## ISS4003 – Data Mining

# Praktikum 3

### **K-Means Clustering**

No. Praktikum	:	3
Minggu / Sesi	:	11/3
Tujuan	:	Memahami penggunaan fungsi kmeans() dalam R untuk melakukan K-Means <i>clustering</i>
Setoran	:	<b>Softcopy</b> : laporan tugas dalam format pdf dan satu program R
Waktu Penyetoran	:	30 November 2017: Pukul 00:00 WIB

#### I. Referensi

1. Gareth James, Daniela Witten, Trevor Hastie, dan Robert Tibshirani, "An Introduction to Statistical Learning with Applications in R," Springer, 2013.

### II. K-Means Clustering

Di praktikum kali ini kita akan mencoba menggunakan fungsi kmeans() untuk melakukan K-Means *clustering* dalam R. Mari kita mulai!

1. Dalam program ini kita menggunakan *simulated dataset*. Untuk itu, buat 2 cluster dengan jumlah masing-masing 25 data. Penting untuk menetapkan *random seed* menggunakan fungsi set.seed(x). Jika tidak menginisialisasi fungsi ini, nilai akan diacak setiap kali program dipanggil. (Catatan: jika tidak percaya, silahkan buktikan dengan menghapus fungsi set.seed(x) atau gunakan set.seed(), dan lihat *dataset* yang dihasilkan.)

```
set.seed(2)
m = matrix (rnorm (50*2), ncol =2)
m[1:25,1] = m[1:25,1] + 3
m[1:25,2] = m[1:25,2] - 4
colnames(m) <- c("x", "y")</pre>
```

2. Gunakan K-Means clustering dengan K=2.

```
km.out = kmeans(x,2,nstart=20)
km.out$cluster
```

Eksekusi kode tersebut akan menghasilkan keluaran sebagai berikut:

Perhatikan baik-baik hasil tersebut! Data yang kita observasi dikelompokkan dengan **sempurna** ke dalam 2 *cluster* meskipun kita tidak memberikan informasi kelompok pada fungsi kmeans().

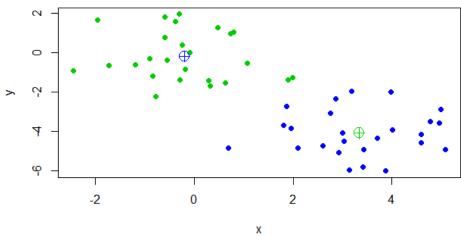
3. *Plot* data yang telah di-*cluster*.

```
plot(m, col=(km.out$cluster+2),
    main="Hasil K-Means Clustering dengan K=2",
    pch=20,
    cex=1.5)

points(km.out$centers,
    col=4:3,
    pch=10,
    cex=2)
```

Eksekusi kode tersebut akan menghasilkan keluaran sebagai berikut:

# Hasil K-Means Clustering dengan K=2



### III. Tugas

1. Tulis ulang program tersebut dengan menggunakan *dummy dataset* (minimal 900 observasi/*tuple*). Contohnya:

observation	X	y
1	1	3
2	3	3
3	4	3 3 3
4	5	3
5	1	2 2
6	4	2
7	1	1
8	2	1
9	2	2
10	5	4
11	4	5
900	5	4

2. Tampilkan dan jelaskan hasil eksekusi dari *line of code* di bawah ini untuk masing-masing inisialisasi nilai K=2, K=3, dan K=4:

km.out\$cluster

- 3. Tampilkan dan jelaskan hasil (plot) K-Means Clustering untuk masing-masing K=2, K=3, dan K-4!
- 4. Berikan pemahaman Anda sendiri untuk setiap baris kode dalam **program yang sudah Anda modifikasi** tersebut (bukan membuat baris komentar dalam program)!

#### IV. Setoran

1. Laporan berisi jawaban untuk soal tugas nomor 1, 2, 3, dan 4, dengan nama dokumen: **praktikum3\_[nim].pdf**. Tulislah **Praktikum 3**, **NIM**, dan **Nama** di sudut kanan atas dokumen. Contoh:



2. Program **kmeans.R** yang sudah dimodifikasi sesuai soal tugas nomor 1. Keduanya disetor ke **ecourse.del.ac.id** pada halaman mata kuliah Data Mining.