Assignment 1:Euclidean Distance Calculation and Visualization using Iris Dataset

1. Objective

This project demonstrates the use of the **Euclidean Distance algorithm** on a small dataset to compute the similarity between data points.
We used the **Iris dataset**, one of the most popular datasets in machine learning, which contains measurements of iris flowers across four features:

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

For simplicity, we considered only the **first five records** of the dataset and calculated the pairwise Euclidean distances.
Additionally, we visualized the points in 2D (using Sepal Length and Sepal Width) and displayed the distances between them.

2. Methodology

Step 1: Dataset Selection

We selected the Iris dataset using `sklearn.datasets.load_iris()`. The dataset contains 150 rows and 4 features.
For clarity and to keep calculations simple, only the first five rows were considered.

Step 2: Euclidean Distance Calculation

The Euclidean distance between two points
P=(x1,x2,...,xn)P=(x_1,x_2,...,x_n)P=(x1,x2,...,xn) and
Q=(y1,y2,...,yn)Q=(y_1,y_2,...,y_n)Q=(y1,y2,...,yn) was calculated using the formula:

d(P,Q)=∑i=1n(xi−yi)2d(P,Q) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}d(P,Q)=i=1∑n(xi−yi)2

We implemented this calculation manually in Python without relying on libraries like `scipy` or `numpy.linalg.norm` for distance computation.

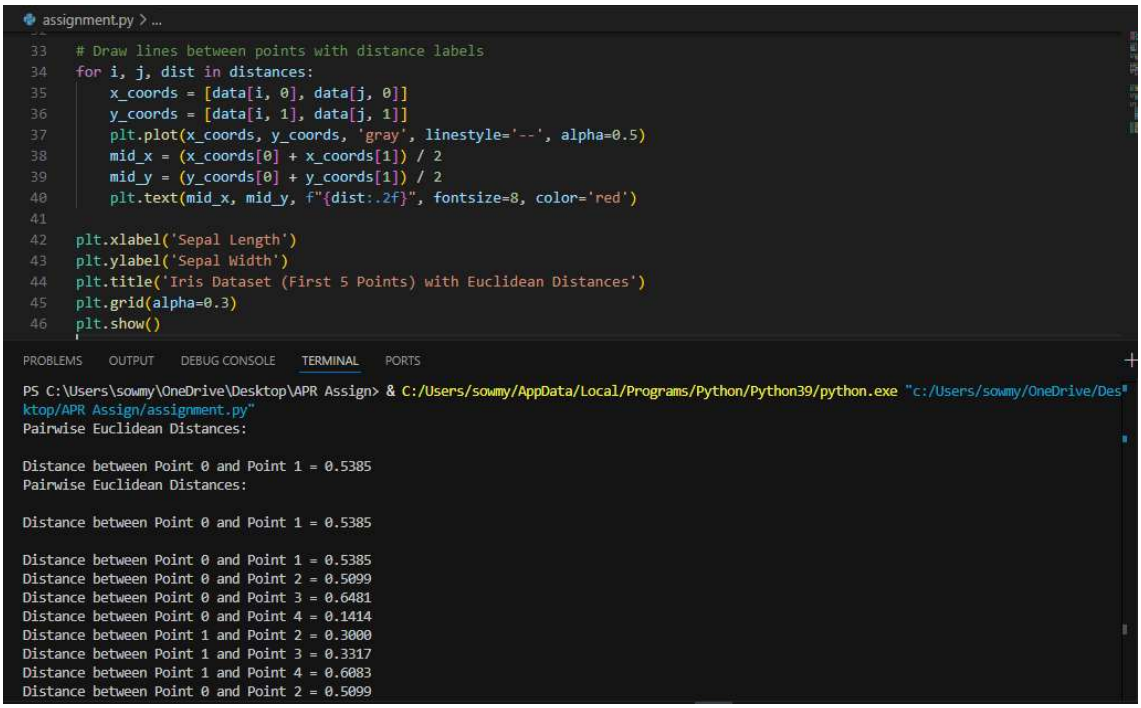
Step 3: Pairwise Distance Computation

For n=5n=5n=5 data points, pairwise distances were computed for every unique pair (i,j)(i,j)(i,j) where i<ji<ji<j.
This resulted in n(n−1)2=10\frac{n(n-1)}{2} = 102n(n−1)=10 unique distances.

