HAC

Dan Burke

10/27/2021

```r
#Using the "classes-to-clusters evaluation," run HierarchicalClusterer on the Iris data using single li
#Which one led to better result? Can you explain why?

iris_Data <- read.csv("C:\\Users\\danbu\\Desktop\\Applied Machine Learning\\Week 4\\iris.csv")
irisNoLabs <- iris_Data[,1:4]

str(irisNoLabs)
```

```
## 'data.frame':    150 obs. of  4 variables:
##  $ sepal_length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ sepal_width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ petal_length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ petal_width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
```

```r
scaledIris <- as.data.frame(scale(irisNoLabs))

summary(scaledIris)
```

```
##   sepal_length       sepal_width       petal_length       petal_width
##  Min.   :-1.86378   Min.   :-2.4308   Min.   :-1.5635   Min.   :-1.4396
##  1st Qu.:-0.89767   1st Qu.:-0.5858   1st Qu.:-1.2234   1st Qu.:-1.1776
##  Median :-0.05233   Median :-0.1245   Median : 0.3351   Median : 0.1328
##  Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.67225   3rd Qu.: 0.5674   3rd Qu.: 0.7602   3rd Qu.: 0.7880
##  Max.   : 2.48370   Max.   : 3.1043   Max.   : 1.7804   Max.   : 1.7052
```

```r
dist_mat <- dist(scaledIris, method='euclidean')

hclust_complete <- hclust(dist_mat, method = 'complete')

ut_com <- cutree(hclust_complete, k = 5)
plot(hclust_complete)
rect.hclust(hclust_complete , k = 5, border = 2:6)
abline(h = 5, col = 'red')
```
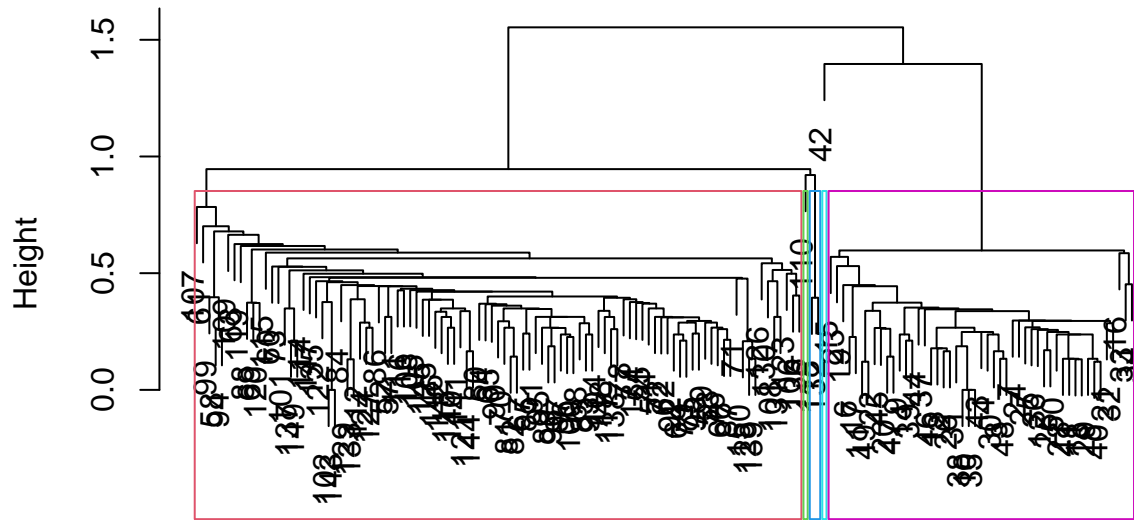
Cluster Dendrogram
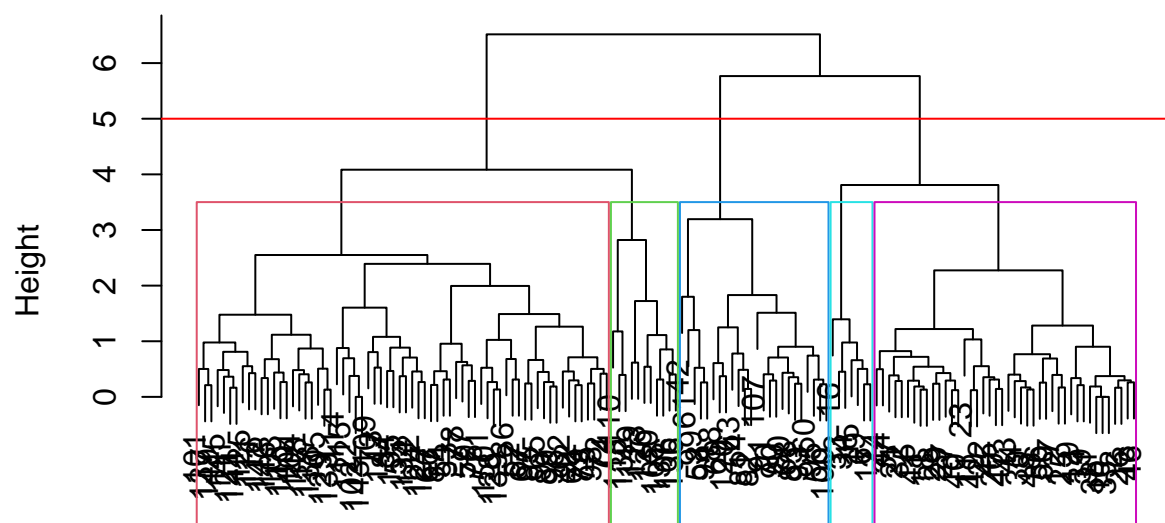
dist_mat
hclust (*, "complete")

```r
hclust_single <- hclust(dist_mat, method='single')
ut_single <- cutree(hclust_single, k = 5)
plot(hclust_single)
rect.hclust(hclust_single , k = 5, border = 2:6)
abline(h = 3, col = 'red')
```

Cluster Dendrogram

Height

dist_mat
hclust (*, "single")

3