HW08

Omkar Sinha

11/11/2021

```
## Load the data
rm(list=ls())
file <- file.choose()


bc_temp <-  read.csv(file)
summary(bc_temp)
```

```
##        id              diagnosis          radius_mean      texture_mean
##  Min.   :     8670   Length:569         Min.   : 6.981   Min.   : 9.71
##  1st Qu.:   869218   Class :character   1st Qu.:11.700   1st Qu.:16.17
##  Median :   906024   Mode  :character   Median :13.370   Median :18.84
##  Mean   : 30371831                      Mean   :14.127   Mean   :19.29
##  3rd Qu.:  8813129                      3rd Qu.:15.780   3rd Qu.:21.80
##  Max.   :911320502                      Max.   :28.110   Max.   :39.28
##  perimeter_mean     area_mean      smoothness_mean   compactness_mean
##  Min.   : 43.79   Min.   : 143.5   Min.   :0.05263   Min.   :0.01938
##  1st Qu.: 75.17   1st Qu.: 420.3   1st Qu.:0.08637   1st Qu.:0.06492
##  Median : 86.24   Median : 551.1   Median :0.09587   Median :0.09263
##  Mean   : 91.97   Mean   : 654.9   Mean   :0.09636   Mean   :0.10434
##  3rd Qu.:104.10   3rd Qu.: 782.7   3rd Qu.:0.10530   3rd Qu.:0.13040
##  Max.   :188.50   Max.   :2501.0   Max.   :0.16340   Max.   :0.34540
##  concavity_mean    concave.points_mean symmetry_mean    fractal_dimension_mean
##  Min.   :0.00000   Min.   :0.00000     Min.   :0.1060   Min.   :0.04996
##  1st Qu.:0.02956   1st Qu.:0.02031     1st Qu.:0.1619   1st Qu.:0.05770
##  Median :0.06154   Median :0.03350     Median :0.1792   Median :0.06154
##  Mean   :0.08880   Mean   :0.04892     Mean   :0.1812   Mean   :0.06280
##  3rd Qu.:0.13070   3rd Qu.:0.07400     3rd Qu.:0.1957   3rd Qu.:0.06612
##  Max.   :0.42680   Max.   :0.20120     Max.   :0.3040   Max.   :0.09744
##    radius_se        texture_se      perimeter_se       area_se
##  Min.   :0.1115   Min.   :0.3602   Min.   : 0.757   Min.   :  6.802
##  1st Qu.:0.2324   1st Qu.:0.8339   1st Qu.: 1.606   1st Qu.: 17.850
##  Median :0.3242   Median :1.1080   Median : 2.287   Median : 24.530
##  Mean   :0.4052   Mean   :1.2169   Mean   : 2.866   Mean   : 40.337
##  3rd Qu.:0.4789   3rd Qu.:1.4740   3rd Qu.: 3.357   3rd Qu.: 45.190
##  Max.   :2.8730   Max.   :4.8850   Max.   :21.980   Max.   :542.200
##  smoothness_se      compactness_se      concavity_se     concave.points_se
##  Min.   :0.001713   Min.   :0.002252   Min.   :0.00000   Min.   :0.000000
##  1st Qu.:0.005169   1st Qu.:0.013080   1st Qu.:0.01509   1st Qu.:0.007638
##  Median :0.006380   Median :0.020450   Median :0.02589   Median :0.010930
##  Mean   :0.007041   Mean   :0.025478   Mean   :0.03189   Mean   :0.011796
```

```
