

Hauptkomponentenanalyse

(Principal Component Analysis (PCA))

Jan Siegemund

**Proseminar „Robuste Signalidentifikation“
WS 2003/04**

Inhalt :

1. Motivation

2. Mathematische Grundlagen

3. Hauptkomponentenanalyse

3.1. Definition der Hauptkomponenten

3.2. Ziel der PCA

3.3. Herleitung der Problemlösung

3.4. Eigenschaften der PCA

(Warum wird die PCA zur Gesichtserkennung eingesetzt ?)

4. Beispiel zur PCA

5. QR-Verfahren zur Eigenwert- und Eigenvektorbestimmung

5.1. Householder-Spiegelung

5.2. Givens-Rotation

5.3. Beschreibung des QR-Algorithmus

6. Quellenangabe

1. Motivation

Ein Problem der Gesichtserkennung, sowie anderer Verfahren der Signalidentifikation, ist die Bewältigung von enormen, hochdimensionalen Datenmengen. Die Gesichtserkennung benutzt Klassifikationsalgorithmen, um gemessene Signale zu verarbeiten und einzuordnen. Klassifikationsalgorithmen arbeiten effizienter, wenn die Dimension der Eingabemenge gering ist. Eingaben hoher Dimension führen meist zu ungenügenden Ergebnissen. Oft sind jedoch nicht wenige Komponenten einer Datenmenge irrelevant, bzw. weniger relevant als andere, da sie nahezu konstant sind.

Beispielsweise unterscheiden sich Gesichter in Nasen-, Augen-, und Mundpartie stärker voneinander als in Ausschnitten der Stirn oder der Wangen, so dass es reicht, diese wichtigen Partien als Unterscheidungsmerkmale zu speichern.

Also ist es sinnvoll, ein Verfahren heranzuziehen, welches die einzelnen Dimensionen der Datenmenge nach ihrer Relevanz, bzw. nach den Abweichungen der Menge in dieser Dimension klassifiziert. Die PCA ist ein Verfahren, das genau dies liefert.

Im nächsten Kapitel werden zunächst einige mathematische Grundbegriffe eingeführt, die den Zusammenhang von Werten beschreiben und diesen messbar machen.

In Kapitel 3 werden dann Hauptkomponenten definiert und das Verfahren hergeleitet, sowie einige wichtige Eigenschaften der PCA erläutert und bewiesen.

Zum besseren Verständnis enthält Kapitel 4 ein einfaches Beispiel zur Durchführung einer Hauptkomponentenanalyse und Kapitel 5 die Beschreibung eines für die PCA relevanten mathematischen Verfahrens.

2.Mathematische Grundlagen:

Die Hauptkomponentenanalyse ist ein Verfahren der multivariaten Statistik. In den Beweisen und Herleitungen der nächsten Kapitel werden diverse Zeichen und Abkürzungen gebraucht, die in dieser Weise nicht allgemein definiert sind und der Vereinfachung der Darstellung dienen sollen. Die Semantik dieser Zeichen soll an dieser Stelle erklärt werden.

Seien x und y Vektoren des \mathbb{R}^n , so entspricht x' dem transponierten Vektor von x . Folglich ist $x'y = xy$ das Skalarprodukt von x und y . \bar{x} ist das Symbol für einen Vektor gleicher Dimension wie x , dessen Komponenten alle aus dem gleichen Wert, nämlich dem arithmetischen Mittel von x bestehen. Sei A eine $(n \times m)$ Matrix und B eine $(m \times p)$ Matrix, so bezeichnet AB das Matrixprodukt. Der Buchstabe E bezeichnet die Einheitsmatrix in geeigneter Dimension.

Es sollen darüber hinaus einige Begriffe aus der Statistik definiert werden, die für die Hauptkomponentenanalyse von zentraler Bedeutung sind. Diese Begriffe sollen im Folgenden erklärt werden.

Eine entscheidende Größe in der Hauptkomponentenanalyse ist die Varianz eines Vektors. Diese ist ein Maß die für Abweichung der Komponenten des Vektors zu ihrem arithmetischen Mittel, also der Abweichung von x zu \bar{x} .

Gegeben sei also ein Vektor $x = (x_1, \dots, x_n)$, so definiert sich seine Varianz als

$$\text{Var}(x) = \frac{1}{n-1} (x - \bar{x})'(x - \bar{x}).$$

Einen ähnlichen Zusammenhang, jedoch mit zwei Vektoren, beschreibt die Kovarianz. Durch die Kovarianz sind zwei Vektoren bezüglich ihrer Varianzen vergleichbar. Sie beschreibt also die Unterschiedlichkeit der Abweichung von x zu \bar{x} zu der Abweichung von y zu \bar{y} . Seien zwei Vektoren $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n)$ gegeben, so definiert sich deren Kovarianz durch

$$\text{Kov}(x, y) = \frac{1}{n-1} (x - \bar{x})'(y - \bar{y}).$$

Es gilt offensichtlich $\text{Kov}(x, x) = \text{Var}(x)$.

