Task 7

Apply hierarchal clustering algorithms for a given dataset. Compare the results and comment on the quality of clustering using evaluation metrics.

Program

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.datasets import load\_iris

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import AgglomerativeClustering

from sklearn.metrics import silhouette\_score, davies\_bouldin\_score

import scipy.cluster.hierarchy as sch

# -----------------------

# 1. Load Dataset

# -----------------------

iris = load\_iris()

X = iris.data

y = iris.target  # Ground truth for reference

# Standardize the data

scaler = StandardScaler()

X\_scaled = scaler.fit\_transform(X)

# -----------------------

# 2. Dendrogram (to choose number of clusters)

# -----------------------

plt.figure(figsize=(10, 4))

dendrogram = sch.dendrogram(sch.linkage(X\_scaled, method='ward'))

plt.title("Hierarchical Clustering Dendrogram (Ward Linkage)")

plt.xlabel("Samples")

plt.ylabel("Distance")

plt.show()

# -----------------------

# 3. Apply Hierarchical Clustering

# -----------------------

# Trying different linkage criteria for comparison

linkage\_methods = ["ward", "complete", "average", "single"]

results = []

for method in linkage\_methods:

    clustering = AgglomerativeClustering(n\_clusters=3, linkage=method)

    labels = clustering.fit\_predict(X\_scaled)

    sil\_score = silhouette\_score(X\_scaled, labels)

    db\_index = davies\_bouldin\_score(X\_scaled, labels)

    results.append([method, sil\_score, db\_index])

    # Plotting the clustering result

    plt.figure(figsize=(6, 4))

    plt.scatter(X\_scaled[:, 0], X\_scaled[:, 1], c=labels, cmap='viridis', s=40)

    plt.title(f"Hierarchical Clustering ({method} linkage)")

    plt.xlabel("Feature 1")

    plt.ylabel("Feature 2")

    plt.show()

# -----------------------

# 4. Display Results

# -----------------------

results\_df = pd.DataFrame(results, columns=["Linkage Method", "Silhouette Score", "Davies-Bouldin Index"])

print("\nClustering Quality Metrics:\n")

print(results\_df)

