ME731 - Métodos em Análise Multivariada
– Análise de Componentes Principais I –

Prof. Carlos Trucíos
ctrucios@unicamp.br
ctruciosm.github.io

Instituto de Matemática, Estatística e Computação Científica,
Universidade Estadual de Campinas

Aula 10

UNICAMP

Agenda I

Introdução

Introdução

ACP

The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the date set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables.
[Jolliffe, Principal Component Analysis, 2nd edition, 2012]

Introdução

- É uma técnica de redução de dimensão.
- Transforma um conjunto de variáveis (variáveis originais) em um conjunto de novas variáveis não correlacionadas (chamadas de componentes).
- As componentes são obtidas de forma que capturem a maior parte da variabilidade total dos dados em apenas poucas componentes.
- O número total de componentes é igual ao número de variáveis originais (porém, já as primeiras componentes capturam a maior parte da variabilidade dos dados).
- Cada componente é uma combinação linear de todas as variáveis originais.

Aplicações

- Geólogia
- Sociologia
- Zoologia
- Administração
- RH
- Economia
- Finanças
- Tratamento ao problema de multicolinearidade
- Passo previo ao análise de agrupamento

Uma visão geral

- As componentes principais são combinações lineares das variáveis originais.
- Os coeficientes das combinações lineares são os elementos dos autovetores associados à matriz de covariância das variáveis originais.
- A primeira componentes principal está associada ao maior autovalor da matriz de covariância das variáveis originais.
- A variância de cada componente é igual ao seu autovalor associado.
- No caso das variáveis originais serem padronizadas, a correlação entre a componente e a variável original é completamente determinada pelo autovalor associado à componente e o elemento do autovetor associado à variavel original.

Análise de Componentes Principais

Análise de Componentes Principais

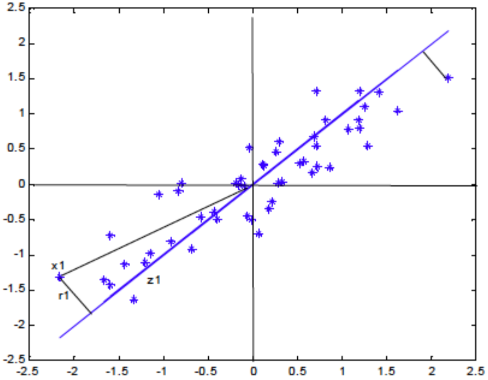
- Queremos encontrar um subespaço de dimensão $r < p$ tal que ao projetar os pontos originais nele, se perca a menor quantidade de informação possível (**equivalentemente**, o subespaço capture a maior parte da variabilidade dos dados).
- Pense em $p = 2$ e um subespaço de dimensão $r = 1$ (uma reta). Queremos que as projeções dos pontos originais sobre a reta preservem a estrutura dos dados originais o máximo possível (*i.e*, que se perca a menor quantidad de informação possível).

Análise de Componentes Principais

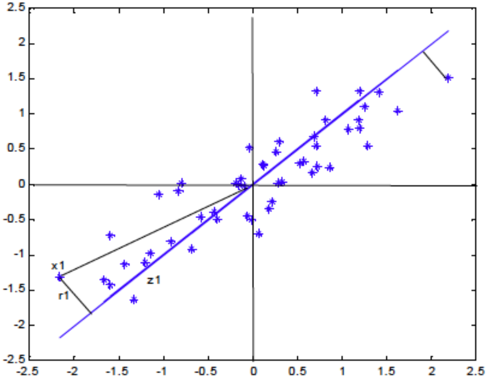
- Queremos encontrar um subespaço de dimensão $r < p$ tal que ao projetar os pontos originais nele, se perca a menor quantidade de informação possível (**equivalentemente**, o subespaço capture a maior parte da variabilidade dos dados).
- Pense em $p = 2$ e um subespaço de dimensão $r = 1$ (uma reta). Queremos que as projeções dos pontos originais sobre a reta preservem a estrutura dos dados originais o máximo possível (*i.e*, que se perca a menor quantidad de informação possível).

