Multivariate Verfahren

MSc Psychologie | MSc Klinische Psychologie und Psychotherapie

WiSe 2023/24

Prof. Dr. Dirk Ostwald

(9) Hauptkomponentenanalyse

Hauptkomponentenanalyse = Principal Component Analysis (PCA).

PCA ist eine Featureselektionsmethode.

- "Features" sind die Komponenten multidimensionaler Zufallsvektoren.
- Korrelierte Features repräsentieren redundante Information.

PCA generiert ein korrelationsfreies Featureset durch lineare Featurekombination.

Die Durchführung einer PCA basiert auf

- einer *Orthonormalzerlegung* der *Stichprobenkovarianzmatrix* und
- einer anschließenden *Vektorkoordinatentransformation*.

PCA kannt zur *Kompression* hochdimensionaler Daten eingesetzt werden.

Vektorraumbasen und Vektorkoordinatentransformationen

Definition und Eigenschaften

Datenkompression

Selbstkontrollfragen

Vektorraumbasen und Vektorkoordinatentransformationen

Definition und Eigenschaften

Datenkompression

Selbstkontrollfragen

Definition (Linearkombination)

$\{v_1, v_2, ..., v_k\}$ sei eine Menge von k Vektoren eines Vektorraums V. Dann ist die *Linearkombination* der Vektoren in $v_1, v_2, ..., v_k$ mit den skalaren Koeffizienten $a_1, a_2, ..., a_k$ definiert als der Vektor

$$w := \sum_{i=1}^{k} a_i v_i \in V. \tag{1}$$

Bemerkung

- Als Beispiel seien in $\mathbb{R}^2$

$$v_1 := \begin{pmatrix} 2 \\ 1 \end{pmatrix}, v_2 := \begin{pmatrix} 1 \\ 1 \end{pmatrix}, v_3 := \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ und } a_1 := 2, a_2 := 3, a_3 := 0. \tag{2}$$

Dann ergibt sich

$$w = a_1 v_1 + a_2 v_2 + a_3 v_3 = 2 \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} + 3 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \end{pmatrix}. \tag{3}$$

Definition (Lineare Hülle und Aufspannen)

V sei ein Vektorraum und es sei $W := \{w_1, ..., w_k\} \subset V$. Dann ist die *lineare Hülle* von W definiert als die Menge aller Linearkombinationen der Elemente von W,

$$\text{Span}(W) := \left\{ \sum_{i=1}^{k} a_i w_i | a_1, ..., a_k \text{ sind skalare Koeffizienten} \right\} \tag{4}$$

Man sagt, dass eine Menge von Vektoren $W \subseteq V$ *einen Vektorraum V aufspannt*, wenn jedes $v \in V$ als eine Linearkombination von Vektoren in W geschrieben werden kann.

Bemerkungen

- Wenn eine Menge W von Vektoren einen Vektorraum aufspannt, dann kann jedes Element des Vektorraums durch eine Linearkombination der Elemente der Menge W gebildet werden.

Definition (Lineare Unabhängigkeit)

V sei ein Vektorraum. Eine Menge $W := \{w_1, w_2, ..., w_k\}$ von Vektoren in V heißt *linear unabhängig*, wenn die einzige Repräsentation des Nullelements $0 \in V$ durch eine Linearkombination der $w \in W$ die triviale Repräsentation

$$0 = a_1 w_1 + a_2 w_2 + \cdots + a_k w_k \text{ mit } a_1 = a_2 = \cdots = a_k = 0 \tag{5}$$

ist. Wenn die Menge W nicht linear unabhängig ist, dann heißt sie *linear abhängig*.

Bemerkungen

- Prinzipiell müsste man für jede Linearkombination der $w \in W$ prüfen, ob sie Null ist.
- Die beiden folgenden Theoreme zeigen, dass es auch einfacher geht.

Theorem (Lineare Abhängigkeit von zwei Vektoren)

V sei ein Vektorraum. Zwei Vektoren $v_1, v_2 \in V$ sind linear abhängig, wenn einer der Vektoren ein skalares Vielfaches des anderen Vektors ist.

Beweis

v_1 sei ein skalares Vielfaches von v_2, also

$$v_1 = \lambda v_2 \text{ mit } \lambda \neq 0. \tag{6}$$

Dann gilt

$$v_1 - \lambda v_2 = 0. \tag{7}$$

Dies aber entspricht der Linearkombination

$$a_1 v_1 + a_2 v_2 = 0 \tag{8}$$

mit $a_1 = 1 \neq 0$ und $a_2 = -\lambda \neq 0$. Es gibt also eine Linearkombination des Nullelementes, die nicht die triviale Repräsentation ist, und damit sind v_1 und v_2 nicht linear unabhängig.

Theorem (Lineare Abhängigkeit einer Menge von Vektoren)

V sei ein Vektorraum und $w_1, ..., w_k \in V$ sei eine Menge von Vektoren in V. Wenn einer der Vektoren w_i mit $i = 1, ..., k$ eine Linearkombination der anderen Vektoren ist, dann ist die Menge der Vektoren linear abhängig.

Beweis

Die Vektoren $w_1, ..., w_k$ sind genau dann linear abhängig, wenn gilt, dass $\sum_{i=1}^{k} a_i w_i = 0$ mit mindestens einem $a_i \neq 0$. Es sei also zum Beispiel $a_j \neq 0$. Dann gilt

$$0 = \sum_{i=1}^{k} a_i w_i = \sum_{i=1, i\neq j}^{k} a_i w_i + a_j w_j \tag{9}$$

Also folgt

$$a_j w_j = -\sum_{i=1, i\neq j}^{k} a_i w_i \tag{10}$$

und damit

$$w_j = -a_j^{-1} \sum_{i=1, i\neq j}^{k} a_i w_i = -\sum_{i=1, i\neq j}^{k} (a_j^{-1} a_i) w_i \tag{11}$$

Also ist w_j eine Linearkombination der w_i für $i = 1, ..., k$ mit $i \neq j$. □

Theorem (Orthonormalität und lineare Unabhängigkeit in $\mathbb{R}^m$)

$q_1, ..., q_m$ seien orthonormale Vektoren in $\mathbb{R}^m$, es gelte also

$$q_i^T q_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \text{ für } 1 \le i, j \le m. \tag{12}$$

Dann sind die $q_1, ..., q_m$ linear unabhängig.

<u>Beweis</u>

Für $i = 1, ..., m$ gilt, dass

$$\begin{aligned} a_1 q_1 + a_2 q_2 + \cdots + a_m q_m &= 0_m \\ \Leftrightarrow (a_1 q_1 + a_2 q_2 + \cdots + a_m q_m)^T &= 0_m^T \\ \Leftrightarrow (a_1 q_1 + a_2 q_2 + \cdots + a_m q_m)^T q_i &= 0_m^T q_i \\ \Leftrightarrow \sum_{j=1}^m a_j q_j^T q_i &= 0 \\ \Leftrightarrow a_i &= 0. \end{aligned} \tag{13}$$

Die einzige Repräsentation des Nullelements 0_m durch eine Linearkombination der $q_1, \ldots, q_m$ ist also die triviale Repräsentation und die $q_1, \ldots, q_m$ sind linear unabhängig.

