***Άσκηση 2.3***

*Ερώτημα B*

A black and red text

Description automatically generated

A screenshot of a computer

Description automatically generated

Βοηθητικός κώδικας:  
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from scipy.cluster.hierarchy import linkage, dendrogram

from scipy.spatial.distance import pdist, squareform

# Data

X = np.array([[1,5],[3,4],[0, 2], [5, 4], [2, 6], [3, 3], [2, 3], [4, 2]])

# D(X)

D = X

# Function to compute the proximity matrix using dc(x,y)=1 - cos(theta\_xy)

def proximity\_matrix(X):

    n = X.shape[0]

    P = np.zeros((n, n))  # Initialize proximity matrix

    for i in range(n):

        for j in range(n):

            if i != j:

                # Compute cosine similarity and transform to proximity

                cos\_theta = np.dot(X[i], X[j]) / (np.linalg.norm(X[i]) \* np.linalg.norm(X[j]))

                P[i, j] = 1 - cos\_theta

    return P

# P(X)

P = proximity\_matrix(X)

# Convert matrices to pandas DataFrames for better formatting

P\_df = pd.DataFrame(P, index=[f"Point {i+1}" for i in range(X.shape[0])],

                    columns=[f"Point {i+1}" for i in range(X.shape[0])])

# Print the matrices in a nicely formatted way

print("\nD(X) - Distance Matrix:")

print(D)

print("\nP(X) - Proximity Matrix:")

print(P\_df.round(3))  # Round to 3 decimal places for better readability

*Ερώτημα Γ*

**Αρχικό Στάδιο**

Ξεκινάμε με 8 μονομελή συμπλέγματα:

{1},{2},{3},{4},{5},{6},{7},{8}.

Αρχικά, εντοπίζουμε την ελάχιστη τιμή εγγύτητας στον πίνακα . Η ελάχιστη εγγύτητα είναι . Συνεπώς, τα σημεία 2 και 7 ενώνονται σε ένα νέο σύμπλεγμα:

{2,7}

Αναζητούμε ξανά τη μέγιστη τιμή εγγύτητας. Εντοπίζουμε ότι η ελάχιστη τιμή είναι . Συνεπώς, τα σημεία 4 και 6 ενώνονται σε ένα νέο σύμπλεγμα:

Η επόμενη ελάχιστη τιμή στον πίνακα εγγύτητας είναι .. Συνεπώς, τα σημεία 1 και 5 ενώνονται:

Τώρα έχουμε τρία διμελή συμπλέγματα: {2,7},{4,6},{1,5} και δύο μονομελή: {3},{8}. Ελέγχουμε τις νέες εγγύτητες και βρίσκουμε την ελάχιστη:

Τα συμπλέγματα {2,7} και {4,6} ενώνονται σε ένα νέο σύμπλεγμα:

Τα συμπλέγματα είναι πλέον: {1,5},{2,4,6,7},{3},{8}. Η ελάχιστη εγγύτητα τώρα βρίσκεται μεταξύ {1,5} και {3}:

Συνεπώς, ενώνονται:

Τα συμπλέγματα είναι τώρα: {1,3,5},{2,4,6,7},{8}. Εντοπίζουμε τη ελάχιστη εγγύτητα μεταξύ {2,4,6,7}και {8}:

Συνεπώς, ενώνονται:

Απομένουν δύο συμπλέγματα: {1,3,5} και {2,4,6,7,8}. Η ελάχιστη εγγύτητα είναι:

Όλα τα σημεία ενώνονται σε ένα τελικό σύμπλεγμα:

Παρακάτω δίνεται ο βοηθητικός κώδικας (έτοιμη βιβλιοθήκη χρησιμοποιήθηκε) και το αντίστοιχο δενδρόγραμμα.

