U. Mortensen

Kurzfassung Vektoren und Matrizen für Multivariate Verfahren

Letzte Änderung: 01. 03. 2013

1. **Vektoren:** Ein Vektor ist eine geordnete Folge von $n > 1$ Zahlen,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}. \tag{1}$$

   Dies ist ein n-dimensionaler Vektor. Die x_i, $i = 1, 2, \ldots, n$ sind die *Komponenten* des Vektors. Der Fall $n = 1$ ist ein Spezialfall, der als *Skalar* bezeichnet wird.

   Vektoren werden als *Spalte* angeschrieben. Ein Vektor kann *gestürzt* oder *transponiert* werden, – dann wird er als Zeile angeschrieben und mit $\mathbf{x}'$ bezeichnet. Es ist dann

$$\mathbf{x}' = (x_1, x_2, \ldots, x_n). \tag{2}$$

   Natürlich ist dann $(\mathbf{x}')' = \mathbf{x}$. Dementsprechend schreibt man auch zur Platzersparnis $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$, d.h. ein Zeilenvektor, der gestürzt wird, ist wieder ein Spaltenvektor.

2. **Multiplikation mit einem Skalar** Es sei $\lambda \in \mathbb{R}$ ein Skalar, und $\mathbf{x}$ sei ein Vektor. Dann ist mit $\lambda\mathbf{x}$ der Vektor

$$\lambda\mathbf{x} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}, \tag{3}$$

   d.h. die Multiplikation mit einem Skalar bedeutet, dass jede Komponente des Vektors mit diesem Skalar multipliziert wird.

3. **Linearkombinationen** Es seien $\mathbf{x}_1$ und $\mathbf{x}_2$ zwei n-dimensionale Vektoren, und es seien λ und μ irgendzwei Skalare. Dann heißt der Vektor $\mathbf{y}$, der durch

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \lambda\mathbf{x}_1 + \mu\mathbf{x}_2 = \begin{pmatrix} \lambda x_{11} + \mu x_{12} \\ \lambda x_{21} + \mu x_{22} \\ \vdots \\ \lambda x_{n1} + \mu x_{n2} \end{pmatrix} \tag{4}$$

   definiert ist, eine *Linearkombination* der Vektoren $\mathbf{x}_1$ und $\mathbf{x}_2$. Damit ist auch erklärt, was unter der Summe und der Differenz zweier Vektoren zu verstehen ist. Sei $\lambda = \mu = 1$. Dann ist $\mathbf{y}$ die Summe der beiden Vektoren, und die Komponenten von $\mathbf{y}$ sind die Summen der Komponenten von $\mathbf{x}_1$ und $\mathbf{x}_2$. Die Differenz erhält man, wenn man $\lambda = 1$ und $\mu = -1$ setzt.

4. **Skalarprodukt** Das *Skalarprodukt* zweier Vektoren $\mathbf{x}$ und $\mathbf{y}$ ist durch

$$\mathbf{x}'\mathbf{y} = (x_1, x_2, \ldots, x_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \sum_{i=1}^{n} x_iy_i \tag{5}$$

definiert. $\mathbf{x}'\mathbf{y}$ ist ein Skalar, also eine "einfache" Zahl (kein Vektor), deswegen der Ausdruck 'Skalarprodukt'. Eine zweite Art von Vektorprodukt, das eine Matrix liefert, wird weiter unten eingeführt (Punkt 14).

5. **Länge und Normierung** Für $\mathbf{x} = \mathbf{y}$ erhält man

$$\mathbf{x}'\mathbf{x} = \sum_{i=1}^{n} x_i^2 = \|\mathbf{x}\|^2 \tag{6}$$

Am Beispiel $n = 2$ sieht man, dass $\|\mathbf{x}\|^2$ das Quadrat der Länge des Vektors $\mathbf{x}$ ist (Satz des Pythagoras). Also ist $\|\mathbf{x}\| = \sqrt{\mathbf{x}'\mathbf{x}}$ die Länge des Vektors.

Es sei $\|\mathbf{x}\| \neq 1$. Dann kann $\mathbf{x}$ *normiert* werden. Dies geschieht durch Multiplikation mit einem geeignet gewählten Skalar λ:

$$\|\lambda\mathbf{x}\| = \lambda\|\mathbf{x}\| = 1.$$

Daraus folgt sofort

$$\lambda = \frac{1}{\|\mathbf{x}\|}. \tag{7}$$

Ein Vektor mit derLänge $\|\mathbf{x}\|$ wird also normiert, indem man seine Komponenten mit $1/\|\mathbf{x}\|$ multipliziert.

6. **Skalarprodukt und Winkel zwischen zwei Vektoren** Es läßt sich zeigen (Anwendung des Kosinussatzes), dass für zwei n-dimensionale Vektoren $\mathbf{x}$ und $\mathbf{y}$, die in einem Winkel θ zueinander liegen,

$$\cos\theta = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{\mathbf{x}'}{\|\mathbf{x}\|}\frac{\mathbf{y}}{\|\mathbf{y}\|} \tag{8}$$

gilt.

Für $\theta = 0$ wird $\cos\theta$ maximal, also $\cos\theta = 1$. Dann folgt

$$\frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = 1 \Rightarrow \mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| \tag{9}$$

Für $\theta \neq 0$ ist $\cos\theta < 1$, so dass

$$\frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} < 1 \Rightarrow \mathbf{x}'\mathbf{y} < \|\mathbf{x}\|\|\mathbf{y}\|. \tag{10}$$

Faßt man (9) und (10) zusammen, hat man

$$\mathbf{x}'\mathbf{y} \leq \|\mathbf{x}\|\|\mathbf{y}\|. \tag{11}$$

Quadriert man beide Seiten, so erhält man

$$(\mathbf{x}'\mathbf{y})^2 \leq \|\mathbf{x}\|^2\|\mathbf{y}\|^2; \tag{12}$$

dies ist die *Cauchy-Scharzsche Ungleichung*, die oft in der Form

$$\left|\sum_{i=1}^n x_i y_i\right|^2 \leq \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 \tag{13}$$

geschrieben wird. Sind $\mathbf{x}$ und $\mathbf{y}$ standardisiert, so ist bekanntlich $\|\mathbf{x}\|^2/n = \|\mathbf{y}\|^2/n = 1$ und $\mathbf{x}'\mathbf{y}/n = r_{xy}$ die Produkt-Momentkorrelation; (12) impliziert dann

$$-1 \leq r_{xy} \leq 1.$$

Für $\theta = \pi/2$ ($= 90^o$) ist $\cos\theta = 0$, dann folgt $\mathbf{x}'\mathbf{y} = 0$. Für den Fall $\theta = \pi/2$ stehen die Vektoren $\mathbf{x}$ und $\mathbf{y}$ senkrecht aufeinander, d.h. sie bilden einen rechten Winkel. Deshalb heißen die beiden Vektoren dann *orthogonal*[1] zueinander.