Na ilustração a seguir, sem perda de generalidade, utilizaremos dados centrados.

Análise de Componentes Principais



Análise de Componentes Principais



- As distâncias entre os pontos originais é preservada (aprox.) nas projeções sobre a reta.

Análise de Componentes Principais

Qual reta escolher?

- De forma que a dispersão dos pontos projetados sobre a reta seja a maior possível.
- De forma a minimizar a distância entre os pontos originais e os pontos nela projetados.

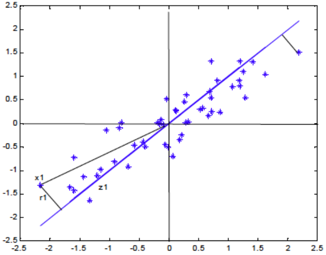
Análise de Componentes Principais

Qual reta escolher?

- De forma que a dispersão dos pontos projetados sobre a reta seja a maior possível.
- De forma a minimizar a distância entre os pontos originais e os pontos nela projetados.

Ambas ideias levam à mesma solução.

Análise de Componentes Principais



Minimizar a distância entre os pontos originais e os pontos nela projetados, ou seja,

$$\text{Minimizar: } \sum_{i=1}^{n} r_i^2.$$

Pelo teorema de Pitágoras, temos que

$$r_i^2 = \underbrace{||x_i||^2}_{x_i' x_i} + z_i^2,$$

em que z_i é distância da origem até a projeção de x_i sobre a reta.

Análise de Componentes Principais

A projeção de x_i sob a reta de direção $a = (a_1, a_2)$ é dada por,

$$Proj_a x_i = \frac{a \cdot x_i}{||a||^2} a = (a \cdot x_i) a$$

Análise de Componentes Principais

A projeção de x_i sob a reta de direção $a = (a_1, a_2)$ é dada por,

$$Proj_a x_i = \frac{a \cdot x_i}{||a||^2} a = (a \cdot x_i) a$$

Então,

$$z_i = \sqrt{(a \cdot x_i)^2 a_1^2 + (a \cdot x_i)^2 a_2^2} = \sqrt{(a \cdot x_i)^2 (a_1^2 + a_2^2)} = \sqrt{(a \cdot x_i)^2}$$

Análise de Componentes Principais

A projeção de x_i sob a reta de direção $a = (a_1, a_2)$ é dada por,

$$Proj_a x_i = \frac{a \cdot x_i}{||a||^2} a = (a \cdot x_i) a$$

Então,

$$z_i = \sqrt{(a \cdot x_i)^2 a_1^2 + (a \cdot x_i)^2 a_2^2} = \sqrt{(a \cdot x_i)^2 (a_1^2 + a_2^2)} = \sqrt{(a \cdot x_i)^2}$$

Então,

$$\text{Minimizar: } \sum_{i=1}^{n} r_i^2 \equiv \text{Minimizar: } \sum_{i=1}^{n} x_i x_i' - \sum_{i=1}^{n} (a \cdot x_i)^2 \equiv \text{Maximizar: } \sum_{i=1}^{n} (a \cdot x_i)^2$$

Análise de Componentes Principais

Note que:

- $a \cdot x_i = x_i a'$ e
- $\bar{x} = 0$ (pois os dados são centrados)

Então, $\sum_{i=1}^{n}(a \cdot x_i)^2 = \sum_{i=1}^{n}(x_i a')(x_i a') = \sum_{i=1}^{n}(x_i a')'(x_i a') = \sum_{i=1}^{n} a x_i' x_i a'$.

Análise de Componentes Principais

Note que:

- $a \cdot x_i = x_i a'$ e
- $\bar{x} = 0$ (pois os dados são centrados)

Então, $\sum_{i=1}^{n}(a \cdot x_i)^2 = \sum_{i=1}^{n}(x_i a')(x_i a') = \sum_{i=1}^{n}(x_i a')'(x_i a') = \sum_{i=1}^{n} a x_i' x_i a'$.