Definition (Basis)

V sei ein Vektorraum und es sei $B \subseteq V$. Dann heißt B eine *Basis von* V, wenn

- die Vektoren in B den Vektorraum V aufspannen und
- die Vektoren in B linear unabhängig sind.

Bemerkung

- Vektorräume haben in der Regel unendlich viele Basen.
- Wir zeigen unten, dass die Forderung der linearen Unabhängigkeit der Vektoren einer Basis impliziert, dass die Darstellung eines Vektors in V durch die Linearkombination der Vektoren in B eindeutig ist.

Theorem (Eigenschaften von Basen)

V sei ein Vektorraum. Dann gelten folgende Eigenschaften für Basen von V.

- Alle Basen von V haben die gleiche Kardinalität und diese wird die *Dimension* von V genannt.
- Jede Menge von m linear unabhängigen Vektoren ist Basis eines m-dimensionalen Vektorraums.

Bemerkung

- Wir verzichten auf einen Beweis des sehr tiefen Theorems.

Definition (Basisdarstellung und Koordinaten)

$B := \{b_1, ..., b_m\}$ sei eine Basis eines m-dimensionalen Vektorraumes V und es sei $v \in V$. Dann heißt die Linearkombination

$$v = \sum_{i=1}^{m} c_i b_i \tag{14}$$

die *Darstellung von v bezüglich der Basis B* und die Koeffizienten $c_1, ..., c_m$ heißen die *Koordinaten von v bezüglich der Basis B*.

Theorem (Eindeutigkeit der Basisdarstellung)

Die Basisdarstellung eines $v \in V$ bezüglich einer Basis B ist eindeutig.

Beweis

Ohne Beschränkung der Allgemeinheit nehmen wir an, dass der Vektorraum von Dimension m ist. Nehmen wir an, dass zwei Darstellungen von v bezüglich der Basis B existieren, also dass

$$
\begin{aligned}
v &= a_1 b_1 + \cdots + a_m b_m \\
v &= c_1 b_1 + \cdots + c_m b_m
\end{aligned}
\tag{15}
$$

Subtraktion der unteren von der oberen Gleichung ergibt

$$
0 = (a_1 - c_1) b_1 + \cdots + (a_m - c_m) b_m \tag{16}
$$

Weil die $b_1, ..., b_m$ linear unabhängig sind, gilt aber, dass $(a_i - c_i) = 0$ für alle $i = 1, ..., m$ und somit sind die beiden Darstellungen von v bezüglich der Basis B identisch.

$\square$

Bemerkung

- Die lineare Unabhängigkeit von Basisvektoren garantiert also, dass ein Vektor zu einer gegebenen Basis nur eine Darstellung, also insbesondere nur ein einziges Set an Koordinaten $c_1, ..., c_m$ hat und nicht etwa mehrere.

Definition (Orthonormalbasis von $\mathbb{R}^m$)

Eine Basis $B := \{q_1, ..., q_m\}$ von $\mathbb{R}^m$ heißt *Orthonormalbasis* von $\mathbb{R}^m$, wenn B eine Menge orthonormaler Vektoren ist, wenn also gilt, dass

$$q_i^T q_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \text{ für } 1 \leq i, j \leq m. \tag{17}$$

Bemerkungen

- Bei einer Orthonormalbasis von $\mathbb{R}^m$ sind die Basisvektoren orthonormal.
- Man beachte, dass eine Basis von $\mathbb{R}^m$ nicht aus orthonormalen Vektoren bestehen muss.

Vektorraumbasen und Vektorkoordinatentransformationen

Beispiel (Orthonormalbasis von $\mathbb{R}^2$)

Es ist

$$B_1 := \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \tag{18}$$

eine Orthonormalbasis von $\mathbb{R}^2$, denn B_1 ist eine Basis und besteht aus orthonormalen Vektoren. Dabei erkennt man die Orthonormalität der Elemente von B_1 zunächst an

$$\begin{aligned}
\begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} &= 1 \cdot 1 + 0 \cdot 0 = 1 + 0 = 1 \\
\begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= 0 \cdot 0 + 1 \cdot 1 = 0 + 1 = 1 \\
\begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= 1 \cdot 0 + 0 \cdot 1 = 0 + 0 = 0
\end{aligned} \tag{19}$$

Da die Orthonormalität zweier Vektoren in $\mathbb{R}^2$ ihre lineare Unabhängkeit impliziert, sind die Elemente von B_1 auch linear unabhängig. Wir wollen schließlich noch nachweisen, dass die Elemente von B_1 den Vektorraum $\mathbb{R}^2$ aufspannen, dass also jedes $v \in \mathbb{R}^2$ durch eine Linearkombination der Elemente von B_1 geschrieben werden kann. Dazu sei $v = (v_1, v_2)^T \in \mathbb{R}^m$ beliebig. Dann aber gilt

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ für } c_1 := v_1 \text{ und } c_2 := v_2 \tag{20}$$

und damit ist B_1 eine Orthonormalbasis von $\mathbb{R}^2$.

Theorem (Kanonische Basis von $\mathbb{R}^m$)

Es sei

$$B := \left\{e_1, ..., e_m | e_{i_j} = 1 \text{ für } i = j \text{ und } e_{i_j} = 0 \text{ für } i \neq j\right\} \subset \mathbb{R}^m. \tag{21}$$

Dann ist B eine Orthonormalbasis von $\mathbb{R}^m$ und wird die *kanonische Basis* von $\mathbb{R}^m$ genannt.

Beweis

Im Sinne der Definition einer Basis müssen wir zeigen, dass die Elemente von B linear unabhängig sind. Dies folgt aber direkt aus der Orthonormalität der Elemente von B. Weiterhin müssen wir zeigen, dass die Elemente von B den Vektorraum $\mathbb{R}^m$ aufspannen, dass also jeder Vektor $v \in \mathbb{R}^m$ als Linearkombination der Elemente von B geschrieben werden kann. Sei also $v \in \mathbb{R}^m$ beliebig. Dann kann $v := (v_1, ..., v_m)^T$ geschrieben werden als

$$v = \sum_{i=1}^{m} c_i e_i \text{ mit } c_i := v_i. \tag{22}$$

□

Bemerkungen

- Die kanonische Basis von $\mathbb{R}^3$ ist zum Beispiel $B := \left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$.

Beispiel (Eine nichtkanonische Orthonormalbasis von $\mathbb{R}^2$)

Neben B_1 ist auch

$$B_2 := \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\} \tag{23}$$

eine Orthonormalbasis von $\mathbb{R}^2$, denn B_2 ist ebenfalls eine Basis von $\mathbb{R}^2$ und besteht ebenfalls aus orthonormalen Vektoren. Dabei erkennt man die Orthonormalität der Elemente von B_2 zunächst an

$$\begin{aligned}
\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} &= \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = \frac{1}{2} + \frac{1}{2} = 1 \\
\begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} &= \left(-\frac{1}{\sqrt{2}}\right) \cdot \left(-\frac{1}{\sqrt{2}}\right) + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = \frac{1}{2} + \frac{1}{2} = 1 \\
\begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} &= -\frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} = -\frac{1}{2} + \frac{1}{2} = 0.
\end{aligned} \tag{24}$$

Da die Orthonormalität zweier Vektoren in $\mathbb{R}^2$ ihre lineare Unabhängkeit impliziert, sind die Elemente von B_2 auch linear unabhängig. Es bleibt also nachzuweisen, dass die Elemente von B den Vektorraum $\mathbb{R}^2$ aufspannen, dass also jedes $v \in \mathbb{R}^2$ durch eine Linearkombination der Elemente von B_2 geschrieben werden kann.

