Q46

John Bute

2024-12-05

R Markdown

Cluster the Wine datasets using DBSCAN, for various algorithm parameters. Validate your results.

```
library(dbscan)

## Warning: package 'dbscan' was built under R version 4.4.2

##
## Attaching package: 'dbscan'

## The following object is masked from 'package:stats':
##
##     as.dendrogram

library(factoextra)

## Warning: package 'factoextra' was built under R version 4.4.2

## Loading required package: ggplot2

## Welcome! Want to learn more? See two factoextra-related books at
https://goo.gl/ve3WBa

data <- read.csv("C:/Users/johnb/Desktop/Machine Learning/data/wine.csv")
data <- scale(data)


set.seed(1)
dbscan_cluster <- dbscan(data, eps = 3, minPts =10)

fviz_cluster(list(data = data, cluster = dbscan_cluster$cluster), geom =
"point")+
  labs(title = "DBSCAN clustering")
```
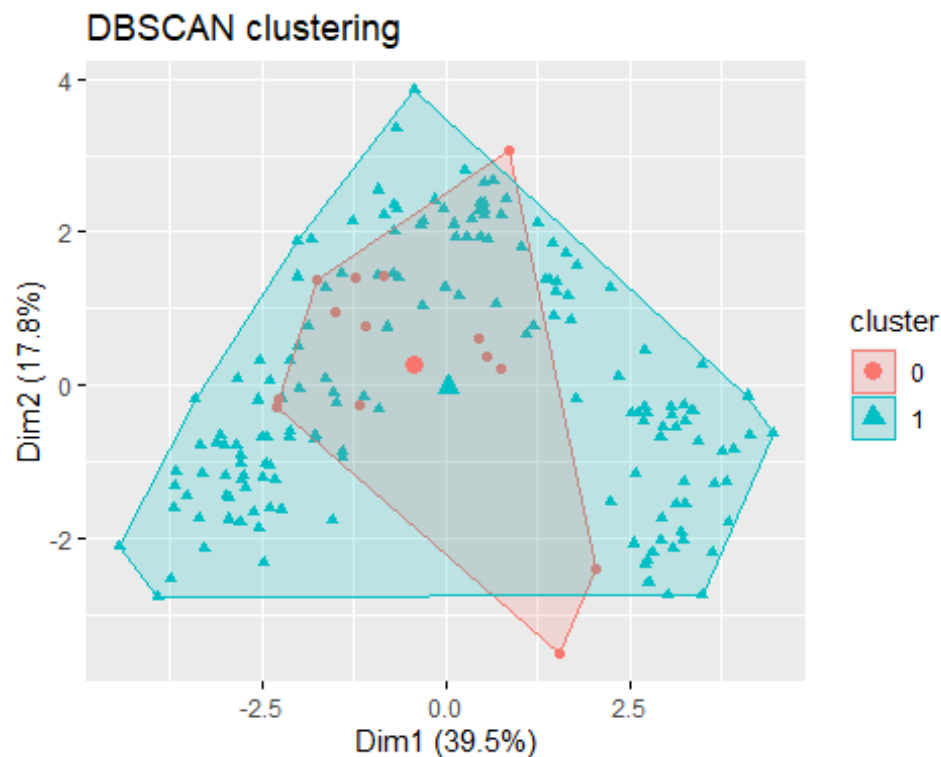
DBSCAN clustering


```r
print(table(dbscan_cluster$cluster))
```

```
##
##   0   1
##  14 164
```

Let us check the silhouette coefficient, to judge the quality of our clusters

```r
library(cluster)
silhouette_score <- silhouette(dbscan_cluster$cluster[dbscan_cluster$cluster
!= -1], dist(data[dbscan_cluster$cluster != -1,]))

avg_silhoueete_score <- mean(silhouette_score[, 3])

cat("Average silhouette score: ", avg_silhoueete_score, "\n")
```

```
## Average silhouette score:  0.1708784
```

Our silhouette score is weak. Furthermore, we do not have many dimensions, so our clustering does not suffer from dimensionality. Therefore, this clustering isn't great.

Cluster the Wine datasets using spectral clustering, for various algorithm parameters. Validate your results.

```r
library(kernlab)
```

```
##
## Attaching package: 'kernlab'
```
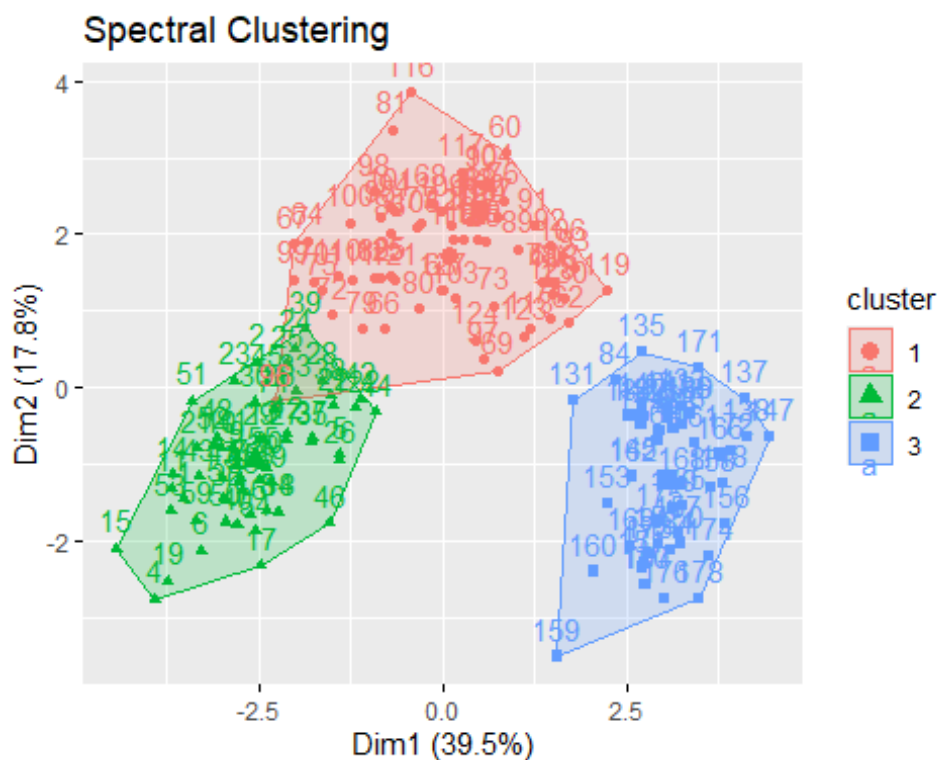
```
## The following object is masked from 'package:ggplot2':
##
##      alpha

library(cluster)
library(factoextra)

set.seed(1)

cluster_spectral <- specc(data, centers = 3)

fviz_cluster(list(data = data, cluster = as.integer(cluster_spectral))) +
  labs(title = "Spectral Clustering")
```


Spectral Clustering

```
print(table(as.integer(cluster_spectral)))

##
##  1  2  3
## 68 61 49
```

We have three clusters now. Interesting

Lets check the silhouette score:

```
cluster_spectral <- as.integer(cluster_spectral)
distance_matrix <- dist(data)
silhouette_score <- silhouette(cluster_spectral, distance_matrix)
```

```r
avg_silhouette_score <- mean(silhouette_score[, 3])
cat("Average silhouette score: ", avg_silhouette_score, "\n")
```

```
## Average silhouette score:  0.3061354
```

We did way better than DBScan.