

## (9) Hauptkomponentenanalyse

Ziel dieses Seminarblatts ist es, die Hauptkomponentenanalyse eines Beispieldatensatzes auf der Implementationsebene nachzuvollziehen und die resultierenden Matrizen zu visualisieren. Untenstehender **R** Code erzeugt dazu zunächst einen Datensatz von  $n = 20$  unabhängigen Beobachtungen eines  $m = 5$ -dimensionalen multivariat normalverteilten Zufallsvektors.

```
# Datensatzsimulation
library(MASS)
set.seed(1)
m = 5
n = 20
mu = rep(0,m)
Sigma = matrix(runif(m^2), nrow = m)
Sigma = 0.5*(Sigma+t(Sigma))
Sigma = Sigma + m*diag(m)
X = t(mvnrnorm(n,mu,Sigma))

# multivariate Normalverteilung (mvnrnorm())
# Reproduzierbare Randomisierung
# Datendimension
# Anzahl Realisierungen
# Erwartungswertparameter
# zufällige Matrix
# symmetrische Matrix
# positiv definite Matrix
# Datensatzgeneration
```

Folgender **R** Code führt dann mithilfe einer Eigenanalyse eine Hauptkomponentenanalyse des Datensatzes durch.

```
# Hauptkomponentenanalyse durch Eigenanalyse
I_n = diag(n)
J_n = matrix(rep(1,n^2), nrow = n)
C = (1/(n-1))*X %*% (I_n-(1/n)*J_n) %*% t(X)
D = diag(1/sqrt(diag(C)))
R = D %*% C %*% D
EA = eigen(C)
lambda = EA$values
Q = EA$vectors
X_tilde = t(Q) %*% X
C_tilde = (1/(n-1))*X_tilde %*% (I_n-(1/n)*J_n) %*% t(X_tilde)
D_tilde = diag(1/sqrt(diag(C_tilde)))
R_tilde = D_tilde %*% C_tilde %*% D_tilde

# Einheitsmatrix I_n
# 1_{nn}
# Stichprobenkovarianzmatrix X
# Kov-Korr-Transformationsmatrix
# Stichprobenkorrelationsmatrix X
# Eigenanalyse von C
# Eigenwerte von C
# Eigenvektoren von C
# Transformierter Datensatz
# Stichprobenkovarianzmatrix \tilde{X}
# Kov-Korr-Transformationsmatrix
# Stichprobenkorrelationsmatrix \tilde{X}
```

Folgender **R** Code visualisiert die Datensatzmatrix  $X$  und die PCA-transformierte Datensatzmatrix  $\tilde{X}$  wie in Abbildung 1 gezeigt.

```
# Visualisierung des Datensatzes X und des PCA-transformierten Datensatzes \tilde{X}
library(plot.matrix)
library(latex2exp)
pdf(
  file = "/9_Abbildungen_S/mv_9_X_tildeX.pdf",
  width = 10,
  height = 7)
par(
  family = "sans",
  mfcol = c(2,1),
  pty = "m",
  bty = "l",
  lwd = 1,
  las = 1,
  mgp = c(2,1,0),
  xaxs = "i",
  yaxs = "i",
  font.main = 1,
  cex = 1,
  cex.main = 1.6)

# Visualisierung X
plot(
  X,
  breaks = c(-5,5),
  col = topo.colors,
  fmt.key = "%.0f",
  polygon.key = NULL,
  axis.key = NULL,
  xlab = "",
  ylab = "",
```

```

main      = TeX("$X$")
# Visualisierung \tilde{X}
plot(
X_tilde,
breaks   = c(-5,5),
col      = topo.colors,
fmt.key  = "%.0f",
polygon.key = NULL,
axis.key = NULL,
xlab     = "",
ylab     = "",
main     = TeX("$\tilde{X}$")
dev.off()

