

# IRIS Dataset

## Descripción del Dataset Iris

El conjunto de datos Iris consta de 150 observaciones de flores de iris, con tres especies diferentes: Iris setosa, Iris versicolor e Iris virginica. Cada observación tiene cuatro características numéricas que describen las mediciones físicas de las flores.

## Características

- Longitud del sépalo (cm)
- Ancho del sépalo (cm)
- Longitud del pétalo (cm)
- Ancho del pétalo (cm)

## Etiquetas

- 0: Iris setosa
- 1: Iris versicolor
- 2: Iris virginica

## Objetivo

El objetivo común al usar este dataset es construir un modelo que pueda clasificar una flor de iris en una de las tres especies basándose en las cuatro características mencionadas.

## Estructura del Dataset

### Características (X):

Una matriz de 150x4, donde cada fila representa una flor y cada columna representa una de las características (longitud del sépalo, ancho del sépalo, longitud del pétalo, ancho del pétalo).

## Etiquetas (y):

Un vector de longitud 150, donde cada entrada es una etiqueta que indica la especie de la flor.

```
from sklearn import datasets

# Cargar el dataset iris
iris = datasets.load_iris()
X = iris.data
y = iris.target

print("Características (X):")
print(X[:5])
print("Etiquetas (y):")
print(y[:5])
```

### Resultado

```
Características (X):
[[5.1 3.5 1.4 0.2]
 [4.9 3.0 1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]
 [5.0 3.6 1.4 0.2]]
Etiquetas (y):
[0 0 0 0 0]
```

## Uso en Aprendizaje Automático

El conjunto de datos Iris es frecuentemente utilizado para:

- Demostraciones y ejemplos: Ideal para explicar conceptos básicos de clasificación y técnicas de aprendizaje supervisado.
- Pruebas de algoritmos: Es pequeño y fácil de manejar, lo que lo convierte en una buena opción para probar nuevos algoritmos o enfoques.

## Visualización

A menudo se utilizan técnicas de visualización como gráficos de dispersión para observar cómo se distribuyen las diferentes especies de flores en función de sus características. Por ejemplo, se pueden

trazar la longitud y el ancho del sépalo para ver cómo se agrupan las especies.

```
import matplotlib.pyplot as plt

# Gráfico de dispersión de dos características
plt.scatter(X[:, 0], X[:, 1], c=y)
plt.xlabel('Longitud del sépalo (cm)')
plt.ylabel('Ancho del sépalo (cm)')
plt.title('Distribución de las Especies de Iris')
plt.show()
```

## Referencias

- [Kaggle: Iris Dataset](#)
- [Scikit-learn: Iris Dataset](#)