Es gelte insbesondere $\mathbf{y} = \lambda\mathbf{x}$, λ ein Skalar, d.h. die beiden Vektoren haben die gleiche Orientierung, aber für $\lambda \neq 1$ verschiedene Längen. Aus (8) folgt dann

$$\cos\theta = \frac{\lambda\|\mathbf{x}\|^2}{\lambda\|\mathbf{x}\|^2} = 1.$$

was natürlich $\theta = 0$ bedeutet, und aus (11) folgt dann, dass für $\mathbf{y} = \lambda\mathbf{x}$, also für

$$\lambda x_i = y_i, \quad i = 1, \ldots, n \tag{14}$$

das Skalarprodukt $\mathbf{x}'\mathbf{y}$ den maximal möglichen Wert annimmt, d.h. es gilt dann

$$\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|, \quad \text{wenn } \mathbf{y} = \lambda\mathbf{x} \tag{15}$$

7. **Lineare Abhängigkeit und Unabhängigkeit von Vektoren:** Gegeben seien p n-dimensionale Vektoren $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Es sei $\mathbf{x}_j$ irgendeiner dieser Vektoren, und es gebe Zahlen $\lambda_1, \ldots, \lambda_{j-1}, \lambda_{j+1}, \ldots, \lambda_p$, die *nicht alle* gleich Null seien, derart, dass

$$\mathbf{x}_j = \lambda_1\mathbf{x}_1 + \cdots + \lambda_{j-1}\mathbf{x}_{j-1} + \lambda_{j+1}\mathbf{x}_{j+1} + \cdots + \lambda_p\mathbf{x}_p \tag{16}$$

gilt, d.h. $\mathbf{x}_j$ kann als Linearkombination der übrigen Vektoren dargestellt werden. Dann heißen die $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ *linear abhängig.*

Andererseits sei angenommen, dass derartige Zahlen *nicht* existieren. Dann kann $\mathbf{x}_j$ *nicht* als Linearkombination der übrigen Vektoren dargestellt werden. Dann heißen die $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ *linear unabhängig.*

Man kann in (16) $\mathbf{x}_j$ auf beiden Seiten subtrahieren und erhält dann, mit $\lambda_j = -1$, eine Darstellung des Nullvektors, also eines Vektors, dessen Komponenten alle gleich Null sind:

$$\vec{0} = \lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + \cdots + \lambda_p\mathbf{x}_p. \tag{17}$$

Wenn die $\mathbf{x}_1, \ldots, \mathbf{x}_p$ linear abhängig sind, gilt (16) mit Zahlen λ_k, die nicht alle gleich Null sind, und dann gilt auch (17) mit diesen Zahlen, d.h. der Nullvektor kann dargestellt werden, ohne dass alle λ_k ($k = 1, \ldots, p$) gleich Null sein müssen. Sind dagegen die $\mathbf{x}_1, \ldots, \mathbf{x}_p$ linear unabhängig, so gilt diese Aussage *nicht*, d.h. der Nullvektor kann in der Form (17) nur dargestellt werden, wenn

$$\lambda_1 = \lambda_2 = \cdots = \lambda_p = 0 \tag{18}$$

[1] von griechisch orthos ($\acute{o}\rho\theta\acute{o}\zeta$) = richtig, recht-, und gonia ($\gamma\omega\nu\iota\alpha$) = Ecke, Winkel

gilt. Schreibt man die Gleichungen (16) und (17) aus, d.h. für jede der Vektorkomponenten an, so sieht man, dass die Gleichungen lineare Gleichungssysteme darstellen. Für (17) erhält man

$$\begin{array}{c} 0 = \lambda_1 x_{11} + \lambda_2 x_{12} + \cdots + \lambda_p x_{1p} \\ 0 = \lambda_1 x_{21} + \lambda_2 x_{22} + \cdots + \lambda_p x_{2p} \\ \vdots \\ 0 = \lambda_1 x_{n1} + \lambda_2 x_{n2} + \cdots + \lambda_p x_{np} \end{array} . \tag{19}$$

Betrachtet man darin die λ_k als Unbekannte, so sieht man, dass die lineare Unabhängigkeit der $\mathbf{x}_k$ nur eine Lösung zulässt, nämlich (18). Im Falle der linearen Abhängigkeit existiert mindestens eine Lösung, bei der nicht alle λ_k gleich Null sind.

8. **Lineare Unabhängigkeit und Orthogonalität:** Sind Vektoren paarweise orthogonal, so sind sie auch linear unabhängig.

   Sind Vektoren linear unabhängig, so sind sie *nicht* notwendig auch orthogonal.

   Beweis der ersten Behauptung: Die n-dimensionalen Vektoren $\mathbf{x}_1, \ldots, \mathbf{x}_p$ seien paarweise orthogonal, d.h. es gelte

$$\mathbf{x}_j' \mathbf{x}_k = 0 \text{ für } j \neq k.$$

   Dann sind die $\mathbf{x}_1, \ldots, \mathbf{x}_p$ auch linear unabhängig. Denn es gelte gemäß (17) die Gleichung

$$\vec{0} = \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \cdots + \lambda_p \mathbf{x}_p.$$

   Um zu zeigen, dass die Vektoren linear unabhängig sind, muß man zeigen, dass die λ_k alle gleich Null sein müssen. Man wähle einen der Vektoren, etwa $\mathbf{x}_j$, und multiplizieren die Gleichung mit $\mathbf{x}_j'$:

$$\mathbf{x}_j' \vec{0} = \lambda_1 \mathbf{x}_j' \mathbf{x}_1 + \lambda_2 \mathbf{x}_j' \mathbf{x}_2 + \cdots + \lambda_p \mathbf{x}_j' \mathbf{x}_p.$$

   Wegen der vorausgesetzten paarweisen Orthogonalität verschwinden auf der rechten Seite alle Skalarprodukte, für die $j \neq k$, und nur für $j = k$ ist $\mathbf{x}_j' \mathbf{x}_j) \|\mathbf{x}_j\|^2 \neq 0$ (es wird vorausgesetzt, dass keiner der Vektoren der Nullvektor ist, also $\mathbf{x}_j \neq \vec{0}$ für alle j). Dann hat man

$$0 = \lambda_j \|\mathbf{x}_j\|^2,$$

   und wegen $\|\mathbf{x}_j\| \neq 0$ folgt $\lambda_j = 0$. Dies gilt für alle j, so dass alle λ_j gleich Null sind, und damit sind die $\mathbf{x}_1, \ldots, \mathbf{x}_p$ linear unabhängig.

9. **Lineare Abhängigkeit und Skalarprodukt:** Es seien $\mathbf{x}$ und $\mathbf{y}$ zwei n-dimensionale Vektoren mit dem Skalarprodukt $\mathbf{x}'\mathbf{y}$. Nach (11) gilt

$$\mathbf{x}'\mathbf{y} \leq \|\mathbf{x}\| \|\mathbf{y}\|.$$

   Nun seien $\mathbf{x}$ und $\mathbf{y}$ linear abhängig, d.h. es möge

$$\lambda_1 \mathbf{x} + \lambda_2 \mathbf{y} = \vec{0}, \quad \lambda_1, \lambda_2 \neq 0$$

   gelten. Dann folgt

$$\mathbf{y} = -\frac{\lambda_2}{\lambda_1} \mathbf{x} = \lambda \mathbf{x}, \quad \lambda = -\lambda_2/\lambda_1.$$

Dies ist aber der in (15) betrachtete Fall, d.h. es folgt $\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|$, d.h. lineare Abhängigkeit impliziert, dass das Skalarprodukt den Maximalwert $\|\mathbf{x}\|\|\mathbf{y}\|$ annimmt. Dann impliziert der Fall $\mathbf{x}'\mathbf{y} < \|\mathbf{x}\|\|\mathbf{y}\|$, dass $\mathbf{x}$ und $\mathbf{y}$ linear unabhängig sind. Umgekehrt folgt aus der linearen Unabhängigkeit, dass diese Ungleichung gilt, denn dann sind die Vektoren nicht parallel und es gilt $\cos\theta < 1$, woraus $\mathbf{x}'\mathbf{y} < \|\mathbf{x}\|\|\mathbf{y}\|$ folgt (vergl. (10)).

10. **Vektorräume und Teilräume:** Eine Menge V von Vektoren heißt *Vektorraum*, wenn alle Linearkombinationen von Vektoren aus V wieder Elemente von V sind. Ist $V_0 \subset V$ eine Teilmenge von Vektoren aus V und gilt, dass alle Linearkombinationen von Vektoren aus V_0 wieder Elemente von V_0 sind, so heißt V_0 Teilvektorraum. Etwas genauer wird der Begriff des Vektorraums durch die folgenden Forderungen gefaßt: V ist ein Vektorraum, wenn die Bedingungen

    (i) Ist $\mathbf{v} \in V$, so auch $-\mathbf{v} \in V$

    (ii) $\lambda, \mu \in \mathbb{R}$, $\mathbf{v}, \mathbf{w} \in V$, $(\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$, $\lambda(\mathbf{v} + \mathbf{w}) = \lambda\mathbf{v} + \mu\mathbf{w}$

    (iii) $\lambda(\mu\mathbf{v}) = (\lambda\mu)\mathbf{v}$

    erfüllt sind. Die Definition eines Teilraums ist analog. Sind die Vektoren aus V n-dimensionale Vektoren, so schreibt man auch V_n für V.