##  3rd Qu.:0.008146   3rd Qu.:0.032450   3rd Qu.:0.04205   3rd Qu.:0.014710
##  Max.   :0.031130   Max.   :0.135400   Max.   :0.39600   Max.   :0.052790
##   symmetry_se       fractal_dimension_se  radius_worst   texture_worst
##  Min.   :0.007882   Min.   :0.0008948    Min.   : 7.93   Min.   :12.02
##  1st Qu.:0.015160   1st Qu.:0.0022480    1st Qu.:13.01   1st Qu.:21.08
##  Median :0.018730   Median :0.0031870    Median :14.97   Median :25.41
##  Mean   :0.020542   Mean   :0.0037949    Mean   :16.27   Mean   :25.68
##  3rd Qu.:0.023480   3rd Qu.:0.0045580    3rd Qu.:18.79   3rd Qu.:29.72
##  Max.   :0.078950   Max.   :0.0298400    Max.   :36.04   Max.   :49.54
##  perimeter_worst    area_worst     smoothness_worst  compactness_worst
##  Min.   : 50.41   Min.   : 185.2   Min.   :0.07117   Min.   :0.02729
##  1st Qu.: 84.11   1st Qu.: 515.3   1st Qu.:0.11660   1st Qu.:0.14720
##  Median : 97.66   Median : 686.5   Median :0.13130   Median :0.21190
##  Mean   :107.26   Mean   : 880.6   Mean   :0.13237   Mean   :0.25427
##  3rd Qu.:125.40   3rd Qu.:1084.0   3rd Qu.:0.14600   3rd Qu.:0.33910
##  Max.   :251.20   Max.   :4254.0   Max.   :0.22260   Max.   :1.05800
##  concavity_worst  concave.points_worst symmetry_worst   fractal_dimension_worst
##  Min.   :0.0000   Min.   :0.00000      Min.   :0.1565   Min.   :0.05504
##  1st Qu.:0.1145   1st Qu.:0.06493      1st Qu.:0.2504   1st Qu.:0.07146
##  Median :0.2267   Median :0.09993      Median :0.2822   Median :0.08004
##  Mean   :0.2722   Mean   :0.11461      Mean   :0.2901   Mean   :0.08395
##  3rd Qu.:0.3829   3rd Qu.:0.16140      3rd Qu.:0.3179   3rd Qu.:0.09208
##  Max.   :1.2520   Max.   :0.29100      Max.   :0.6638   Max.   :0.20750
```

```
summary(bc_temp$diagnosis)
```

```
##    Length     Class      Mode
##       569 character character
```

```
table(bc_temp$diagnosis)
```

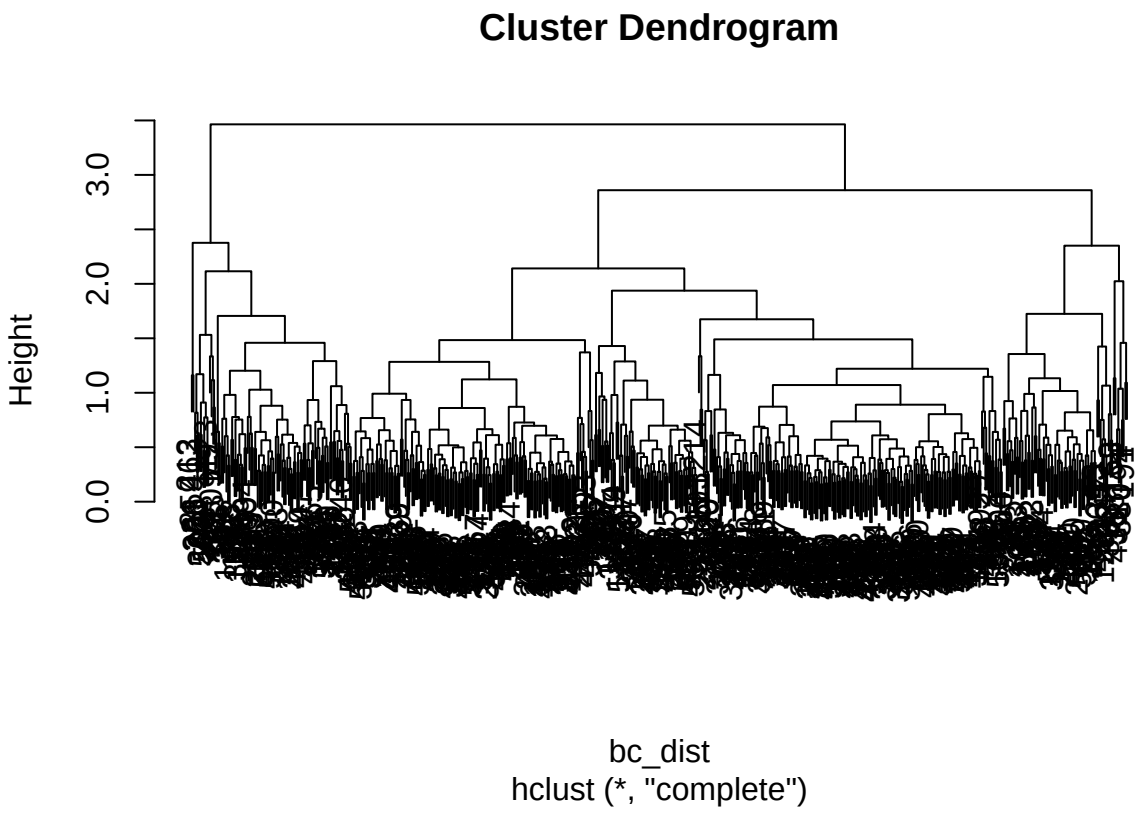
```
##
##   B   M
## 357 212
```

```
mmnorm2 <- function(x)
{ z <- ((x-min(x))/(max(x)-min(x)))
return(z)
}

myvector <- 1:20
mmnorm2(myvector)
```

```
##  [1] 0.00000000 0.05263158 0.10526316 0.15789474 0.21052632 0.26315789
##  [7] 0.31578947 0.36842105 0.42105263 0.47368421 0.52631579 0.57894737
## [13] 0.63157895 0.68421053 0.73684211 0.78947368 0.84210526 0.89473684
## [19] 0.94736842 1.00000000
```

```
bc <- data.frame(id=as.character(bc_temp$id)
                 ,diagnosis=as.factor(bc_temp$diagnosis)
                 ,radius_mean =mmnorm2( bc_temp$radius_mean)
```

```
                  ,texture_mean=mmnorm2(bc_temp$texture_mean)
                  ,perimeter_mean=mmnorm2(bc_temp$perimeter_mean)
                  ,area_mean=mmnorm2(bc_temp$area_mean)
                  ,smoothness_mean=mmnorm2(bc_temp$smoothness_mean)
                  ,compactness_mean=mmnorm2(bc_temp$compactness_mean)
                  ,concavity_mean=mmnorm2(bc_temp$concavity_mean)
                  ,concave.points_mean=mmnorm2(bc_temp$concave.points_mean)
                  ,symmetry_mean=mmnorm2(bc_temp$symmetry_mean)
                  ,fractal_dimension_mean=mmnorm2(bc_temp$fractal_dimension_mean)
                  ,radius_se=mmnorm2(bc_temp$radius_se)
                  ,perimeter_se=mmnorm2(bc_temp$perimeter_se)
                  ,texture_se=mmnorm2(bc_temp$texture_se)
