

Bloque 2

Aprendizaje Automático

Práctica 2:

Aplicación de los algoritmos de Perceptrón y Regresión Logística a varios datasets

DOCENCIA VIRTUAL



Responsable del Tratamiento: Universitat Politècnica de València (UPV)

Finalidad: Prestación del servicio público de educación superior en base al interés público de la UPV (Art. 6.1.e del RGPD).

Ejercicio de derechos y segunda capa informativa: Podrán ejercer los derechos reconocidos en el RGPD y la LOPDGDD de acceso, rectificación, oposición, supresión, etc., escribiendo al correo dpd@upv.es.

Para obtener más información sobre el tratamiento de sus datos puede visitar el siguiente enlace: <https://www.upv.es/contenidos/DPD>.

Propiedad Intelectual: Uso exclusivo en el entorno del aula virtual.

Queda prohibida la difusión, distribución o divulgación de la grabación de las clases y particularmente su compartición en redes sociales o servicios dedicados a compartir apuntes.

La infracción de esta prohibición puede generar responsabilidad disciplinaria, administrativa y/o civil.

Objetivo

Generar y evaluar un clasificador sobre datasets (conjuntos de datos)

- Sesiones

- S1: G1:22/11 G2:19/11
- S2: G1:29/11 G2:26/11
- S3: G1: 5/12 G2: 3/12
- Examen P2: G1:13/12 G2:10/12

Sesiones de la práctica 2

Sesión 1:

- Familiarizarse con el entorno de trabajo (Google Colab)
- Analizar conjuntos de datos (datasets): **iris**, **digits**, **olivetti**, **openml**

Sesión 2:

- Aplicación del algoritmo del Perceptron a tareas de clasificación: dataset **iris**.
- **Ejercicio:** Aplicar Perceptrón a **digits** y **olivetti**

Sesión 3:

- Aplicación de Regresión Logística a tareas de clasificación: dataset **iris**.
- **Ejemplo de examen:**
 - Aplicación de Perceptrón y Regresión Logística a un dataset de OpenML.

Sesión 4 (**examen**):

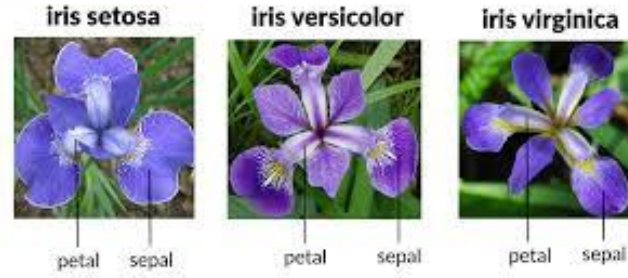
- Se pedirá la aplicación de Regresión Logística para una tarea diferente de OpenML
- Hay que subir la **solución del Ejercicio**

Sesión 1: entorno de trabajo y conjuntos de datos

- Entorno de trabajo: **Google Colab** (<https://colab.research.google.com>)
 - Se trabaja con Notebooks (código + texto)
 - Similar a un Notebook de Jupyter
- **Dataset:** colección de datos que se utiliza para entrenar, probar y validar modelos de aprendizaje automático.
- Un dataset típicamente consiste en varias "instancias" o "ejemplos", cada uno de los cuales incluye varias "características" o "atributos".
- Cada instancia en el dataset también tiene una "etiqueta" o "target", que es el valor que el modelo de aprendizaje intenta predecir

Sesión 1: entorno de trabajo y conjuntos de datos

- Ej dataset: iris

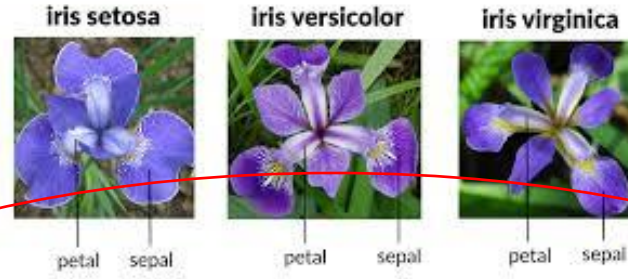


	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...

species: es la variable que el **clasificador** basado en un modelo de aprendizaje automático intentará predecir.

Sesión 1: entorno de trabajo y conjuntos de datos

- Ej dataset: iris



Todas las filas (individuos) quitando la columna de clase se usan para entrenar el modelo y validarlo.

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...

species: es la variable que el **clasificador** basado en un modelo de aprendizaje automático intentará predecir.

