

MACHINE LEARNING

El Conjunto de Datos Iris

PhD. César Astudillo | Facultad de Ingeniería

¿Qué es el Iris Dataset?

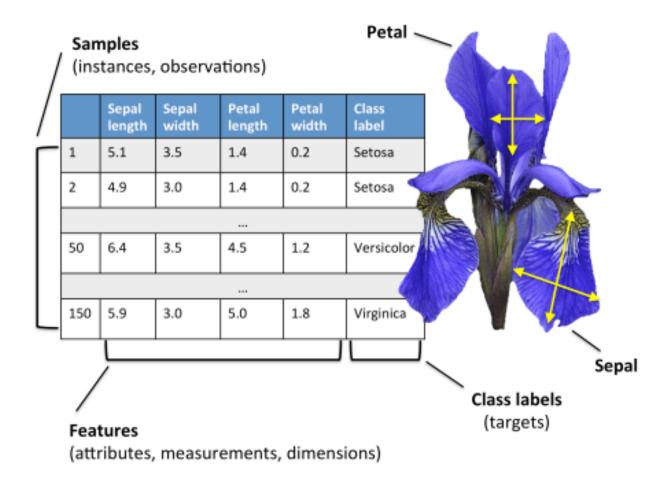
- El Iris dataset es un conjunto de datos de flores introducido por Ronald A. Fisher en 1936. [Paper Original]
- Es ampliamente utilizado en estadística, machine learning, y AI para experimentación y enseñanza.





Características del Dataset

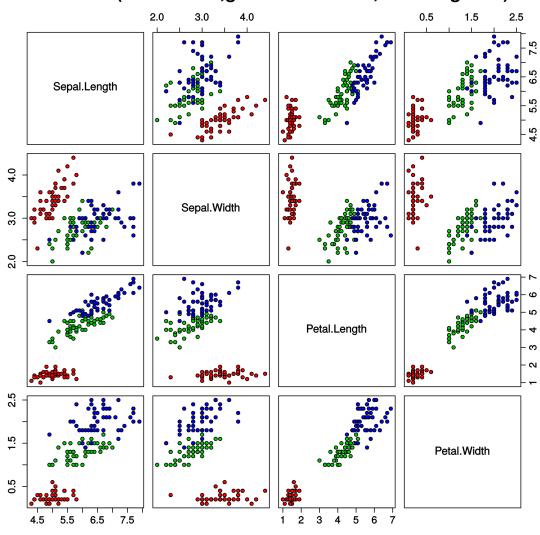
- El dataset incluye tres especies de flores:
- 1. Iris setosa
 - 2. Iris versicolor
 - 3. Iris virginica
- Cada muestra se describe mediante cuatro características numéricas:
 - Largo del sépalo (cm)
 - Ancho del sépalo (cm)
 - Largo del pétalo (cm)
 - Ancho del pétalo (cm)



Distribución de Clases en el Dataset

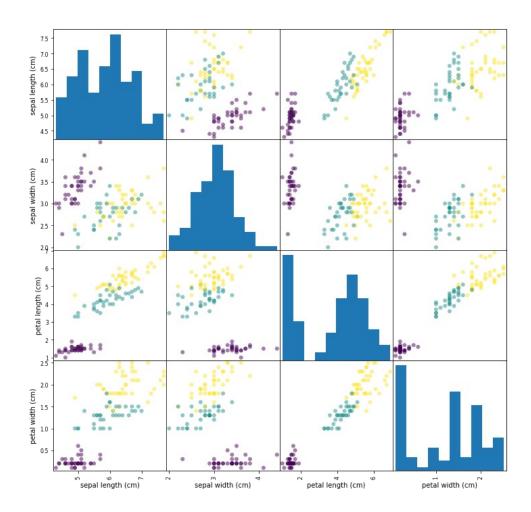
- El dataset contiene 150 muestras, distribuidas equitativamente entre las tres especies:
- 50 muestras de Iris setosa
 - 50 muestras de Iris versicolor
 - 50 muestras de Iris virginica

Iris Data (red=setosa,green=versicolor,blue=virginica)



Aplicaciones Comunes

- Ejemplos de uso del Iris dataset incluyen:
- - Clasificación de especies
 - Visualización de datos
 - Pruebas de algoritmos de machine learning
- Es considerado el 'Hello World' de machine learning.

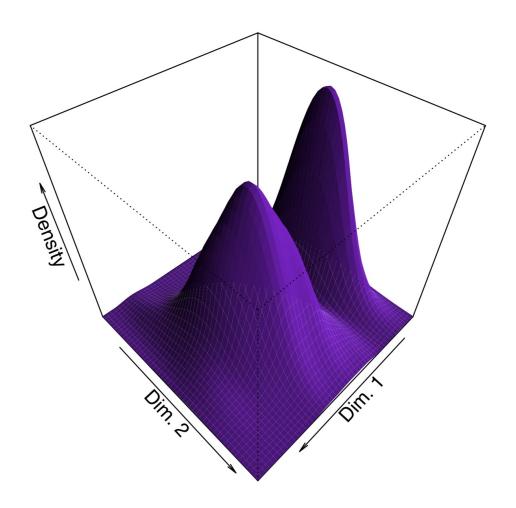


Visualización del Iris Dataset

• Gráficos de dispersión se utilizan para mostrar relaciones entre características como largo y ancho del pétalo, coloreados por clase.

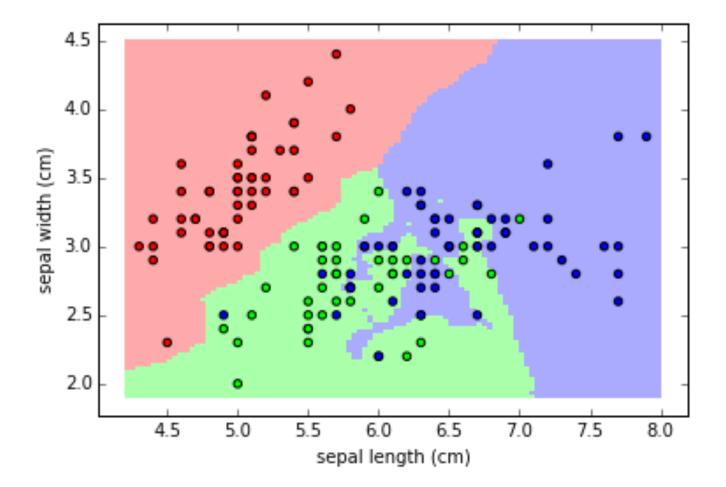
Análisis Estadístico Básico

 Se pueden calcular estadísticas básicas como media, mediana, y desviación estándar para entender mejor los datos.



Implementación en Machine Learning

- Algoritmos comunes aplicados al dataset incluyen K-Nearest Neighbors, Decision Trees, y SVM.
- Es fácil entrenar un modelo de clasificación con el Iris dataset.



Conclusión

- El Iris dataset es fundamental en la enseñanza y experimentación en machine learning.
- Explora y experimenta con el dataset por cuenta propia.



MACHINE LEARNING

El Conjunto de Datos Iris

PhD. César Astudillo | Facultad de Ingeniería