Assignment6

Haojin Li (Declan)

2020/7/29

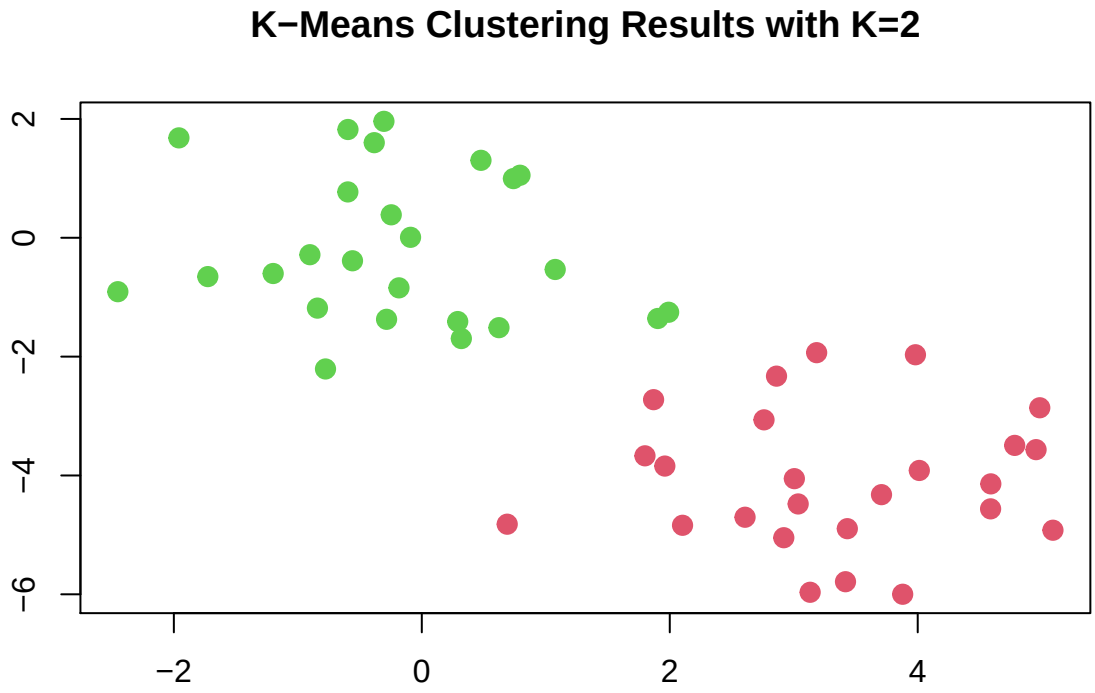
Contents

10.5.1 K-Means Clustering

```
set.seed(2)
x=matrix(rnorm (50*2), ncol=2)
x[1:25,1]=x[1:25,1]+3
x[1:25,2]=x[1:25,2]-4
km.out=kmeans (x,2, nstart =20)
km.out$cluster
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2
## [39] 2 2 2 2 2 2 2 2 2 2 2 2
```

