Homework 6: k-means clustering

TDA231 - Algorithms for Machine Learning & Inference

Theodor Åstrand, theast@student.chalmers.se, 931109-9114
Linnéa Otterlind, linott@student.chalmers.se, 921126-3620

Chalmers University of Technology

March, 2016

Problem 1.1

(b)

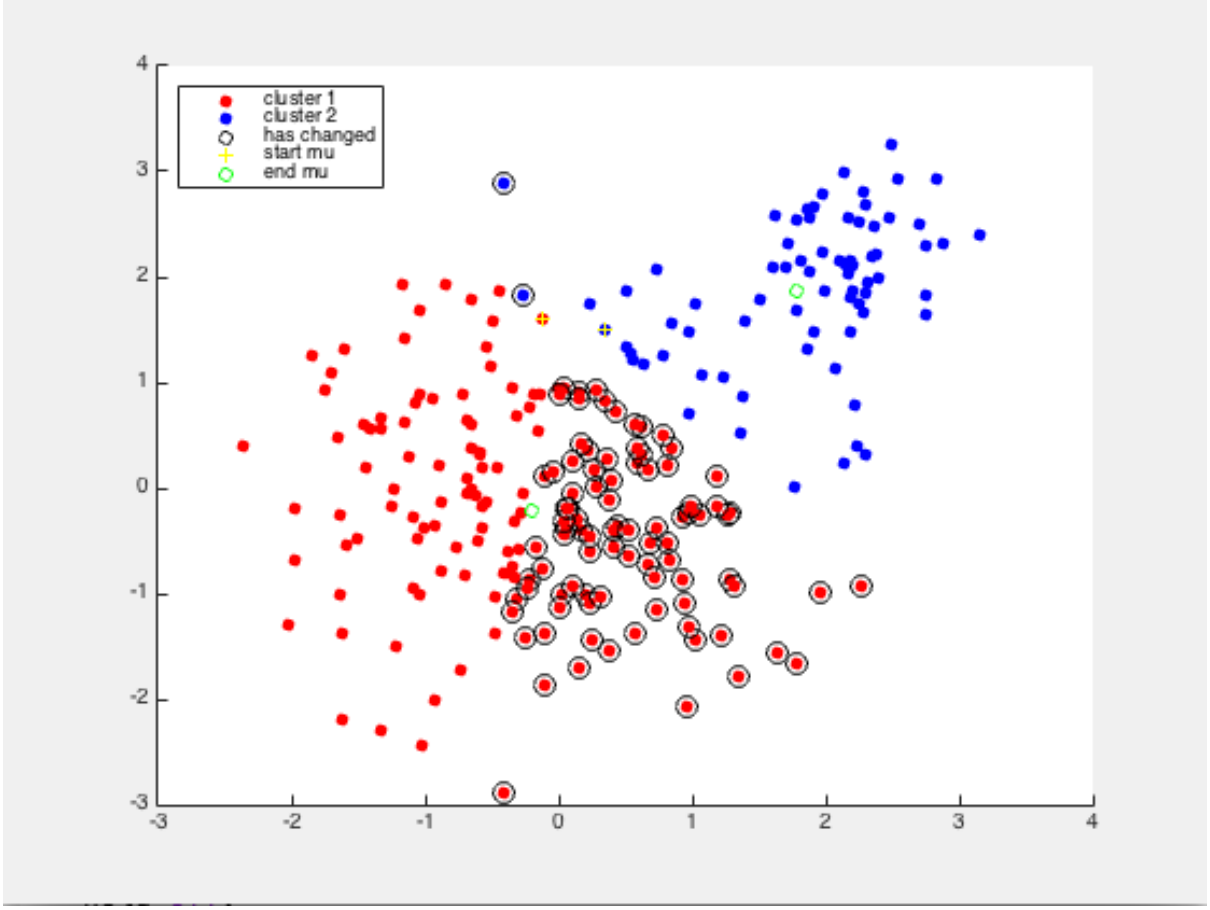

Figure 1: k-means algorithm on hw6_p1a.mat with k = 2

(d)

