

Homework Assignment 6

Due date: 17th of December 2013

Please note that even though the homework assignments are optional, you're still highly encouraged to answer them, as they will help you prepare for your final exam. You can work in a group of two or alone. Solutions can be dropped off at the institute's homework mailbox located on the 2nd floor, next to room 238. In that case, please make sure both your name and matriculation number is noted down. If your answers span more than one sheet, kindly staple them together. Another alternative is to send your homework via email to: elmaarry@ifis.cs.tu-bs.de

LECTURE 7: DOCUMENT CLUSTERING

Using a K-Means algorithm on a data collection, four clusters were generated, namely: x,y,z, and w. Given the below distance matrix between the different points in space. Answer the following exercises.

Note: Cluster x have two members: x_1, x_2 .

	y_1	y_2	y_3	y_4	w_1	w_2	z_1	z_2	x_1	x_2
y_1	0	3	4	5	12	12.5	10	13	6	6
y_2		0	5	4	12.5	13	9	12	8	8.5
y_3			0	3	10	12	5.1	8	13	13.5
y_4				0	8.5	9.5	5.4	8	12	12.5
w_1					0	2	6	5	19	19
w_2						0	7.5	7	19	19
z_1							0	3	17	17.5
z_2								0	20	20.5
x_1									0	1
x_2										0

EXERCISE 6.1

Compute the quality of the K-Means clustering.

EXERCISE 6.2

Draw the dendrogram corresponding to the agglomerative hierarchical clustering when the similarity between clusters is based on:

- i. Single-link clustering
- ii. Complete-link clustering
- iii. Centroid clustering
- iv. Group average clustering