Vektorraumbasen und Vektorkoordinatentransformationen

Beispiel (Eine nichtkanonische Orthonormalbasis von $\mathbb{R}^2$ fortgeführt)

Sei also $v \in \mathbb{R}^m$ beliebig. Dann gilt, dass

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = c_1 \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} + c_2 \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \text{ für } c_1 := \frac{\sqrt{2}}{2}(v_1 + v_2) \text{ und } c_2 := \frac{\sqrt{2}}{2}(v_2 - v_1), \tag{25}$$

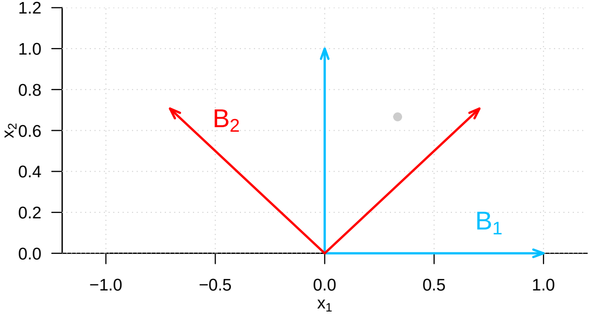
denn dann ist

$$\begin{aligned} c_1 \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} + c_2 \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} &= \frac{\sqrt{2}}{2}(v_1 + v_2) \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} + \frac{\sqrt{2}}{2}(v_2 - v_1) \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \\ &= \frac{1}{2}(v_1 + v_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{2}(v_2 - v_1) \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2}(v_1 + v_2) - \frac{1}{2}(v_2 - v_1) \\ \frac{1}{2}(v_1 + v_2) + \frac{1}{2}(v_2 - v_1) \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2}v_1 + \frac{1}{2}v_2 - \frac{1}{2}v_2 + \frac{1}{2}v_1 \\ \frac{1}{2}v_1 + \frac{1}{2}v_2 + \frac{1}{2}v_2 - \frac{1}{2}v_1 \end{pmatrix} \\ &= \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \\ &= v. \end{aligned} \tag{26}$$

Damit ist B_2 eine Orthonormalbasis von $\mathbb{R}^2$.

Vektorraumbasen und Vektorkoordinatentransformationen

Kanonische und nichtkanonische Orthonormalbasen B_1 und B_2 von $\mathbb{R}^2$



Im Rahmen der Hauptkomponentenanalyse sind wir daran interessiert, basierend auf den Koordinaten eines Vektors bezüglich einer Basis die Koordinaten desselben Vektors bezüglich einer anderen Basis zu berechnen. Dazu führen wir im Folgenden die Begriffe der *Orthogonalprojektion*, der *Vektorkoordinaten bezüglich einer Orthogonalbasis* und der *Vektorkoordinatentransformation* ein.

Definition (Orthogonalprojektion)

x und q seien Vektoren im Euklidischen Vektorraum $\mathbb{R}^m$. Dann ist die *Orthogonalprojektion von* x *auf* q definiert als der Vektor

$$\tilde{x} = aq \text{ mit } a := \frac{q^T x}{q^T q}, \tag{27}$$

wobei der Skalar a *Projektionsfaktor* genannt wird.

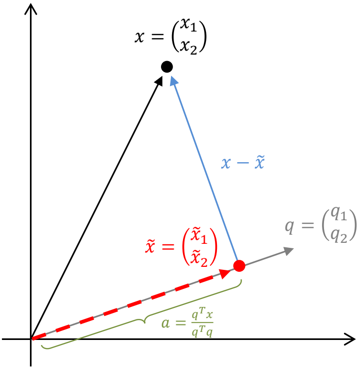
Bemerkungen

- Per definition ist $\tilde{x} = aq$ mit $a \in \mathbb{R}$ der Punkt in Richtung von q der x am nähesten ist.
- Diese minimierte Distanzeigenschaft impliziert die Orthogonalität von q und $x - \tilde{x}$.
- Die Formel von a folgt direkt aus der Orthogonalität von $x - \tilde{x}$ und q, da gilt

$$q^T(x - \tilde{x}) = 0 \Leftrightarrow q^T(x - aq) = 0 \Leftrightarrow q^T x - aq^T q = 0 \Leftrightarrow a = \frac{q^T x}{q^T q}. \tag{28}$$

- Wenn q die Länge $||q|| = \sqrt{q^T q} = 1$ hat, dann gilt $a = \frac{q^T x}{||q||^2} = q^T x$.
- Die Orthogonalprojektion von x auf q wird manchmal auch einfach als *Projektion von* x *auf* q bezeichnet

Orthogonalprojektion



Theorem (Vektorkoordinaten bezüglich einer Orthogonalbasis)

Es sei $x \in \mathbb{R}^m$ und es sei $B := \{q_1, ..., q_m\}$ eine Orthonormalbasis von $\mathbb{R}^m$. Dann ergeben sich für $i = 1, ..., m$ die Koordinaten c_i in der Basisdarstellung von x bezüglich B als die Projektionsfaktoren

$$c_i = x^T q_i \tag{29}$$

in der Orthogonalprojektion von x auf q_i. Äquivalent ist die Basisdarstellung von x bezüglich B gegeben durch

$$x = \sum_{i=1}^{m} (x^T q_i) q_i. \tag{30}$$

<u>Beweis</u>

Für $i = 1, ..., m$ gilt

$$x = \sum_{j=1}^{m} c_j q_j \Leftrightarrow q_i^T x = q_i^T \sum_{j=1}^{m} c_j q_j \Leftrightarrow q_i^T x = \sum_{j=1}^{m} c_j q_i^T q_j \Leftrightarrow q_i^T x = c_i \Leftrightarrow c_i = x^T q_i. \tag{31}$$

Bemerkung

- Hinsichtlich der kanonischen Basis von $\mathbb{R}^m$ ergibt sich offenbar $c_i = x^T e_i = x_i$ für $i = 1, ..., m$.