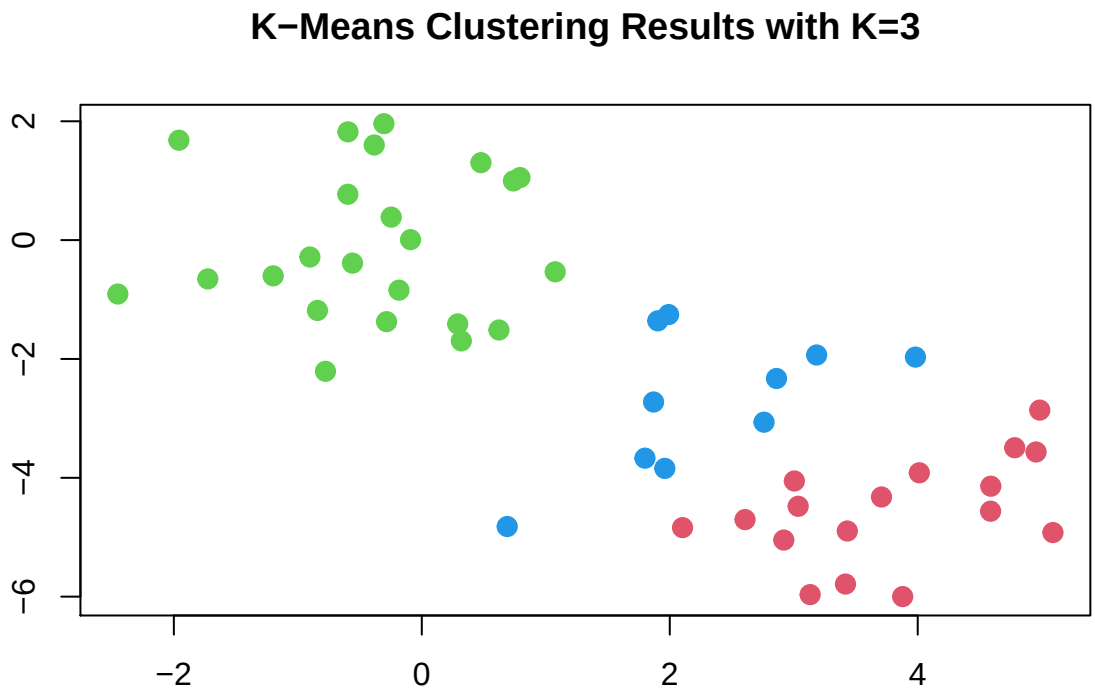
```
plot(x, col=(km.out$cluster +1), main="K-Means Clustering Results with K=2",
xlab="", ylab="", pch=20, cex=2)
```




```
set.seed(4)
km.out=kmeans (x,3, nstart =20)
km.out
```

```
## K-means clustering with 3 clusters of sizes 17, 23, 10
##
## Cluster means:
##         [,1]        [,2]
## 1  3.7789567 -4.56200798
## 2 -0.3820397 -0.08740753
## 3  2.3001545 -2.69622023
##
## Clustering vector:
##  [1] 1 3 1 3 1 1 1 3 1 3 1 3 1 3 1 1 1 1 1 3 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2
## [39] 2 2 2 2 2 3 2 3 2 2 2 2
##
## Within cluster sum of squares by cluster:
## [1] 25.74089 52.67700 19.56137
##  (between_SS / total_SS =  79.3 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
plot(x, col=(km.out$cluster +1), main="K-Means Clustering Results with K=3",
xlab="", ylab="", pch=20, cex=2)
```




```
set.seed(3)
km.out = kmeans(x, 3, nstart = 1)
km.out$tot.withinss
```

```
## [1] 97.97927
```

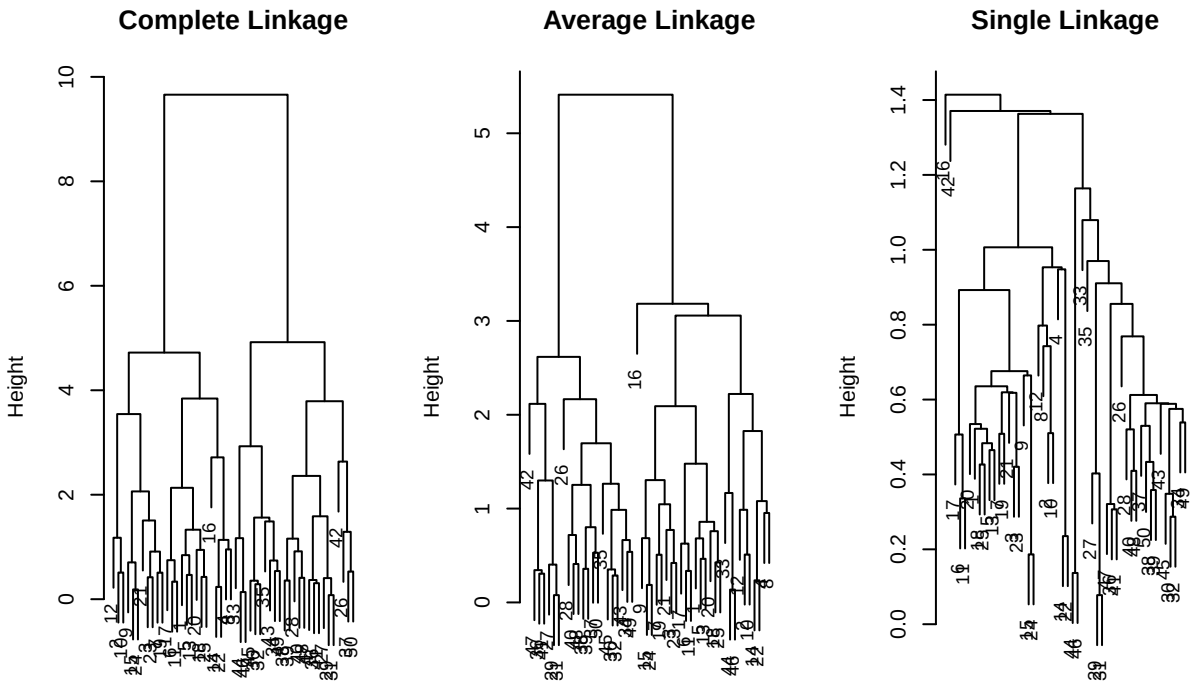
```
km.out = kmeans(x, 3, nstart = 20)
km.out$tot.withinss
```

```
## [1] 97.97927
```