    Anmerkung: Ein Teilraum ist nicht einfach eine Teilmenge von Vektoren aus einem Vektorraum. So betrachte man die Menge der 2-dimensionalen Vektoren $\mathbf{x} = (x_1, x_2)'$ mit $x_1^2 + x_2^2 = 1$ (das sind Vektoren, deren Endpunkte auf dem Kreis mit dem Radius 1 liegen, wenn die Anfangspunkte in den Koordinatenursprung gelegt werden). Die Menge dieser Vektoren ist kein Vektorraum, weil die beliebige Linearkombination zweier Vektoren nicht mehr Element dieser Menge ist. $\square$

    Zur Veranschaulichung: Es seien $\mathbf{v}, \mathbf{w}$ irgendzwei 2-dimensionale Vektoren. Dann sind alle Linearkombinationen $\mathbf{u} = \lambda\mathbf{v} + \mu\mathbf{w}$ ebenfalls 2-dimensionale Vektoren, egal, welche Werte man für die Skalare λ und μ wählt, – durch die Addition der Vektoren kann ja keine dritte Komponente entstehen, d.h. die Bildung von Linearkombinationen führt nicht aus der Menge aller möglichen 2-dimensionalen Vektoren heraus, die Menge dieser Vektoren bildet einen V_2. Für den Spezialfall $\mathbf{v} = \mathbf{w}$ allerdings entsteht ein Vektor $\mathbf{u}$, der die gleiche Orientierung wie $\mathbf{v}$ und $\mathbf{w}$ hat. In diesem Fall liegen alle $\mathbf{u}$ auf einer Geraden im 2-dimensionalen Vektorraum. So lange nur Vektoren $\mathbf{v}$ und $\mathbf{w}$ aus dieser Geraden gewählt werden, kann $\mathbf{u}$ nie eine andere Orientierung als eben die $\mathbf{v}$ und $\mathbf{w}$ haben, die Gerade ist also ein Teilraum V_1 des 2-dimensionalen Vektorraums V_2. Da es beliebig viele Orientierungen für eine Gerade im V_2 gibt, gibt es beliebig viele 1-dimensionale Teilräume des V_2.

    Es sei V_3 ein 3-dimensionaler Vektorraum, seine Elemente sind 3-dimensionale Vektoren. Jede Linearkombination von 3-dimensionalen Vektoren ist wieder ein 3-dimensionaler Vektor, denn durch die Bildung einer Linearkombination gehen keine Komponenten verloren noch kommen neue hinzu. Nun werden 2 3-dimensionale Vektoren $\mathbf{v}$ und $\mathbf{w}$ aus V_3 gewählt, die nicht parallel seien, die also verschiedene orientierungen im V_3 haben mögen. Diese beiden Vektoren definieren eine Ebene im V_3. Denn es sei $\mathbf{n} = (n_1, n_2, n_3)'$ ein 3-dimensionaler Vektor, der senkrecht auf den beiden Vektoren $\mathbf{v}$ und $\mathbf{w}$ steht; dann steht $\mathbf{n}$ auf auf jeder Linearkombination von $\mathbf{v}$ und $\mathbf{w}$ senkrecht: ist

    $$\mathbf{u} = \lambda\mathbf{v} + \mu\mathbf{w}$$

irgendeine Linearkombination von $\mathbf{v}$ und $\mathbf{w}$, und bildet man nun das Skalarprodukt von $\mathbf{n}$ und $\mathbf{u}$, so folgt, dass es gleich Null ist:

$$\mathbf{n}'\mathbf{u} = \lambda\mathbf{n}'\mathbf{v} + \mu\mathbf{n}'\mathbf{w} = 0,$$

denn $\mathbf{n}'\mathbf{v} = \mathbf{n}'\mathbf{w} = 0$ nach der Orthogonalitätsvoraussetzung. $\mathbf{n}$ steht also senkrecht auf der durch die Vektoren $\mathbf{v}$ und $\mathbf{w}$ definierten Ebene und gibt damit die Orientierung der Ebene im V_3 an. Die Ebene ist ein Teilraum $V_2 \subset V_3$, da eben jede Linearkombination von $\mathbf{v}$ und $\mathbf{w}$ in dieser Ebene liegt, wie eben gezeigt wurde. Haben $\mathbf{v}$ und $\mathbf{w}$ die gleiche Orientierung, generieren sie einen 1-dimensionalen Teilraum $V_1 \subset V_3$. Damit also ein 2-dimensionaler Teilraum erzeugt werden kann, müssen die beiden Vektoren $\mathbf{v}$ und $\mathbf{w}$ verschiedene Orientierungen haben, und das heißt, dass sie linear unabhängig (l.u.) sein müssen, denn genau dann kann man etwa $\mathbf{w}$ nicht als Linearkombination von $\mathbf{v}$ darstellen: es gibt kein λ derart, dass $\lambda\mathbf{v}$ eine andere Orientierung als $\mathbf{v}$ hat. Ebenso müssen die drei 3-dimensionalen Vektoren $\mathbf{u}$, $\mathbf{v}$ und $\mathbf{w}$ isngesamt linear unabhängig sein, damit man *alle* Vektoren des V_3 aus ihnen durch geeignete linearkombination darstellen kann. Denn angenommen, es sei $\mathbf{u}$ eine Linearkombination von $\mathbf{v}$ und $\mathbf{w}$. Dann liegt $\mathbf{u}$ – wie eben gezeigt – notwendig in der durch $\mathbf{v}$ und $\mathbf{w}$ definierten Ebene und damit auch jeder Vektor

$$\mathbf{z} = \lambda_1\mathbf{u} + \lambda_2\mathbf{v} + \lambda_3\mathbf{w}$$

Setzt man für $\mathbf{u}$ den Ausdruck $\lambda\mathbf{v} + \mu\mathbf{w}$ ein, ergibt sich

$$\mathbf{z} = \lambda_1(\lambda\mathbf{v} + \mu\mathbf{w}) + \lambda_2\mathbf{v} + \lambda_3\mathbf{w} = (\lambda_1\lambda + \lambda_2)\mathbf{v} + (\lambda_1\mu + \lambda_3)\mathbf{w},$$

und also ist $\mathbf{z}$ wieder eine Linearkombination von $\mathbf{v}$ und $\mathbf{w}$ und liegt damit in der durch die beiden Vektoren aufgespannten Ebene V_2.

Drei l.u. 3-dimensionale Vektoren aus V_3 bilden ein *Erzeugendensystem* oder auch *Basis* des V_3, weil man eben alle Vektoren des V_3 mit ihnen durch Linearkombination generieren kann. Zwei l.u. 3-dimensionale Vektoren aus V_3 bilden eine *Teilbasis* des V_3, die einen Teilraum V_2 generieren. Ein 3-dimensionaler Vektor bildet eine Basis für einen 1-dimensionalen Teilraum des V_3. Da für eine Basis oder Teilbasis stets nur die lineare Unabhängigkeit der Vektoren der Basis gefordert wird, gibt es beliebig viele Basen bzw. Teilbasen des V_3.

Allgemein seien $\mathbf{v}_1, \ldots, \mathbf{v}_r$ r linear unabhängige n-dimensionale Vektoren. Für $r = n$ bilden sie eine Basis des V_n, für $r < n$ bilden sie eine Teilbasis eines Teilraums $V_r \subset V_n$.