Sesión 1: entorno de trabajo y conjuntos de datos

- Ej dataset: iris



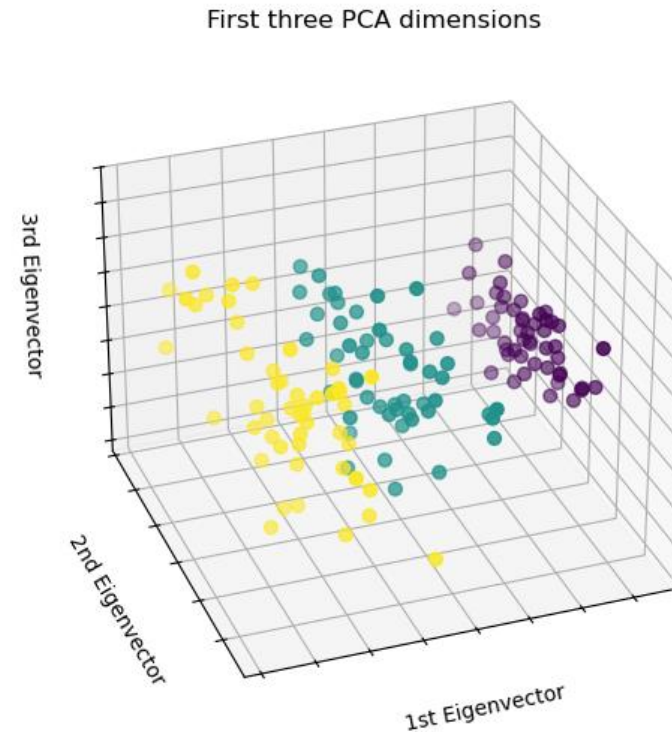
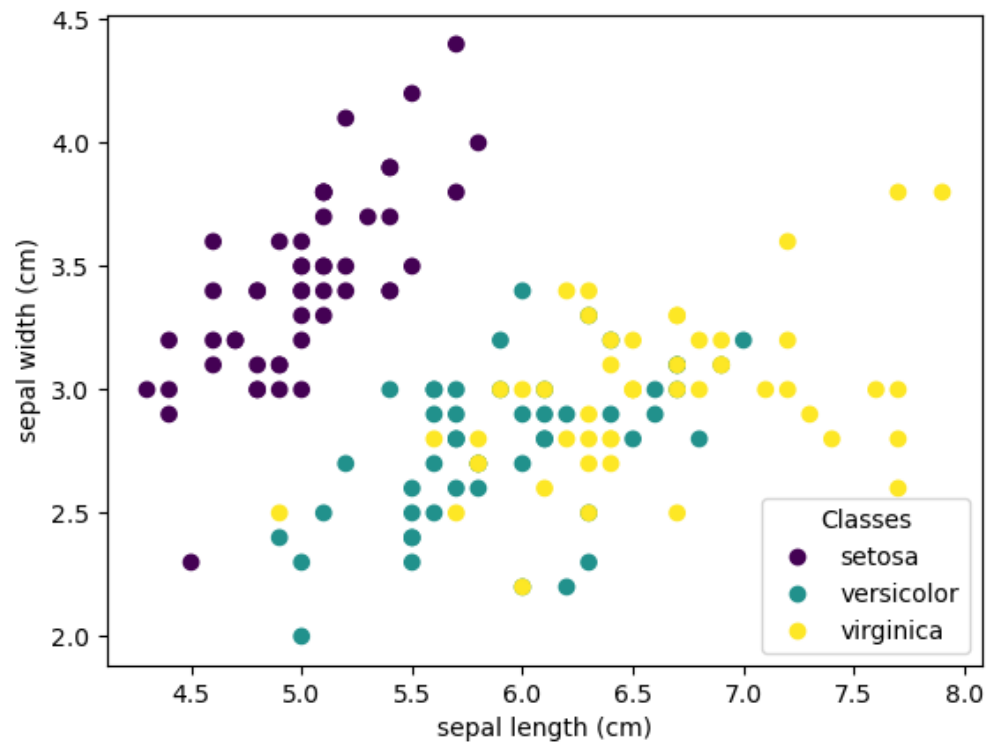
La columna de clase determinar el valor a predecir por el clasificador dado un nuevo individuo.

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...

species: es la variable que el **clasificador** basado en un modelo de aprendizaje automático intentará predecir.