Erweitern wir diese Definitionen von Vektoren auf Matrizen, erhalten wir die Kovarianzmatrix. Es sei also eine $(m \times n)$ Matrix A gegeben. Die zugehörige Kovarianzmatrix errechnet sich als

$$\text{Kov}(A)_{ij} = \text{Kov}(a_i, a_j) = \left(\frac{1}{n-1} A' A \right)_{ij}.$$

Die Kovarianzmatrix ist quadratisch und symmetrisch und daher, falls sie nur reelle Einträge hat, diagonalisierbar.

Um ein Maß für die Stärke des linearen Zusammenhangs zweier Messgrößen $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n)$ zu erhalten, definieren wir deren Korrelationskoeffizient durch

$$\text{Kor}(x, y) = \frac{\text{Kov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}}.$$

Die Werte des Korrelationskoeffizienten liegen im Bereich $[-1, 1]$. Beträgt das Ergebnis von $\text{Kor}(x, y)$ Null, dann sind x und y unkorreliert, also unabhängig. Ist der Wert positiv, implizieren hohe Werte von x hohe Werte von y und andersherum. Ein negativer Wert beschreibt einen entgegengesetzten Zusammenhang. Hohe Werte von x entsprechen niedrigen Werten von y und umgekehrt.

Um die Korrelationen innerhalb einer (Messdaten-) Matrix zu berechnen, definieren wir die Korrelationsmatrix zu einer gegebenen Matrix A . Diese errechnet sich analog zur Kovarianzmatrix als

$$\text{Kor}(A)_{ij} = \text{Kor}(a_i, a_j) .$$

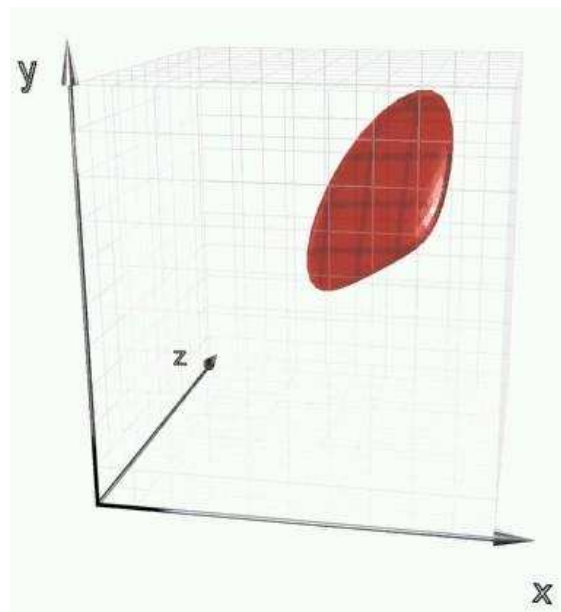
Die Korrelationsmatrix ist quadratisch und symmetrisch.

3. Hauptkomponentenanalyse (Principal Component Analysis PCA)

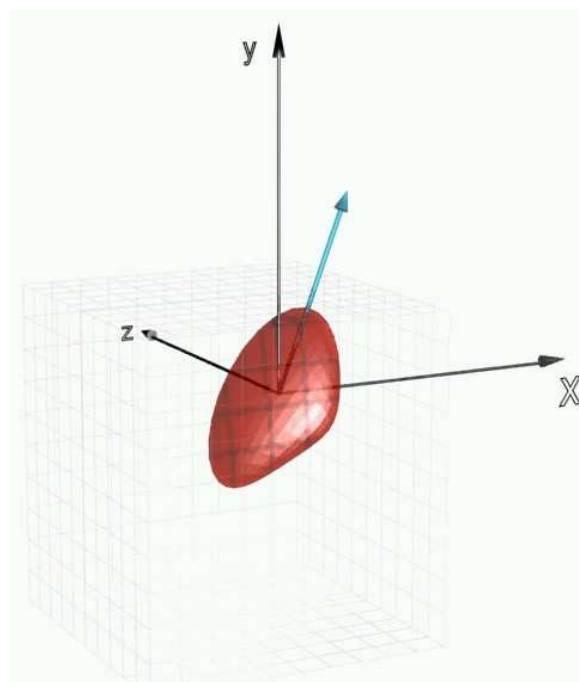
3.1. Definition (Interpretation) der Hauptkomponenten :

Zur Anschauung betrachten wir zunächst das Beispiel einer dreidimensionalen Datenmenge, welche im Sinne der PCA transformiert und anschließend auf zwei Dimensionen reduziert wird.