Theorem (Vektorkoordinatentransformation)

$B_v := \{v_1, ..., v_m\}$ und $B_w := \{w_1, ..., w_m\}$ seien zwei Orthonormalbasen eines Vektorraums. $A \in \mathbb{R}^{m \times m}$ sei die Matrix, die durch die spaltenweise Konkatenation der Koordinaten der Vektoren in B_w in der Basisdarstellung bezüglich der Basis B_v ergibt. Dann können die Koordinaten $x_i, i = 1, ..., m$ eines Vektors x bezüglich der Basis B_v in die Koordinaten $\tilde{x}_1, ..., \tilde{x}_m$ des Vektors bezüglich der Basis B_w durch

$$\tilde{x} = A^T x \tag{32}$$

transformiert werden. Analog können die Koordinaten $\tilde{x}_1, ..., \tilde{x}_m$ des Vektors hinsichtlich der Basis B_w in die Koordinaten $x_1, ..., x_m$ des Vektors hinsichtlich B_v durch

$$x = A\tilde{x}. \tag{33}$$

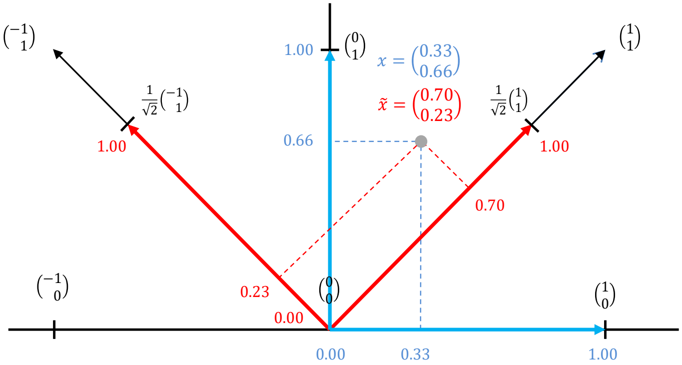
transformiert werden.

Bemerkungen

- Das Theorem erlaubt die Berechnung von Vektorkoordinaten bezüglich einer anderen Orthonormalbasis.
- Für die Berechnung muss zunächst die Matrix A gebildet und dann (nur) entsprechend multipliziert werden.
- Wir verzichten auf einen Beweis und demonstrieren das Theorem an einem Beispiel.

Ein Vektor wird hier als fester Punkt in $\mathbb{R}^m$ betrachtet; die Komponenten (Zahlen) des Vektors werden dagegen nur als Koordinaten bezüglich einer spezifischen Basis interpretiert.

Beispiel



Man beachte, dass x und $\tilde{x}$ am selben Ort in $\mathbb{R}^2$ liegen.

Beispiel

Wir nehmen an, dass wir die Koordinaten von $x = (1/3, 2/3)^T \in \mathbb{R}^2$ hinsichtlich der kanonischen Orthonormalbasis $B_v := \{e_1, e_2\}$ in die Koordinaten bezüglich der Basis

$$B_w := \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\} \tag{34}$$

transformieren wollen. Die Basisdarstellungen der in Vektoren B_w bezüglich der Basisvektoren in B_v sind

$$\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = a_{11} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_{21} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ und } \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = a_{12} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_{22} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{35}$$

Die Projektionsfaktoren der Orthogonalprojektionen der Vektoren in B_w auf die Vektoren in B_v sind

$$a_{11} = \frac{1}{\sqrt{2}}, a_{21} = \frac{1}{\sqrt{2}}, a_{12} = -\frac{1}{\sqrt{2}}, a_{22} = \frac{1}{\sqrt{2}}. \tag{36}$$

Die Transformationsmatrix $A \in \mathbb{R}^{m \times m}$ in obigem Theorem ergibt sich also zu

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}. \tag{37}$$

Die Vektorkoordinatentransformation von $x \in \mathbb{R}^2$ ergibt sich also zu

$$\tilde{x} = A^T x = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix} \approx \begin{pmatrix} 0.70 \\ 0.23 \end{pmatrix}. \tag{38}$$

Vektorraumbasen und Vektorkoordinatentransformationen

Definition und Eigenschaften

Datenkompression

Selbstkontrollfragen

Definition (Hauptkomponentenanalyse)

$\mathbb{C}(\xi)$ sei die Kovarianzmatrix eines m-dimensionalen Zufallsvektors ξ. Dann heißt die Orthonormalzerlegung

$$\mathbb{C}(\xi) = Q\Lambda Q^T, \tag{39}$$

wobei

- $Q \in \mathbb{R}^{m\times m}$ die Matrix der spaltenweisen Konkatenation der Eigenvektoren von $\mathbb{C}(\xi)$ und
- $\Lambda \in \mathbb{R}^{m\times m}$ die Diagonalmatrix der zugehörigen Eigenwerte bezeichnen,

die *Hauptkomponentenanalyse* von $\mathbb{C}(\xi)$ und die Spalten von Q heißen die *Hauptkomponenten* von $\mathbb{C}(\xi)$. Der m-dimensionale Zufallsvektor

$$\tilde{\xi} = Q^T\xi \tag{40}$$

heißt *Hauptkomponenten-transformierter oder Hauptkomponenten-projizierter Zufallsvektor*.

Bemerkungen

- Man spricht auch von der Hauptkomponentenanalyse/den Hauptkomponenten von ξ.

Theorem (Basiseigenschaften der Hauptkomponentenanalyse)

$\mathbb{C}(\xi) \in \mathbb{R}^{m \times m}$ sei die Kovarianzmatrix eines m-dimensionalen Zufallsvektors ξ, es sei $\mathbb{E}(\xi) = 0_m$ und es sei

$$\mathbb{C}(\xi) = Q\Lambda Q^T, \tag{41}$$

die Hauptkomponentenanalyse von $\mathbb{C}(\xi)$. Dann gelten:

(1) Die Spalten von Q, also die Hauptkomponenten von $\mathbb{C}(\xi)$, bilden eine Orthonormalbasis von $\mathbb{R}^m$.

(2) Multiplikation mit Q^T transformiert die Koordinaten von ξ bezüglich der kanonischen Basis von $\mathbb{R}^m$ in Koordinaten bezüglich der Hauptkomponenten von $\mathbb{C}(\xi)$.

Bemerkungen

- Werte von ξ werden üblicherweise zunächst hinsichtlich der kanonischen Basis von $\mathbb{R}^m$ verstanden.
- Die Hauptkomponentenanalyse resultiert in einer alternativen Orthonormalbasis für möglichen Werte von ξ.
- $\tilde{\xi}$ entspricht den Koordinaten von ξ bezüglich dieser alternativen Orthonormalbasis.

Definition und Eigenschaften

Beweis

(1) Mit dem Theorem zu den Eigenschaften von Basen gilt, dass jede Menge von m linear unabhängigen Vektoren Basis eines m-dimensionalen Vektorraums ist. Weiterhin gilt mit dem Theorem zu Orthonormalität und linearer Unabhängigkeit, dass m orthonormale Vektoren in $\mathbb{R}^m$ auch m linear unabhängige Vektoren in $\mathbb{R}^m$ sind. Damit sind die Spalten von Q m linear unabhängige Vektoren in $\mathbb{R}^m$ und folglich eine Basis von $\mathbb{R}^m$.