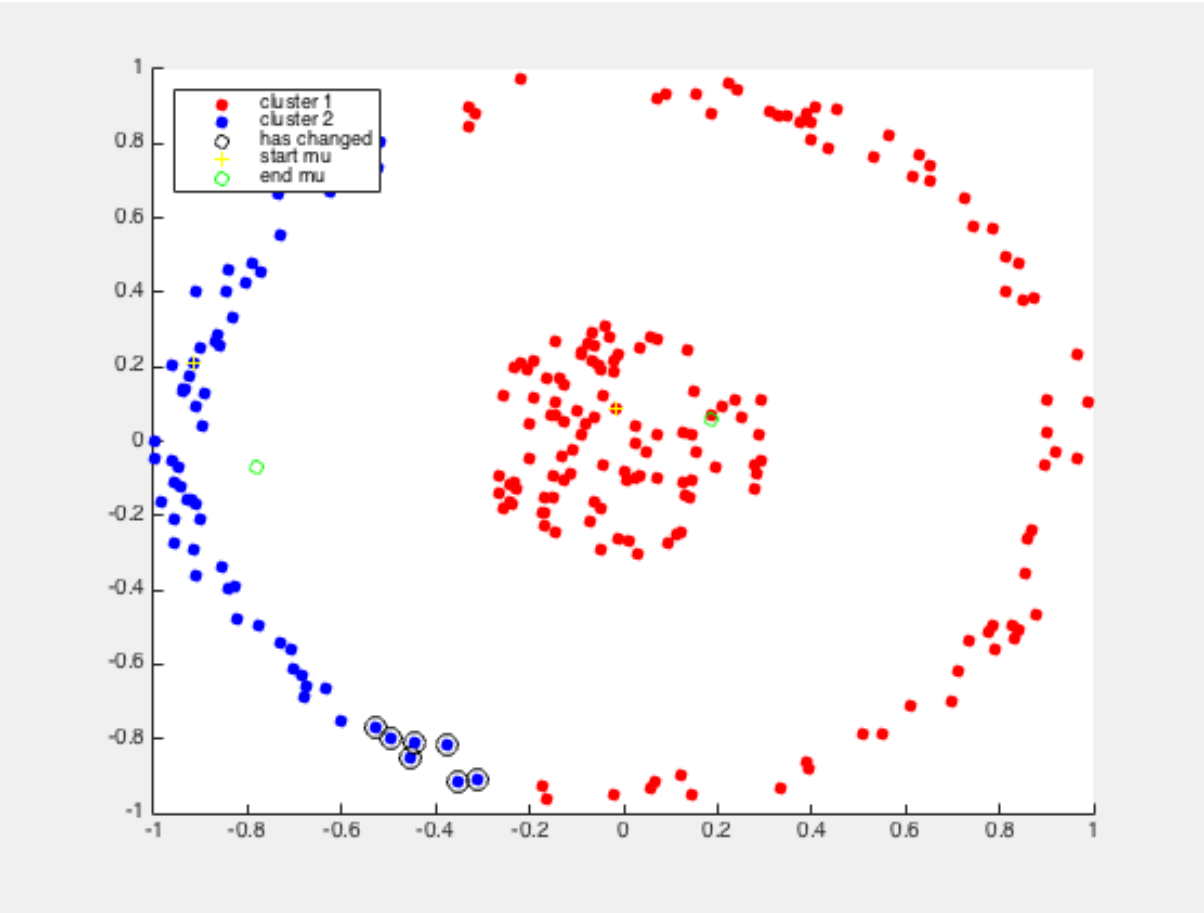

Figure 2: k-means algorithm on hw6_p1b.mat with k = 2

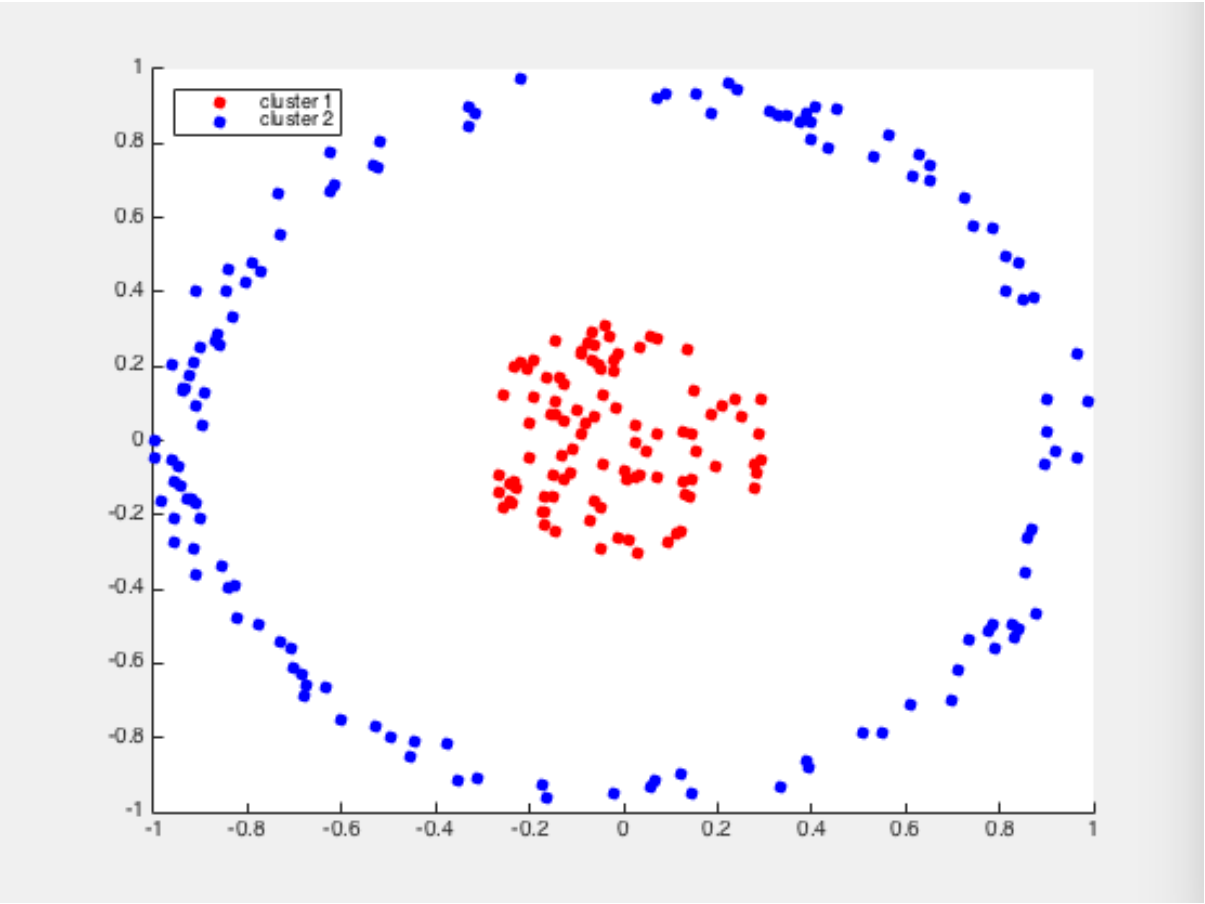

Figure 3: k-means kernel algorithm on hw6_p1b.mat with k = 2 and $\sigma = 0.2$

Problem 1.2

(a)

c 1	**c 2**	**c 3**	**c 4**	**c 5**	**c 6**	**c 7**	**c 8**	**c 9**	**c 10**
something	be	concerned	team	country	put	area	business	curtis	another
come	that	political	ever	on	brought	now	making	oliver	up
nothing	more	that	start	to	finally	near	new	smith	making
own	some	acknowledging	final	well	after	nearby	financial	allen	back
what	an	seeking	next	at	to	today	for	scott	with
happy	similarly	government	last	from	eventually	part	public	james	be
kind	even	nonetheless	going	making	soon	still	limited	frank	to
gone	to	regard	making	for	before	it	full	george	it
did	instance	to	coming	with	bringing	from	corporate	leslie	turn
supposed	well	regarding	chance	first	immediately	close	private	miller	one

(b)

Each run separately yielded the following values:

0.4916
0.6623
0.5058
0.6453
0.7139
0.5471
0.7577
0.4739
0.6270
0.5485

The average fraction of word pairs that remains in the same cluster is thus: 0.5973. So the chance of a word pair remaining in a pair is quite high.