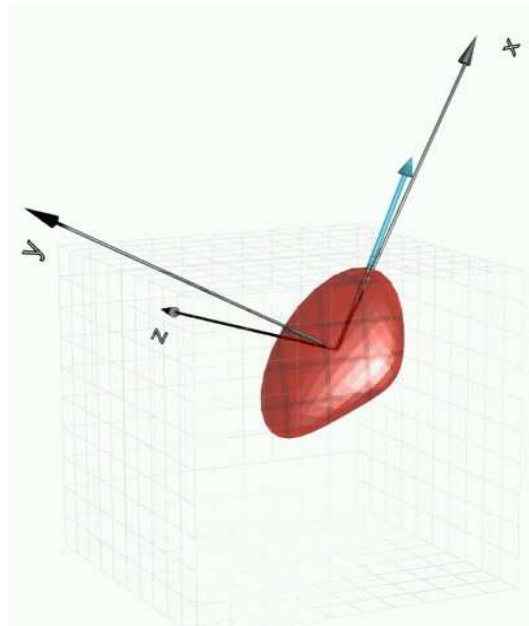
Gegeben sei eine Reihe mehrdimensionaler Messungen (Datenmenge), die folgende Punktwolke (roter Körper) bilden.



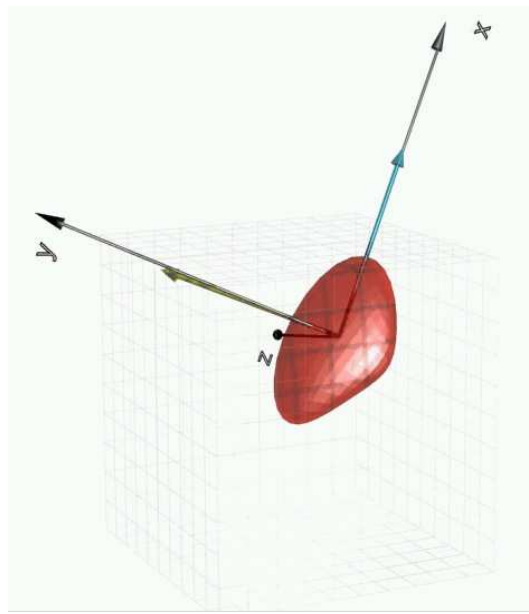
Als erstes wird der Ursprung des Koordinatensystems in den Schwerpunkt der Punktwolke gesetzt.



Als nächstes wird das Koordinatensystem gedreht, so dass die erste Achse in Richtung der größten Abweichung, bzw. der größten Varianz ausgerichtet ist (blauer Pfeil).



Der nächste Schritt dreht die zweite Achse in Richtung der größtmöglichen Varianz unkorreliert zur ersten Achse (gelber Pfeil). D.h., die Drehung des Koordinatensystems richtet die zweite Achse in Richtung der größten Varianz aus, die möglich ist, ohne die Richtung der ersten Achse zu verändern (also Drehung des Systems um die erste Achse).



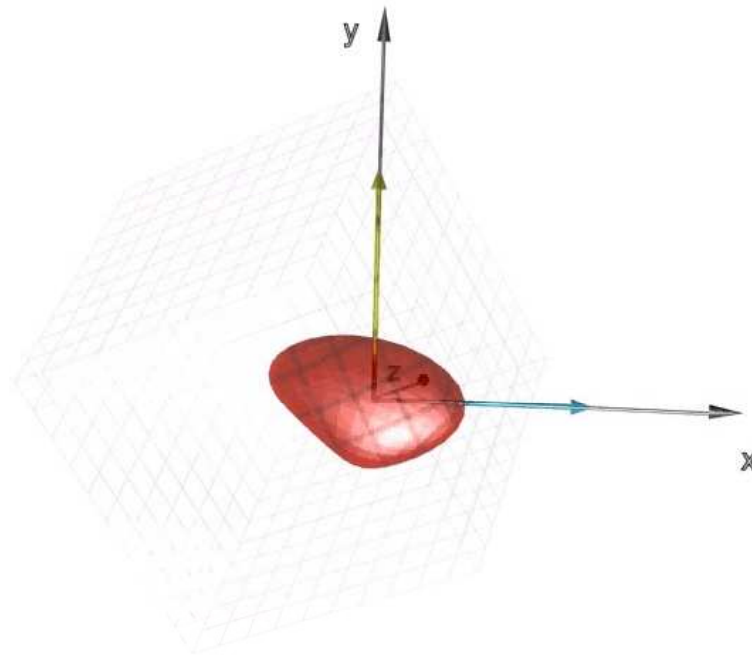
Das Verfahren wird fortgesetzt, bis die k -te Achse in Richtung der größten Varianz ausgerichtet ist, unkorreliert zu den ersten $k-1$ Achsen. Die k -te Achse bezeichnet so die k -te Hauptkomponente.

Dies führt zur geometrischen Interpretation der Hauptkomponenten als Hauptachsen eines Ellipsoiden (Punktwolke).