(2) Wir betrachten das Theorem zur Vektorkoordinatentransformation und setzen $B_v := \{e_1, ..., e_m\}$ und $B_w := \{q_1, ..., q_m\}$ mit den Spalten $q_1, ..., q_m \in \mathbb{R}^m$ von Q. Dann gilt, dass $Q \in \mathbb{R}^{m \times m}$ die Matrix ist, die sich durch die spaltenweise Konkatenation der Koordinaten der Vektoren in B_w in der Basisdarstellung bezüglich der Basis B_v ergibt, denn für $i = 1, ..., m$ gilt, dass die Basisdarstellung von q_i bezüglich der kanonischen Basis B_v gegeben ist durch

$$q_i = \sum_{j=1}^{m} (q_i^T e_j) e_j = \sum_{j=1}^{m} q_{i_j} e_j = q_i. \tag{42}$$

Äquivalent ist natürlich jeder Vektor $q \in \mathbb{R}^m$ schon immer identisch mit der Basisdarstellung von q bezüglich der kanonischen Basis. Damit folgt aber mit dem Theorem zur Vektorkoordinatentransformation direkt, dass der Hauptkomponenten-projizierte Zufallsvektor

$$\tilde{\xi} = Q^T \xi \tag{43}$$

aus den Koordinaten des Vektors bezüglich der Hauptkomponenten von $\mathbb{C}(\xi)$ besteht.

Theorem (Kovarianzeigenschaften der Hauptkomponentenanalyse)

$\mathbb{C}(\xi) \in \mathbb{R}^{m\times m}$ sei die Kovarianzmatrix eines m-dimensionalen Zufallsvektors ξ, es sei $\mathbb{E}(\xi) = 0_m$ und es sei

$$\mathbb{C}(\xi) = Q\Lambda Q^T, \tag{44}$$

die Hauptkomponentenanalyse von $\mathbb{C}(\xi)$. Dann gelten

(1) Die Kovarianzmatrix des Hauptkomponenten-transformierten Zufallsvektors ist die Diagonalmatrix Λ, es gilt also insbesondere $\mathbb{V}(\tilde{\xi}_i) = \lambda_i$ für $i = 1, ..., m$ und $\mathbb{C}(\tilde{\xi}_i, \tilde{\xi}_j) = 0$ für $i \neq j, 1 \leq i, j \leq m$.

(2) Die Summen der Varianzen der Komponenten von ξ und der Komponenten von $\tilde{\xi}$ sind identisch,

$$\sum_{i=1}^{m} \mathbb{V}(\xi_i) = \sum_{i=1}^{m} \mathbb{V}(\tilde{\xi}_i). \tag{45}$$

Bemerkungen

- Bei Annahme von $\lambda_1 > \lambda_2 > \cdots > \lambda_m$ mit zugehörigen Eigenvektoren $q_1, ..., q_m$ gilt

$$\mathbb{V}(\tilde{\xi}_1) > \mathbb{V}(\tilde{\xi}_2) > \cdots > \mathbb{V}(\tilde{\xi}_m) \Leftrightarrow \mathbb{V}(q_1^T\xi) > \mathbb{V}(q_2^T\xi) > \cdots > \mathbb{V}(q_m^T\xi). \tag{46}$$

  und $q_1^T\xi$ maximiert die Varianz der unkorrelierten Linearkombinationen der Komponenten von ξ mit orthonormalen Koeffizientenvektoren.
- Die paarweise nicht-identischen Kovarianzen der Komponenten von $\tilde{\xi}$ sind Null, die Komponenten von $\tilde{\xi}$ sind also unkorreliert und repräsentieren keine redundante Information.
- Die Gesamtvarianz der Komponenten von ξ und $\tilde{\xi}$ sind identisch, wobei Gesamtvarianz hier im Sinne der Summe der Varianzen der jeweiligen Komponenten zu verstehen ist.

Definition und Eigenschaften

Beweis

(1) Wir erinnern zunächst daran, dass die inverse Matrix einer orthogonalen Matrix Q durch Q^T gegeben ist. Mit $QQ^T = Q^TQ = I_m$ gilt dann, dass

$$\mathbb{C}(\xi) = Q\Lambda Q^T \Leftrightarrow Q^T\mathbb{C}(\xi)Q = Q^TQ\Lambda Q^TQ \Leftrightarrow Q^T\mathbb{C}(\xi)Q = \Lambda. \tag{47}$$

Weiterhin gilt, dass mit $\mathbb{E}(\xi) = 0_m$ die Kovarianzmatrix von ξ gegeben ist durch

$$\mathbb{C}(\xi) = \mathbb{E}\left((\xi - \mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))^T\right) = \mathbb{E}\left(\xi\xi^T\right). \tag{48}$$

Damit ergibt sich für die Kovarianzmatrix des PCA-transformierte Vektors $\tilde{\xi} = Q^T\xi$ aber, dass

$$\begin{aligned}
\mathbb{C}(\tilde{\xi}) &= \mathbb{E}\left(\left(\tilde{\xi} - \mathbb{E}(\tilde{\xi})\right)\left(\tilde{\xi} - \mathbb{E}(\tilde{\xi})\right)^T\right) \\
&= \mathbb{E}\left(\left(Q^T\xi - \mathbb{E}(Q^T\xi)\right)\left(Q^T\xi - \mathbb{E}(Q^T\xi)\right)^T\right) \\
&= \mathbb{E}\left(\left(Q^T\xi - Q^T\mathbb{E}(\xi)\right)\left(Q^T\xi - Q^T\mathbb{E}(\xi)\right)^T\right) \\
&= \mathbb{E}\left((Q^T\xi)(Q^T\xi)^T\right) = Q^T\mathbb{E}\left(\xi\xi^T\right)Q = Q^T\mathbb{C}(\xi)Q = \Lambda.
\end{aligned} \tag{49}$$

(2) Wir erinnern daran, dass schon für jede symmetrische Matrix mit verschiedenen Eigenwerten, also auch für positiv-semidefinite Matrizen mit verschiedenen Eigenwerten die Spur gleich der Summe der Eigenwerte ist. Also gilt hier insbesondere für die Eigenwerte $\lambda_1, ..., \lambda_m$ von $\mathbb{C}(\xi)$

$$\sum_{i=1}^m \mathbb{V}(\xi_i) = \mathrm{tr}(\mathbb{C}(\xi)) = \sum_{i=1}^m \lambda_i = \sum_{i=1}^m \mathbb{V}(\tilde{\xi}_i) \tag{50}$$

Theorem (Projektionseigenschaften der ersten Hauptkomponente)

$\mathbb{C}(\xi) \in \mathbb{R}^{m \times m}$ sei die Kovarianzmatrix eines m-dimensionalen Zufallsvektors ξ, es sei $\mathbb{E}(\xi) = 0_m$ und es sei

$$\mathbb{C}(\xi) = Q \Lambda Q^T, \tag{51}$$

die Hauptkomponentenanalyse von $\mathbb{C}(\xi)$. $q_1 \in \mathbb{R}^m$ sei die erste Spalte von Q, also der erste Eigenvektor von $\mathbb{C}(\xi)$, mit zugehörigen Diagonalelementen λ_1 von Λ, also dem größten Eigenwert von $\mathbb{C}(\xi)$. Dann gelten

(1) Unter allen Orthogonalprojektion von ξ auf Vektoren $v \in \mathbb{R}^m$ der Länge $\|v\| = 1$ hat der Projektionsfaktor

$$\alpha := q_1^T \xi \tag{52}$$

der Orthogonalprojektion von ξ auf q_1 die größte Varianz.