Since the probability of a word pair to remain is dominant (about 60%), we can conclude that it's likely that words stick together in the same cluster. Although, there is still a quite large probability (about 40%) that words change cluster. Since the probability that the words change cluster is so high we can conclude that the clusters are not particularly stable.

(c)

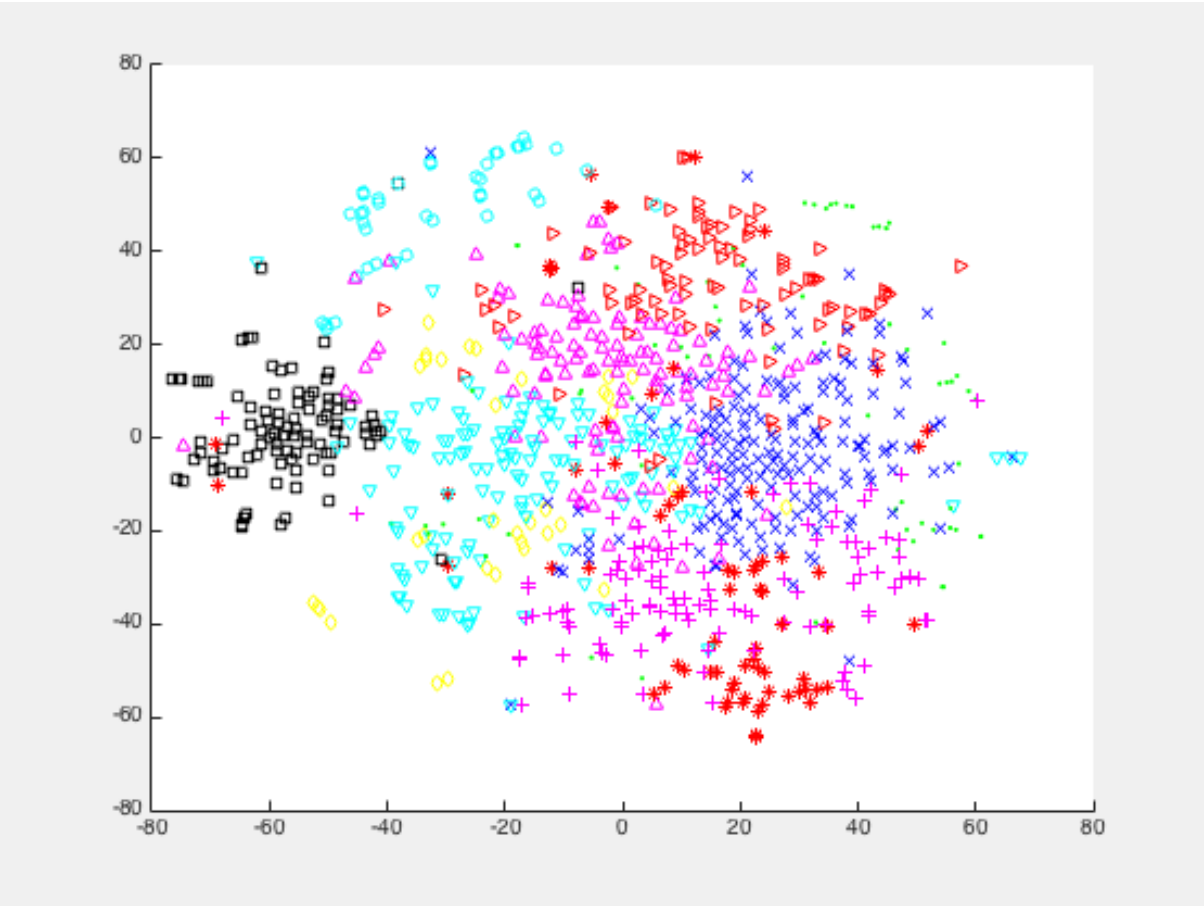

Figure 4: Visualisation of 1000 random sampled words.