Motiviert durch diese anschauliche Betrachtung können wir nun die Aufgabenstellung formal definieren. Die Datenmenge sei gegeben durch n p -dimensionale Messungen in Form einer $(p \times n)$ Matrix $X' = (x_1, \dots, x_p)$. Die Suche nach den p Hauptkomponenten der Datenmenge entspricht der Suche nach p unkorrelierten Linearkombinationen der Vektoren x_1, \dots, x_p mit maximaler Varianz. D.h., es müssen p p -dimensionale Vektoren a_i gefunden werden, so dass die Varianz von y_i , mit $y_i = a_1 x_1 + \dots + a_p x_p = a_i' X$, maximal wird. Diese y_i sind die Hauptkomponenten der Datenmenge.

3.2 Ziel der Hauptkomponentenanalyse:

Das Ziel der PCA ist, wie in der Motivation beschrieben, die Hilfe bei der Interpretation einer Datenmenge durch Bestimmung der Komponenten mit dem größten/kleinsten Einfluss. Wenn diese Komponenten bekannt sind, kann eine kleinere Anzahl von Basisvektoren für die Menge gefunden werden (also eine Dimensionsreduzierung vorgenommen werden), so dass nur ein möglichst kleiner Teil der in den Daten enthaltenen Informationen verloren geht.



3.3. Herleitung der Problemlösung:

Wir betrachten die Aufgabenstellung die in Kapitel 3.1 definiert wurde. Gegeben sei eine Datenmenge, welche aus n p -elementigen Beobachtungen in Form einer $(p \times n)$ Matrix X besteht. Sei $S = \text{Kov}(X)$ die $(p \times p)$ Kovarianzmatrix zu X . Gesucht ist der p -dimensionale Vektor a_1 für den gilt, dass $\text{Var}(a_1' X)$ maximal wird.

Diese Bedingung entspricht, laut Definition von Varianz und Kovarianz, dem Problem $a_1' S a_1$ zu maximieren.

Da jedoch der Ausdruck für beliebige a_1 beliebig groß wird, benötigt man eine Schrankenbedingung.

Wir benutzen die Bedingung $a_1' a_1 = 1$.

(möglich sind auch andere Bedingungen z.B.: $\max |a_1^{(j)}| = 1$)

Das Problem ist also nun die Maximierung eines Ausdrucks mit Nebenbedingung. Für die Lösung verwenden wir den Lagrange-Multiplikator λ in der Gleichung $a_1' S a_1 - \lambda(a_1' a_1 - 1)$, welche Ausdruck und Nebenbedingung in einer Formel zusammenfasst. Wir suchen also den Vektor a_1 , der das Ergebnis dieser Gleichung maximiert.

Wie gewohnt differenzieren wir dazu nach a_1 , um einen Extremwert zu erhalten. Die Ableitung liefert

$$S a_1 - \lambda a_1 = 0 \Rightarrow (S - \lambda E) a_1 = 0.$$

Offensichtlich ist dies ein Eigenwertproblem von S , wobei λ ein Eigenwert (EW) ist und a_1 zu λ gehörender Eigenvektor (EV).

Aus $S a_1 - \lambda a_1 = 0$ folgt $S a_1 = \lambda a_1$. Wenn wir diese Erkenntnis in das ursprüngliche Problem, welches durch die Maximierung von $a_1' S a_1$ gegeben war, einsetzen, erhalten wir die Umformung

$$\begin{aligned} \max \{a_1' S a_1 \mid a_1' a_1 = 1\} &= \max \{a_1' \lambda a_1 \mid a_1' a_1 = 1 \wedge \lambda \text{ ist EW von } S\} \\ &= \max \{\lambda a_1' a_1 \mid a_1' a_1 = 1 \wedge \lambda \text{ ist EW von } S\} = \max \{\lambda \mid \lambda \text{ ist EW von } S\}. \end{aligned}$$

Gesucht ist daher der größte EW von S .

Nun suchen wir den p-dimensionalen Vektor a_2 , für den gilt: $\text{Var}(a_2' X)$ wird maximal, $a_2' a_2 = 1$ und a_1 ist unkorreliert zu a_2 .

D.h., es muss gelten

$$\begin{aligned} 0 &= \text{Kov}(a_1' X, a_2' X) = a_1' S a_2 = \lambda_1 a_1' a_2 = \lambda_1 a_2' a_1 \\ \Rightarrow \quad a_1 \text{ unkorreliert zu } a_2 &\Leftrightarrow a_1' S a_2 = a_2' S a_1 = a_1' a_2 = a_2' a_1 = 0. \end{aligned}$$

Es tritt eine ähnliche Situation wie im ersten Schritt auf, was zur Anwendung einer erweiterten Lagrange-Multiplikatorgleichung führt, welche zwei Multiplikatoren verwendet:

$$a_2' S a_2 - \lambda(a_2' a_2 - 1) - \phi a_2' a_1 = \max. \quad (\text{A})$$

Ableiten nach a_2 liefert

$$Sa_2 - \lambda a_2 - \phi a_1 = 0 .$$

Multiplizieren mit a_1 ergibt

$$a_1' Sa_2 - \lambda a_1 a_2 - \phi a_1 a_1 = 0$$

$$\Rightarrow \quad \phi = 0 . \quad (B)$$

(A) und (B) liefern dann

$$Sa_2 - \lambda a_2 = 0 \Rightarrow (S - \lambda E)a_2 = 0 .$$

Gesucht ist also der zweitgrößte EW.