Sesión 1: entorno de trabajo y conjuntos de datos

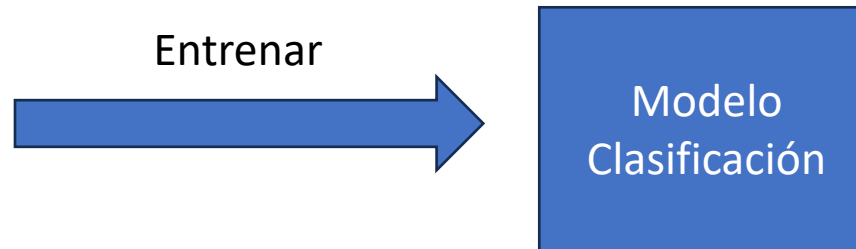
- Ej dataset: iris



Sesión 1: entorno de trabajo y conjuntos de datos

- Ej dataset: iris

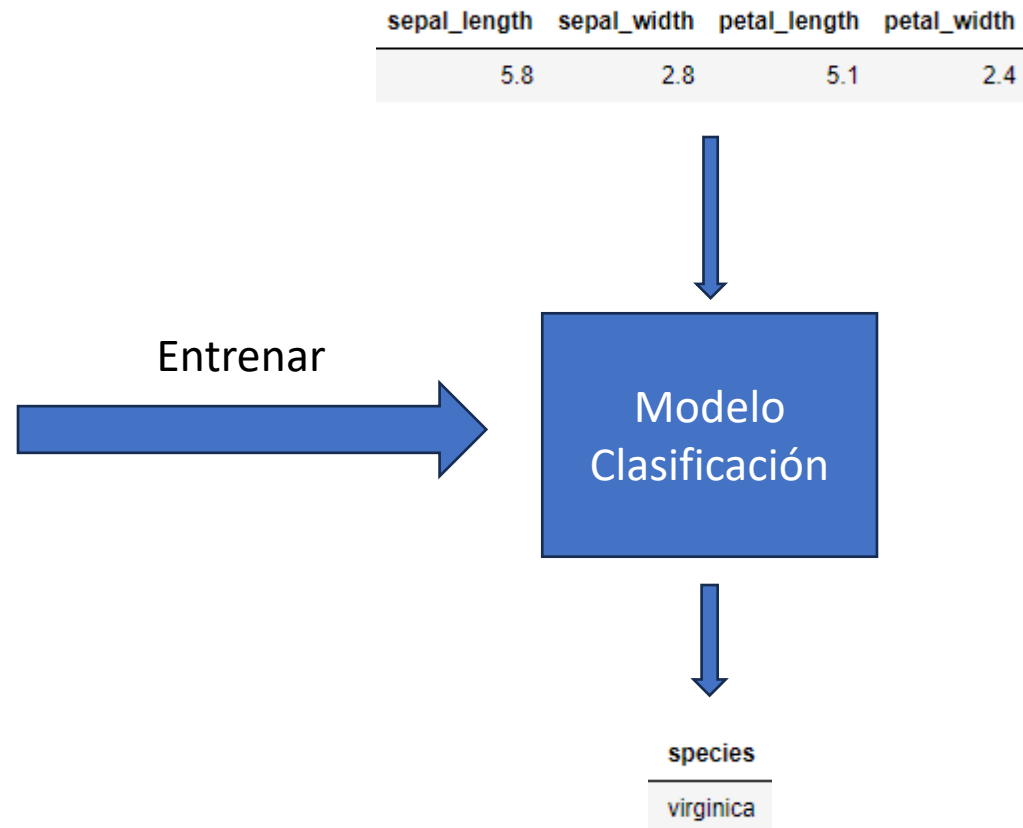
Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5	3.4	1.5	0.2	Iris-setosa
9	4.4	2.9	1.4	0.2	Iris-setosa
10	4.9	3.1	1.5	0.1	Iris-setosa
11	5.4	3.7	1.5	0.2	Iris-setosa
12	4.8	3.4	1.6	0.2	Iris-setosa
13	4.8	3	1.4	0.1	Iris-setosa
14	4.3	3	1.1	0.1	Iris-setosa
15	5.5	4	1.2	0.2	Iris-setosa
16	5.7	4.6	1.5	0.4	Iris-setosa
17	5.4	3.9	1.3	0.4	Iris-setosa
18	5.1	3.5	1.4	0.3	Iris-setosa
19	5.7	3.8	1.7	0.4	Iris-setosa



