PCA case study

Abhirup Sen

20/05/2021

PCA Case Study

Importing the dataset

```
dataset = read.csv('Wine.csv')
```

Applying PCA using caret package

```
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
?preProcess
```

```
## starting httpd help server ...
```

```
##  done
```

```
mypca = preProcess(x = dataset[-14], method = 'pca', thresh = 0.9, verbose = TRUE)
```

```
## Calculating 13 means for centering
## Calculating 13 standard deviations for scaling
## Computing PCA loadings for 13 predictors
```

```
names(mypca)
```

```
##  [1] "dim"              "bc"               "yj"
##  [4] "et"               "invHyperbolicSine" "mean"
##  [7] "std"              "ranges"           "rotation"
## [10] "method"           "thresh"           "pcaComp"
## [13] "numComp"          "ica"              "wildcards"
## [16] "k"                "knnSummary"       "bagImp"
## [19] "median"           "data"             "rangeBounds"
```

```
mypca$rotation
```

```
##                                PC1          PC2         PC3         PC4
## Alcohol               -0.144329395  0.483651548 -0.20738262  0.01785630
## Malic_Acid             0.245187580  0.224930935  0.08901289 -0.53689028
## Ash                    0.002051061  0.316068814  0.62622390  0.21417556
## Ash_Alcanity           0.239320405 -0.010590502  0.61208035 -0.06085941
## Magnesium             -0.141992042  0.299634003  0.13075693  0.35179658
## Total_Phenols         -0.394660845  0.065039512  0.14617896 -0.19806835
## Flavanoids            -0.422934297 -0.003359812  0.15068190 -0.15229479
## Nonflavanoid_Phenols   0.298533103  0.028779488  0.17036816  0.20330102
## Proanthocyanins       -0.313429488  0.039301722  0.14945431 -0.39905653
## Color_Intensity        0.088616705  0.529995672 -0.13730621 -0.06592568
## Hue                   -0.296714564 -0.279235148  0.08522192  0.42777141
## OD280                 -0.376167411 -0.164496193  0.16600459 -0.18412074
## Proline               -0.286752227  0.364902832 -0.12674592  0.23207086
##                                PC5         PC6         PC7         PC8
## Alcohol               -0.26566365  0.21353865 -0.05639636  0.39613926
## Malic_Acid             0.03521363  0.53681385  0.42052391  0.06582674
## Ash                   -0.14302547  0.15447466 -0.14917061 -0.17026002
## Ash_Alcanity           0.06610294 -0.10082451 -0.28696914  0.42797018
## Magnesium              0.72704851  0.03814394  0.32288330 -0.15636143
## Total_Phenols         -0.14931841 -0.08412230 -0.02792498 -0.40593409
## Flavanoids            -0.10902584 -0.01892002 -0.06068521 -0.18724536
## Nonflavanoid_Phenols  -0.50070298 -0.25859401  0.59544729 -0.23328465
## Proanthocyanins        0.13685982 -0.53379539  0.37213935  0.36822675
## Color_Intensity       -0.07643678 -0.41864414 -0.22771214 -0.03379692
## Hue                   -0.17361452  0.10598274  0.23207564  0.43662362
## OD280                 -0.10116099  0.26585107 -0.04476370 -0.07810789
## Proline               -0.15786880  0.11972557  0.07680450  0.12002267
```

```
mypca
```

```
## Created from 178 samples and 13 variables
##
## Pre-processing:
##   - centered (13)
##   - ignored (0)
##   - principal component signal extraction (13)
##   - scaled (13)
##
## PCA needed 8 components to capture 90 percent of the variance
```

```
dataset = predict(mypca, dataset)
dim(dataset)
```

```
## [1] 178   9
```

Fitting SVM to the Training set

```
library(e1071)
classifier = svm(formula = Customer_Segment ~ .,
                 data = dataset,
                 type = 'C-classification',
                 kernel = 'linear')
```

Predicting the Test set results

```
y_pred = predict(classifier, newdata = dataset[-1])
```

Making the Confusion Matrix

```
cm = table(dataset[, 1], y_pred)
library(caret)
confusionMatrix(cm)
```

```
## Confusion Matrix and Statistics
##
##    y_pred
##      1  2  3
##   1 59  0  0
##   2  1 70  0
##   3  0  1 47
##
## Overall Statistics
##
##                Accuracy : 0.9888
##                  95% CI : (0.96, 0.9986)
##     No Information Rate : 0.3989
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9829
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: 1 Class: 2 Class: 3
## Sensitivity            0.9833   0.9859   1.0000
## Specificity            1.0000   0.9907   0.9924
## Pos Pred Value         1.0000   0.9859   0.9792
## Neg Pred Value         0.9916   0.9907   1.0000
## Prevalence             0.3371   0.3989   0.2640
## Detection Rate         0.3315   0.3933   0.2640
## Detection Prevalence   0.3315   0.3989   0.2697
## Balanced Accuracy      0.9917   0.9883   0.9962
```