11. **Matrizen** Matrizen entstehen, wenn man n jeweils m-dimensionale Vektoren nebeneinander schreibt, etwa

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ & & \vdots & \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}, \tag{20}$$

oder indem man m n-dimensionale Vektoren als Zeilen untereinander schreibt. X heißt auch $(m \times n)$-Matrix. Für den Fall $m = n$ heißt X *quadratisch*. Die Zeilen einer Matrix heißen auch *Zeilenvektoren*, die Spalten *Spaltenvektoren*.

Auch eine Matrix läßt sich "stürzen": man schreibt dann X'. Die Zeilenvektoren von X' sind die Spaltenvektoren von X, und die Spaltenvektoren von X' sind die Zeilenvektoren von X.

12. **Multiplikation mit einem Vektor** Es sei X eine $(m \times n)$-Matrix und $\mathbf{u} = (u_1, \ldots, u_n)'$ ein n-dimensionaler Vektor. Es werden die Skalarprodukte der *Zeilenvektoren* von X mit $\mathbf{u}$ gebildet. Dann entsteht ein neuer (Spalten-)Vektor $\mathbf{v}$, dessen Komponenten eben die genannten Skalarprodukte sind. Man schreibt

$$\mathbf{y} = X\mathbf{u} = \begin{pmatrix} \sum_{j=1}^{n} x_{1j}u_j \\ \sum_{j=1}^{n} x_{2j}u_j \\ \vdots \\ \sum_{j=1}^{n} x_{nj}u_j \end{pmatrix}. \tag{21}$$

Sind $\mathbf{x}_1, \ldots, \mathbf{x}_n$ die m-dimensionalen Spaltenvektoren von X, so sieht man leicht, dass (21) einen Spaltenvektor $\mathbf{y}$ liefert, der sich als Linearkombination der $\mathbf{x}_j$ ergibt:

$$\mathbf{y} = X\mathbf{u} = \sum_{j=1}^{n} u_j \mathbf{x}_j, \tag{22}$$

u_j die Komponenten von $\mathbf{u}$. Eine Matrix, multipliziert mit einem (Spalten-)Vektor *von rechts*, liefert wieder einen Spaltenvektor.

Nun sei $\mathbf{u}$ ein m-dimensionaler Vektor, und es werden die Skalarprodukte des Zeilenvektors $\mathbf{u}$ mit den Spaltenvektoren von X gebildet. Dann ist

$$\mathbf{z}' = \mathbf{u}'X = \left(\sum_{i=1}^{m} u_i x_{i1}, \sum_{i=1}^{m} x_{i2}u_i, \ldots, \sum_{i=1}^{m} x_{in}u_i \right). \tag{23}$$

Ein Zeilenvektor, multipliziert *von links* mit einer Matrix, liefert wieder einen Zeilenvektor ($\mathbf{z}'$). Der Vektor $\mathbf{z}$ ergibt sich als Linearkombination der *Zeilenvektoren* von X. $\mathbf{z}$ ergibt sich, wenn man die rechte Seite von (23) als Spalte anschreibt:

$$\mathbf{z} = \sum_{i=1}^{m} u_i \mathbf{x}^{(i)}, \tag{24}$$

wenn $\mathbf{x}^{(i)}$ der i-te Zeilenvektor von X ist. u_i ist die i-te Komponente von $\mathbf{u}$.

13. **Multiplikation von Matrizen** Es sei A eine $(m \times n)$-Matrix und B sei eine $(n \times r)$-Matrix. Dann kann man das *Matrixprodukt* AB bilden. Es entsteht wieder eine Matrix C, deren Elemente die Skalarprodukte der Zeilenvektoren von A mit den Spaltenvektoren von B sind:

$$\begin{aligned} AB &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ & & \vdots & \\ b_{n1} & b_{n2} & \cdots & b_{nr} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^{n} a_{1j}b_{j1} & \sum_{j=1}^{n} a_{1j}b_{j2} & \cdots & \sum_{j=1}^{n} a_{1j}b_{jr} \\ & & \vdots & \\ \sum_{j=1}^{n} a_{mj}b_{j1} & \sum_{j=1}^{n} a_{mj}b_{j2} & \cdots & \sum_{j=1}^{n} a_{mj}b_{jr} \end{pmatrix} = C \end{aligned} \tag{25}$$

Notwendige Voraussetzung für die Berechnung eines Produkts zweier Matrizen A und B ist, dass die Anzahl der Spalten von A gleich der Anzahl von Zeilen von B ist.

Die Matrixmultiplikation ist *nicht kommutativ*, d.h. es gilt im Allgemeinen

$$AB \neq BA \tag{26}$$

(falls BA überhaupt berechnet werden kann, – dazu muß ja die Anzahl der Spalten von B auch gleich der Anzahl der Zeilen von A sein).

Weiter gilt

$$(AB)' = B'A'. \tag{27}$$

Dies zeigt man durch Nachrechnen.

Die folgende Anmerkung erweist sich als nützlich bei Anwendungen der Matrixrechnung bei Methoden wie der multiplen Regression, der Faktorenanalyse etc.

Gilt wie in (26) $AB = C$, so sind die Spaltenvektoren von C Linearkombinationen der Spaltenvektoren von A. So seien $\mathbf{a}_1, \ldots, \mathbf{a}_n$ die m-dimensionalen Spaltenvektoren von A, $\mathbf{b}_1, \ldots, \mathbf{b}_r$ seien die n-dimensionalen Spaltenvektoren von B und $\mathbf{c}_1, \ldots, \mathbf{c}_r$ seien die m-dimensionalen Spaltenvektoren von C. Aus der letzten Zeile von Gleichung (26) läßt sich direkt ablesen, dass

$$\mathbf{c}_1 = \sum_{j=1}^{n} b_{j1}\mathbf{a}_j,\ \mathbf{c}_2 = \sum_{j=1}^{n} b_{j2}\mathbf{a}_j,\ \ldots,\ \mathbf{c}_r = \sum_{j=1}^{n} b_{jr}\mathbf{a}_j,$$

oder allgemein

$$\mathbf{c}_k = \sum_{j=1}^{n} b_{jk}\mathbf{a}_j, \quad k = 1, \ldots, r \tag{28}$$

Hierin ist b_{jk} die j-te Komponente von $\mathbf{b}_k$. Da wegen (27) $C' = B'A'$ folgt aus dem vorangegangenen sofort, dass die Spaltenvektoren von C' Linearkombinationen der Spaltenvektoren von B' sind. Dies heißt aber nichts anderes als dass die *Zeilenvektoren* von C sich als Linearkombinationen der *Zeilenvektoren* von B ergeben.

14. **Das Produkt $\mathbf{xy}'$** Es seien $\mathbf{x}$ und $\mathbf{y}$ zwei Vektoren. Das Produkt $\mathbf{xy}'$ ist definiert als die Matrix

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} (y_1, y_2, \ldots, y_m) = \mathbf{xy}' = \begin{pmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_m \\ x_2y_1 & x_2y_2 & \cdots & x_2y_m \\ & & \vdots & \\ x_ny_1 & x_ny_2 & \cdots & x_ny_m \end{pmatrix} \tag{29}$$

Diese Regel ergibt sich aus der für die Matrixmultiplikation, wenn man zu $\mathbf{x}$ eine Anzahl von Nullvektoren $\vec{0}$ hinzufügt und zu $\mathbf{y}'$ eine entsprechende Zahl von Zeilenvektoren $\vec{0}'$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} (y_1, y_2, y_3) = \begin{pmatrix} x_1 & 0 \\ x_2 & 0 \end{pmatrix} \begin{pmatrix} y_1 & y_2 & y_3 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} x_1y_1 & x_1y_2 & x_1y_3 \\ x_2y_1 & x_2y_2 & x_2y_3 \end{pmatrix}$$

15. **Einheitsmatrix** Die Einheitsmatrix enthält nur Nullen, bis auf die Zahlen in der Diagonalen der Matrix, die alle gleich 1 sind:

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{30}$$

I (von Identität) spielt die Rolle der 1 beim Rechnen mit Skalaren. Demnach ist $AI = A$, $IA = A$.

