problem4-4

Mehrab Atighi

11/19/2021

Chemical Features of Wine.
Table 4.13 shows the PCA output on data non normalized in which the variables represent chemical characteristics of wine, and each case is a different wine.

TABLE 4.13 **PRINCIPAL COMPONENTS OF NON-NORMALIZED WINE DATA**

code for running PCA on the wine data

```
wine.df <- read.csv("Wine.csv")
pcs.cor <- prcomp(wine.df[,-1])
summary(pcs.cor)
pcs.cor$rot[,1:4]
```

Output

```
> summary(pcs.cor)

importance of components:
                            PC1      PC2     PC3     PC4     PC5
Standard deviation     314.9632 13.13527 3.07215 2.23409 1.10853
Proportion of Variance   0.9981  0.00174 0.00009 0.00005 0.00001
Cumulative Proportion    0.9981  0.99983 0.99992 0.99997 0.99998
                           PC6     PC7     PC8     PC9    PC10
Standard deviation     0.91710 0.5282 0.3891 0.3348 0.2678
Proportion of Variance 0.00001 0.0000 0.0000 0.0000 0.0000
Cumulative Proportion  0.99999 1.0000 1.0000 1.0000 1.0000
                          PC11    PC12    PC13
Standard deviation     0.1938 0.1452 0.09057
Proportion of Variance 0.0000 0.0000 0.00000
Cumulative Proportion  1.0000 1.0000 1.00000

> pcs.cor$rot[,1:4]

                               PC1           PC2           PC3           PC4
Alcohol              -0.0016592647 -1.203406e-03 -0.016873809  0.141446778
Malic_Acid            0.0006810156 -2.154982e-03 -0.122003373  0.160389543
Ash                  -0.0001949057 -4.593693e-03 -0.051987430 -0.009772810
Ash_Alcalinity        0.0046713006 -2.645039e-02 -0.938593003 -0.330965260
Magnesium            -0.0178680075 -9.993442e-01  0.029780248 -0.005393756
Total_Phenols        -0.0009898297 -8.779622e-04  0.040484644 -0.074584656
Flavanoids           -0.0015672883  5.185073e-05  0.085443339 -0.169086724
Nonflavanoid_Phenols  0.0001230867  1.354479e-03 -0.013510780  0.010805561
Proanthocyanins      -0.0006006078 -5.004400e-03  0.024659382 -0.050120952
Color_Intensity      -0.0023271432 -1.510035e-02 -0.291398464  0.878893693
Hue                  -0.0001713800  7.626731e-04  0.025977662 -0.060034945
OD280_OD315          -0.0007049316  3.495364e-03  0.070323969 -0.178200254
Proline              -0.9998229365  1.777381e-02 -0.004528682 -0.003112916
```

problem

a. The data are in the file Wine.csv. Consider the rows labeled Proportion of Variance. Explain why the value for PC1 is so much greater than that of any other column.
b. Comment on the use of normalization (standardization) in part (a).

solution

a)

According to output (above table) we can see summary and the rotation outputs.
the summary proprotaion of variance for PC1 is the biggest and is equal to 0.9981, it means the first components coverage about 99.81% of all variables variance and we can reduce our dimantion to one dimention with PC1 rotation and new values.
the cofficient of each variable for PC1 is equal to first column of second output(pcscorrot[,1:4])

b)

i think that we should use principal components with normalaized data cuse we have alot of variables and they are with diffrent measure and scale and when we normalizing them we are using scale and center on them and the its better for doing principal component method. for doing these code we just need to :

```
#wine.df <- read.csv("Wine.csv")
#pcs.cor.norm <- prcomp(wine.df[,-1] ,
#center = TRUE, scale. = FALSE )
#summary(pcs.cor.norm)
#pcs.cor.norm$rot[,1:4]
```