10.5.2 Hierarchical Clustering

```
hc.complete =hclust(dist(x), method="complete")
hc.average =hclust(dist(x), method ="average")
hc.single=hclust(dist(x), method ="single")
```

```
par(mfrow=c(1,3))
plot(hc.complete ,main="Complete Linkage", xlab="", sub="", cex=.9)
plot(hc.average , main="Average Linkage", xlab="", sub="", cex=.9)
plot(hc.single , main="Single Linkage", xlab="", sub="", cex=.9)
```




```
cutree(hc.complete , 2)
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2
## [39] 2 2 2 2 2 2 2 2 2 2 2 2
```

```
cutree(hc.average , 2)
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 1 2 2 2 2 2
## [39] 2 2 2 2 2 2 1 2 1 2 2 2 2
```

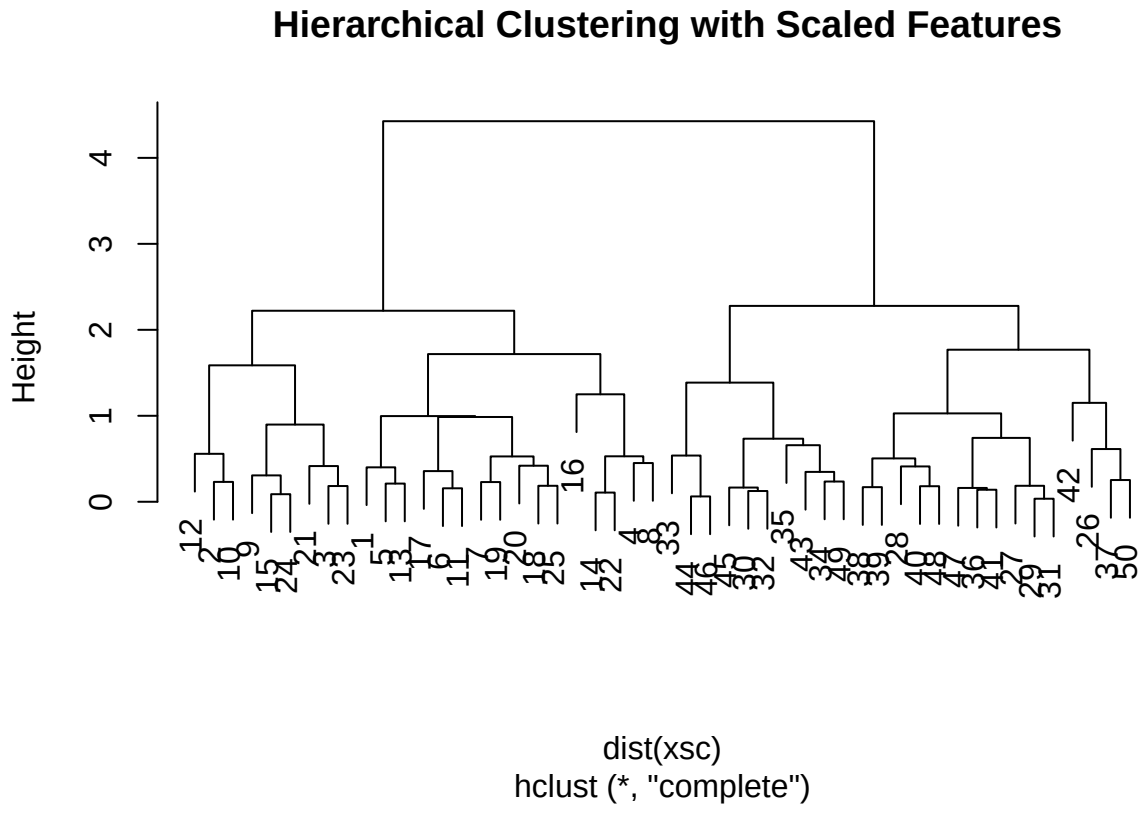
```
cutree(hc.single , 2)
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 1 1 1 1 1 1 1 1 1 1 1 1
```