Sesión 1: entorno de trabajo y conjuntos de datos

- Ej dataset: iris

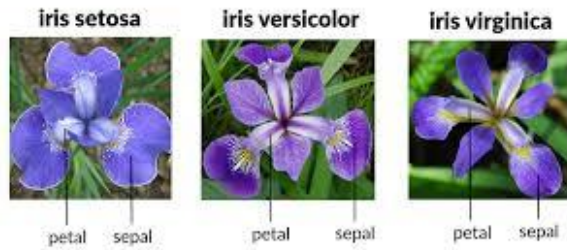
id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5	3.4	1.5	0.2	Iris-setosa
9	4.4	2.9	1.4	0.2	Iris-setosa
10	4.9	3.1	1.5	0.1	Iris-setosa
11	5.4	3.7	1.5	0.2	Iris-setosa
12	4.8	3.4	1.6	0.2	Iris-setosa
13	4.8	3	1.4	0.1	Iris-setosa
14	4.3	3	1.1	0.1	Iris-setosa
15	5.5	4	1.2	0.2	Iris-setosa
16	5.7	4.6	1.5	0.4	Iris-setosa
17	5.4	3.9	1.3	0.4	Iris-setosa
18	5.1	3.5	1.4	0.3	Iris-setosa
19	5.7	3.8	1.7	0.3	Iris-setosa



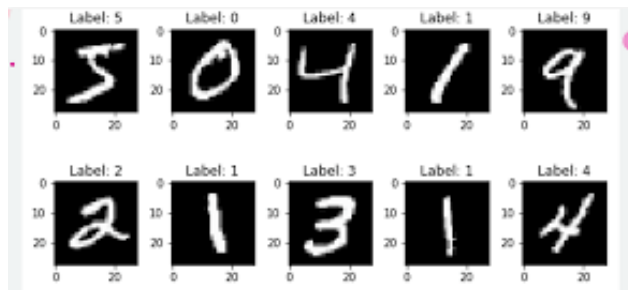
Sesión 1: entorno de trabajo y conjuntos de datos

- Datasets:

- IRIS



- DIGITS



- OLIVETTI



- OPENML (repositorio)



Uso de Google Colab

Paso 1: Crear una Cuenta de Google (si aún no tienes una)

- Para usar Google Colab, necesitas una cuenta de Google. Si ya tienes una cuenta de Gmail, YouTube, o cualquier otro servicio de Google, puedes usar esa misma cuenta. Si no, sigue estos pasos para crear una nueva:
 1. Ve a accounts.google.com.
 2. Sigue las instrucciones para crear una nueva cuenta de Google.

Paso 2: Acceder a Google Colab

- Una vez que tengas tu cuenta de Google, puedes acceder a Google Colab:
 1. Ve a [Google Colab](https://colab.google).
 2. Inicia sesión con tu cuenta de Google.

Uso de Google Colab

Paso 3: Crear un Nuevo Notebook

- Después de iniciar sesión en Google Colab, puedes comenzar a crear notebooks:
 1. Haz clic en Nuevo Notebook en la esquina inferior derecha de la pantalla. Esto abrirá un nuevo notebook en una nueva pestaña del navegador.
 2. Puedes cambiar el nombre del notebook haciendo clic en el título (por defecto, algo como "Untitled0.ipynb") en la parte superior de la página y escribiendo el nombre que desees.

Paso 4: Escribir y Ejecutar Código

- Google Colab te permite escribir y ejecutar código Python de manera interactiva:
 1. Escribe tu código Python en una celda.
 2. Para ejecutar el código en esa celda, presiona Shift + Enter o haz clic en el botón de reproducción (triángulo) en la esquina superior izquierda de la celda.
 3. El resultado de la ejecución del código aparecerá debajo de la celda.

Uso de Google Colab

Paso 5: Guardar y Compartir tu Notebook

- Google Colab guarda automáticamente tus notebooks en Google Drive:
 1. Para guardar manualmente, ve a Archivo > Guardar.
 2. Para compartir tu notebook, haz clic en el botón Compartir en la esquina superior derecha y sigue las instrucciones para compartirlo como lo harías con cualquier otro archivo de Google Drive.

Paso 6: Importar y Exportar Notebooks

- Puedes importar notebooks existentes o exportar los tuyos:
 - **Importar:** Archivo > Abrir notebook > elige desde GitHub, Google Drive o sube un archivo.
 - **Exportar:** Archivo > Descargar > elige el formato que prefieras (por ejemplo, .ipynb para Jupyter Notebook o .py para un script de Python).

Sesión 1: entorno de trabajo y conjuntos de datos

Probar los siguientes notebooks en Google colab:

- iris.ipynb
- digits.ipynb
- olivetti.ipynb

- openml.ipynb

Sesión 1: entorno de trabajo y conjuntos de datos

¿Qué hay que saber de la sesión de hoy?

- Saber qué es y que formato tiene un dataset
- Saber interpretar un dataset: nº individuos, nº clases, ...
- Saber cargar un dataset de openml
- Saber extraer las características y la columna objetivo de un dataset