OMC 406 Data Science Laboratory

Name: Deepankar Sharma
Student ID: 23351203
Course: MCA - ODL
Semester: 04

Ques 1 Hierarchical clustering on the Iris Dataset

Algorithm

1. load data
2. standardise the features to zero mean and unit variance
3. compute Euclidean distance matrix
4. apply hclust with method = "ward.D2"
5. plot the dendogram
6. cut off tree at k ← decide
8. plot clusters

code

```
library(ostata)
library(cluster)
library(ggplot2)

X ← scale(iris[, 1:4])
d ← dist(X, method = "euclidean")
hc ← hclust(d, method = "ward.D2")

plot(as.dendrogram(hc), main = "practical", ylab = "Height",
     Xlab = "Samples")

k ← 3

cl ← cutree(hc, k = k)
table(cl)
```

```
sil_k <- sapply(2:6, function(kk){
        ss <- silhouette(cutree(hc, k=kk), d)
        mean(ss[, "sil_width"])
    })

pca <- prcomp(X, center=FALSE, scale.=FALSE)
pc <- as.data.frame(pca$x[, 1:2])
pc$cluster <- factor(cl)

ggplot(pc, aes(PC1, PC2, cluster)) +
        geom_point(size=2) +
        labs(title="Iris") +
        theme_minimal()
```

Output

```
cl
 1  2  3
49 30 71
```

Dendogram