

Assignment 4

In order to illustrate the behavior of the single linkage and complete linkage cluster algorithms we use the following sample consisting of 6 points and their values for two variables:

Point	Variable 1	Variable 2
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

Use the Euclidean distance as similarity measure and assume that the two variables are already standardized. The Euclidean distance matrix for these six points is shown below:

	P1	P2	P3	P4	P5	P6
P1	0.00	0.24	0.22	0.37	0.34	0.23
P2	0.24	0.00	0.15	0.20	0.14	0.25
P3	0.22	0.15	0.00	0.15	0.28	0.11
P4	0.37	0.20	0.15	0.00	0.29	0.22
P5	0.34	0.14	0.28	0.29	0.00	0.39
P6	0.23	0.25	0.11	0.22	0.39	0.00

For the following tasks DO NOT use SPSS or any other kind of software package. Use just paper, pencil and a ruler.

1. Plot the six points in a coordinate system where the two variables represent the orthogonal axes.
2. Use Single Linkage and plot the dendrogram. The height at which 2 clusters are merged in the dendrogram reflects the distance of the 2 clusters.
3. Use Complete Linkage and plot the dendrogram. The height at which 2 clusters are merged in the dendrogram reflects the distance of the 2 clusters.