el preprocesamiento para crear un set de entrenamiento y otro de validación que permitan crear un clasificador de transacciones con tarjeta atendiendo sólo a las características necesarias. Explica qué has hecho y porqué. (0,5 puntos)
- 2) Prueba con distintas configuraciones de las dos métricas principales. La primera métrica corresponde al número de individuos que usarás para clasificar una nueva instancia y la segunda cómo vas a medir la cercanía de esa nueva instancia con el resto. ¿Qué decisiones has tomado? ¿Por qué? (1 punto)
- 3) Elige la mejor configuración entre las anteriores. Para ello crea una tabla para ver cómo evoluciona la clasificación. Dibuja los resultados que se obtienen con ambas configuraciones elegidas como las mejores. (1 punto)
- 4) Utiliza el clasificador para saber qué ocurre con los audios de un dataset que se llame "iris2clasif.csv" que obtendremos del dataset proporcionado. (0,5 puntos)

## Entregable 2 – Clasificador de Género

*Este entregable vale 3 puntos de la nota final de la práctica 3.*

El género es una construcción social. La manera en la que los chicos y las chicas son tratados desde que nacen moldean su comportamiento y sus preferencias personales según lo que la sociedad espera para sus géneros.

Desde el Observatorio de Igualdad quieren hacer un estudio de hasta qué punto se puede identificar el género de una persona si sabemos sus gustos o preferencias sobre algunos temas. Para eso se quiere implementar un clasificador que proporcione una serie de reglas teniendo en cuenta las preferencias de una persona se pueda saber si se trata de un chico o de una chica.

Elige el clasificador más adecuado de entre los vistos en clase y usa scikit-learn junto con las librerías que necesites para resolver las siguientes cuestiones. Muestra todos los resultados del algoritmo paso a paso.

- 1) Crea un clasificador en el que uses al menos dos criterios de división distintos. Calcula el error de cada uno de ellos y elige el que mejor clasifique (1 punto).
- 2) Dibuja el modelo elegido en el punto anterior (0.5 puntos).
- 3) Selecciona tres reglas que sean las que generalicen lo menos posible y otras tres que especialicen lo menos posible. Interpreta las. (0,5 puntos).
- 4) Usa tu clasificador para decidir qué género corresponde a las preferencias de cada uno de los miembros del equipo de prácticas (al menos estudia 4 preferencias) (1 punto)



## Entregable 3 – Clasificador de Residuos

*Este entregable vale 2 puntos de la nota final de la práctica 3.*

La gestión de residuos es un problema grande a nivel nacional. Muchos de los residuos terminan en vertederos donde se hace difícil o casi imposible controlar su impacto en el medio ambiente. Para ayudar y fomentar el reciclaje, la Comunidad de Madrid quiere desarrollar una aplicación que ayude al usuario a saber en qué contenedor tiene que depositar cada residuo que genera. Para ello se va a crear un clasificador que realizando una foto a un desperdicio pueda decidir a qué tipo de contenedor de reciclaje debe ir. Dicho clasificador funcionará mediante un set de entrenamiento donde se buscará un plano que divida las diferentes clases dispuesta en un espacio  $n$ -dimensional dependiendo de sus características. Muestra todos los resultados del algoritmo paso a paso.

Para crear el clasificador vamos a trabajar con el dataset `garbage.zip` que contiene diferentes fotos de residuos organizadas según la clase de residuo a la que pertenecen: cartón (cardboard), vidrio (glass), metal (metal), papel (paper), plástico (plastic), basura (trash). Podréis encontrar una breve descripción del dataset en el archivo `garbage.info`.

Elige el clasificador más adecuado de entre los vistos en clase y usa `scikit-learn` junto con las librerías que necesites para resolver las siguientes cuestiones. Muestra todos los resultados del algoritmo paso a paso.

- 1) Crea un clasificador que permita saber qué tipo de residuo aparece en una imagen. Realiza al menos dos configuraciones y dibuja una tabla donde se muestre la precisión del modelo. (1 punto)
- 2) Elige 5 imágenes de diferentes residuos que no hayas usado ni para entrenar ni para validar el modelo y clasifícalas. Utiliza el modelo que consideres que clasifica mejor obtenido en el punto anterior. Indica con qué error ha funcionado el clasificador (1 punto).

## Entregable 4 – Clasificar Tipo de Música

*Este entregable vale 2 puntos de la nota final de la práctica 3.*

Una aplicación que ofrece un servicio de música en streaming quiere facilitar el acceso al contenido de su catálogo atendiendo al tipo de música de una canción incluyendo un clasificador de tipo de música. De esta manera, ahorrarán costes porque el clasificador ayudará a la automatización de la categorización de la música de la aplicación. En particular están interesados en crear un clasificador basado en las diferentes características de los audios y la probabilidad de que pertenezca a un tipo de música u otro.

Para crear el clasificador vamos a trabajar con el dataset `musica.csv` que contiene diferentes características de un archivo de audio (canciones) etiquetadas con el tipo de música al que pertenece (blues, classical, country, disco, etc.). Podréis encontrar una breve descripción del dataset en el archivo `música.info`.

Elige el clasificador más adecuado de entre los vistos en clase y usa `scikit-learn` junto con las librerías que necesites para resolver las siguientes cuestiones. Muestra todos los resultados del algoritmo paso a paso.

- 1) Realiza todo el preprocesamiento necesario para poder entrenar el clasificador. (1 punto)
- 2) Crea un clasificador e indica su error. Úsalo para saber a qué tipo corresponden al menos 5 audios/canciones que no hayas usado para entrenar el modelo. (1 punto)