(2) Unter allen Orthogonalprojektionen $\tilde{\xi}_v := (v^T \xi) v$ von ξ auf Vektoren $v \in \mathbb{R}^m$ der Länge $\|v\| = 1$ mit *erwartetem quadratischem Fehler*

$$E(\xi_v) := \mathbb{E}\left(\|\xi - \tilde{\xi}_v\|^2\right) \tag{53}$$

hat die Orthogonalprojektion von ξ auf die erste Hauptkomponente q_1, $\tilde{\xi}_1 := (q_1^T \xi) q$ den geringsten erwarteten quadratischen Fehler.

Bemerkungen

- Man sagt auch oft etwas ungenau, dass die erste Hauptkomponente/der erste Eigenvektor "die Varianz maximiert und den (Rekonstruktions)fehler minimiert".

Beweis

(1) Die erste Aussage entspricht der Lösung des restringierten Optimierungsproblems

$$\max_{v \in \mathbb{R}^m} \mathbb{V}\left(v^T\xi\right) \text{ unter der Nebenbedingung } v^Tv = 1. \tag{54}$$

Zu seiner Lösung geben wir zunächst eine Repräsentation der Varianz des Projektionsfaktors $v^T\xi$ durch die Kovarianzmatrix von ξ an und lösen das Problem dann analytisch mithilfe seiner Lagrangefunktion.

Es gilt zunächst aufgrund von $\xi = 0_m$ und damit $\mathbb{C}(\xi) = \mathbb{E}\left(\xi\xi^T\right)$, dass

$$\begin{aligned}
\mathbb{V}(\alpha) &= \mathbb{V}\left(v^T\xi\right) \\
&= \mathbb{E}\left(\left(v^T\xi - \mathbb{E}(v^T\xi)\right)^2\right) \\
&= \mathbb{E}\left(\left(v^T\xi - v^T\mathbb{E}(\xi)\right)^2\right) \\
&= \mathbb{E}\left(\left(v^T\xi\right)^2\right) \\
&= \mathbb{E}\left((v^T\xi)(v^T\xi)^T\right) \\
&= \mathbb{E}\left(v\xi\xi^Tv\right) \\
&= v^T\mathbb{E}\left(\xi\xi^T\right)v \\
&= v^T\mathbb{C}\left(\xi\right)v
\end{aligned} \tag{55}$$

Definition und Eigenschaften

Beweis (fortgeführt)

Das restringierte Optimierungsproblem hat also die Form

$$\max_{v \in \mathbb{R}^m} v^T \mathbb{C}(\xi)\, v \text{ unter der Nebenbedingung } v^T v = 1. \tag{56}$$

Die Lagrangefunktion dieses Problems hat entsprechend die Formel

$$L : \mathbb{R}^m \times \mathbb{R}, (v, l) \mapsto L(v, l) := v^T \mathbb{C}(\xi)\, v - l(v^T v - 1) \tag{57}$$

wobei wir mit $l \in \mathbb{R}$ den Lagrangemultiplikator für die Nebenbedingung $v^T v - 1 = 0$ bezeichnen. Pragmatisches Ableiten der Lagrangefunktion hinsichtlich v und Nullsetzen ergibt dann

$$\begin{aligned}
\frac{\partial}{\partial v} L(v, l) &= 0 \\
\Leftrightarrow \frac{\partial}{\partial v} \left(v^T \mathbb{C}(\xi)\, v - l(v^T v - 1) \right) &= 0 \\
\Leftrightarrow 2\mathbb{C}(\xi)\, v - 2lv &= 0 \\
\Leftrightarrow \mathbb{C}(\xi)\, v &= lv.
\end{aligned} \tag{58}$$

Also ist an der Stelle der notwendigen Bedingung für ein Maximum von L, dass v ein Eigenvektor von $\mathbb{C}(\xi)$ mit Eigenwert l ist. Weiterhin gilt an der Stelle der notwendigen Bedingung für ein Maximum aufgrund der Nebenbedingung, dass

$$\mathbb{C}(\xi)\, v = lv \Leftrightarrow v^T \mathbb{C}(\xi)\, v = v^T lv \Leftrightarrow \mathbb{V}\left(v^T \xi\right) = lv^T v \Leftrightarrow \mathbb{V}\left(v^T \xi\right) = l \tag{59}$$

Der größte Eigenwert von $\mathbb{C}(\xi)$ ist aber λ_1 und der entsprechende Eigenvektor, der das Optimierungsproblem löst damit q_1.

Beweis (fortgeführt)

(2) Die zweite Aussage entspricht der Lösung des restringierten Optimierungsproblems

$$\min_{v\in\mathbb{R}^m} E(\xi_v) \text{ unter der Nebenbedingung } v^Tv = 1. \tag{60}$$

Zu seiner Lösung geben wir zunächst eine Repräsentation der erwarteten quadratischen Fehlers mithilfe der Kovarianzmatrix von ξ an führen die Lösung dann auf die Lösung des unter (1) betrachteten Problems zurück. Es gilt aufgrund von $\xi = 0_m$ und damit $\mathbb{C}(\xi) = \mathbb{E}\left(\xi\xi^T\right)$, dass

$$
\begin{aligned}
E(\xi_v) &= \mathbb{E}\left(\|\xi - \tilde{\xi}_v\|^2\right) \\
&= \mathbb{E}\left((\xi - \tilde{\xi}_v)^T(\xi - \tilde{\xi}_v)\right) \\
&= \mathbb{E}\left(\xi^T\xi - 2\xi^T\tilde{\xi}_v + \tilde{\xi}_v^T\tilde{\xi}v\right) \\
&= \mathbb{E}\left(\xi^T\xi - 2\xi^T(v^T\xi)v + ((v^T\xi)v)^T(v^T\xi)v\right) \\
&= \mathbb{E}\left(\xi^T\xi - 2(v^T\xi)(\xi^Tv) + (v^T(\xi^Tv))(v^T\xi)v\right) \\
&= \mathbb{E}\left(\xi^T\xi - 2(v^T\xi)(\xi^Tv) + (v^T\xi)(\xi^Tv)(v^Tv)\right) \\
&= \mathbb{E}\left(\xi^T\xi - 2(v^T\xi)(\xi^Tv) + (v^T\xi)(\xi^Tv)\right) \\
&= \mathbb{E}\left(\xi^T\xi - (v^T\xi)(\xi^Tv)\right) \\
&= \mathbb{E}\left(\xi^T\xi - v^T(\xi\xi^T)v\right) \\
&= \mathbb{E}\left(\xi^T\xi\right) - v^T\mathbb{E}\left(\xi\xi^T\right)v \\
&= \mathbb{E}\left(\xi^T\xi\right) - v^T\mathbb{C}(\xi)v
\end{aligned}
\tag{61}
$$

<u>Beweis (fortgeführt)</u>

Mit $\mathbb{E}\left(\xi^T\xi\right) \geq 0$ wird das Minimum von

$$E(\xi_v) = \mathbb{E}\left(\xi^T\xi\right) - v^T\mathbb{C}(\xi)v \text{ unter der Nebenbedingung } v^Tv = 1 \tag{62}$$

dann aber für das Maximum von

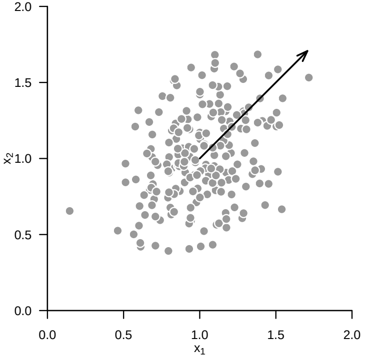
$$\mathbb{V}\left(v^T\xi\right) \text{ unter der Nebenbedingung } v^Tv = 1. \tag{63}$$

angenommen und ist durch Wahl von $v = q_1$ gegeben.

