

Componentes Principales

Juan F. Pérez

Departamento MACC
Matemáticas Aplicadas y Ciencias de la Computación
Universidad del Rosario

juanferna.perez@urosario.edu.co

2019

Contenidos

- 1 Introducción
- 2 PCA en R
- 3 Análisis de Correspondencias

Introducción

Introducción

- Búsqueda de patrones
- Gran número de variables/descriptores/atributos
- ¿Podemos usarlos todos? ¿Es factible? ¿Interpretable?
- ¿Cuáles escoger?
- Reducción de dimensionalidad
- PCA: principal component analysis

Componentes Principales

- Resumir atributos en tan pocos como sea posibles
- Al seleccionar m atributos (de D disponibles), deben ser los mejores
- Mejores: capturan la variabilidad de los datos al máximo
- Descomposición de la matriz de covarianza (eigenvalores)

PCA en R

PCA en R

```
names(USArrests )  
apply(USArrests , 2, mean)  
  
apply(USArrests , 2, var)  
apply(USArrests , 2, sd)  
  
pc =prcomp (USArrests , scale =TRUE)  
names(pc)  
  
#media vars orig  
pc$center  
# sd vars orig  
pc$scale
```

PCA en R

```
names(USArrests )  
apply(USArrests , 2, mean)  
apply(USArrests , 2, var)  
apply(USArrests , 2, sd)  
  
pc =prcomp (USArrests , scale =TRUE)  
names(pc)
```


PCA en R

```
#media vars orig  
pc$center  
# sd vars orig  
pc$scale  
  
#pesos  
pc$rotation  
  
#pcs  
dim(pc$x )  
pc$x
```

PCA en R

```
biplot (pc, scale =0)
```

```
#pc sd  
pc$sdev  
# varianza explicada  
pc.var =pc$sdev ^2  
pc.var  
# % varianza explicada  
pc.var =pc$sdev ^2  
pc.var  
pc.pve=pc.var/sum(pc.var )  
pc.pve
```

PCA en R

```
plot(pc.pve , xlab="Componente Principal " ,  
      ylab=" Proporción de Varianza expl" ,  
      ylim=c(0,1) , type='b')
```

```
plot(cumsum (pc.pve) ,  
      xlab="Componente Principal" ,  
      ylab=" Prop, Varianza Expl. Acum." ,  
      ylim=c(0,1) , type='b')
```

PCA en R

```
correl <- cor(USArrests)  
correl
```

```
install.packages("ggcorrplot")
```

```
library(ggcorrplot)  
ggcorrplot(correl)
```

PCA en R - FactoMineR

```
install.packages("FactoMineR")
```

```
library(FactoMineR)
```

```
pca <- PCA(USArrests)
```

```
names(pca)
```

```
pca$eig
```

```
names(pca$var)
```

```
pca$var$coord
```

```
pca$var$contrib
```

PCA en R

Considere el siguiente set de datos

```
library(ISLR)  
x.labs <- NCI60$labs  
x.data <- NCI60$data
```

```
dim(x.data)  
table(x.labs)
```

Determine un conjunto de componentes principales que permitan resumir el conjunto de atributos

PCA en R

```
pca <- prcomp(x.data , scale=TRUE)  
summary(pca)  
plot(pca)
```

PCA en R

```
pve = 100 * pca$sdev ^2 / sum(pca$sdev ^2)
par(mfrow = c(1,2))
plot(pve , type = "o" , ylab="PVE " ,
     xlab="Componente Principal" ,
     col = " blue")
plot(cumsum ( pve ), type="o" ,
     ylab = "PVE Acumulada" ,
     xlab = "Principal Component" ,
     col = " brown3 ")
```


PCA en R

```
library(FactoMineR)
pca <- PCA(x.data)
names(pca)
pca$eig
pca$eig[1:10,]
```

Análisis de Correspondencias

Análisis de Correspondencias

- Datos categóricos
- Múltiples categorías
- Reducción de dimensionalidad
- Tablas de contingencia

Tablas de contingencia

```
library(MASS)
```

```
?Cars93
```

```
Cars93$Type
```

```
table(Cars93$Type)
```

```
prop.table(table(Cars93$Type))
```

```
Cars93$Origin
```

```
table(Cars93$Origin)
```

```
prop.table(table(Cars93$Origin))
```

```
tabla <- table(Cars93[,c("Type", "DriveTrain")])
```

Tablas de contingencia

```
rowSums(tabla , dims=3)  
colSums(tabla , dims=3)  
class(tabla)
```

```
install.packages(" gplots" )  
library(gplots)  
?balloonplot  
balloonplot(tabla , main =" Carros" , xlab ="" , ylab="" ,  
            label = FALSE, show.margins = FALSE)
```

Análisis de correspondencias

```
library(FactoMineR)
ca <- CA(tabla , graph = TRUE)
names(ca)
ca$eig
ca$col$coord
ca$col$contrib
```

Análisis de correspondencias

```
library(ca)
x <- author
ca <- CA(x, graph = TRUE)
names(ca)
ca$eig
ca$col$coord
ca$col$contrib
```