Fortsetzung bis p liefert folgende Werte:

- $\{a_1, \dots, a_p\}$ als Hauptvektoren und damit $\{a_1 I, \dots, a_p I\}$ als Hauptkomponenten mit $I =$ Matrix aus den Basisvektoren des Ausgangssystems
- $\{\lambda_1, \dots, \lambda_p\}$ als deren Varianzen
- $\frac{\lambda_k}{\lambda_1 + \dots + \lambda_p}$ als ein Maß für den Anteil der k-ten Hauptkomponente an der Gesamtvarianz
- $\{z_1, \dots, z_n\}$, mit $Z = A'X$ sind die Positionen der Messungen im Hauptkomponentensystem

3.4. Eigenschaften der PCA:

In diesem Kapitel soll deutlich werden, warum die PCA zur Gesichtserkennung eingesetzt wird. Eine wichtige Eigenschaft der Hauptkomponentenanalyse ist die optimale Rekonstruktion im Sinne der kleinsten Fehlerquadrate. Das bedeutet, dass das Verfahren eine Dimensionsreduzierung ermöglicht, bei der der Informationsverlust minimal ist. Eine formale Beschreibung und der Beweis werden unter Eigenschaft 2 angegeben. Um den Beweis zu führen ist jedoch Eigenschaft 1 erforderlich, welche im Folgenden definiert und bewiesen wird.

Eigenschaft 1:

Sei $y = B'x$ eine orthonormale Abbildung (mit $B = (p \times q)$, $1 \leq q \leq p$), also $S_y = B'S_xB$, wobei $S_y = \text{Kov}(y)$ und $S_x = \text{Kov}(x)$ gilt.

Dann gilt

$\text{Spur}(S_y)$ wird maximal genau dann, wenn $B = A_q$ ist,
wobei A_q aus den ersten q Hauptkomponenten besteht.

Beweis:

Sei $B = AC$ (C ist eine $(p \times q)$ Transformationsmatrix),

dann folgt

$$\text{Spur}(B'SB) = \text{Spur}(C' \underbrace{A'SA}_{D = \text{diag}(\lambda_1, \dots, \lambda_p)} C) = \text{Spur}(C'DC) = \sum_{j,k=1}^p \lambda_j c_{jk}^2 . \quad (1)$$

Da die Spalten von A und B orthonormal sind, folgt Orthonormalität für die Spalten von C :

$$C'C = B'AA'B = B'B = E_q$$
$$\Rightarrow \text{Spur}(C'C) = \sum_{j,k=1}^p c_{jk}^2 = \text{Spur}(E_q) = q .$$

Für die Zeilen von C gilt:

$$c_j'c_j \leq 1 \quad (\text{da } C \text{ Teil einer Orthogonalmatrix ist})$$
$$\Rightarrow \sum_{j,k=1}^p c_{jk}^2 \leq 1 . \quad (2)$$

Aus 1) und 2) folgt also

$$\sum_{j,k=1}^p \lambda_j c_{jk}^2 \text{ wird maximal, falls gilt: } \sum_{k=1}^p c_{jk}^2 = \begin{cases} 1 & j=1, \dots, q \\ 0 & j=q+1, \dots, p \end{cases} .$$

Dies wird erfüllt durch $C = E_q$, also $B = A_q$.

q.e.d.

Bemerkung: Umgekehrt wird der Wert minimal, falls $B = \tilde{A}_q$ gilt.

Dabei werden die Spalten von \tilde{A}_q aus EV zu den q kleinsten EW gebildet.

Mit Hilfe von Eigenschaft 1 kann nun Eigenschaft 2 bewiesen werden.

Eigenschaft 2: Optimale Rekonstruktion (im Sinne der kleinsten Fehlerquadrate)

Gegeben sei eine Punktwolke $\{x_1, \dots, x_n\}$ (Messung) in einem p-dimensionalen Raum, derart wie sie in Kapitel 3.1 beschrieben wurde.

Durchzuführen sei eine Projektion auf einen q-dimensionalen Unterraum $y_i = Bx_i$.

Angestrebt ist ein kleinstmöglicher Datenverlust durch diese Projektion.

Ein Maß für den Datenverlust ist die Summe der quadrierten Abstände \sum der Punkte zum Unterraum.