Definition und Eigenschaften

Visualisierung der Projektionseigenschaften der ersten Hauptkomponente für

$m = 2$, $\xi \sim N(\mu, \Sigma)$, • Realisierungen von ξ, $\rightarrow$ Erste Hauptkomponente q_1



Vektorraumbasen und Vektorkoordinatentransformationen

Definition und Eigenschaften

Datenkompression

Selbstkontrollfragen

Überblick

- Datenkompression entspricht der Reduktion der Dimensionalität m eines Datensatzes.
- Ziel der Datenkompression im Rahmen der Datenpräprozessierung bei prädiktiver Modellierung ist es, dem Undersampling hochdimensionaler Datenräume entgegenzuwirken.
- Inspiriert von den Eigenschaften der Hauptkomponentenanalyse möchte man mit dem unten skizzierten Verfahren den Fehler zwischen einem Datensatz und einem dimensionsreduziertem Datensatz bei gleichzeitiger Retention maximal variabler Datenaspekte minimieren.
- Prinzipiell gibt es viele weitere Möglichkeiten der Datenkompression.

Definition (Hauptkomponentenanalyse eines Datensatzes)

$C \in \mathbb{R}^{m \times m}$ sei die Stichprobenkovarianzmatrix eines Datensatzes $X \in \mathbb{R}^{m \times n}$ bestehend aus unabhängigen Realisierungen $x_j \in \mathbb{R}^m$ mit $j = 1, ..., n$ eines m-dimensionalen Zufallsvektors ξ. Dann heißt die Orthonormalzerlegung

$$C = Q\Lambda Q^T \tag{64}$$

wobei

- $Q \in \mathbb{R}^{m \times m}$ die spaltenweise Konkatenation der Eigenvektoren von C und
- $\Lambda \in \mathbb{R}^{m \times m}$ die Diagonalmatrix der zugehörigen Eigenwerten bezeichnen,

die *Hauptkomponentenanalyse von* C und die Spalten von Q heißen die *Hauptkomponenten von* C. Für $j = 1, ..., n$ heißt der m-dimensionale Vektor

$$\tilde{x}_j := Q^T x_j \tag{65}$$

Hauptkomponenten-transformierter oder *Hauptkomponenten-projizierter Datenvektor* und der $m \times n$-dimensionale Datensatz

$$\tilde{X} := Q^T X \tag{66}$$

heißt *Hauptkomponenten-transformierter* oder *Hauptkomponenten-projizierter Datensatz*.

Bemerkungen

- Man spricht auch von der Hauptkomponentenanalyse/den Hauptkomponenten des Datensatzes X.

Theorem (Basiseigenschaften der HKA eines Datensatzes)

$C \in \mathbb{R}^{m \times m}$ sei die Stichprobenkovarianzmatrix eines Datensatzes $X \in \mathbb{R}^{m \times n}$, es sei $\bar{X} = 0_m$ und es sei

$$C = Q\Lambda Q^T, \tag{67}$$

die Hauptkomponentenanalyse von C. Dann gelten:

(1) Die Spalten von Q, also die Hauptkomponenten von C, bilden eine Orthonormalbasis von $\mathbb{R}^m$.

(2) Multiplikation mit Q^T transformiert die Koordinaten der Datenvektoren $x_j, j = 1, ..., n$ bezüglich der kanonischen Basis von $\mathbb{R}^m$ in Koordinaten bezüglich der Hauptkomponenten von C.

Bemerkung

- Ein Beweis ergibt sich analog zum Beweis des Theorems zu den Basiseigenschaften der HKA.

Theorem (Stichprobenkovarianzeigenschaften der HKA)

$C \in \mathbb{R}^{m \times m}$ sei die Kovarianzmatrix eines Datensatzes $X \in \mathbb{R}^{m \times n}$, es sei $\bar{X} = 0_m$ und es sei

$$C = Q\Lambda Q^T, \tag{68}$$

die Hauptkomponentenanalyse von C. Dann gelten

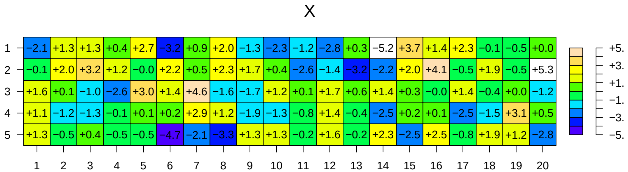
(1) Die Kovarianzmatrix des Hauptkomponenten-transformierten Datensatzes ist die Diagonalmatrix Λ. Es gilt also insbesondere, dass die Stichprobenvarianz der $i$ten Komponente der $\tilde{x}_j$ gegeben ist durch λ_i für $i = 1, ...m$ und dass die Stichprobenkovarianz der $i$ten und $k$ten Komponenten der $\tilde{x}_j$ gleich 0 ist.

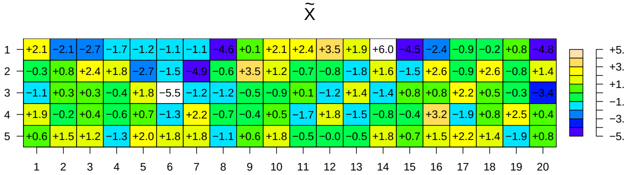
(2) Die Summen der Stichprobenvarianzen der Komponenten der x_j und der Komponenten der $\tilde{x}_j$ sind identisch.

Bemerkung

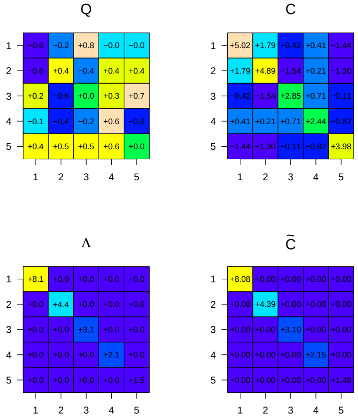
- Ein Beweis ergibt sich analog zum Beweis des Theorems zu den Kovarianzeigenschaften der HKA.

Hauptkomponentenanalyse eines Datensatzes mit $m = 5$ und $n = 20$





Hauptkomponentenanalyse eines Datensatzes mit $m = 5$ und $n = 20$



Die Summe der Diagonalelemente von C und $\tilde{C}$ ist hier 19.2.