                  ,area_se=mmnorm2(bc_temp$area_se)
                  ,smoothness_se=mmnorm2(bc_temp$smoothness_se)
                  ,compactness_se=mmnorm2(bc_temp$compactness_se)
                  ,concavity_se=mmnorm2(bc_temp$concavity_se)
                  ,concave.points_se =mmnorm2(bc_temp$concave.points_se)
                  ,symmetry_se=mmnorm2(bc_temp$symmetry_se)
                  ,fractal_dimension_se=mmnorm2(bc_temp$fractal_dimension_se)
                  ,radius_worst=mmnorm2(bc_temp$radius_worst)
                  ,texture_worst=mmnorm2(bc_temp$texture_worst)
                  ,perimeter_worst=mmnorm2(bc_temp$perimeter_worst)
                  ,area_worst=mmnorm2(bc_temp$area_worst)
                  ,smoothness_worst=mmnorm2(bc_temp$smoothness_worst)
                  ,compactness_worst=mmnorm2(bc_temp$compactness_worst)
                  ,concavity_worst=mmnorm2(bc_temp$concavity_worst)
                  ,concave.points_worst=mmnorm2(bc_temp$concave.points_worst)
                  ,symmetry_worst=mmnorm2(bc_temp$symmetry_worst)
                  ,fractal_dimension_worst=mmnorm2(bc_temp$fractal_dimension_worst)
)


bc_dist <- dist(bc[,c(-1,-2)])
hclust_resutls <- hclust(bc_dist )
plot(hclust_resutls)
```




```
hclust_2 <- cutree(hclust_resutls,2)
table(Cluster=hclust_2,Actual=bc[,2])
```

```
##         Actual
## Cluster   B   M
##       1   0  98
##       2 357 114
```

```
kmeans_2 <- kmeans(bc[,c(-1,-2)],2,nstart = 10)
kmeans_2$cluster
```

```
##   [1] 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [38] 2 2 1 2 2 1 1 2 1 2 1 2 2 2 2 2 1 2 2 1 1 2 2 2 2 1 2 1 1 2 2 1 2 1 2 1 2
##  [75] 2 1 2 1 1 2 2 1 1 1 2 1 2 1 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 1 2 2 1 2 2
## [112] 2 1 2 2 2 2 1 1 2 2 1 1 2 2 2 2 1 1 1 2 1 1 2 1 2 2 2 1 2 2 1 2 2 2 2 1 2
## [149] 2 2 2 2 1 2 2 2 1 2 2 2 2 1 1 2 1 2 2 1 1 2 2 2 1 2 2 2 2 1 2 2 1 1 1 2 2
## [186] 2 1 2 2 2 1 2 2 1 1 2 1 2 1 1 2 1 1 1 2 2 2 2 2 2 1 2 1 1 1 1 2 2 1 1 2 2
## [223] 2 1 2 2 2 2 2 1 1 2 2 1 2 2 1 1 2 1 2 2 2 2 1 2 2 2 2 2 1 2 1 1 1 2 1 1 1
## [260] 1 1 2 1 2 1 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 1 1 2 2 2 2 2 2 1 2 2 2 2 2
## [297] 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 1 2 2 2 2 1 1 1 2 2
## [334] 2 2 1 2 1 2 1 2 2 2 1 2 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2 1 1 2 1 1
## [371] 1 2 1 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 2 2 2 2 2 1 2 2 2 2 2 2
## [408] 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 1 2 1 1 2 1 2 2 2 2 2 1 2 2
## [445] 1 2 1 2 2 1 2 1 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2
## [482] 2 2 2 2 1 2 1 2 2 2 2 1 2 2 2 2 2 2 1 1 2 1 2 1 2 2 2 2 2 2 1 2 2 1 2 2 2 1 1
```

```
## [519] 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [556] 2 2 2 2 2 2 2 1 1 1 1 1 1 2
```

```
table(kmeans_2$cluster,Actual=bc[,2])
```

```
##    Actual
##       B   M
##   1   9 180
##   2 348  32
```