Heinrich Heine Universität Düsseldorf

HEIDERLAB

Exercise 4 on Machine Learning WS 2023/24
Prof. Dr. Dominik Heider
M. Sc. Arsam (Mohammad) Tajabadi
Submission: Until 22.11.2023, 23:59 on Ilias.

Task 1. K-means (5 points)

The following data points are given in Euclidean space:

Point	X	Y
A	1.2	0.8
B	-0.6	-1.3
C	-0.8	0.2
D	0.2	0.3

Point A and C are initially assigned to cluster centroid C1, points B and D are initially assigned to cluster centroid C2.

a) Determine the minimum-distance partition using the K-Means method and the Euclidean distance. Start by first determining the respective cluster centroids. (2.5 P.)

b) Calculate the Silhouette Coefficients in Python and visualize them. (1.5 P.)

c) Briefly describe what silhouettes are calculated for, explain the meaning of negative values, and indicate what the silhouette coefficient represents. (1 P.)

Task 2. Clustering (5 points)

a) Compute a Hierarchical Clustering from the distance matrix (*distancematrix.csv*) using the Average Linkage method. Draw the corresponding dendrogram (either in Python or by hand). (2 P.)

b) How do partitioning methods differ from hierarchical methods? (1 P.)

c) Explain hard clustering and soft clustering. (1 P.)

d) Briefly describe when density-based clustering techniques are used and the advantages and disadvantages. (1 P.)