Definition (Dimensionsreduzierter transformierte Datensatz)

$X \in \mathbb{R}^{m \times n}$ sei ein Datensatz, C sei seine Stichprobenkovarianzmatrix,

$$C = Q \Lambda Q^T \tag{69}$$

sei die Hauptkomponentenanalyse von C und es gelte $\lambda_1 > \lambda_2 > \cdots > \lambda_m$ für die Diagonalelemente von Λ. Schließlich sei für $k \leq m$ Q_k die Matrix, die aus Q durch Streichen der Spalten $k+1, ..., m$ entsteht. Dann nennt man

$$\tilde{X}_k = Q_k^T X \in \mathbb{R}^{k \times n} \tag{70}$$

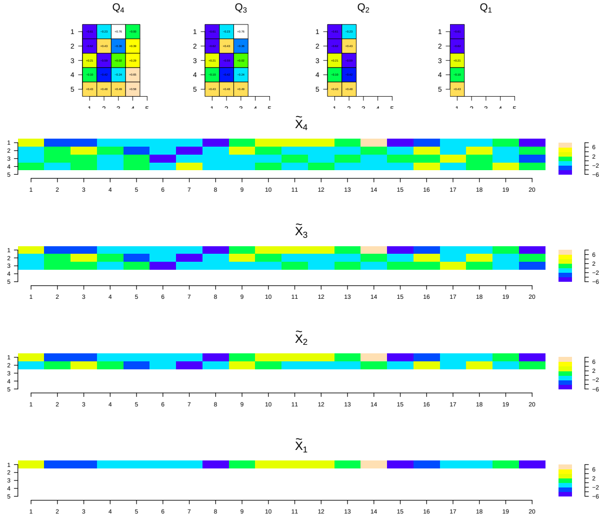
einen *dimensionsreduzierten transformierten Datensatz*.

Bemerkung

- Die Definition ist durch das Theorem zu den Projektionseigenschaften der HKA motiviert.
- Man kann ein analoges Theorem für die HKA eines Datensatzes formulieren und beweisen.
- $\tilde{X}_k = Q_k^T X$ entspricht einer $(k \times n) = (k \times m) \cdot (m \times n)$ Matrixmultiplikation.
- $\tilde{X}_k$ ist der Datensatz, der aus $\tilde{X}$ durch Streichen der ($k+1$)-ten bis m-ten Zeile ensteht.

Datenkompression

PCA-dimensionsreduzierte Datensätze



Definition (Rekonstruierter Datensatz, Datenrekonstruktionsfehler)

$X \in \mathbb{R}^{m \times n}$ sei ein Datensatz und für $k \leq m$ sei

$$\tilde{X}_k = Q_k^T X \in \mathbb{R}^{k \times n} \tag{71}$$

ein dimensionsreduzierter Datensatz. Dann heißt

$$X_k = Q_k \tilde{X}_k \in \mathbb{R}^{m \times n} \tag{72}$$

rekonstruierter Datensatz und

$$e = \|\text{vec}(X - X_k)\|^2 \geq 0 \tag{73}$$

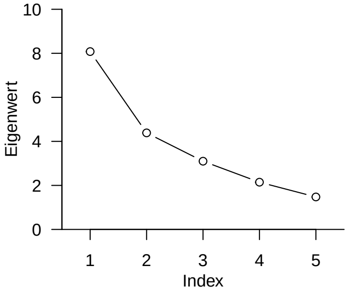
heißt *Datenrekonstruktionsfehler*.

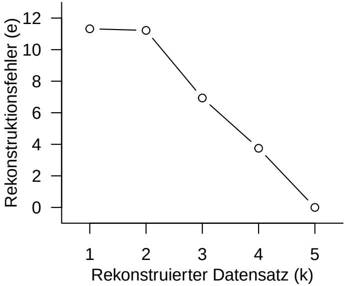
Bemerkungen

- $X_k = Q_k \tilde{X}_k$ entspricht einer $(m \times n) = (m \times k) \cdot (k \times n)$ Matrixmultiplikation
- Für $M \in \mathbb{R}^{m \times n}$ ist $\text{vec}(M) \in \mathbb{R}^{mn}$ der Vektor, der durch Stapeln der Spalten von M entsteht.
- Für $k = m$ gilt $Q\tilde{X}_k = QQ^T X = X$ und damit $e = 0$.
- Der Datenrekonstruktionsfehler ist das Stichprobenanalogon zum erwarteten quadratischen Fehler.

Datensatzrekonstruktion und Daterekonstruktionsfehler

Eigenwerte ("Scree-Plot") und Rekonstruktionsfehler

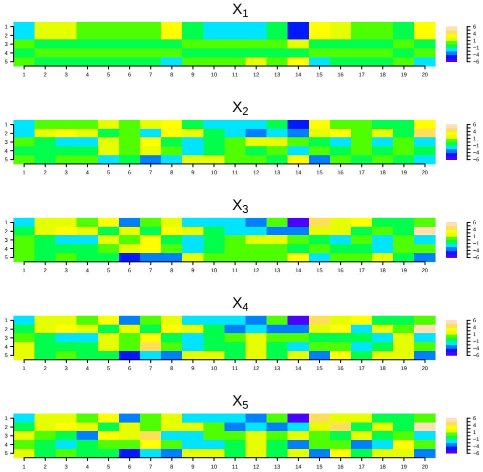




Datenkompression

Datensatzrekonstruktion und Daterekonstruktionsfehler

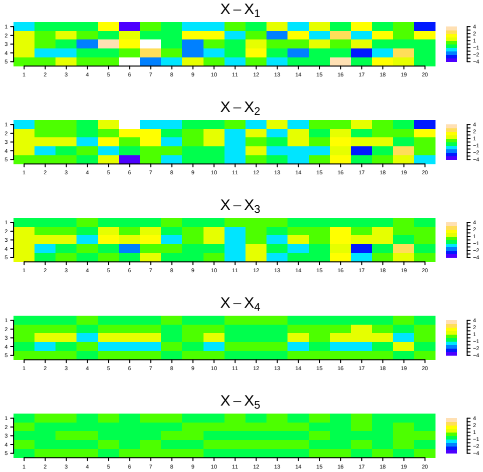
Rekonstruierte Datensätze



Datenkompression

Datensatzrekonstruktion und Daterekonstruktionsfehler

Rekonstruktionsfehler



Vektorraumbasen und Vektorkoordinatentransformationen

Definition und Eigenschaften

Datenkompression

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition der Linearen Unabhängigkeit von Vektoren wieder.
2. Geben Sie das Theorem zur Linearen Abhängigkeit von zwei Vektoren wieder.
3. Geben Sie die Definition der Basis eines Vektorraums wieder.
4. Geben Sie die Definition von Basisdarstellung und Koordinaten wieder.
5. Geben Sie das Theorem zur kanonischen Basis von $\mathbb{R}^m$ wieder.
6. Geben Sie die Definition des Begriffs der Orthogonalprojektion wieder.
7. Geben Sie das Theorem zu Vektorkoordinaten bezüglich einer Orthogonalbasis wieder.
8. Geben Sie das Vektorkoordinatentransformationstheorem wieder.
9. Erläutern Sie das Vektorkoordinatentransformationstheorem.
10. Geben Sie die Definition der Hauptkomponentenanalyse wieder.
11. Geben Sie das Theorem zu den Basiseigenschaften der Hauptkomponentenanalyse wieder.
12. Geben Sie das Theorem zu den Kovarianzeigenschaften der Hauptkomponentenanalyse wieder.
13. Geben Sie das Theorem zu den Projektionseigenschaften der ersten Hauptkomponente wieder.
14. Erläutern Sie das Prinzip der Datenkompression durch Hauptkomponentenanalyse.