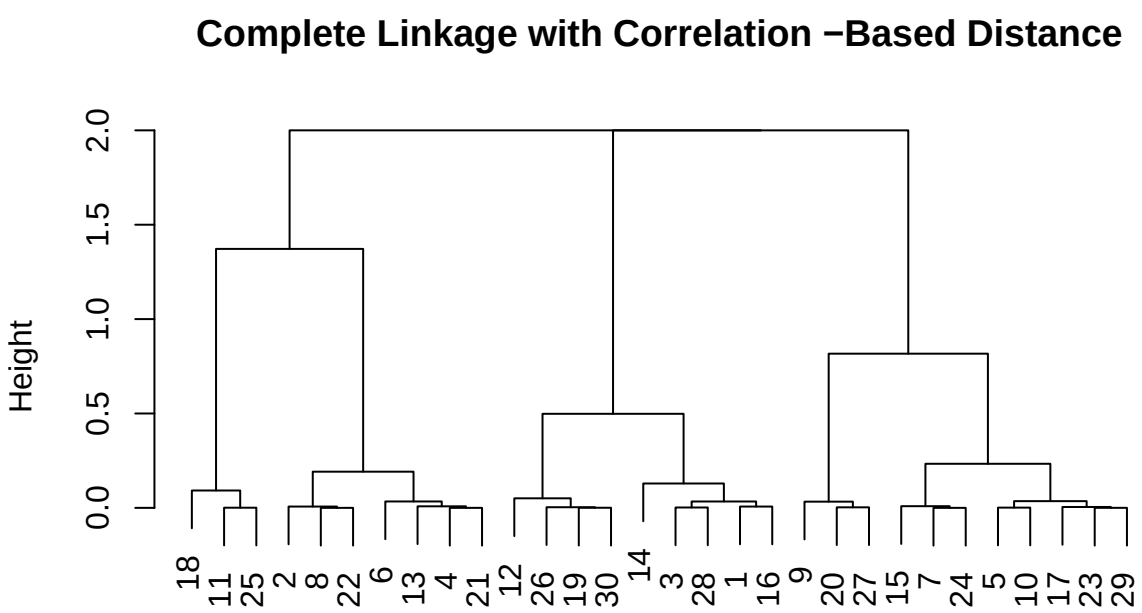
```
cutree(hc.single , 4)
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 3 3 3 3 3 3 3 3 3 3 3 3 3
## [39] 3 3 3 4 3 3 3 3 3 3 3 3
```

```
xsc=scale(x)
plot(hclust(dist(xsc), method ="complete"),
main=" Hierarchical Clustering with Scaled Features")
```




```
x=matrix(rnorm (30*3), ncol=3)
dd=as.dist(1-cor(t(x)))
plot(hclust(dd, method ="complete"),
main="Complete Linkage with Correlation -Based Distance", xlab="", sub ="")
```


Complete Linkage with Correlation −Based Distance
Height
0.0
0.5
1.0
1.5
2.0
18
11
25
2
8
22
6
13
4
21
12
26
19
30
14
3
28
1
16
9
20
27
15
7
24
5
10
17
23
29