Name: (bitte Druckbuchstaben)[1]

Matrikelnr.:

Probeklausur zur Vorlesung Multivariate Methoden, SoSe 2013

.

Psychologisches Institut der Johannes Gutenberg-Universität Mainz
17. 06. 2013

1. **Aufgabe:** Die Multiplikation eines Vektors mit einem Skalar bedeutet (1) die Reduktion des Vektors auf einen Skalar, (2) die Veränderung der Länge des Vektors, (3) das Verbot der Addition eines Skalars. Welche der Alternativen trifft zu?

   Antwort: Die Veränderung der Länge, wenn der Skalar $\neq 1$ ist.

2. **Aufgabe:** Was versteht man unter einer Linearkombination der Vektoren $\mathbf{x}_1, \ldots, \mathbf{x}_n$? Wann gilt die Gleichung

   $$\lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n = \vec{0},$$

   λ_j, $j = 1, \ldots, n$ Skalare ? Unter welchen Bedingungen gilt insbesondere $\lambda_1 = \cdots = \lambda_n = 0$?

   Antwort: Linearkombination: Jeder Ausdruck der Form

   $$\mathbf{x} = \lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n.$$

   Die Gleichung

   $$\lambda_1\mathbf{x}_1 + \cdots + \lambda_n\mathbf{x}_n = \vec{0},$$

   gilt immer, so lange keine speziellen Bedingungen für die λ_j gestellt werden; man muß ja nur $\lambda_1 = \cdots = \lambda_n = 0$ setzen, dann gilt die Gleichung trivialerweise. Dieser Fall ist allerdings die *einzige Lösung*, wenn die Vektoren $\mathbf{x}_j$, $j = 1, \ldots, n$ linear unabhängig sind.

3. **Aufgabe:** In welcher Beziehung steht das Skalarprodukt zweier Vektoren zum Kosinus des Winkels zwischen den Vektoren? Unter welchen Bedingungen ist das Skalarprodukt gleich dem Produkt-Moment-Korrelationskoeffizienten?

   Antwort: Es ist $\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$, – dies ist die gefragte Beziehung. Andererseits ist

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

   Wenn die Komponenten von $\mathbf{x}$ und $\mathbf{y}$ zentrierte Werte sind (dh Abweichungen vom Mittelwert sind, $x_i = X_i - \bar{x}$, $y_i = Y_i - \bar{y}$), so ist

   $$\frac{1}{m}\mathbf{x}'\mathbf{y} = \frac{1}{m}\sum_{i=1}^{m} x_i y_i$$

[1] File: Klausur17062013.tex

gerade die Kovarianz, und

$$s_x = \frac{1}{\sqrt{m}}\|\mathbf{x}\|, \quad s_y = \frac{1}{\sqrt{m}}\|\mathbf{y}\|$$

sind die Standardabweichungen. Der Faktor $1/m$ kürzt sich heraus und

$$\cos\theta = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = r_{xy}$$

ist der Korrelationskoeffizient. Für

$$s_x = \frac{1}{\sqrt{m}}\|\mathbf{x}\| = s_y = \frac{1}{\sqrt{m}}\|\mathbf{y}\| = 1$$

erhält man $\mathbf{x}'\mathbf{y} = r_{xy}$; dies ist der Fall, wenn die Komponenten x_i und y_i *standardisierte* Werte sind.

4. **Aufgabe:** Eine Möglichkeit, die Korrelation r_{jk} zwischen zwei Variablen V_j und V_k zu interpretieren, besteht darin, die Korrelation auf die Wirkung einer (oder mehrerer) anderer ("latenter") Variablen $L_1, L_2, \ldots$ sowohl auf V_j wie auf V_k zurückzuführen. Man habe n Variablen bei m Personen gemessen, $m > n$. Wieviele latente Variable können maximal zur Interpretation der r_{jk} postuliert werden?

   Antwort: Maximal n. Bedenken Sie, dass die Methode der Hauptachsentransformation zunächst einmal nur eine Rotation der Konfiguration bzw. der Koordinatenachsen ist, und dass wegen der Messfehler irgendzwei Vektoren stets linear unabhängig sind, dh die gemessenen Vektoren sind alle linear unabhängig (zumindest numerisch).