Assim,

Maximizar: $\sum_{i=1}^{n} a x_i' x_i a' \equiv$ Maximizar: $\frac{1}{n}\sum_{i=1}^{n} a x_i' x_i a' \equiv$ Maximizar: aSa'

Análise de Componentes Principais

Note que:

- $a \cdot x_i = x_i a'$ e
- $\bar{x} = 0$ (pois os dados são centrados)

Então, $\sum_{i=1}^{n} (a \cdot x_i)^2 = \sum_{i=1}^{n} (x_i a')(x_i a') = \sum_{i=1}^{n} (x_i a')'(x_i a') = \sum_{i=1}^{n} a x_i' x_i a'$.

Assim,

Maximizar: $\sum_{i=1}^{n} a x_i' x_i a' \equiv$ Maximizar: $\frac{1}{n} \sum_{i=1}^{n} a x_i' x_i a' \equiv$ Maximizar: aSa'

A mesma ideia continua valendo para $p > 2$!

Análise de Componentes Principais

Seja $x_i' \in \mathbb{R}^p$ um ponto no hiperplano $(x_i = (x_{i1}, \cdots, x_{ip})')$e seja $a_1 = (a_{11}, \cdots, a_{1p})'$ um vetor direção.

Análise de Componentes Principais

Seja $x_i' \in \mathbb{R}^p$ um ponto no hiperplano $(x_i = (x_{i1}, \cdots, x_{ip})')$e seja $a_1 = (a_{11}, \cdots, a_{1p})'$ um vetor direção.

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} x_i' x_i - \sum_{i=1}^{n} (a_1' x_i)(a_1' x_i)'$$

Análise de Componentes Principais

Seja $x_i' \in \mathbb{R}^p$ um ponto no hiperplano ($x_i = (x_{i1}, \cdots, x_{ip})'$)e seja $a_1 = (a_{11}, \cdots, a_{1p})'$ um vetor direção.

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} x_i' x_i - \sum_{i=1}^{n} (a_1' x_i)(a_1' x_i)'$$

$$\text{Minimizar: } \sum_{i=1}^{n} r_i^2 \equiv \text{Maximizar: } \sum_{i=1}^{n} (a_1' x_i)(a_1' x_i)' \equiv \text{Maximizar:} a_1' S a_1$$

Análise de Componentes Principais

Seja $x_i' \in \mathbb{R}^p$ um ponto no hiperplano $(x_i = (x_{i1}, \cdots, x_{ip})')$e seja $a_1 = (a_{11}, \cdots, a_{1p})'$ um vetor direção.

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} x_i' x_i - \sum_{i=1}^{n} (a_1' x_i)(a_1' x_i)'$$

$$\text{Minimizar:} \sum_{i=1}^{n} r_i^2 \equiv \text{Maximizar:} \ \sum_{i=1}^{n} (a_1' x_i)(a_1' x_i)' \equiv \text{Maximizar:} a_1' S a_1$$

Note que $a_1' S a_1$ cresce sem limite a medida que multiplicamos a_1 por uma constante qualquer. Então, para que a maximização tenha solução impomos a restrição de que $a_1' a_1 = 1$

Análise de componentes principais

Queremos maximizar $a_1' S a_1$ sujeito a $a_1' a_1 = 1$. Então,

$$L = a_1' S a_1 - \lambda(a_1' a_1 - 1)$$

Derivando w.r.t a_1

$$\frac{\partial L}{\partial a_1} = 2Sa_1 - 2\lambda a_1$$

Igualando a zero

$$S a_1 = \lambda a_1. \tag{1}$$

o que implica que a_1 é o autovetor associado ao autovalor λ

Para que

$$a_1' S a_1 = a_1' \lambda a_1 = \lambda a_1' a_1 = \lambda$$

seja máximo, λ deve ser o maior autovalor. Logo, a_1 é o autovetor associado ao maior autovalor

