

Tugas 1

Kode program dengan menggunakan 900 dummy dataset.

```
2 set.seed(2)
3 m=matrix(rnorm(900*2), ncol=2)
4 m[1:450,1]=m[1:450,1]+5
5 m[1:450,2]=m[1:450,2]-6
6 colnames(m) <- c("x","y")
```

Tugas 2

Kode program K-Means Clustering dengan K = 2

```
1 # K-Means clustering dengan k=2
2 set.seed(2)
3 m=matrix(rnorm(900*2), ncol=2)
4 m[1:450,1]=m[1:450,1]+5
5 m[1:450,2]=m[1:450,2]-6
6 colnames(m) <- c("x","y")
7
8 km.out=kmeans(m,2,nstart = 20)
9 km.out$cluster
10
11 plot(m, col=(km.out$cluster+5),
12      main = "Hasil K-Means Clustering dengan k = 2",
13      pch=18,
14      cex=1)
15
16 points(km.out$centers,
17        col=4:3,
18        pch=8,
19        cex=2)
```

Kode program K-Means Clustering dengan K = 3

```
21 # K-Means clustering dengan k=3
22 set.seed(3)
23 m=matrix(rnorm(900*2), ncol=2)
24 m[1:450,1]=m[1:450,1]+5
25 m[1:450,2]=m[1:450,2]-6
26 colnames(m) <- c("x","y")
27
28 km.out=kmeans(m,3,nstart = 20)
29 km.out$cluster
30
31 plot(m, col=(km.out$cluster+5),
32      main = "Hasil K-Means Clustering dengan k = 3",
33      pch=20,
34      cex=1.5)
35
36 points(km.out$centers,
37        col=4:3,
38        pch=8,
39        cex=2)
```

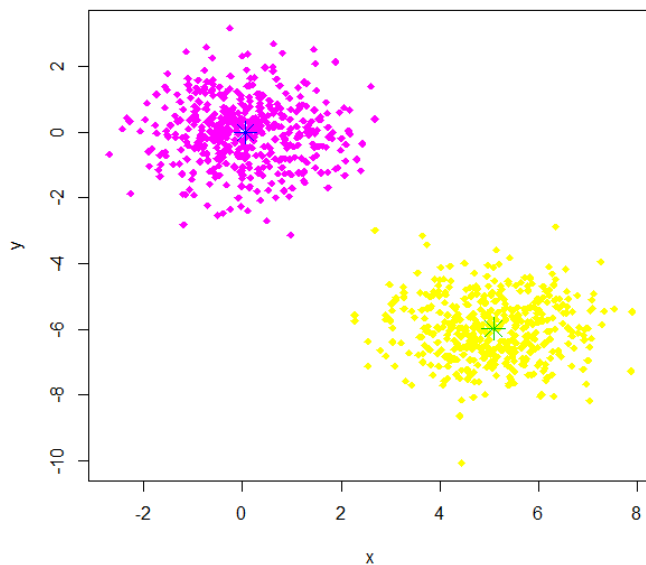
Kode program K-Means Clustering dengan $K = 4$.

```
42 # K-Means clustering dengan K=4
43 set.seed(4)
44 m=matrix(rnorm(900*2), ncol=2)
45 m[1:450,1]=m[1:450,1]+5
46 m[1:450,2]=m[1:450,2]-6
47 colnames(m) <- c("x","y")
48
49 km.out=kmeans(m,4,nstart = 20)
50 km.out$cluster
51
52 plot(m, col=(km.out$cluster+4),
53      main = "Hasil K-Means Clustering dengan K = 4",
54      pch=20,
55      cex=1)
56
57 points(km.out$centers,
58        col=2:1,
59        pch=8,
60        cex=2)
```