16. **Inverse Matrix** Es sei A eine Matrix und B eine Matrix derart, dass $AB = I$. Dann heißt B die zu A *inverse* Matrix. Man schreibt dann $B = A^{-1}$. Notwendige Voraussetzung für die Existenz einer zu A inversen Matrix A^{-1} ist, dass A quadratisch ist. Die inverse Matrix existiert *nicht notwendig* für eine gegebene quadratische Matrix A. Existiert sie nicht, so heißt A *singulär.*

    Sind A und B Matrizen mit existierenden Inversen A^{-1} und B^{-1}, so gilt

    $$(AB)^{-1} = B^{-1}A^{-1}. \tag{31}$$

    Natürlich gilt

    $$A^{-1}A = AA^{-1} = I, \tag{32}$$

    I die in (30) eingeführte Einheitsmatrix.

17. **Kreuzprodukte** Es sei X eine $(m \times n)$-Matrix. Man möchte alle Skalarprodukte zwischen den Spaltenvektoren von X berechnen. Dazu berechnet man die Matrix

    $$U = X'X. \tag{33}$$

    Man möchte die Skalarprodukte zwischen allen Zeilenvektoren von X berechnen. Dann berechnet man die Matrix

    $$V = XX'. \tag{34}$$

    Die Matrizen $X'X$ und XX' sind Beispiele für *symmetrische* Matrizen. Dies folgt aus der Definition des Skalarprodukts: für irgendzwei n-dimensionale Vektoren $\mathbf{x}$ und $\mathbf{y}$ gilt ja $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$.

    Spezialfall: Es sei Z die Matrix der (spalten-)standardisierten Messwerte x_{ij}, d.h. sei $z_{ij} = (x_{ij} - \bar{x}_j)/s_j$. Dann ist die Matrix Z der Korrelationen zwischen den Variablen, die die Spalten von X bzw. Z definieren, durch das Kreuzprodukt

    $$R = \frac{1}{m} Z'Z \tag{35}$$

    gegeben. Der Faktor $1/m$ bedeutet, dass jedes Element von $Z'Z$ mit ihm multipliziert werden soll.

18. **Orthogonale Matrizen** Es sei X eine beliebige $(m \times n)$-Matrix, und die Spaltenvektoren seien paarweise orthogonal. Dann gilt

    $$X'X = D = \begin{pmatrix} \|\mathbf{x}_1\|^2 & 0 & \cdots & 0 \\ 0 & \|\mathbf{x}_2\|^2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \|\mathbf{x}_n\|^2 \end{pmatrix} \tag{36}$$

    D ist eine *Diagonalmatrix.* Sie resultiert, wenn die Spaltenvektoren von X paarweise orthogonal sind, denn die Skalarprodukte verschiedener Vektoren müssen dann gleich Null sein. In den Diagonalzellen stehen die quadrierten Längen der Spaltenvektoren.

19. **Eigenvektoren und Eigenwerte einer quadratischen Matrix** Es sei M eine $(n \times n)$-Matrix und $\mathbf{x}$ ein n-dimensionaler Vektor. Dann ist

    $$M\mathbf{x} = \mathbf{y}, \tag{37}$$

und $\mathbf{y}$ ist wieder ein n-dimensionaler Vektor. Im Allgemeinen haben $\mathbf{x}$ und $\mathbf{y}$ verschiedene Längen und verschiedene Orientierungen. Es gelte nun der Spezialfall

$$M\mathbf{x} = \mathbf{y} = \lambda\mathbf{x}, \quad \lambda \in \mathbb{R}, \tag{38}$$

d.h. der Vektor $\mathbf{y}$ habe die gleiche Orientierung wie $\mathbf{x}$ und unterscheide sich nur in der Länge, vorausgesetzt $\lambda \neq 1$. Dann heißt $\mathbf{x}$ *Eigenvektor* von M und λ ist der zugehörige Eigenwert.

M sei $(n \times n)$ und symmetrisch. Dann gelten die folgenden Aussagen (hier ohne Beweis, vergl. Skriptum 'Vektoren und Matrizen'):

(a) Es gibt maximal n verschiedene Eigenwerte λ_j, $j = 1, \ldots, n$, $\lambda_j \geq 0$. (M ist *positiv-semidefinit.*)

(b) Alle Eigenvektoren können als auf die Länge 1 normiert betrachtet werden,

(c) Je zwei verschiedene Eigenvektoren sind orthogonal; wegen der Normiertheit sind sie *orthonormal*.

(d) Die $(n \times n)$-Matrix M sei symmetrisch und P sei die $(n \times n)$-Matrix, deren Spalten die Eigenvektoren von M sind, und $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ sei die Diagonalmatrix der Eigenwerte. Dann gelten die Gleichungen

$$\begin{aligned} P' &= P^{-1} && (39)\\ MP &= P\Lambda && (40)\\ M &= P\Lambda P' && (41)\\ P'MP &= \Lambda && (42) \end{aligned}$$

(39) drückt aus, dass P *orthonormal* ist. Die Gleichungen (40) bis (42) folgen dann aus (39).

Es sei $\mathbf{P}_j$ der j-te Spaltenvektor (= Eigenvektor) in der Matrix P. Unter Berücksichtigung der Vektormultiplikation (29) läßt sich (41) in der Form

$$M = PDP' = \sum_{j=1}^{n} \lambda_j \mathbf{P}_j \mathbf{P}'_j \tag{43}$$

darstellen. Die zu M inverse Matrix M^{-1} ist durch

$$M^{-1} = PD^{-1}P' = \sum_{j=1}^{n} \frac{\mathbf{P}_j \mathbf{P}'_j}{\lambda_j} \tag{44}$$

gegeben. Dies folgt sofort aus (31) und der Tatsache, dass P orthonormal ist, so dass $P' = P^{-1}$, und (29), denn dann ist

$$M^{-1} = (PDP')^{-1} = (P')^{-1}D^{-1}P^{-1} = PD^{-1}P',$$

da ja $P' = P^{-1}$ und $(P^{-1})^{-1} = P$. Von dieser Darstellung wird bei der Diskussion von Eigenschaften von Parameterschätzungen Gebrauch gemacht.

20. **Rang einer Matrix** Es sei X eine $(m \times n)$-Matrix und $m > n$. Läßt sich keiner der Spaltenvektoren von X als Linearkombination der übrigen darstellen, so ist der Rang r von X gleich n, $r = n$, andernfalls ist $r < n$.

    Alle Vektoren lassen sich als Linearkombination irgendwelcher anderer Vektoren darstellen. Es seien $\mathbf{L}_1, \ldots, \mathbf{L}_r$ linear unabhängige, m-dimensionale Vektoren, d.h.

keiner dieser Vektoren lasse sich als Linearkombination der übrigen darstellen; die $\mathbf{L}_1, \ldots, \mathbf{L}_r$ bilden eine Teilbasis des V_m, wenn $r < m$. Es sei $r \leq n$. Lassen sich alle Spaltenvektoren von X als Linearkombination dieser r Basisvektoren darstellen, so ist der Rang von X gleich r. Dann lassen sich auch alle m n-dimensionalen Zeilenvektoren durch r linear unabhängige, n-dimensionale Basisvektoren $\tilde{\mathbf{L}}, \ldots, \tilde{\mathbf{L}}_r$ darstellen (Zeilenrang gleich Spaltenrang, vergl. unten die Interpretation der Singularwertzerlegung).

Hat X den Rang r, so hat auch $M = X'X$ den Rang r. Es gibt dann genau r von Null verschiedene Eigenwerte von M. Der Rang von M gibt also an, wieviele "latente" Vektoren $\mathbf{L}_1, \ldots, \mathbf{L}_r$ man benötigt, um alle Spaltenvektoren von X als Linearkombination dieser Vektoren darzustellen.

