Machine Learning Seminar – Clustering

1. *k*-means

Given the following data:

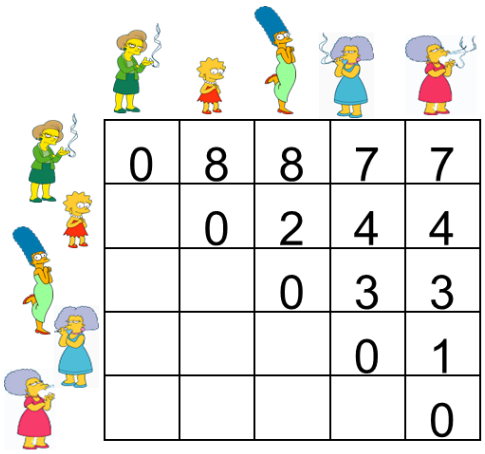
x_1	x_2
96	8
121	9
89	10
87	11
112	11
9	22
6	24
8	23
11	22
12	22

Cluster the data using *k*-means clustering (k=2). You should normalise the data, and then initialise the two centroids to the positions of the first two data (remember: k-means doesn't use medoids – we are just starting the centroids in the same place).

Complete three iterations of the k-means algorithm (or stop early if membership doesn't change for an iteration)

2. PAM

Given the following data:



0	8	8	7	7
	0	2	4	4
		0	3	3
			0	1
				0

Cluster this data using PAM and k=2. Use the first two data as your initial medoids (Selma and Lisa)

3. Hierarchical:

Given the following data:

					
	0	3	7	10	8
		0	18	6	15
			0	5	8
				0	8
					0

Cluster this data using bottom-up average-linkage clustering

4. Density Peaks

Given the following data:

id	x	y
1	5	20
2	19	7
3	16	9
4	17	5
5	7	25
6	6	22
7	4	24

Use the density peaks algorithm, with a cut-off kernel and d_c=4, to cluster the data with k=2. You do **not** need to normalise this data