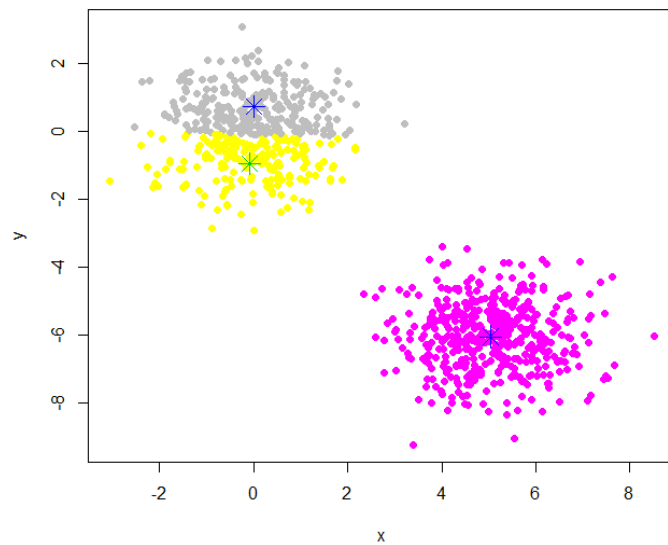
Tugas 3

Visualisasi *plots* dan *points* hasil *run* kode program.

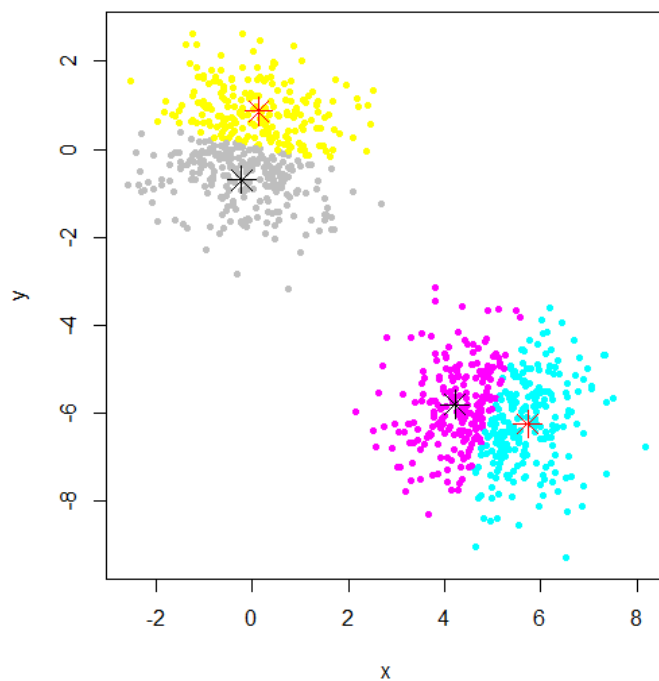
Hasil K-Means Clustering dengan $K = 2$



Hasil K-Means Clustering dengan K = 3



Hasil K-Means Clustering dengan K = 4



Tugas 4

Penjelasan kode program *clustering* dengan jumlah klaster K=2.

Menetapkan random seed. Kita harus menetapkan set seed setiap saat kita ingin mendapatkan hasil acak yang dapat dihasilkan kembali.

```
set.seed(2)
```

Mendefinisikan objek `m` sebagai matrix dengan jumlah dataset 900 dan meng-*generate* vector dari 2 *pseudo-random normal* dengan mean 0 dan varians 1, dan terdiri dari 2 kolom matrix.

```
m=matrix(rnorm(900*2), ncol=2)
```

Mendistribusikan jumlah data sebanyak 450 untuk masing masing cluster. Konstanta 5 dan 6 merupakan nilai yang akan menentukan jarak antar cluster.

```
m[1:450,1]=m[1:450,1]+5
m[1:450,2]=m[1:450,2]-6
```

Membuat nama kolom objek m dengan nama x dan y.

```
colnames(m) <- c("x","y")
```

Fungsi untuk melakukan clustering terhadap objek m, dengan jumlah cluster 2, dan jumlah iterasi 20.

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

Hasil yang diperoleh setelah menjalankan program diatas adalah sebagai berikut.