Análise de componentes principais

Assim, a primeira componente principal z_1 é aquela combinação linear que maximiza a variânca de z_1 e é dada por,

$$z_1 = \mathbf{x}_{np} a_1.$$

Análise de componentes principais

Assim, a primeira componente principal z_1 é aquela combinação linear que maximiza a variânca de z_1 e é dada por,

$$z_1 = \mathbf{x}_{np} a_1.$$

Suponha agora que não queremos calcular apenas a primeira, mas a duas primeiras componentes.

Análise de componentes principais

Assim, a primeira componente principal z_1 é aquela combinação linear que maximiza a variânca de z_1 e é dada por,

$$z_1 = \mathbf{x}_{np} a_1.$$

Suponha agora que não queremos calcular apenas a primeira, mas a duas primeiras componentes.

Queremos maximizar $Var(z_1 + z_2)$ sujeito às restrições que $a_1' a_1 = 1$, $a_2' a_2 = 1$ e z_1 e z_2 são não correlacionados.

Análise de componentes principais

Assim, a primeira componente principal z_1 é aquela combinação linear que maximiza a variânca de z_1 e é dada por,

$$z_1 = \mathbf{x}_{np} a_1.$$

Suponha agora que não queremos calcular apenas a primeira, mas a duas primeiras componentes.

Queremos maximizar $Var(z_1 + z_2)$ sujeito às restrições que $a_1' a_1 = 1$, $a_2' a_2 = 1$ e z_1 e z_2 são não correlacionados.

Maximizar: $a_1' S a_1 + a_2' S a_2 - \lambda_1 (a_1' a_1 - 1) - \lambda_2 (a_2' a_2 - 1)$

Análise de componentes principais

Função Objetivo: $L = a_1' S a_1 + a_2' S a_2 - \lambda_1(a_1' a_1 - 1) - \lambda_2(a_2' a_2 - 1)$

Análise de componentes principais

Função Objetivo: $L = a_1' S a_1 + a_2' S a_2 - \lambda_1(a_1' a_1 - 1) - \lambda_2(a_2' a_2 - 1)$

Derivando w.r.t a_1 e a_2:

$$\frac{\partial L}{\partial a_1} = 2Sa_1 - 2\lambda_1 a_1 \quad e \quad \frac{\partial L}{\partial a_2} = 2Sa_2 - 2\lambda_1 a_2,$$

Análise de componentes principais

Função Objetivo: $L = a_1' S a_1 + a_2' S a_2 - \lambda_1 (a_1' a_1 - 1) - \lambda_2 (a_2' a_2 - 1)$

Derivando w.r.t a_1 e a_2:

$$\frac{\partial L}{\partial a_1} = 2Sa_1 - 2\lambda_1 a_1 \quad e \quad \frac{\partial L}{\partial a_2} = 2Sa_2 - 2\lambda_1 a_2,$$

Igualando a zero, $Sa_1 = \lambda_1 a_1 \quad e \quad Sa_2 = \lambda_1 a_2,$

O que indica que a_1 e a_2 devem ser os autovetores associados aos autovalores λ_1 e λ_2.

Análise de componentes principais

Para que

$$a_1' S a_1 + a_2' S a_2 = a_1' \lambda_1 a_1 + a_2' \lambda_2 a_2 = \lambda_1 a_1' a_1 + \lambda_2 a_2' a_2 = \lambda_1 + \lambda_2$$

seja máximo, λ_1 e λ_2 devem ser os dois maiores autovalores e a_1 e a_2 seus autovetores associados.

$\vdots$

Análise de componentes principais

Para que

$$a_1' S a_1 + a_2' S a_2 = a_1' \lambda_1 a_1 + a_2' \lambda_2 a_2 = \lambda_1 a_1' a_1 + \lambda_2 a_2' a_2 = \lambda_1 + \lambda_2$$

seja máximo, λ_1 e λ_2 devem ser os dois maiores autovalores e a_1 e a_2 seus autovetores associados.