5. **Aufgabe:** Sie haben die Messwerte für die V_j in m-dimensionalen Vektoren $\mathbf{x}_1, \ldots, \mathbf{x}_n$ zusammengefasst und möchten wissen, ob sich $\mathbf{x}_1$ in Form einer multiplen Regression durch die übrigen Vektoren darstellen läßt. Welche Parameter müssen Sie dann schätzen und welche Methode zur Schätzung der Parameter würden Sie benützen?

   Antwort: Dem Ansatz der multiplen Regression zufolge soll also

   $$\mathbf{x}_1 = \lambda_2\mathbf{x}_2 + \cdots + \lambda_n\mathbf{x}_n + \mathbf{e}_1$$

   gelten. Dann müssen die Parameter $\lambda_2, \ldots, \lambda_n$ geschätzt werden; man verwendet im Allgemeinen die Methode der Kleinsten Quadrate.

6. **Aufgabe:** Was bedeutet der Begriff der Multikollinearität im Zusammenhang mit der vorigen Aufgabe, und worin besteht der Nachteil ausgeprägter Multikollinearitäten?

   Antwort: Dass einige der Vektoren sich nach denm Prinzip der multiplen Regression aus den jeweils anderen voraussagen lassen. Dieser Sachverhalt impliziert die Existenz von Korrelationen zwischen den Vektoren. Werden diese Vektoren als Prädiktoren zur Vorhersage anderer Variablen verwendet, so haben die geschätzten Regressionskoeffizienten eine hohe Varianz und sind i. A. negativ korreliert; sie sind deshalb schwer interpretierbar und versagen bei der Vorhersage neuer Fälle.

7. **Aufgabe:** Unrealistischerweise werde einmal angenommen, dass Sie die Variablen V_j und V_k (Galvanischer Hautwiderstand und Herzrate) "fehlerfrei" messen können, und Sie finden $\mathbf{x}_j = \lambda \mathbf{x}_k$, $\lambda \in \mathbb{R}$ eine Konstante (Skalar). Dann sind $\mathbf{x}_j$ und $\mathbf{x}_k$ linear abhängig, – warum?. Nun sind Sie aber in der Realität und finden, dass Ihre Messungen zufällige "Fehler" enthalten, und $\mathbf{x}_j$ und $\mathbf{x}_k$ sind linear unabhängig. Warum, und welche Implikation hat diese lineare Unabhängigkeit für den Begriff des Fehlers?

   Antwort: Die Vektoren sind linear abhängig, denn es gilt nun

   $$\lambda_j \mathbf{x}_j + \lambda_k \mathbf{x}_k = \vec{0}$$

   mit $\lambda_j \neq 0$, $\vec{\lambda}_k \neq 0$. Dies folgt aus der Beziehung $\mathbf{x}_j = \lambda \mathbf{x}_k$, denn dann folgt ja

   $$\mathbf{x}_j - \lambda \mathbf{x}_k = \vec{0}$$

   mit $\lambda_j = 1$, $\lambda_k = -\lambda$. Enthalten die Messungen aber Fehler, so kann man

   $$\mathbf{x}_j = \mathbf{x}_{0j} + \xi_j, \quad \mathbf{x}_k = \mathbf{x}_{0k} + \xi_k$$

   schreiben, wobei ξ_j und ξ_k zufällige Vektoren sind, deren Komponenten nicht proportional zu den Komponenten von $\mathbf{x}_{0j}$ bzw. $\mathbf{x}_{0k}$ sind. Dann sind aber auch die $\mathbf{x}_j$ und $\mathbf{x}_k$ nicht mehr parallel, weshalb sie linear unabhängig sind. Die Fehlervektoren dürfen nicht parallel zu den "wahren" Vektoren $\mathbf{x}_{0j}$ und $\mathbf{x}_{0k}$ sein, – wären sie es, würden sie in $\mathbf{x}_{0j}$ und $\mathbf{x}_{0k}$ "absorbiert", dh sie würden die Orientierung dieser Vektoren nicht ändern und wären also keine "Fehler".

8. **Aufgabe:** Auf welchem Sachverhalt beruht die Tatsache, dass Sie für jede Menge von n empirisch gewonnenen Vektoren $\mathbf{x}_j$, $j = 1, \ldots, n$ von Daten $r \leq n$ orthogonale Vektoren ("Basisvektoren") finden können derart, dass sich die $\mathbf{x}_j$ als Linearkombinationen der Basisvektoren dargestellt werden können. Ist die Wahl der Basisvektoren eindeutig?