[illegible]

Menampilkan hasil cluster dan center points. Pada kode program, col digunakan untuk menentukan warna visualisasi plot dan points, pch untuk menentukan bentuk plot dan points, dan cex digunakan untuk menentukan ukuran plot dan points.

```
plot(m, col=(km.out$cluster+5),
     main = "Hasil K-Means Clustering dengan K = 2",
     pch=18,
     cex=1)
```

```
points(km.out$centers,
       col=4:3,
       pch=8,
       cex=2)
```

Penjelasan kode program *clustering* dengan jumlah kluster K=3.

Menetapkan random seed. Kita harus menetapkan set seed setiap saat kita ingin mendapatkan hasil acak yang dapat dihasilkan kembali. Kemudian endefenisikan objek m sebagai matrix dengan jumlah dataset 900 dan meng-*generate* vector dari 2 *pseudo-random normal* dengan mean 0 dan varians 1, dan terdiri dari 2 kolom matrix.

Data sebanyak 450 didistribusikan untuk masing masing cluster. Konstanta 5 dan 6 merupakan nilai yang akan menentukan jarak antar cluster.

Membuat nama kolom objek m dengan nama x dan y.

```
set.seed(3)
m=matrix(rnorm(900*2), ncol=2)
m[1:450,1]=m[1:450,1]+5
m[1:450,2]=m[1:450,2]-6
colnames(m) <- c("x","y")
```

Fungsi untuk melakukan clustering terhadap objek m, dengan jumlah cluster 3, dan jumlah iterasi 20.

```
km.out=kmeans(m,3,nstart = 20)
km.out$cluster
```

Hasil yang diperoleh setelah menjalankan program diatas adalah sebagai berikut.

[illegible]

Menampilkan hasil cluster dan center points. Pada kode program, col digunakan untuk menentukan warna visualisasi plot dan points, pch untuk menentukan bentuk plot dan points, dan cex digunakan untuk menentukan ukuran plot dan points.

```
plot(m, col=(km.out$cluster+5),
     main = "Hasil K-Means Clustering dengan K = 3",
     pch=20,
     cex=1.5)
```

```
points(km.out$centers,
       col=4:3,
       pch=8,
       cex=2)
```

Penjelasan kode program *clustering* dengan jumlah kluster K=4

Menetapkan random seed. Kita harus menetapkan set seed setiap saat kita ingin mendapatkan hasil acak yang dapat dihasilkan kembali. Lalu mendefinisikan objek m sebagai matrix dengan jumlah dataset 900 dan meng-*generate* vector dari 2 *pseudo-random normal* dengan mean 0 dan varians 1, dan terdiri dari 2 kolom matrix. Mendistribusikan jumlah data sebanyak 450 untuk masing masing cluster. Konstanta 5 dan 6 merupakan nilai yang akan menentukan jarak antar cluster dan membuat nama kolom objek m dengan nama x dan y.

```
# K-Means clustering dengan K=4
set.seed(4)
m=matrix(rnorm(900*2), ncol=2)
m[1:450,1]=m[1:450,1]+5
m[1:450,2]=m[1:450,2]-6
colnames(m) <- c("x","y")
```

Fungsi untuk melakukan clustering terhadap objek m, dengan jumlah cluster 4, dan jumlah iterasi 20.

```
km.out=kmeans(m,4,nstart = 20)
km.out$cluster
```

Hasil yang diperoleh setelah menjalankan program diatas adalah sebagai berikut.

[illegible]

Menampilkan hasil cluster dan center points. Pada kode program, col digunakan untuk menentukan warna visualisasi plot dan points, pch untuk menentukan bentuk plot dan points, dan cex digunakan untuk menentukan ukuran plot dan points.

```
plot(m, col=(km.out$cluster+4),
     main = "Hasil K-Means Clustering dengan K = 4",
     pch=20,
     cex=1)

points(km.out$centers,
       col=2:1,
       pch=8,
       cex=2)
```

Reference:

<https://stackoverflow.com/questions/13605271/reasons-for-using-the-set-seed-function>

<http://www.stat.umn.edu/macanova/htmlhelp/node1.htm>