Anàlisi de components principals

Vicent Caselles Ballester

2024-04-04

Exercici 1

a) Escribir la función de densidad $f(x1, x2)$ del vector x y representarla en tres dimensiones.

La primera part feta a mà. No ho mostro, però m'he basat en: https://www.ime.unicamp.br/~cnaber/mvnprop.pdf.

Segona part:

First I'm gonna create a function that converts a covariance matrix into a correlation matrix:

```
cov2cor_vc <- function(cov){
  # from:
  # https://math.stackexchange.com/questions/186959/correlation-matrix-from-covariance-matrix
  D <- diag(sqrt(diag(cov)))
  cor <- solve(D) %*% cov %*% solve(D)
  cor
}
```

Let's try it out:

```
cov1 <- matrix(c(8,5,5,4), ncol=2)
cov2cor_vc(cov1)
```

```
##           [,1]      [,2]
## [1,] 1.0000000 0.8838835
## [2,] 0.8838835 1.0000000
```

```
# We see that if we use the "built-in" function from R we get the same
cov2cor(cov1)
```

```
##           [,1]      [,2]
## [1,] 1.0000000 0.8838835
## [2,] 0.8838835 1.0000000
```

Cool.

Bivariate density function given x1, x2, cov matrix.

```
bivariate_df_from_cov <- function(x1, x2, cov){
  # assumes mu = 0
  cor <- cov2cor_vc(cov)
  p12 <- cor[1,2]
  x1_stuff <- x1/(sqrt(cov[1,1]))
  x2_stuff <- x2/sqrt(cov[2,2])
  exp_stuff <- (1/(1-p12**2)) * ((x1_stuff**2) + (x2_stuff**2) - 2*p12*x1_stuff*x2_stuff)
```

```
  exp_stuff <- -exp_stuff/2

  pre_exp <- 1/(2*pi*sqrt(det(cov)))
  fx1x2 <- pre_exp * exp(exp_stuff)
  fx1x2
}
```

Define variables:

```
x1 <- seq(-10, 10, length=41)
x2 <- x1
z_vc <- outer(x1,x2,FUN = bivariate_df_from_cov, cov1) # calculating the density values
```

Plot density function:

```
# Commented because I cannot plot it on pdf.
# library(plotly)
# plot_ly() %>% add_surface(x = x1, y = x2, z = t(z_vc))
```

Now I'm just checking I've gotten the z values correctly:

```
mu1<-0 # setting the expected value of x1
mu2<-0 # setting the expected value of x2
s11 <- 8 # setting the variance of x1
s12 <- 5 # setting the covariance between x1 and x2
s22 <- 4 # setting the variance of x2
rho <- 5/sqrt(8*4) # setting the correlation coefficient between x1 and x2
x1 <- seq(-10, 10, length=41) # generating the vector series x1
x2 <- x1 # copying x1 to x2

f<-function(x1,x2){
term1 <- 1/(2*pi*sqrt(s11*s22*(1-rho^2)))
term2 <- -1/(2*(1-rho^2))
term3 <- (x1-mu1)^2/s11
term4 <- (x2-mu2)^2/s22
term5 <- -2*rho*((x1-mu1)*(x2-mu2))/(sqrt(s11)*sqrt(s22))
term1*exp(term2*(term3+term4+term5))
} # setting up the function of the multivariate normal density >#
z <- outer(x1,x2,f) # calculating the density values

all.equal(z, z_vc)
```

```
## [1] TRUE
```

Cool

b)

COV is a 2x2 matrix symmetric matrix, therefore it's eigenvectors form an orthogonal matrix. I has two real eigenvalues.

```
V <- eigen(cov1)$vectors; D <- eigen(cov1)$values
a1 <- V[,1]
a2 <- V[,2]

t(a1)%*%cov1%*%a1
```

```
##          [,1]
```

```
## [1,] 11.38516
```

```r
D[1]
```

```
## [1] 11.38516
```

```r
t(a2)%*%cov1%*%a2
```

```
##           [,1]
## [1,] 0.6148352
```

```r
D[2]
```

```
## [1] 0.6148352
```

Veiem que la primera component explicaria un % elevat de la variança de les dades originals. Podríem reduir de $p = 2$ a $m = 1$ variables.

c)

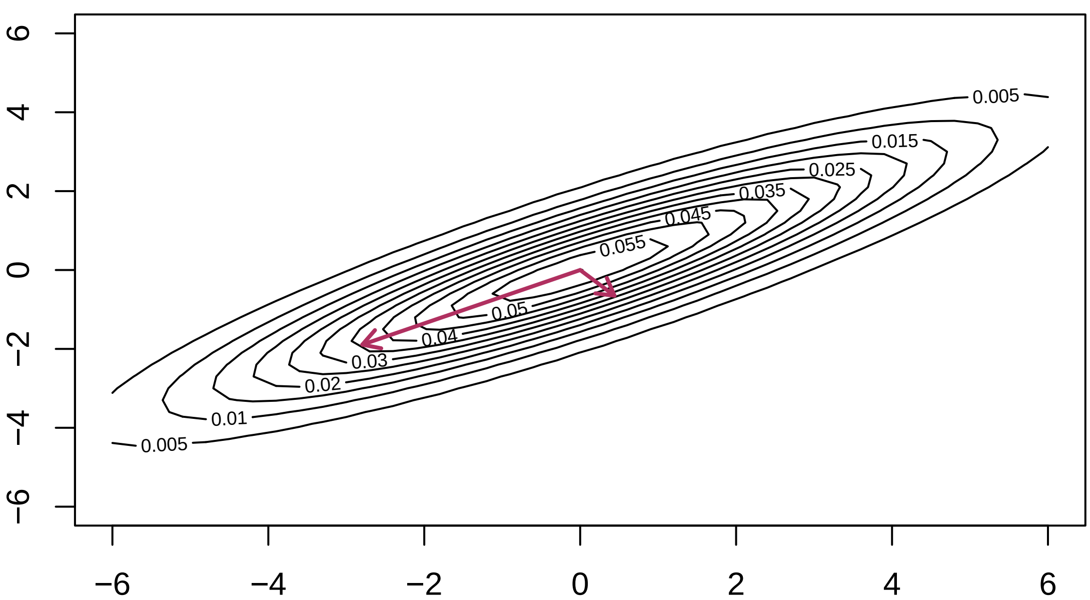
Very much copied from the solution...

Question: Why are the vectors scaled by $\sqrt{\lambda}$?

```r
x1 <- seq(-6, 6, length=41); x2 <- x1
z <- outer(x1, x2, FUN=bivariate_df_from_cov, cov1)

contour(x1,x2,z, nlevels=20)
arrows(0, 0,
       a1[1]*sqrt(D[1]),
       a1[2]*sqrt(D[1]),
       len=0.1,lwd=2,col="maroon")

arrows(0,0,
       a2[1]*sqrt(D[2]),
       a2[2]*sqrt(D[2]),
       len=0.1,lwd=2,col="maroon")
```



Clearly the eigenvectors of Σ are the shortest and longest radius of the ellipse.

TODO: Think about how the directions of maximum variance relate to the axis of the pdf of a multivariate normal distribution. Some resources:

- https://fkorona.github.io/ATML/2017_2/Lecture_notes/03C_Normal.pdf
- https://www.cs.princeton.edu/courses/archive/fall10/cos513/notes/2010-11-15.pdf
- https://www.cs.columbia.edu/~djhsu/coms4771-f20/lectures/06-multivariate_gaussians_and_pca.pdf

Exercici 2