   Antwort: Der Sachverhalt ist einfach ein Ergebnis der Linearen Algebra: hat man r m-dimensionale Vektoren, und ist $r < m$, so liegen die Vektoren in einem höchstens r-dimensionalen Teilraum des m-dimensionalen Vektorraums. Man kann dann irgendeine Menge von r m-dimensionalen, linear unahängigen Vektoren aus diesem Teilraum wählen und die gegebenen Vektoren als Linearkombinationen dieser gewählten Vektoren darstellen. Das kann man sich klarmachen, wenn man die entsprechenden linearen Gleichungssysteme betrachtet, aber darauf muß nicht eingegangen werden.

   Die Wahl ist *nicht* eindeutig, was schon daraus folgt, dass man eben *irgendeine* Menge von r m-dimensionalen l.u. Vektoren wählen kann, – hat man eine gewählt, kann man auch wieder eine andere wählen und alle Vektoren als Linearkombinationen eben dieser Vektoren darstellen.

9. **Aufgabe:** Die Vektoren $\mathbf{x}_j$ werden in einer $m \times n$-Matrix X zusammengefasst. Es werde angenommen, dass es $r \leq n < m$ m-dimensionale, linear unabhängige Vektoren $\mathbf{L}_1, \ldots, \mathbf{L}_r$ gibt, deren Linearkombinationen die $\mathbf{x}_j$ der Matrix X sind. Geben Sie eine Matrixgleichung an, die diesen Sachverhalt ausdrückt (Sie müssen dazu eine weitere Matrix einführen). Welches zusätzliche Postulat bezüglich der $\mathbf{L}_k$ müssen Sie machen, damit diese tatsächlich berechenbar werden? Ist dieses Postulat eine Einschränkung der Allgemeinheit dieses Ansatzes? Wie sieht die Gleichung aus, derzufolge Sie die $\mathbf{L}_k$ aus der Matrix X berechnen können?

Antwort: Die Gleichung ist $X = LP'$, wobei L die $\mathbf{L}_k$ als Spaltenvektoren enthält. Fordert man zusätzlich, dass die $\mathbf{L}_k$ orthogonal sind, so folgt $X'X = PL'LP' = P\Lambda P$, Λ eine Diagonalmatrix, und diese Gleichung impliziert, dass P die orthonormale Matrix der Eigenvektoren von $X'X$ sein muß. Also folgt auch $XP = LP'P = L$, und dies bedeutet, dass die $\mathbf{L}_k$ als Linearkombinationen der Spaltenvektoren von X berechnet werden können, sobald man die Matrix P berechnet hat (das macht dann der Computer).

10. **Aufgabe:** Die Anzahl der zur Darstellung der m-dimensionalen Spaltenvektoren der Matrix X benötigten latenten Variablen heißt Spaltenrang der Matrix. Die Anzahl der zur Darstellung der n-dimensionalen Zeilenvektoren benötigten latenten Variablen heißt Zeilenrang der Matrix. Welche Beziehung besteht zwischen dem Zeilen- und dem Spaltenrang?

    Antwort: Der Spaltenrang ist stets gleich dem Zeilenrang einer Matrix, weshalb man auch nur vom Rang einer Matrix spricht.

11. **Aufgabe:** Die Betrachtungen in der Aufgabe 9 zur Interpretation der Datenmatrix X sind rein formal, d.h. sie sind nur eine Anwendung von Ergebnissen der linearen Algebra. Kann man folgern, dass die untersuchten psychologischen Beziehungen zwischen den betrachteten Variablen stets auch adäquat durch diese formalen Strukturen abgebildet werden?

    Antwort: Nein, das kann man nicht. Wie die psychologischen Beziehungen zwischen den betrachteten Variablen tatsächlich sind wissen oft nur die Götter.

    Anmerkung: Hier fragen manche, warum man denn überhaupt derartige Analysen durchführt, wenn doch gar nicht klar ist, ob die Resultate der Wahrheit entsprechen. Die Antwort ist, dass man immerhin eine *mögliche* Interpretation der Daten erhält, die man im Prinzip in weiteren Untersuchungen überprüfen kann. Wissenschaftstheoretische Analysen zeigen, dass es zumindest in den interessanten Fällen keine Beschreibung ohne damit einhergehende theoretische Annahmen gibt[2]. Die in manchen Statistiklehrbüchern zu findende Behauptung, die PCA liefere "nur" eine Beschreibung, macht in dieser Allgemeinheit und ohne weitere Spezifikation keinen Sinn.