(Die Darstellung der Spaltenvektoren einer Datenmatrix durch eine möglichst kleine Anzahl solcher latenter Vektoren ist das Ziel der Faktorenanalyse!)

21. **Darstellung durch latente Vektoren** Gegeben sei eine $(m \times n)$-(Daten-)Matrix X und die Spaltenvektoren sollen durch $r \leq n$ linear unabhängige Vektoren $\mathbf{L}_j$, $j = 1, \ldots, r$ dargestellt werden. Gesucht ist demnach eine Matrix L mit den Vektoren $\mathbf{L}_j$, $j = 1, \ldots, r \leq n$ als Spaltenvektoren, die linear unabhängig sind.

    Dazu werde angenommen, dass $\mathbf{L}_j$ paarweise orthogonal sind, – dann sind sie ja auch linear unabhängig. Da nur X gegeben ist, müssen die $\mathbf{L}_j$ auf der Basis von X berechnet werden, d.h. man muß eine Transformationsmatrix P finden derart, dass

    $$XP = L \tag{45}$$

    gilt. Wenn L als orthogonal vorausgesetzt wird, muß $L'L = D$ eine Diagonalmatrix sein, also

    $$L'L = D = P'X'XP. \tag{46}$$

    Daraus folgt, dass die gesuchte Matrix P gerade die orthonormalen Eigenvektoren von $X'X$ enthält, und D die zugehörigen Eigenwerte.

22. **Singularwertzerlegung (SVD)**[2] Da P orthonormal, folgt aus (45)

    $$X = LP'. \tag{47}$$

    $L'L = D$ bedeutet $\mathbf{x}_j'\mathbf{x}_j = \|\mathbf{L}_j\|^2 = d_j$, d.h. das Element in der j-ten Diagonalzelle von D ist gerade gleich dem Quadrat der Länge von $\mathbf{L}_j$. Man normiert $\mathbf{L}_j$ auf die Länge 1, indem man die Komponenten von $\mathbf{L}_j$ mit $1/\sqrt{d_j} = 1/\|\mathbf{L}_j\|$ multipliziert. In Matrixform geschrieben bedeutet dies, dass die Spaltenvektoren von

    $$Q = LD^{-1/2}, \tag{48}$$

    orthonormal sind. Die $\sqrt{d_j}$, also die Diagonalelemente von $D^{1/2}$, sind die Wurzeln aus den Eigenwerten λ_j von $X'X$; die $\sqrt{\lambda_j}$ heißen auch *Singularwerte*. Dann kann (47) in der Form

    $$X = QD^{1/2}P' \quad \text{(Singularwertzerlegung)} \tag{49}$$

    geschrieben werden. Weiter folgt

    $$XX' = QD^{1/2}P'PD^{1/2}Q' = QDQ', \tag{50}$$

    d.h. Q enthält die Eigenvektoren der Matrix XX', und die von Null verschiedenen Eigenwerte von XX' sind identisch mit denen von $X'X$.

    Interpretation

[2] Singular Value Decomposition

(a) Nach (47 sind die Spaltenvektoren von X Linearkombinationen der Spaltenvektoren von L bzw. Q, wenn von (49) ausgegangen wird. die i-te Komponente des j-ten Spaltenvektors von X enthält die Messung des i-ten "Falles" (Person, Objekt) für die j-Variable. Die i-te Komponente von $\mathbf{L}_k$ repräsentiert dann den i-ten Fall auf der k-ten latenten Variablen (analog für die Spaltenvektoren von Q). Wegen $X' = PL'$ sind die Spaltenvektoren von X' – also die Zeilenvektoren von X – Linearkombinationen der Spalten von P. p_{jk} repräsentiert dann die Ausprägung der k-ten latenten Variablen bei der j-ten gemessenen Variablen.

(b) Je nach Fokus – entweder auf die Fälle oder die Variablen – werden diese Koordinaten aber mit den Wurzeln aus den Eigenwerten, also mit $D^{1/2}$, skaliert, so dass man als "Fall'koordinaten $L = QD^{1/2}$ bzw als Variablenkoordinaten $A = PD^{1/2}$ (dies sind dann die "Ladungen" der Variablen) betrachtet.

(c) Alle n Spaltenvektoren von X werden als Linearkombination von r l. u. (wegen der Orthogonalität) Spaltenvektoren, etwa von L, dargestellt, also ist der Spaltenrang von X gleich r. Alle m Zeilenvektoren von X werden als Linearkombination von r l.u. (wegen der Orthogonalität) Spaltenvektoren von P dargestellt, – also ist der Zeilenrang von X gleich r. Also ist der Zeilenrang gleich dem Spaltenrang, d.h. es gibt nur einen Rang r.

Von der folgenden Eigenschaften der SVD wird oft Gebrauch gemacht:

(a) **Zentrierung** Die Spalten von X seinen *zentriert*, d.h. es gelte $x_{ij} = X_{ij} - \bar{x}_j$, $\bar{x}_j$ der Mittelwert der j-ten Spalte. Dann folgt $\sum_i x_{ij} = 0$ für alle j. Nun folgt $L = XP$ wegen der Orthonormalität von P, und deshalb

$$L_{ij} = \sum_k x_{ik} p_{jk} \Rightarrow \sum_i L_{ij} = \sum_k p_{jk} \underbrace{\sum_i x_{ik}}_{=0} = 0 \tag{51}$$

für alle k, so dass auch $\sum_i L_{ik} = 0$ für alle Spalten $k = 1, \ldots, r$ von L, d.h. auch die Spalten von L sind zentriert (d.h. sie haben den Mittelwert gleich Null).

(b) **Bedeutung eines Eigenwertes:** Es sei

$$\mathbf{L}_j = (L_{1j}, L_{2j}, \ldots, L_{mj})'$$

der j-te Spaltenvektor von L. Da $\sum_i L_{ij} = 0$, kann

$$\|\mathbf{L}_j\|^2 = \lambda_j \tag{52}$$

als Varianz der Komponenten von $\mathbf{L}_j$ auf gefasst werden, d.h. der Eigenwert λ_j ist gleich der Varianz der Komponenten von $\mathbf{L}_j$.

(c) **Varianz einer Variablen:** Es ist

$$s_j^2 = \frac{1}{m} \sum_{i=1}^{m} x_{ij}^2 = \frac{1}{m} \mathbf{x}_j' \mathbf{x}_j = \frac{1}{m} \|\mathbf{x}_j\|^2$$

die (Stichproben-)Varianz der j-ten gemessenen Variablen (der Komponenten der j-ten Spalte von X). Andererseits gilt für den j-ten Spaltenvektor $\mathbf{x}_j$ von X

$$\mathbf{x}_j = L\mathbf{P}_j,$$

$\mathbf{P}_j$ der j-te Spaltenvektor von P'. Es sei wieder $A = PD^{1/2}$, so dass $a_{jk} = p_{jk}\sqrt{\lambda_j}$ die "Ladung" der j-ten Variablen auf der k-ten latenten Dimension ist. Dann folgt

$$s_j^2 = \frac{1}{m}\mathbf{P}'_j L' L \mathbf{P}_j = \frac{1}{m}\mathbf{P}'_j D\mathbf{P}_j = \frac{1}{m}\sum_{k=1}^n \lambda_k p_{kj}^2 = \frac{1}{m}\sum_{k=1}^n a_{jk}^2 \qquad (53)$$

Hat man die Messungen standardisiert, d.h. ist man von den X_{ij} zu den $z_{ij} = (X_{ij} - \bar{x}_j)/s_j$ übergegangen, so wird s_j durch die Varianz der z_{ij}-Werte in einer Spalte ersetzt, – und die ist gleich 1, so dass

$$1 = \frac{1}{m}\sum_{k=1}^n a_{jk}^2 \qquad (54)$$

folgt. Die Variablen werden dann durch Punkte repräsentiert, die auf der Oberfläche einer n-dimensionalen Hyperkugel liegen. Genügen drei latente Dimensionen, liegen sie auf einer Kugel, genügen zwei Dimensionen, liegen alle Variablen auf einem Kreis.

