Assignment 21.1

Name : Y Vasudev

Batch : DA with R , Excel and Tableau

1. Use the below given data set

Data Set

2. Perform the below given activities:

a. Apply PCA to the dataset and show proportion of variance

b. Perform PCA using SVD approach

c. Show the graphs of PCA components

#a),b),c)

#reading the dataset

#using iris dataset

data.iris= read.csv("D:/BIG DATA/DATA ANALYTICS WITH R, EXCEL & TABLEAU/21 MODEL DEPLOYMENT/iris.csv") # read iris dataset

View(data.iris)

str(data.iris)

# Checking the data set with boxplot

boxplot(data.iris[,-5])

plot(SepalLengthCm~SepalWidthCm,data.iris)

# Normalization

# data.iris[,-5] , here -5 is just to remove species variable which is a factor.

m<-apply(data.iris[,-5],2,mean) #generating mean for all the variables

sd<-apply(data.iris[,-5],2,sd) # generating standard deviation for all the variables

z<-scale(data.iris[,-5],m,sd) # Scaling it

boxplot(z,horizontal=T) # boxplot to check data variability within variables after normilization

#PCA USING SPECTRAL DECOMPOSITION IN R

pc<- PC(data.iris[,-5],method="eigen",scaled=T,graph=F,rm.na=T,print.results=T)

pc1.cor<-princomp(data.iris[,-5], cor=TRUE) #PCA performed with correlation matrix

pc1.cor

#PCA USING SINGULAR VALUE DECOMPOSITION IN R method

library(factoextra)

pca1<- prcomp(data.iris[,-5],scale=T)

pca1

#this command show most useful info

summary(pca1)

#outputs the mean of variables

pca1$center

#scale

pca1$scale

#The rotation measure provides the principal component loading

pca1$rotation

#standard deviation

pca1$sdev

#the principal component score vector

pca1$x

#Proportion of Variance

summary(pca1)$importance[2,]

#Cumulative Proportion

summary(pca1)$importance[3,]

#The variance explained by each principal component

VE <- pca1$sdev^2

VE

#the proportion of variance explained by each principal component

PVE <- VE / sum(VE)

PVE

#c)

#answers

#graphs

#scree plot

fviz\_eig(pca1)

fviz\_screeplot(pca1)

#more

plot(pca1, type = "l")

#plot the resultant principal components

biplot(pca1, scale = 0)

#ggbiplot

library(ggbiplot)

g <- ggbiplot(pca1, obs.scale = 1, var.scale = 1,

groups = ir.species, ellipse = TRUE,

circle = TRUE)

g <- g + scale\_color\_discrete(name = '')

g <- g + theme(legend.direction = 'horizontal',

legend.position = 'top')

print(g)

#individuals pca

fviz\_pca\_ind(pca1,

col.ind = "cos2", # Color by the quality of representation

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE # Avoid text overlapping

)

#variables pca

fviz\_pca\_var(pca1,

col.var = "contrib", # Color by contributions to the PC

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE # Avoid text overlapping

)

#biplot pca

fviz\_pca\_biplot(pca1, repel = TRUE,

col.var = "#2E9FDF", # Variables color

col.ind = "#696969" # Individuals color

)