```

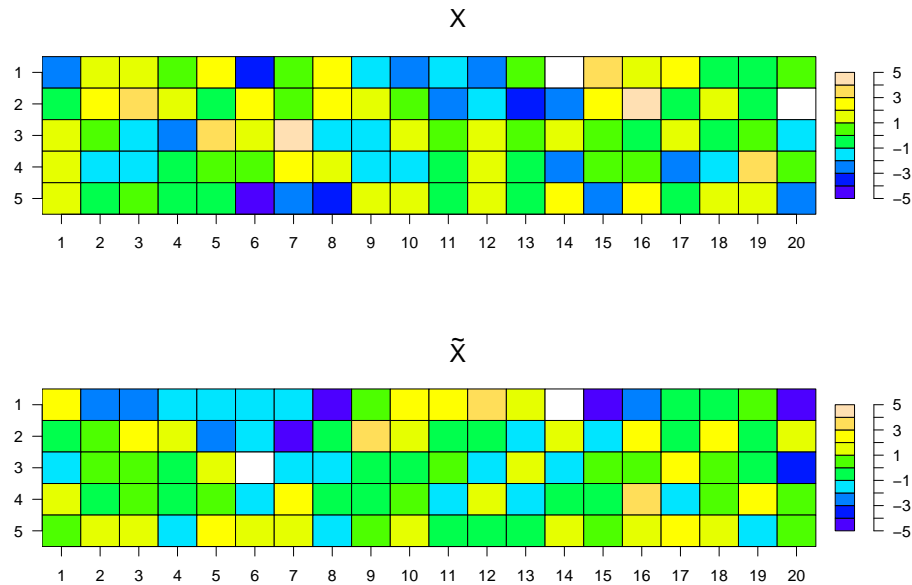


Abbildung 1. Datensatz  $X$  und PCA-transformierter Datensatz  $\tilde{X}$

Folgender **R** Code schließlich visualisiert die zentralen Matrizen der Hauptkomponentenanalyse des Datensatzes wie in Abbildung 2 gezeigt.

```

# Visualisierung von Q, \Lambda, C, \tilde{C}, R, \tilde{R}
pdf(
file      = "../9_Abbildungen_S/mv_9_Q_Lambda_C.pdf",
width     = 11,
height    = 8)
par(
family    = "sans",
mfcol     = c(2,3),
pty       = "m",
bty       = "l",
lwd       = 1,
las       = 1,
mgp       = c(2,1,0),
xaxs     = "i",
yaxs     = "i",
font.main = 1,
cex       = 1,
cex.main  = 2)

```

```

# Q und \Lambda
plot(
Q,
col      = topo.colors,
digits  = 2,
key      = NULL,
cex      = 0.8,
polygon.key = NULL,
axis.key = NULL,
xlab     = "",
ylab     = "",
main     = TeX("$Q$"))
plot(
diag(lambda),
col      = topo.colors,
digits  = 2,
key      = NULL,
cex      = 0.8,
polygon.key = NULL,
axis.key = NULL,
xlab     = "",
ylab     = "",
main     = TeX("$\\Lambda$"))

# C und \tilde{C}
C_tilde[C_tilde < 0.001] = 0
plot(
C,
col      = topo.colors,
digits  = 2,
key      = NULL,
cex      = 0.8,
polygon.key = NULL,
axis.key = NULL,
xlab     = "",
ylab     = "",
main     = TeX("$C$"))
plot(
C_tilde,
col      = topo.colors,
digits  = 2,
key      = NULL,
cex      = 0.8,
polygon.key = NULL,
axis.key = NULL,
xlab     = "",
ylab     = "",
main     = TeX("$\\tilde{C}$"))

# R und \tilde{R}
R_tilde[abs(C_tilde) < 0.001] = 0
plot(
R,
col      = topo.colors,
digits  = 2,
key      = NULL,
cex      = 0.8,
polygon.key = NULL,
axis.key = NULL,
xlab     = "",
ylab     = "",
main     = TeX("$R$"))
plot(
R_tilde,
col      = topo.colors,
digits  = 2,
key      = NULL,
cex      = 0.8,
polygon.key = NULL,
axis.key = NULL,
xlab     = "",
ylab     = "",
main     = TeX("$\\tilde{R}$"))
dev.off()

```

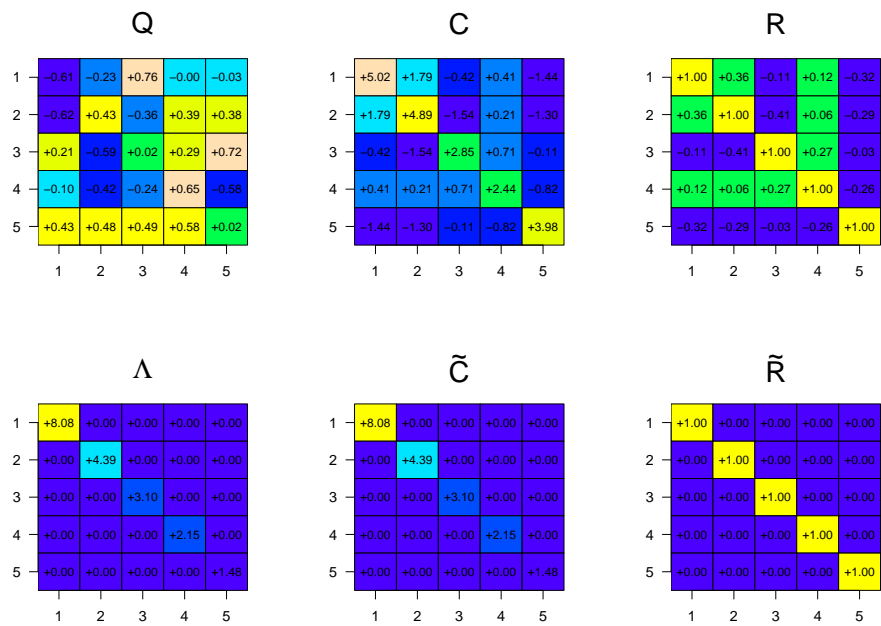


Abbildung 2. Zentrale Matrizen der Hauptkomponentenanalyse