# Convert the proximity matrix to a condensed form

P\_condensed = squareform(P)

# Hierarchical Clustering with Single Linkage

linkage\_matrix\_single = linkage(P\_condensed, method='single')

# Plot dendrogram for Single Linkage

plt.figure(figsize=(10, 5))

dendrogram(linkage\_matrix\_single, labels=np.arange(1, X.shape[0] + 1))

plt.title('Dendrogram (Single Linkage)')

plt.xlabel('Data Points')

plt.ylabel('Proximity')

plt.show()

A diagram of a single linkage

Description automatically generated

*Ερώτημα Δ*

**Αρχικό Στάδιο**

Ξεκινάμε με 8 μονομελή συμπλέγματα:

{1},{2},{3},{4},{5},{6},{7},{8}.

Αρχικά, εντοπίζουμε την ελάχιστη τιμή εγγύτητας στον πίνακα . Η ελάχιστη εγγύτητα είναι . Συνεπώς, τα σημεία 2 και 7 ενώνονται σε ένα νέο σύμπλεγμα:

{2,7}

Αναζητούμε ξανά τη μέγιστη τιμή εγγύτητας. Εντοπίζουμε ότι η ελάχιστη τιμή είναι . Συνεπώς, τα σημεία 4 και 6 ενώνονται σε ένα νέο σύμπλεγμα:

Η επόμενη ελάχιστη τιμή στον πίνακα εγγύτητας είναι . Συνεπώς, τα σημεία 1 και 5 ενώνονται:

Τώρα έχουμε τρία διμελή συμπλέγματα: {2,7},{4,6},{1,5} και δύο μονομελή: {3},{8}. Ελέγχουμε τις νέες εγγύτητες και βρίσκουμε την μέγιστη:

Τα συμπλέγματα {2,7} και {4,6} ενώνονται σε ένα νέο σύμπλεγμα:

Τα συμπλέγματα είναι πλέον: {1,5},{2,4,6,7},{3},{8}. Η μέγιστη εγγύτητα τώρα μεταξύ {1,5} και {3}:

Συνεπώς, ενώνονται:

Τα συμπλέγματα είναι τώρα: {1,3,5},{2,4,6,7},{8}. Εντοπίζουμε τη μέγιστη εγγύτητα μεταξύ {2,4,6,7}και {8}:

Συνεπώς, ενώνονται:

Απομένουν δύο συμπλέγματα: {1,3,5} και {2,4,6,7,8}. Η ελάχιστη εγγύτητα είναι:

Όλα τα σημεία ενώνονται σε ένα τελικό σύμπλεγμα:

Παρακάτω δίνεται ο βοηθητικός κώδικας (έτοιμη βιβλιοθήκη χρησιμοποιήθηκε) και το αντίστοιχο δενδρόγραμμα.

# Hierarchical Clustering with Complete Linkage

linkage\_matrix\_complete = linkage(P\_condensed, method='complete')

# Plot dendrogram for Complete Linkage

plt.figure(figsize=(10, 5))

dendrogram(linkage\_matrix\_complete, labels=np.arange(1, X.shape[0] + 1))

plt.title('Dendrogram (Complete Linkage)')

plt.xlabel('Data Points')

plt.ylabel('Proximity')

plt.show()

A graph with colored lines

Description automatically generated

*Ερώτημα Ε*

Παρατηρούμε ότι τα δύο δενδρογράμματα που προκύπτουν από τις μεθόδους απλού και πλήρους δεσμού είναι πανομοιότυπα. Και στις δύο περιπτώσεις, οι ομαδοποιήσεις που προκύπτουν είναι οι εξής: ένα cluster για τα σημεία {3,1,5} και ένα για τα σημεία {2,7,4,6,8}. Το γεγονός ότι οι δύο μέθοδοι οδηγούν στο ίδιο αποτέλεσμα μπορεί να αποδοθεί στη φύση των δεδομένων. Πιο συγκεκριμένα, οι αποστάσεις μεταξύ των σημείων πρέπει να είναι τέτοιες ώστε η ελάχιστη και η μέγιστη απόσταση εντός των ομάδων να είναι παρόμοιες, ώστε τόσο ο απλός όσο και ο πλήρης δεσμός να καταλήγουν σε παρόμοια ή ταυτόσημα αποτελέσματα.