(d) **Matrixapproximation** X sei eine $(m \times n)$-Matrix, $m > n$. Enthält X fehlerbehaftete Messwerte, so hat X üblicherweise den Rang n. Berücksichtigt man in $X = QD^{1/2}P'$ nur die erste $r < n$ Spaltenvektoren in Q, die dazu korrespondierenden r Eigenwerte und die ersten r Spaltenvektoren in P, so erhält man eine Approximation $\hat{X}_r = Q_r D_r^{1/2} P'_r$. Es läßt sich zeigen, dass für gegebenen Wert von r die Matrix $\hat{X}_r$ die beste Approximation für X im Sinne des Kriteriums der Kleinsten Quadrate ist.[3]

Die Gleichungen (52) und (53) sowie die Matrix $\hat{X}_r$ spielen eine wichtige Rolle bei der Approximation der Faktorenanalyse.

23. **Eigenwerte und Korrelationen** Es sei Z eine spaltenstandardisierte $(m \times n)$-Datenmatix; dann ist

$$R = \frac{1}{m}Z'Z$$

die Matrix der Korrelationen zwischen den Variablen, die die Spalten von Z definieren. Zwischen den Eigenwerten von R und den Korrelationen r_{jk} aus R besteht eine Bziehung, die anhand zweier Spezialfälle illustriert werden soll. Zuerst werde der Fall

$$r_{jk} = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases}$$

betrachtet. Alle Variablen sind perfekt unkorreliert. Für Z hat man die Singularwertzerlegung $Z = Q\Lambda^{1/2}P'$, so dass

$$R = \frac{1}{m}Z'Z = \frac{1}{m}I = \frac{1}{m}P\Lambda P',$$

I die $(n \times n)$-Einheitsmatrix, woraus wegen der Orthonormalität von P

$$P = P\Lambda$$

[3]Eckart, C., Young, G. (1936) The aproximation of one matrix by another of lower rank. *Psychometrika* 1, 211–218; Johnson, R.M. (1963) On a theorem stated by Eckart and Young. *Psychometrika*, 28 (3), 259–263

folgt. Bezeichnet man mit $\mathbf{P}_j$ die Spaltenvektoren von P, so impliziert diese Gleichung

$$[\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_n] = [\lambda_1 \mathbf{P}_1, \lambda_2 \mathbf{P}_2, \ldots, \lambda_n \mathbf{P}_n],$$

also

$$\mathbf{P}_j = \lambda_j \mathbf{P}_j, \quad j = 1, \ldots, n$$

was nur möglich ist, wenn $\lambda_j = 1$ für alle j.

Es sei umgekehrt $r_{jk} = 1$ für alle j, k. Dies bedeutet $\mathbf{Z}_j' \mathbf{Z}_k / m = 1$ für alle j, k, oder $\mathbf{Z}_j = \mathbf{Z}_k$. Generell gilt wieder $Z = Q\Lambda^{1/2} P' = LP'$. Die Spaltenvektoren von Z sind Linearkombinationen der Spaltenvektoren von L, insbesondere gelten dann die Beziehungen

$$\mathbf{Z}_j = L\mathbf{P}_j, \quad \mathbf{Z}_k = L\mathbf{P}_k, \tag{55}$$

$\mathbf{P}_j$ und $\mathbf{P}_k$ Spaltenvektoren von P' bzw. Zeilenvektoren von P. Wegen $\mathbf{Z}_j = \mathbf{Z}_k$ folgt aber

$$\mathbf{Z}_j - \mathbf{Z}_k = \vec{0} = L(\mathbf{P}_j - \mathbf{P}_k) = (p_{j1} - p_{k1})\mathbf{L}_1 + \cdots (p_{jn} - p_{kn})\mathbf{L}_n.$$

Wegen der linearen Unabhängigkeit der $\mathbf{L}_j$ ist diese Gleichung aber nur möglich, wenn die Koeffizienten $p_{ji} - p_{ki} = 0$ für alle i, so dass $\mathbf{P}_j = \mathbf{P}_k$ für alle j, k. Das ist aber nicht möglich, weil die $\mathbf{P}_j$, $\mathbf{P}_k$ orthonormal sein müssen. Daraus folgt, dass Q und damit L nur aus einem Spaltenvektor bestehen kann, der proportional zu den $\mathbf{Z}_j$ ist. Korrelationen $r_{jk} \approx 1$ für alle j, k legen also nahe, dass es nur eine latente Variable gibt (der Rang von Z ist 1), während $r_{jk} \approx 0$ bedeutet, dass es n latente Variable gibt (Rang von Z ist n).

24. **Korrelationen und die Repräsentation von Merkmalen:** Es sei

$$Z = Q\Lambda^{1/2} P' = QA', \quad A = P\Lambda^{1/2} \tag{56}$$

die SVD der spaltenstandardisierten Matrix Z. Die Korrelation zwischen den Merkmalen M_j und M_k ist dann

$$r_{jk} = \frac{1}{m} \vec{Z}_j' \vec{Z}_k = \frac{1}{m} (L\vec{a}_j)'(L\vec{a}_k) = \frac{1}{m} \vec{a}_j' \vec{a}_k, \tag{57}$$

denn $L'L = \Lambda$ die Diagonalmatrix der Eigenwerte. Stehen die Zeilen von Z für Vpn, so sieht man, dass die Scores q_{ik} (Score der i-ten Vpn auf der k-ten latenten Variablen) nicht in die Korrelation eingehen, die Korrelation hängt nur von den "Ladungen" $a_{ju} = p_{ju}\sqrt{\lambda_u}$ des j-ten (und analog dazu des k-ten) Merkmals ab, wobei $u = 1, \ldots, n$ die latenten Merkmale sind. Die Ladungen reflektieren aber nicht die Fähigkeiten oder die Ausprägungen der Merkmale bei den Personen, sondern Ausmaße, mit denen die gemessenen Variablen latenten Variablen erfassen.

Da $r_{jj} = 1$ für alle j folgt

$$r_{jj} = 1 = \frac{1}{m} \sum_{u=1}^{n} a_{ju}^2 = 1. \tag{58}$$

Für den Spezialfall $n = 2$ liegen die Endpunkte der $\vec{a}_j$ auf einem Kreis der Länge 1; allgmein liegen sie auf der Oberfläche einer Hyperkugel mit dem Radius 1. Für $n = 2$ kann man die beiden Merkmale in der Ebene durch die 2-dimensionalen Vektoren $\vec{a}_j = (a_{1j}, a_{2j})'$ und $\vec{a}_k = (a_{1k}, a_{2k})'$ repräsentieren, wobei man die Anfangspunkte der Vektoren in den Ursprung legt. Wegen $r_{jk} = \cos\theta_{jk}$ reflektiert der Winkel θ_{jk}

zwischen den Vektoren die Korrelation zwischen den Merkmalen. Dies gilt allgemein auch für $n > 2$ latente Variablen, nur muß man dann den Winkel zwischen 3-, 4-, ...-dimensionalen Vektoren betrachten.