$\vdots$

> Obter as componentes principais é um problema de calcular autovalores-autovetores da matriz de covariância

Calcular componentes principais equivale a multiplicar a matriz X por uma matriz ortogonal (matriz de autovetores) para obter novas variáveis $(Z_1, \ldots, Z_N)$ não correlacionadas

Análise de componentes principais

Algoritmo

1. Calcular a matriz de covariância **S**
2. Decompor **S** em autovalores e autovetores
$$\mathbf{S}M = M\Lambda$$
em que Λ é uma matriz diagonal de autovalores ordenados de maior a menor e M a matriz de autovetores assoaciados aos autovalores
3. As componentes serão
$$\mathbf{z} = \mathbf{x}M$$

Análise de componentes principais

Algoritmo

1. Calcular a matriz de covariância $\mathbf{S}$
2. Decompor $\mathbf{S}$ em autovalores e autovetores
$$\mathbf{S}M = M\Lambda$$
em que Λ é uma matriz diagonal de autovalores ordenados de maior a menor e M a matriz de autovetores assoaciados aos autovalores
3. As componentes serão
$$\mathbf{z} = \mathbf{x}M$$

Observação: se $\mathbf{x}$ não for centrada, as componentes serão obtidas como

$$\mathbf{z} = (\mathbf{x} - \bar{x})M$$

Análise de componentes principais: Exemplo

```
acp_me731 <- function(x) {
  S <- cov(x)
  M <- eigen(S)$vectors
  z <- scale(x, center = TRUE, scale = FALSE) %*% M
  return(z)
}
```

Análise de componentes principais: Exemplo

```
acp_me731 <- function(x) {
  S <- cov(x)
  M <- eigen(S)$vectors
  z <- scale(x, center = TRUE, scale = FALSE) %*% M
  return(z)
}
```

Utilizaremos o *data set* `penguins` do pacote `palmerpenguins`.

```
library(dplyr)
library(palmerpenguins)
dados <- penguins %>%
  select(ends_with("_mm"), "body_mass_g") %>%
  na.omit()
```

Análise de Componentes Principais: Exemplo

```
glimpse(dados)
```

```
## Rows: 342
## Columns: 4
## $ bill_length_mm    <dbl> 39.1, 39.5, 40.3, 36.7, 39.3, 38.9,
## $ bill_depth_mm     <dbl> 18.7, 17.4, 18.0, 19.3, 20.6, 17.8,
## $ flipper_length_mm <int> 181, 186, 195, 193, 190, 181, 195,
## $ body_mass_g       <int> 3750, 3800, 3250, 3450, 3650, 3625,
```

Análise de Componentes Principais: Exemplo

```r
componentes_home_made <- acp_me731(dados)
componentes_princomp <- princomp(dados)$scores
componentes_prcomp <- prcomp(dados)$x
round(head(componentes_home_made, 3), 4)
```

```
##          [,1]    [,2]    [,3]    [,4]
## [1,] -452.0232 13.3366  1.1480 -0.3535
## [2,] -401.9500  9.1527 -0.0904 -1.0483
## [3,] -951.7409 -8.2615 -2.3518  0.8418
```

Análise de Componentes Principais: Exemplo

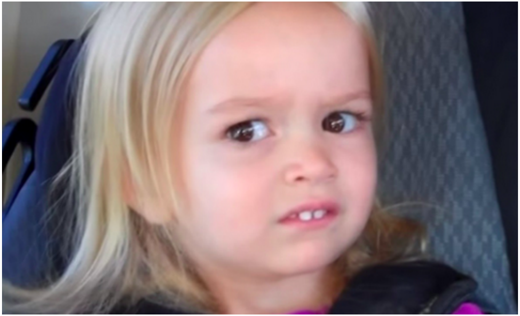
```
round(head(componentes_princomp, 3), 4)
```

```
##           Comp.1   Comp.2  Comp.3  Comp.4
## [1,] -452.0232 -13.3366  1.1480  0.3535
## [2,] -401.9500  -9.1527 -0.0904  1.0483
## [3,] -951.7409   8.2615 -2.3518 -0.8418
```

```
round(head(componentes_prcomp, 3), 4)
```