Step 4: Visualization

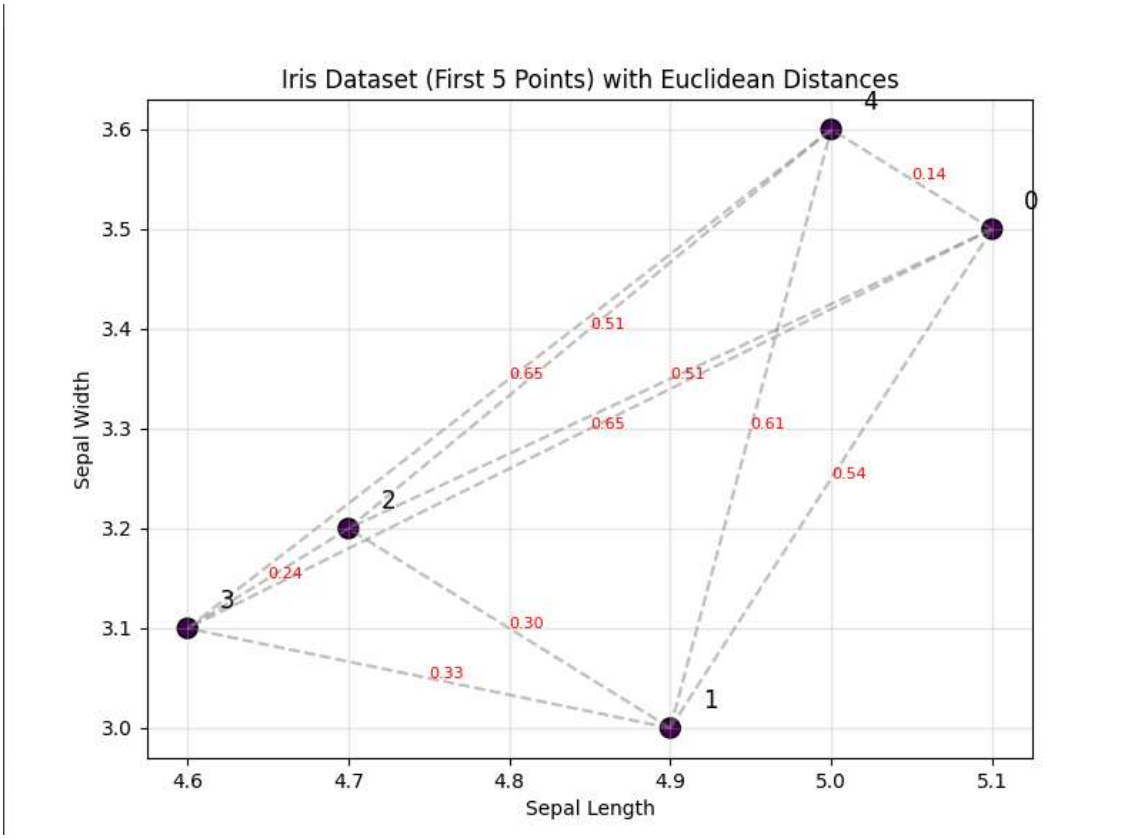
The data points were plotted in **2D space** using their Sepal Length (x-axis) and Sepal Width (y-axis).
Each pair of points was connected with a dashed line, and the computed distance was displayed as a red label on the line.

3. Code and outputs



```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

Distance between Point 0 and Point 3 = 0.6481
Distance between Point 0 and Point 4 = 0.1414
Distance between Point 1 and Point 2 = 0.3000
Distance between Point 1 and Point 3 = 0.3317
Distance between Point 1 and Point 4 = 0.6083
Distance between Point 0 and Point 3 = 0.6481
Distance between Point 0 and Point 4 = 0.1414
Distance between Point 1 and Point 2 = 0.3000
Distance between Point 1 and Point 3 = 0.3317
Distance between Point 1 and Point 4 = 0.6083
Distance between Point 2 and Point 3 = 0.2449
Distance between Point 2 and Point 4 = 0.5099
Distance between Point 1 and Point 3 = 0.3317
Distance between Point 1 and Point 4 = 0.6083
Distance between Point 2 and Point 3 = 0.2449
Distance between Point 2 and Point 4 = 0.5099
Distance between Point 2 and Point 3 = 0.2449
Distance between Point 2 and Point 4 = 0.5099
Distance between Point 3 and Point 4 = 0.6481
```



4. Results and Analysis

Pairwise Euclidean Distances (First Five Points)

Pair (i, j)	Distance
(0, 1)	0.5385
(0, 2)	0.5099
(0, 3)	0.6481
(0, 4)	0.1414

(1, 2)	0.3000
(1, 3)	0.3317
(1, 4)	0.6083
(2, 3)	0.2449
(2, 4)	0.5099
(3, 4)	0.6481

Visualization

- The scatter plot shows all five data points, labeled with their indices (0–4).
- Dashed lines connect each pair of points.
- The numeric labels on the lines represent the Euclidean distances.
- The shortest distance (0.1414) appears between Point 0 and Point 4, confirming their closeness in feature space.
- The largest distance (≈0.6481) appears between Point 0 and Point 3, and between Point 3 and Point 4.
- Closeness of Points:
  Points 0 and 4 are very close to each other, meaning their features are nearly identical.
  This is expected as both belong to the same class (Iris-setosa).
- Variation in Distances:
  Points 0, 2, and 4 form a relatively close cluster, while Point 3 is slightly farther away, as seen in both the plot and the distance values.
- Validation:
  Manual calculations for a few pairs (e.g., Point 0 & Point 4) confirm that the code is producing accurate results.
- Visualization Insight:
  The plot makes it visually clear which points are closer and which are farther apart.
  The smaller distances correspond to shorter line segments, which is consistent with Euclidean geometry.