

Unsupervised Learning with the Iris Dataset

Objective

To understand **unsupervised learning** by applying **clustering** to the Iris dataset and discovering natural groupings **without using labels or evaluation metrics**.

Dataset

Use the **Iris dataset** with the following features:

- Sepal Length
- Sepal Width
- Petal Length
- Petal Width

Do NOT use the species/target column at any stage.

Concepts Covered

- Unsupervised Learning
- Exploratory Data Analysis
- Feature Scaling
- K-Means Clustering
- Elbow Method
- Cluster Interpretation

TASKS

Task 1: Load and Explore the Dataset

1. Load the Iris dataset using `sklearn`.
2. Convert it into a Pandas DataFrame.
3. Display:
 - First 5 rows
 - Shape of the dataset
 - Summary statistics

Question:

Why does unsupervised learning not require labeled data?

Task 2: Visual Exploration

1. Create scatter plots for:
 - Sepal Length vs Sepal Width
 - Petal Length vs Petal Width
2. Observe how data points are distributed.

Question:

Which pair of features seems to show natural groupings?

Task 3: Feature Scaling

1. Apply `StandardScaler` to all features.

2. Show a comparison of original vs scaled values.

Question:

Why is feature scaling important when using K-Means?

Task 4: Apply K-Means Clustering

1. Train a K-Means model with:
 - $k = 3$
2. Assign cluster labels to each data point.
3. Display:
 - First 10 rows with cluster labels
 - Cluster centroids

Question:

What does a cluster centroid represent in simple terms?

Task 5: Cluster Visualization

1. Plot the clustered data using:
 - Petal Length vs Petal Width
2. Color points by cluster.
3. Mark cluster centroids.

Question:

Do the clusters appear clearly separated? Explain your observation.

Task 6: Choosing the Number of Clusters (Elbow Method)

1. Run K-Means for values of $k = 1$ to 10 .
2. Plot the **Elbow Curve (Within-Cluster Sum of Squares vs K)**.
3. Identify the value of k where the curve bends.

Question:

Why is the elbow point a good choice for the number of clusters?

Task 7: Interpretation & Reflection

Write a short paragraph answering:

- What patterns did the clustering algorithm discover?
- How many clusters seem reasonable based on the plots?
- What challenges did you notice when clustering the data?

DELIVERABLES

Students should submit:

1. Jupyter Notebook (`.ipynb`)
2. All plots clearly labeled
3. Written answers to all questions
4. Clean, well-commented code