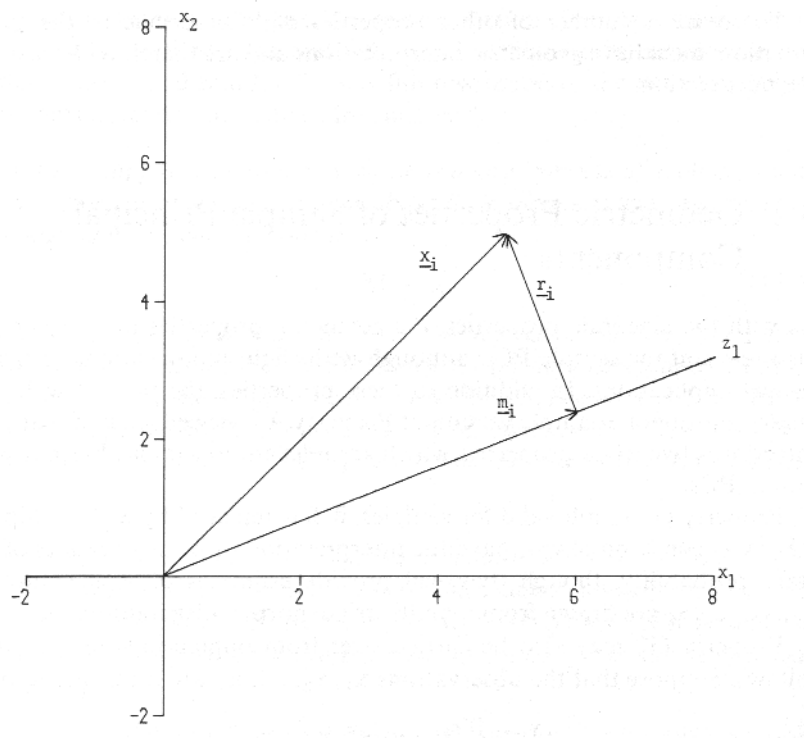
Es gilt, \sum wird minimal genau dann, wenn $B = A_q$,
wobei A_q die Matrix aus den EV der größten q EW von S ist.

(d.h.: $y_i = A_q x_i$ beschreibt diejenige Projektion mit dem geringsten Datenverlust)

Beweis:

Wir betrachten das Beispiel mit p=2 und q=1.

Dabei gilt:



z_1 = Unterraum,

x_i = i-ter Messwert,

m_i = Projektion von x_i
auf z_1 , also Position
von y_i im alten
Koordinatensystem,

$r_i = x_i - m_i$
= Abstand x_i zu m_i .

$\sum_{i=1}^n r_i' r_i$ ist also die Summe der quadrierten Fehler.

Die Benennung und Anschauung ist im Fall höherdimensionaler p und q, wobei $p > q$ gelten muss, analog. Der Beweis kann also für den allgemeinen Fall geführt werden.

Es gilt

$$x_i' x_i = (m_i + r_i)'(m_i + r_i) = m_i' m_i + r_i' r_i + 2r_i' m_i = m_i' m_i + r_i' r_i ,$$

da m_i orthogonal zu r_i ist, also

$$\sum_{i=1}^n r_i' r_i = \sum_{i=1}^n x_i' x_i - \sum_{i=1}^n m_i' m_i .$$

Um $\sum_{i=1}^n r_i' r_i$ zu minimieren, muss man $\sum_{i=1}^n m_i' m_i$ maximieren,

also $\sum_{i=1}^n y_i' y_i$ maximieren.

Durch einfache Umformungsschritte erhalten wir ein Maximierungsproblem, das wir gemäß Eigenschaft 1 lösen können.

$$\begin{aligned} \sum_{i=1}^n y_i' y_i &= \sum_{i=1}^n x_i' B B' x_i = \text{Spur} \left(\sum_{i=1}^n x_i' B B' x_i \right) = \sum_{i=1}^n \text{Spur}(x_i' B B' x_i) \stackrel{\equiv}{=} \text{Spur}(AB) = \text{Spur}(BA) \\ \sum_{i=1}^n \text{Spur}(B' x_i x_i' B) &= \text{Spur} \left[B' \left(\sum_{i=1}^n x_i x_i' \right) B \right] = \text{Spur}[B' X' X B] = \\ (n-1) \text{Spur}(B' S B) &= \max . \end{aligned}$$

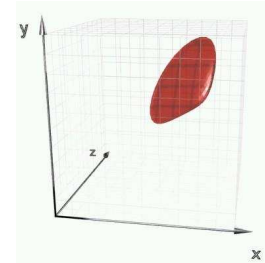
Dies ist laut Eigenschaft 1 genau dann erfüllt, wenn gilt $B = A_q$.

q.e.d.

4. Beispiel zur Hauptkomponentenanalyse

Gegeben seien sechs (= n) Messungen von Monatsmitteltemperaturen in vier (= p) Städten als Datenmenge.

| Tabelle 1: Verwendete Datenreihen (Monatsmitteltemperaturen in .1°C). | | | | | | | |
|---|--------|---------|------|-------|-----|------|--------|
| Station | Januar | Februar | März | April | Mai | Juni | Mittel |
| Athen | 84 | 109 | 140 | 166 | 205 | 247 | 158.5 |
| Gibraltar | 138 | 157 | 161 | 165 | 197 | 223 | 173.5 |
| Luqa | 132 | 141 | 147 | 167 | 197 | 239 | 170.5 |
| Palma | 101 | 135 | 128 | 128 | 181 | 219 | 148.66 |



Führen wir nun nacheinander die oben beschriebenen Schritte aus.