12. **Aufgabe:** Die Singularwertzerlegung (SVD) der Matrix X erlaubt es stets, für die Probanden oder Versuchspersonen Scores auf hypothetischen latenten Dimensionen zu finden und gleichermaßen Ladungen der Variablen auf denselben latenten Dimensionen. Geben Sie an, in welcher Weise der Messwert x_{ij} der i-ten Person für die j-te Variable sich als Skalarprodukt der Faktorscores auf den latenten Variablen und den Ladungen auf den latenten Variablen darstellen läßt. Unter welchen Bedingungen nimmt x_{ij} dann den maximal möglichen Wert an? (Hinweis: Cauchy-Schwarzsche Ungleichung).

[2] Für Interessierte: diese Diskussion wurde u.a. geführt, nachdem der zum Wiener Kreis (1922 – 1936) gehörende Philosoph und Mathematiker Rudolf Carnap in Anlehnung an eine These L. Wittgensteins aus dessen Tractatus Logico-Philosophicus sogenannte Protokollsätze, das sind rein beschreibende Sätze ohne theoretischen Hintergrund, als Basis für die Bildung von Theorien postulierte. Die Diskussion mit dem ebenfalls zum Wiener Kreis gehörenden Otto Neurath (die sog. Protokollsatzdebatte) zeigte, dass es Protokollsätze im Carnapschen Sinn nicht gibt. Andere Philosophen (zB der Physiker und Philosoph Pierre Duhem, der Mathematiker, Physiker und Philosoph Henri Poincaré) hatten sich schon vor der Bildung des Wiener Kreises ähnlich kritisch über "rein" beschreibende Aussagen geäußert. Analoge Betrachtungen gelten auch für hermeneutische Aussagen der "Verstehenden", geisteswissenschaftlichen Psychologie, die dort u.a. behauptete Möglichkeit der "Wesensschau" existiert nicht in der postulierten Form.

Antwort: In der Form

$$x_{ij} = q_{i1}a_{j1} + \cdots + q_{in}a_{jn}.$$

Der Wert x_{ij} wird maximal (relativ zur Länge der Vektoren), wenn $q_{ik} = \alpha a_{jk}$ für alle k.

13. **Aufgabe:** Die SVD ergibt sich aus der Annahme der Orthogonalität der latenten Vektoren $\mathbf{L}_k$ (wie?). Durch Rotation der Vektoren können Sie zu alternativen "Lösungen" des Problems, latente Variable zur Erklärung der beobachteten Variablen zu finden gelangen. Welche wichtige Eigenschaft zeichnet die Vektoren, die sich aus der SVD ergeben, vor den rotierten Lösungen aus?[3]

    Antwort: Der Ansatz ist $X = LP'$, L orthogonal; Normierung der $\mathbf{L}_k$ in L: $Q = L\Lambda^{-1/2}$, also $L = Q\Lambda^{1/2}$. Die Koordinaten L sind die einzigen, die unkorreliert sind, – werden sie rotiert, ist die Punktekonfiguration in den neuen, rotierten Achsen nicht mehr unkorreliert. Die wichtige Eigenschaft ist, dass die Varianz der Koordinaten auf der ersten Achse maximal ist, etc.

14. **Aufgabe:** In einem Versuch werden Messerte x_{ij} (i-te Person, j-te Variable) erhoben, die von zwei latenten Variablen in der Form

    $$x_{ij} = p_{j1}L_{i1} + p_{j2}L_{i2} + p_{j3}L_{i1}L_{i2} + e_{ij}$$

    abhängen; bemerken Sie, dass L_{i1} und L_{i2} hier keine Vektoren, sondern Variablennamen sind. Gibt es eine Möglichkeit, die Produktterme $L_{i1}L_{i2}$ durch Komponenten eines Vektors zu repräsentieren? Ist die Repräsentation der Daten anhand der SVD mit diesem Sachverhalt verträglich?

    Antwort: Man kann einen Vektor $\mathbf{L}_3$ definieren, dessen Komponenten aus den Produkten der Komponenten von $\mathbf{L}_1$ und $\mathbf{L}_2$ bestehen:

    $$\mathbf{L}_3 = \begin{pmatrix} L_{11}L_{12} \\ L_{21}L_{22} \\ \vdots \\ L_{m1}L_{m2} \end{pmatrix}.$$

    Dieses Modell ist aber nicht mit der SVD komptibel, weil $\mathbf{L}_3$ nicht mehr notwendig orthogonal zu $\mathbf{L}_1$ und $\mathbf{L}_2$ ist.

