problems-3-and-4.R

lukemcevoy

2021-12-15

```
#### Problem 3

# clear the environment
rm(list=ls())

# select the data
filename<-'/Users/lukemcevoy/Develop/stevens/f21/dataMining/final/p3-4/k-means-data.csv'
data<-read.csv(filename)

kmeans_2<- kmeans(data[,2:4],2,nstart = 1)
kmeans_2$cluster
```

```
## [1] 2 1 1 1 2 1 2
```

```
kmeans_2$centers
```

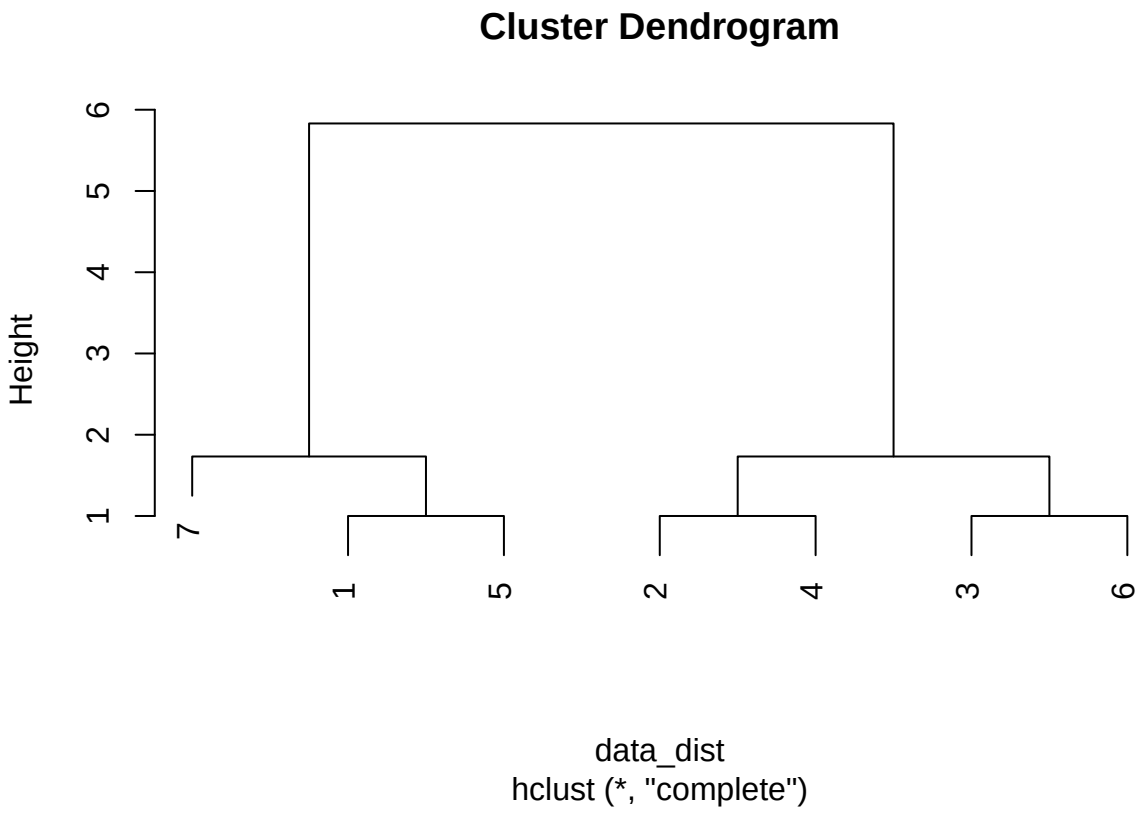
```
##          X        Y        Z
## 1 4.250000 3.500000 4.250000
## 2 1.333333 1.333333 1.333333
```

```
table(kmeans_2$cluster,data[,1])
```

```
##
##     a b c d e f g
##   1 0 1 1 1 0 1 0
##   2 1 0 0 0 1 0 1
```

```
# Part A
#   Members of each cluster are
#     Cluster 1 = (a,e,g)
#     Cluster 2 = (b,c,d,f)

# Part B
#   Coordinates for each cluster center
#                    X        Y        Z
#   Cluster 1     1.333333 1.333333 1.333333
#   Cluster 2     4.250000 3.500000 4.250000



# Problem 4

# clear the environment
rm(list=ls())
```

```r
# select the data
filename<-'/Users/lukemcevoy/Develop/stevens/f21/dataMining/final/p3-4/k-means-data.csv'
data<-read.csv(filename)

data_dist<-dist(data[2:4])
hclust_resutls<-hclust(data_dist)
plot(hclust_resutls)
```




```r
dev.off()
```

```
## null device
##           1
```

```r
hclust_2<-cutree(hclust_resutls,2)
table(hclust_2,data[,1])
```

```
##
## hclust_2 a b c d e f g
##        1 1 0 0 0 1 0 1
##        2 0 1 1 1 0 1 0
```

```r
round(hclust_resutls$height,2)
```

```
## [1] 1.00 1.00 1.00 1.73 1.73 5.83
```

```r
# Part A
#   Members of each cluster are
#     Cluster 1 = (a,e,g)
#     Cluster 2 = (b,c,d,f)
```

```
# Part B
#   Coordinates for each cluster center
#     1.00 1.00 1.00 1.73 1.73 5.83
```