

El título que queráis

Vuestros nombres¹

Abstract. Resumen del trabajo realizado en 100 palabras.

1 Introducción

El proyecto que hemos realizado trata sobre el análisis de un conjunto de datos recogidos por 8 sensores de gas, un sensor de temperatura y un sensor de humedad. Estos sensores fueron expuestos a estímulos por la presencia de vino y plátanos. Además, se han recogido datos de la respuesta a la no presencia del vino y los plátanos. El objetivo del proyecto es la clasificación de las respuestas de los sensores a los estímulos previamente dichos. Para ello primero analizaremos los datos del dataset mediante técnicas de preprocesamiento para ver que nos pueden ofrecer esos datos a la hora de clasificar las respuestas. Después, elegiremos los modelos de aprendizaje automático supervisado que mejor nos convengan para dicho dataset. Por último, compararemos los resultados de los distintos modelos y razonaremos si son resultados aceptables y qué modelos han dado mejores resultados.

2 Descripción del dataset

El dataset proporcionado de los datos de los sensores está compuesto por dos archivos: el archivo HT_Sensor_dataset.dat; que contiene el identificador de la inducción, instantes de tiempo para cada inducción y los datos de los sensores para cada inducción en esos instantes de tiempo, y el archivo HT_Sensor_metadata.dat; que contiene para cada inducción su identificador, el día en que fue realizada, el estímulo usado para la inducción, y el intervalo de tiempo de la inducción, que es dividido en tiempo inicial y duración de la misma. En este dataset hay 3 clases distintas para clasificar, que se corresponden con los estímulos que reciben los sensores: vino (wine), plátano (banana) y ningún estímulo (background). En total se realizan 100 inducciones, donde 36 se realizan con vino, 33 con plátano y 31 son background. De esas 100 inducciones se recogen 928991 datos en distintos instantes de tiempo. Al analizar los datos podemos comprobar que no hay datos para la inducción con el identificador 95, por lo que la descartamos del análisis. Además de los datos recogidos durante la inducción, se nos ofrecen los datos de los instantes previos y posteriores de la inducción. Esta división del tiempo la podemos calcular fácilmente para cada inducción con la ayuda de los datos del tiempo del archivo de los metadatos, por lo que podemos distinguir para cada experimento estos tres periodos diferenciados. Comprobamos que para los experimentos con identificadores 14 y 76 no tenemos datos posteriores a la inducción, por lo que también los descartamos del análisis. Los atributos del dataset son numéricos y reales, por lo que podemos utilizar las medidas características de la distribución de los atributos para analizarlos. Estas medidas pueden ser de

centralización, como la media, o de dispersión, como la varianza. Con estas medidas podemos ver como se comporta cada sensor frente a los estímulos. Más adelante vamos a ver cómo analizar los atributos y ver qué medidas nos pueden ayudar realmente a conseguir mejores resultados.

3 Memoria

La longitud máxima serán *8 caras* sin incluir referencias. Se valorará la capacidad de síntesis por lo que superar las 8 páginas tendrá penalización. La memoria se deben tratar, de forma orientativa, los siguientes aspectos:

- Introducción [1pt]: breve introducción al problema a analizar, descripción del dataset y objetivos.
- Análisis exploratorio de los datos [1pt]: Descripción estadística de los datos: Número de clases, distribución de las clases, otras estadísticas y análisis.
- Descripción de los distintos atributos propuestos y cómo se obtienen [2pt] Modelos utilizados, descripción del protocolo experimental, estimación de parámetros, etc [2pt]: En esta sección se debe especificar toda la información necesaria para que otra persona, sin acceso a vuestro código, pueda reproducir los experimentos que habéis hecho. Debe quedar claro en la descripción que no se usan los datos de test para entrenar los modelos.
- Resultados obtenidos en forma tabular y/o usando gráficas [1pt]. Se debe describir que muestra cada tabla o gráfica.
- Discusión de los resultados obtenidos y conclusiones [2pt] Esta sección es la más importante del documento ya que es dónde se pone en valor el trabajo realizado. Debéis responder a preguntas tipo ¿Qué atributos y métodos han dado mejores resultados? ¿Por qué creéis que es así? ¿Son resultados aceptables? ¿Qué modelos recomendaríais bajo qué condiciones? Tal vez un modelo funcione mejor cuando se entrena con pocos datos o funcione mejor para clasificar una de las clases y peor para otras, etc.
- Además se deben utilizar al menos dos de las técnicas descritas a lo largo del curso por vuestros compañeros [1pt]

Se valorará la correcta redacción del documento.

4 Presentación

Debéis entregar un ppt o pdf con el resumen del trabajo. Debe ser una presentación para presentar en 12 minutos. El tiempo de presentación será estricto y se parará a los que se pasen de tiempo.

ID	age	weight
1	15	65
2	24	74
3	18	69
4	32	78

Table 1: Age and weight of people.

5 Tablas y figuras

Las tablas y figuras hay que referenciarlas desde el texto y describir qué muestran. Por ejemplo, en la Figura ?? se muestra el logo de scikit-learn. En la Tabla 1 se muestra un ejemplo de tabla.

6 Citas

Es una buena práctica referenciar los trabajos en los que se ha basado vuestro análisis. Por ejemplo, si los experimentos se han realizado utilizando scikit-learn habría que citarlo [?].

7 Cómo obtener este pdf usando L^AT_EX

Podéis seguir los siguientes pasos para obtener el pdf junto con sus referencias desde una consola linux:

```
pdflatex fuente.tex
bibtex fuente
pdflatex fuente.tex
pdflatex fuente.tex
```

Estos comandos generan el fichero pdf que incluye las referencias. Las referencias se guardan en un fichero aparte (en esta caso mi_bibliografia.bib).