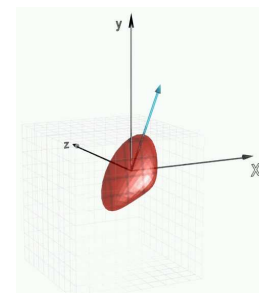
1. Schritt:

Transponierung zur (n x p)-Matrix.

2. Schritt:

Zentrierung der Spaltenwerte (Zentrierung der Punktwolke) durch Subtrahieren des Spaltenmittelwertes liefert

$$\bar{X} = \begin{pmatrix} -74.5 & -35.5 & -38.5 & -47.66 \\ -49.5 & -16.5 & -29.5 & -13.66 \\ -18.5 & -12.5 & -23.5 & -20.66 \\ 7.5 & -8.5 & -3.5 & -20.66 \\ 46.5 & 23.5 & 26.5 & 32.33 \\ 88.5 & 49.5 & 68.5 & 70.33 \end{pmatrix}$$



3. Schritt:

Bilden der Kovarianzmatrix

$$S = \begin{pmatrix} 3678.701 & 1820.498 & 2406.300 & 2436.602 \\ 1820.498 & 952.698 & 1238.098 & 1318.599 \\ 2406.300 & 1238.098 & 1662.299 & 1694.200 \\ 2436.602 & 1318.599 & 1694.200 & 1861.068 \end{pmatrix}$$

4. Schritt:

Die Berechnung der Eigenwerte und Eigenvektoren ergibt die folgenden Werte.

EW:

$$\lambda_1 = 7937.275$$

$$\lambda_2 = 180.048$$

$$\lambda_3 = 37.224$$

$$\lambda_4 = 0.219$$

EV:(schon normiert)

$$a_1 = -.6733; +.3448; +.4535; +.4713)$$

$$a_2 = -.6539; +.2059; +.0604; +.7255)$$

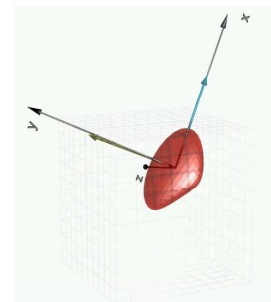
$$a_3 = -.3041; -.1625; +.8888; -.3020)$$

$$a_4 = -.1630; +.9013; -.0271; -.4005).$$

5. Schritt:

Überführen wir die Messungen gemäß der Formel $\bar{Z} = \bar{X}A$ ins neue Koordinatensystem, erhalten wir die neue Matrix

$$\bar{Z} = \begin{pmatrix} -102,3 & -16,4 & 8,6 & 0,4 \\ -58,8 & 1,2 & -4,4 & -0,4 \\ -37,2 & -19,7 & -7,0 & 0,7 \\ -9,2 & -23,8 & 2,2 & -0,5 \\ 66,7 & 13,9 & -4,2 & -0,2 \\ 140,7 & 44,7 & 4,7 & -0,0 \end{pmatrix}$$

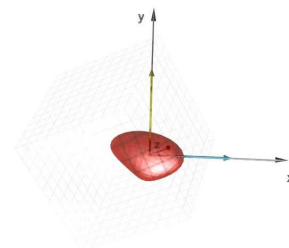


6. Schritt:

Um die Dimension sinnvoll reduzieren zu können, muss zunächst für jede einzelne Hauptkomponente der Anteil der zugehörigen Varianz an der Gesamtvarianz berechnet werden.

Die Anteile errechnen sich durch die Formel $r_l = \frac{\lambda_l}{\sum_{i=1}^p \lambda_i}$.

$r_1 = 97.3330 \%$ Durch die Kenntnis über die Wichtung der Hauptkomponenten kann nun die Dimension, je nach Anforderung in der jeweiligen Situation, reduziert werden.
 $r_2 = 2.2079 \%$
 $r_3 = 0.4565 \%$ (je nachdem, ob die Komponenten mit der größten oder der kleinsten Varianz gesucht werden)
 $r_4 = 0.0027 \%$



5. QR-Verfahren zur Eigenwert- und Eigenvektorbestimmung

Das QR-Verfahren ist ein sehr leistungsfähiges, iteratives Verfahren zur Berechnung der Eigenwerte und Eigenvektoren einer gegebenen Matrix A. Ein zentraler Bestandteil des Algorithmus ist die QR-Zerlegung. Die QR-Zerlegung einer Matrix A ist ein Verfahren, welches zu A eine orthogonale Matrix Q und eine obere Dreiecksmatrix R findet, so dass $A = QR$ gilt.

In der Beschreibung des QR-Algorithmus werden Householder-Spiegelungen und Givens-rotationen angewendet. Diese beiden Verfahren sollen im Folgenden beschrieben werden.