25. **Hauptachsentransformation** Die $(n \times n)$-Matrix M sei symmetrisch und $\mathbf{x}$ sei ein n-dimensionaler Vektor. Dann ist

$$\mathbf{x}' M \mathbf{x} = k, \tag{59}$$

k ein Skalar. Ausdrücke dieser Art heißen *quadratische Form*, multipliziert man sie aus, sieht man, dass sie Gleichungen für Ellipsoide darstellen, im 2-dimensionalen Fall für Ellipsen. Die Endpunkte aller Vektoren $\mathbf{x}$, die dieser Gleichung genügen, liegen demnach auf einem Ellipsoid (für $n = 2$ auf einer Ellipse). N sei eine $(n \times n)$-Diagonalmatrix, $\mathbf{y}$ sei ebenfalls ein n-dimensionaler Vektor; dann ist das Ellipsoid

$$\mathbf{y}' N \mathbf{y} = k \tag{60}$$

achsenparallel. Die Vektoren $\mathbf{x}$ sollen so transformiert (= rotiert) werden, dass das Ellipsoid (59) in das achsenparallele Ellipsoid (60) übergeht. Also ist eine *Transformationsmatrix* T gesucht derart, dass $T\mathbf{x} = \mathbf{y}$. Dann gilt also

$$\mathbf{y}' N \mathbf{y} = \mathbf{x}' T' N T \mathbf{x} = k. \tag{61}$$

Für die $\mathbf{x}$ gilt aber (59), so dass

$$M = T' N T \tag{62}$$

folgt. Da Rotationen auch rückgängig gemacht werden können, muß

$$T^{-1}\mathbf{y} = T^{-1} T \mathbf{x} = \mathbf{x}$$

gelten, d.h. T muß orthonormal sein. Dann folgt wegen (41), dass T die Matrix der Eigenvektoren von M sein muß, und N die Matrix der zugehörigen Eigenwerte. Die Transformation der Vektoren $\mathbf{x}$ mit der Matrix der Eigenvektoren von M bewirkt also die Transformation des orientierten Ellipsoids in ein achsenparalleles Ellipsoid. Damit werden auch die Hauptachsen des Ellipsoids transformiert, – deshalb heißt die Transformation auch *Hauptachsentransformation*. Oft wird die Abkürzung PCA benützt; sie steht für den englischen Ausdruck *Principal Component Analysis*, der wiederum für 'Hauptachsentransformation' steht.

26. **Vektordifferentiation und die Methode der Kleinsten Quadrate:** Man kann einen Vektor nach seinen Komponenten differenzieren, indem man jede Komponente nach einer bestimmten Komponente differenziert. Ist also $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$, so ist

$$\frac{\partial \mathbf{x}}{\partial x_j} = \left(\frac{\mathbf{x}_1}{dx_j}, \frac{dx_2}{dx_j}, \ldots, \frac{dx_n}{dx_j} \right)'.$$

Nun hängen aber die Komponenten $x_1, x_2, \ldots$ gar nicht von x_j ab, nur x_j selbst hängt von x_j ab, so dass die $dx_k/dx_j = 0$ sind für alle $k \neq j$, und $dx_j/dx_j = 1$.

Daraus folgt, dass

$$\frac{\partial \mathbf{b}}{\partial x_j} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{e}_j \tag{63}$$

gilt, wobei die 1 an der j-ten Stelle steht, $\mathbf{e}_j$ ist also der j-te Einheitsvektor. Damit läßt sich die Methode der Kleinsten Quadrate (MKQ) leicht durchführen: es gelte

$$\mathbf{y} = X\mathbf{b} + \mathbf{e},$$

wobei X eine Matrix von Prädiktoren ist und $\mathbf{b}$ ein unbekannter Vektor von Regressinosgewichten. $\mathbf{e}$ ist ein Fehlervektor. Der MKQ zufolge sollen die Komponenten von $\mathbf{b}$ so bestimmt werden, dass die Summe der Fehlerquadrate minimal wird. Diese Summe ist aber gleich $\mathbf{e}'\mathbf{e}$, und damit soll

$$Q(\mathbf{e}) = (\mathbf{y} - X\mathbf{b})'(\mathbf{y} - X\mathbf{b}) = \mathbf{e}'\mathbf{e} \stackrel{!}{=} \min \tag{64}$$

werden. Ausmultipliziert ergibt sich

$$Q(\mathbf{e}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'X\mathbf{b} - \mathbf{b}'X'\mathbf{y} + \mathbf{b}'X'X\mathbf{b}$$

Dann ist

$$\frac{\partial Q(\mathbf{e})}{\partial b_j} = -\mathbf{y}'X\mathbf{e}_j - \mathbf{e}_j'X'\mathbf{y} + \mathbf{e}_j'X'X\hat{\mathbf{b}} + \hat{\mathbf{b}}'X'X\mathbf{e}_j = 0,$$

denn die partiellen Ableitungen nach b_j werden ja gleich Null gesetzt, um die KQ-Schätzung $\hat{b}_j$ für b_j zu bekommen. $\hat{\mathbf{b}}$ ist der Lösungsvektor für diese Gleichung. Nun ist $\mathbf{y}'X\mathbf{e}_j$ ein Skalarprodukt ($X\mathbf{e}_j$ ist ein Spaltenvektor) und $\mathbf{e}_j'X'\mathbf{y}$ ist ebenfalls ein Skalarprodukt ($X'\mathbf{y}$ ist ebenfalls ein Spaltenvektor), und es folgt $\mathbf{y}'X\mathbf{e}_j = \mathbf{e}_j'X'\mathbf{y}$

$$2\mathbf{e}_j'X\mathbf{y} = 2\mathbf{e}_j'X'X\hat{\mathbf{b}}.$$

Diese Gleichungen kann man zusammenfassen, wobei die $\mathbf{e}_j$ insgesamt gerade die Einheitsmatrix I ergeben, so dass $X'\mathbf{y} = X'X\hat{\mathbf{b}}$ und damit

$$\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y} \tag{65}$$

folgt. Man erhält für $\mathbf{y}$ die Schätzung

$$\hat{\mathbf{y}} = X\hat{\mathbf{b}} + \hat{\mathbf{e}}, \tag{66}$$

und

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}}.$$

Setzt man für $\hat{\mathbf{b}}$ den Ausdruck (65) ein, so erhält man

$$\mathbf{y} = X(X'X)^{-1}X'\mathbf{y} + \hat{\mathbf{e}},$$

wobei

$$P = X(X'X)^{-1}X' \tag{67}$$

eine *Projektionsmatrix* ist. Wie man sofort nachrechnet ist P (i) symetrisch, und (ii) *idempotent*, denn

$$P^2 = X(X'X)^{-1}X'X(X'X)^{-1}X = X(X'X)^{-1}X'.$$

Der Ausdruck 'Projektionsmatrix' ergibt sich aus der Tatsache, dass einerseits $\hat{\mathbf{y}}$ eine Linearkombination der Spaltenvektoren von X ist, denn $\hat{\mathbf{y}} = X\hat{\mathbf{b}}$, und andererseits $\hat{\mathbf{e}}$ orthogonal zu $\hat{\mathbf{y}}$ ist. Denn

$$\hat{\mathbf{e}}'\hat{\mathbf{y}} = (\mathbf{y} - \hat{\mathbf{y}})'X\hat{\mathbf{b}} = \mathbf{y}'X\hat{\mathbf{b}} - \hat{\mathbf{y}}'X\hat{\mathbf{b}} = \mathbf{y}'X(X'X)^{-1}X'\mathbf{y} - \mathbf{y}'(X'X)^{-1}X'\mathbf{y} = 0.$$

Andererseits ist

$$\hat{\mathbf{y}} = X(X'X)^{-1}X'\mathbf{y} = P\mathbf{y}.$$

Dies heißt, das $\hat{\mathbf{y}}$ eine Transformation von $\mathbf{y}$ ist, und insbesondere erweist sich $\hat{\mathbf{y}}$ als eine Projektion des Vektors $\mathbf{y}$ auf den Vektorraum der Linearkombinationen der Spaltenvektoren von X; P ist gerade die Transformationsmatrix, die diese Projektion bewirkt. Der Vektor $\hat{\mathbf{e}}$ ist gewissermaßen das Lot vom Endpunkt des Vektors $\mathbf{y}$ auf den Raum der Linearkombinationen von X.

27. **Gauß-Markov-Theorem:** Mithilfe der Projektionsmatrix läßt sich leicht das *Gauß-Markov-Theorem* beweisen, nach dem $\hat{\mathbf{b}}$ die beste lineare unverzerrte (biasfreie) Schätzung für $\mathbf{b}$ ist, die sich finden läßt.