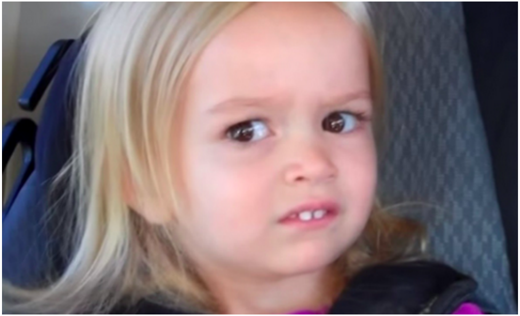
```
##              PC1     PC2     PC3     PC4
## [1,] -452.0232 13.3366 -1.1480 -0.3535
## [2,] -401.9500  9.1527  0.0904 -1.0483
## [3,] -951.7409 -8.2615  2.3518  0.8418
```

Análise de Componentes Principais: Exemplo



Os valores das componentes apenas coincidem em módulo!

Análise de Componentes Principais: Exemplo



Os valores das componentes apenas coincidem em módulo!

Se v é autovetor associado a λ, então $-v$ também é,

$$Av = \lambda v = A(-v) = \lambda(-v)$$

Análise de Componentes principais: exemplo

```r
round(eigen(cov(dados))$vectors, 4)
```

```
##         [,1]    [,2]    [,3]    [,4]
## [1,]  0.0041 -0.3085  0.9448 -0.1101
## [2,] -0.0012  0.0904  0.1443  0.9854
## [3,]  0.0153 -0.9468 -0.2941  0.1300
## [4,]  0.9999  0.0158  0.0008 -0.0004
```

```r
round(prcomp(dados)$rotation, 4)
```

```
##                       PC1     PC2     PC3     PC4
## bill_length_mm     0.0041 -0.3085 -0.9448 -0.1101
## bill_depth_mm     -0.0012  0.0904 -0.1443  0.9854
## flipper_length_mm  0.0153 -0.9468  0.2941  0.1300
## body_mass_g        0.9999  0.0158 -0.0008 -0.0004
```

Propriedades

Propriedades

Note que por construção $Var(z_i) = \lambda_1$. Então,

- A soma das variâncias das variáveis originais é igual à soma das variâncias das componentes.

$$Var(x_1) + \cdots + Var(x_N) = Tr(S) = \sum_{i=1}^{N} \lambda_i = Var(z_1) + \cdots + Var(z_N)$$

- A variância generalizada das variáveis originais é igual à variância generalizada das componentes $|S_X| = \prod_{i=1}^{N} \lambda_i = \prod_{i=1}^{N} Var(z_i) = |S_Z|$

Propriedades

Note que por construção $Var(z_i) = \lambda_1$. Então,

- A soma das variâncias das variáveis originais é igual à soma das variâncias das componentes.

  $$Var(x_1) + \cdots + Var(x_N) = Tr(S) = \sum_{i=1}^{N} \lambda_i = Var(z_1) + \cdots + Var(z_N)$$

- A variância generalizada das variáveis originais é igual à variância generalizada das componentes $|S_X| = \prod_{i=1}^{N} \lambda_i = \prod_{i=1}^{N} Var(z_i) = |S_Z|$

A variância total $(tr(S))$ e a variância generalizada $(|S|)$ das variáveis originais e das componentes principais é a mesma!.

Propriedades

Qual a proporção da variância total que é explicada pela primeira componente?