5.1 Die Householder-Spiegelung:

Gegeben sei ein Vektor $y = (y_1, \dots, y_n)'$.

Daraus berechnen wir den Wert $\alpha = \text{sign}(y_1) \|y\|$,

und bilden
$$v = \begin{pmatrix} y_1 + \alpha \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Wir definieren
$$Q_v = E - \frac{2}{v'v} vv'$$
,

woraus folgt, dass gilt $Q_v y = -\alpha e_1$.

Wir wenden dieses Verfahren auf eine (m x n)-Matrix $A = \begin{pmatrix} * & * & * & * \\ y_1 & * & * & * \\ \vdots & * & * & * \\ y_n & * & * & * \end{pmatrix}$ an.

Q hat dann die Form
$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & \tilde{Q} & & \\ 0 & & & \end{pmatrix},$$

daraus folgt
$$Q_v' A = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix}.$$

5.2 Die Givens-Rotation:

Gegeben sei eine (m x n)-Matrix A. Wir nehmen ein Element a_{ii} von der Diagonalen, sowie ein Element b_{ik} aus derselben Spalte und führen die folgende Fallunterscheidung durch. Dabei entsprechen c und s dem Kosinus und Sinus des Rotationswinkels φ , also $c = \cos(\varphi)$ und $s = \sin(\varphi)$.

Falls $b = a$,setze $c = 1, s = 0$.
 Falls $|b| \geq |a|$,setze $r = \frac{a}{b}, s = \frac{1}{\sqrt{1+r^2}}, c = s r$.
 Sonst ,setze $r = \frac{b}{a}, c = \frac{1}{\sqrt{1+r^2}}, s = c r$.

Damit definieren wir $G_{ik} =$

$$\begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & c & 0 & \cdots & 0 & s \\ & & & 0 & 1 & & & 0 \\ & & & \vdots & & \ddots & & \vdots \\ & & & 0 & & & 1 & 0 \\ & & & -s & 0 & \cdots & 0 & c \\ & & & & & & & 1 & & \ddots & \\ & & & & & & & & & & 1 \end{pmatrix} .$$

Sei $A = \begin{pmatrix} * & * & * & * \\ *_{ik} & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix}$, so ist $G_{ik}A = \begin{pmatrix} * & * & * & * \\ 0_{ik} & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix}$.

5.3 Beschreibung des QR-Algorithmus:

Die beiden soeben beschriebenen Verfahren sind wichtige Bestandteile des QR-Algorithmus zur Lösung eines Eigenwertproblems, welcher im Folgenden beschrieben wird.

Gegeben sei eine symmetrische-(n x n) Matrix A. Wir bringen A per Householder-Spiegelungen auf die Tridiagonalform

$$Q_k \dots Q_1 A Q_1' \dots Q_k' = \begin{pmatrix} * & * & 0 & 0 \\ * & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & * \\ 0 & 0 & * & * \end{pmatrix} = \tilde{A} .$$

Per Givens-Rotationen bilden wir dann rekursiv

$$A_0 = \tilde{A}, \quad A_k = Q_k R_k, \quad A_{k+1} = R_k Q_k.$$

Für diese Rekursion gilt:

$$A_k \text{ approximiert } \text{diag}(\lambda_1, \dots, \lambda_n) \quad , \text{mit } \lambda_i = \text{EW von } A.$$

$$G_l \dots G_1 Q_k \dots Q_1 \text{ approximiert } (a_1, \dots, a_n) \quad , \text{mit } a_i = \text{EV von } A.$$

Die Konvergenzgeschwindigkeit hängt von der Trennung der Beträge der Eigenwerte ab. Ist die Trennung schlecht, also die Abstände zwischen den Eigenwertbeträgen sehr klein, so ist das Verfahren sehr langsam. Es existieren diverse Verfahren (z.B. Shift-Techniken) um die Trennung zu verbessern und so die Konvergenzgeschwindigkeit zu erhöhen. Diese Verfahren sollen jedoch hier nicht erläutert werden.

6. Quellenangabe:

- I. T. Jolliffe, *Principal Component Analysis*, Springer, 1986
- <http://www.iam.ubc.ca/~norris/research/amythesis.pdf>
- <http://www.rz.uni-frankfurt.de/~grieser/pca/pca.html>
- <http://www.informatik.uni-leipzig.de-cgip-lehre-ws0203-sm11.pdf>
- <http://www.fbn-dummerstorf.de/fb2/rudolph/webfa/21/21.htm>
- http://rcswww.urz.tu-dresden.de/~am12/kor_reg.htm
- <http://www.digimusik.de/PCA/math.html>
- <http://www.digimusik.de/PCA/bsp3d.html>
- <http://www.dr-preuss.com/>
- <http://www.math.fu-berlin.de/~bioinf/download/mvsec4.pdf>
- <http://www.ubka.unikarlsruhe.de/cgi-bin/psview?document=1997/mathematik/12&format=1&page=139>