Exp 6

Prof Meenakshi Garg

8/2/2020

Implementation and analysis of clustering algorithms like

1. K-Means

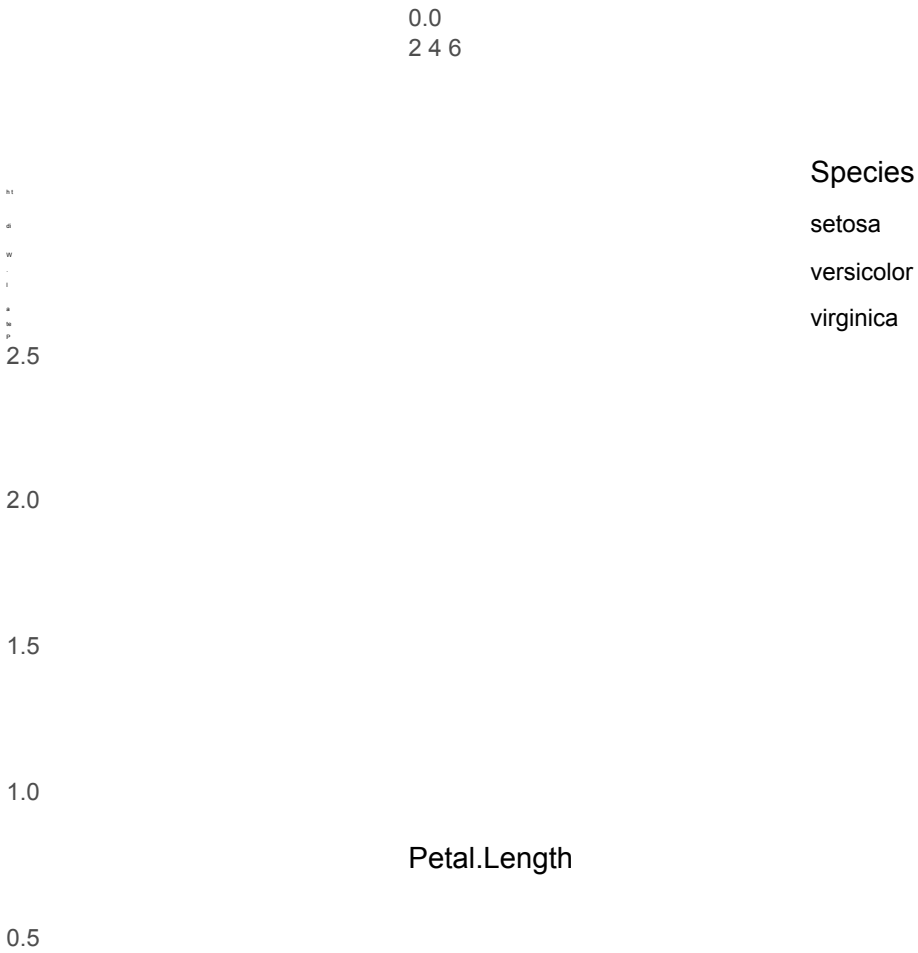
2. Agglomerative

** K Means Clustering** setwd("E:/R Orientation")

```
head(iris)
```

Sepal.Length Sepal.Width Petal.Length Petal.Width Species ## 1 5.1 3.5 1.4 0.2 setosa ## 2 4.9 3.0 1.4 0.2 setosa ## 3 4.7 3.2 1.3 0.2 setosa ## 4 4.6 3.1 1.5 0.2 setosa ## 5 5.0 3.6 1.4 0.2 setosa ## 6 5.4 3.9 1.7 0.4 setosa

```
library(ggplot2)
ggplot(iris, aes(Petal.Length, Petal.Width, color = Species)) + geom_point() 1
```




```
set.seed(20)
irisCluster <- kmeans(iris[, 3:4], 3, nstart = 20)
irisCluster
```

```
## K-means clustering with 3 clusters of sizes 52, 48, 50
##
## Cluster means:
## Petal.Length Petal.Width
## 1 4.269231 1.342308
## 2 5.595833 2.037500
```

```
## 3 1.462000 0.246000
##
## Clustering vector:
## [1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 ## [38] 3 3 3 3 3 3 3 3 3 3 3 3 3 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 ## [75] 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 1 2 2 2 2 ## [112] 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 ## [149] 2 2
##
## Within cluster sum of squares by cluster:
## [1] 13.05769 16.29167 2.02200
## (between_SS / total_SS = 94.3 %)
##
## Available components:
##
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss" ## [6] "betweenss" "size" "iter" "ifault"
```

2

```r
table(irisCluster$cluster, iris$Species)
```

```
##
## setosa versicolor virginica
## 1 0 48 4
## 2 0 2 46
## 3 50 0 0
```

```r
irisCluster$cluster <- as.factor(irisCluster$cluster)
ggplot(iris, aes(Petal.Length, Petal.Width, color = irisCluster$cluster)) + geom_point()
```

2.5

2.0

irisCluster$cluster 1

2

3

ht di W . l a te P

0.0

2 4 6

1.5

1.0

0.5

Petal.Length

#Agglomerative Clustering

{ r Agglomerative} head(iris) clusters <- hclust(dist(iris[, 3:4])) plot(clusters) clusterCut <- cutree(clusters, 3) table(clusterCut, iris$Species) clusters <- hclust(dist(iris[, 3:4]), method = 'average') plot(clusters) clusterCut <- cutree(clusters, 3) table(clusterCut, iris$Species) ggplot(iris, aes(Petal.Length, Petal.Width, color = iris$Species)) + geom_point(alpha = 0.4, size = 3.5) + geom_point(col = clusterCut) + scale_color_manual(values = c('black', 'red', 'green'))