Propriedades

Qual a proporção da variância total que é explicada pela primeira componente?

$$\frac{\lambda_1}{\lambda_1 + \cdots + \lambda_p}.$$

Propriedades

Qual a proporção da variância total que é explicada pela primeira componente?

$$\frac{\lambda_1}{\lambda_1 + \cdots + \lambda_p}.$$

Qual a proporção da variância total que é explicada pelas primeiras k **componentes?**

Propriedades

Qual a proporção da variância total que é explicada pela primeira componente?

$$\frac{\lambda_1}{\lambda_1 + \cdots + \lambda_p}.$$

Qual a proporção da variância total que é explicada pelas primeiras k **componentes?**

$$\frac{\lambda_1 + \cdots + \lambda_k}{\lambda_1 + \cdots + \lambda_p}.$$

Propriedades: Exemplo

```
prcomp(dados)
```

```
## Standard deviations (1, .., p=4):
## [1] 802.055230   7.179472   4.004453   1.530847
##
## Rotation (n x k) = (4 x 4):
##                             PC1         PC2           PC3
## bill_length_mm      0.004051279 -0.30848927 -0.9448307702 -0.1
## bill_depth_mm      -0.001162051  0.09044334 -0.1443173596  0.9
## flipper_length_mm   0.015275204 -0.94678621  0.2940520764  0.1
## body_mass_g         0.999874445  0.01581922 -0.0008317408 -0.0
```

Propriedades: Exemplo

```
lambdas <- prcomp(dados)$sdev^2
cumsum(lambdas) / sum(lambdas)
```

```
## [1] 0.9998913 0.9999714 0.9999964 1.0000000
```

A primeira componente explica mais do 99% da variância total!

Propriedades: Exemplo

```
lambdas <- prcomp(dados)$sdev^2
cumsum(lambdas) / sum(lambdas)
```

```
## [1] 0.9998913 0.9999714 0.9999964 1.0000000
```

A primeira componente explica mais do 99% da variância total!

Propriedades:

Cuidado com aplicar ACP quando as variáveis não estão na mesma escala!.

Propriedades:

Cuidado com aplicar ACP quando as variáveis não estão na mesma escala!.

Para calcular a primeira componente maximizamos, sujeito à restrição que $a'a = 1$,

$$aSa = \sum_{i=1}^{p} a_i^2 s_i^2 + \sum_{i \neq j} a_i a_j s_{ij}.$$

Propriedades:

Cuidado com aplicar ACP quando as variáveis não estão na mesma escala!.

Para calcular a primeira componente maximizamos, sujeito à restrição que $a'a = 1$,

$$aSa = \sum_{i=1}^{p} a_i^2 s_i^2 + \sum_{i \neq j} a_i a_j s_{ij}.$$

Se uma das variáveis (digamos x_1) tiver uma variância (s_1^2) muito maior do que as outras, para maximizar a expressão acima, fazemos a_1 tão grande quanto pudermos (em casos extremos, a primeira componente será, basicamente, x_1).

Isto significa que, se aplicarmos ACP em variáveis com diferentes escalas, a solução dependerá das escalas e variáveis com valores grandes terão maior peso nas componentes (o que não é desejável).

Propriedades:

Na próxima aula veremos como lidar com este problema, apresentaremos outras propriedades e discutiremos desafios práticos do dia a dia do estatístico/cientista de dados.

Referências

Referências

- Härdle, W. K., & Simar, L. (2019). Applied Multivariate Statistical Analysis. Fifth Editon. Springer Nature. Capítulo 11.
- Johnson, R. A., & Wichern, D. W. (2007). Applied multivariate statistical analysis. Sixth Edition. Pearson Prentice Hall. Capítulo 4.
- Mardia, K. V., Kent, J. T., & Bibby, J, M. (1979). Multivariate Analysis. Academic Press. Capítulo 8.
- Peña, D. (2002). Análisis de Datos Multivariantes. Mc Graw Hill. Capítulo 5.