15. **Aufgabe:** Numerisch benötigt man für eine $m \times n$-Datenmatrix ($m < n$) stets n latente Dimensionen, – dies ist ein Resultat der Tatsache, dass man so gut wie nie fehlerfreie Daten hat (warum?). Wie können Sie feststellen, ob $r < n$ latente Dimensionen genügen, um die Daten zu erklären? Nennen Sie bitte einen Test und erläutern ihn.

    Antwort: Das Problem fehlerbehafteter Messungen ist oben bereits besprochen worden. Der Test ist der Scree-Test.

16. **Aufgabe:** Der zu einer latenten Dimension korrespondierende Eigenwert λ_k ergibt sich ursprünglich als Quadrat der Länge (= Skalarprodukt $\mathbf{L}_k'\mathbf{L}_k$) des

[3] Erinnern Sie sich daran, dass man gerne ein Koordinatensystem hätte, in Bezug auf das alle Korrelationen gleich Null sind.

Vektors, der die k-te latente Dimension repräsentiert. In welchem Sinne repräsentiert λ_k eine Varianz?

Antwort: Für zentrierte Daten ist $\mathbf{L}_k'\mathbf{L}_k$ gerade die Summe der Quadrate der Abweichungen vom Mittelwert (der hier gleich Null ist). Damit ist $\mathbf{L}_k'\mathbf{L}_k$ proportional zur Varianz der Koordinaten (der Faktor ist $1/m$ bzw $1/(m-1)$).

17. **Aufgabe:** Die Die Ladung der j-ten Variablen auf der k-ten latenten Dimension ist durch $a_{jk} = p_{jk}\sqrt{\lambda_k}$ gegeben. Woher kennt man den Faktor p_{jk}, und unter welchen Bedingungen liegen alle Variablen auf einem Kreis?

    Antwort: p_{jk} ist die j-te Komponente des k-ten Eigenvektors von $X'X$. Es ist

    $$r_{jj} = \sum_{k=1}^{r} a_{jk}^2 = 1.$$

    Für $r = 2$ ist dies die Gleichung eines Kreises; die Variablen liegen auf einem Kreis, wenn man genau 2 latente Dimensionen zur Darstellung der Daten benötigt.

18. **Aufgabe:** n Experten beurteilen m Bewerber um eine leitende Position in einem großen Konzern. Wie können Sie durch Anwendung der SVD abschätzen, dass die Experten ihre Urteile objektiv, d.h. ohne Wechselwirkung mit irrelevanten Merkmalen der Bewerber abgeben? Beziehen Sie Ihr Argument auf die aus der Testtheorie bekannten Begriffe der Sensitivität und Spezifität.

    Antwort: Die Ladungen müssen unabhängig von der Stichprobe sein.

    Zur Sensitivität (eines Tests/Beobachters): sie ist die Wahrscheinlichkeit, ein Merkmal $\mathcal{M}$ zu entdecken oder zu bemerken, wenn es tatsächlich vorhanden ist.

    Zur Spezifität (eines Tests/Beobachters): sie ist die Wahrscheinlichkeit, $\mathcal{M}$ *nicht* zu diagnostizieren, wenn es tatsächlich auch *nicht* vorhanden ist (also die Wahrscheinlichkeit, keinen 'falschen Alarm' zu geben).

    Sowohl die Sensitivität als auch die Spezifität eines Experten, der Personen beurteilt, können von der beurteilten Person (= BewerberIn) abhängen: die gepflegte Sprache einer/s BewerberIn können manche Experten dazu verführen, einen gehobenen Bildungsstand oder eine deutlich überdurchschnittliche Intelligenz zu diagnostizieren, auch wenn weder das eine noch das andere Merkmal vorhanden ist (geringe Spezifität), und umgekehrt können eher ungepflegte sprachliche Umgangsformen dazu verleiten, einen Mangel an Bildung und/oder Intelligenz zu diagnostizieren (geringe Sensitivität). die p_{jk}-Werte für diese beiden Merkmale (Bildung, Intelligenz) hängen dann von der jeweils beurteilten Person ab.