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?

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 $\lambda_1 = \cdots = \lambda_n = 0$?

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?

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 zu postulieren. Man habe n Variablen bei m Personen gemessen, $m > n$. Wieviele latente Variable können maximal zur Interpretation der r_{jk} postuliert werden?

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?

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

7. **Aufgabe:** Unrealistischer Weise 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 $\mathbf{x}_j$

[1] File: Klausur17062013.tex

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?

8. **Aufgabe:** Auf welchem Sachverhalt beruht die Tatsache, dass Sie für jede Menge von n empirisch gewonnenen Vektoren $\mathbf{x}_j$, $j = 1, m \ldots, n$ 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?

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 ausdrücken (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?

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-dimensinalen Zeilenvektoren benötigten latenten Variablen heißt Zeilenrang der Matrix. Welche Beziehung besteht zwischen dem Zeilen- und dem Spaltenrang?

11. **Aufgabe:** Die Betrachtungen in der vorangegangenen Aufgabe 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?

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).

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?[2]

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

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

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

abhängen; bemerken Sie, dass $_{i1}$ 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?

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.

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

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?

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.