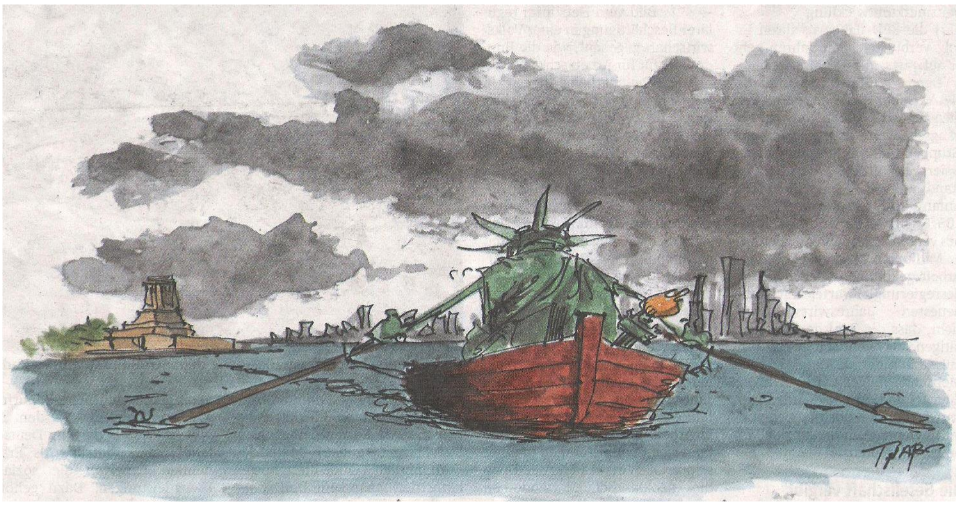
The day after tomorrow



· 05. 02. 2017

Zusammenfassung – Klausur + Frage (Folien)

1. **Vektoren:** Linearkombinationen, Skalarprodukt, dyadisches Produkt:

Linearkombination: $\mathbf{y}, \mathbf{x}_1, \ldots, \mathbf{x}_n$ m-dimensional. Dann ist

$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = X\mathbf{a}$$

eine Linearkombination der $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

Skalarprodukt: $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} = x_1y_1 + x_2y_2 + \cdots + x_ny_n \in \mathbb{R}$,

Dyadisches Produkt: $\mathbf{x}\mathbf{y}' = (x_iy_j)$ ((m,n)-Matrix).

Fragen:

(a) Linearkombinationen können nur dann berechnet werden, wenn die Komponenten der Vektoren mindestens Intervallskalenniveau haben!

falsch

(b) Skalarprodukte können nur dann berechnet werden, wenn sichergestellt ist, dass $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$

Nonsense

(c) Das dyadische Produkt ist eine Matrix mit Rang 1!

korrekt

(d) Das dyadische Produkt kann nur berechnet werden, wenn $\mathbf{x}$ und $\mathbf{y}$ dieselbe Anzahl von Komponenten haben!

falsch

Es gilt $\mathbf{x}'\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$ nur dann, wenn $\cos\theta > 0$.

falsch

(e) Zwei Vektoren sind kollinear, wenn sie beide Linearkombinationen derselben Grundvektoren sind.

falsch

(f) Zwei Vektoren sind kollinear, wenn sie dieselbe Orientierung haben.

korrekt

2. **Lineare Unabhängigkeit:** $\mathbf{x}_1, \ldots, \mathbf{x}_n$ sind linear unabhängig dann, wenn $a_1 = \cdots = a_n = 0$ für $a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n = \vec{0}$. Andernfalls sind sie linear abhängig.

Fragen:

(a) Die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_n$ können nur linear abhängig sein, wenn sie miteinander korrelieren.

"falsch" ist falsch, "korrekt" ist korrekt.

Kommentar: Alle Skalarprodukte ("Korrelationen") seien gleich Null. Dann sind die Vektoren paarweise orthogonal und damit sind sie auch linear unabhängig (l.u.). Sind sie l.u. so folgt *noch nicht*, dass sie auch paarweise orthogonal sind; es genügt für die lin. Unabhängigkeit, dass $|\cos\theta_{jk}| < 1$ ist, d.h. dass sie nicht alle parallel zueinander sind, dass also $|\cos\theta_{jk} = 1$ *nicht* gilt[1]. Der Fall $0 < |\cos\theta_{jk}| < 1$ für alle $\mathbf{x}_j, \mathbf{x}_k$ ist mit linearer Unabhängigkeit kompatibel. Eine "Korrelation" ($\mathbf{x}_j'\mathbf{x}_k \neq 0$) ist nur *notwendig*, aber nicht *hinreichend* für lineare Abhängigkeit; der Fall $\mathbf{x}_j'\mathbf{x}_k \neq 0$ schließt ja die *Möglichkeit* $\cos\theta_{jk} = 1$ ein. Hinreichend ist $|\cos\theta_{jk}| = 1$ für die lin. Abhängigkeit der beiden Vektoren, nur der Fall $\mathbf{x}_j'\mathbf{x}_k = 0$ schließt lin. Abhängigkeit aus. Dementsprechend *können* sie nur lin. abhängig sein, wenn sie "korrelieren", wenn also das Skalarprodukt von Null verschieden ist, sie *müssen* aber nicht lin. abh. sein ...

(b) Wenn die Vektoren $\mathbf{x}_j$ linear unabhängig sind, sind sie auch orthogonal.

falsch

(c) Kollineare Vektoren sind linear abhängig.

korrekt

3. **Vektorräume und -teilräume:** Menge V von Vektoren derart, dass jede Linearkombination von Vektoren aus V wieder Element aus V ist (Kurzdefinition!). Ein Teilraum V_t ist eine Teilmenge aus V derart, dass alle Linearkombinationen von Vektoren aus V_t wieder in V_t sind.

   Beispiel: Die 5-dimensionalen Vektoren $\mathbf{x}_1, \mathbf{x}_2$ aus dem Vektorraum V aller 5-dimensionalen Vektoren. Teilraum: die lineare Hülle $\mathcal{L}(\mathbf{x}_1, \mathbf{x}_2)$. Alle Linearkombinationen von $\mathbf{x}_1, \mathbf{x}_2$ liegen in einer Ebene des Raums V (die Ebene ist $\mathcal{L}$). $\mathcal{L}^c$ ist der *Komplementärraum* von $\mathcal{L}$. Es existiert 5-dimensionaler Vektor $\mathbf{n} \in \mathcal{L}^c$, mit $\mathbf{n}$ orthogonal zu allen $\mathbf{x} \in \mathcal{L}$; dies ist der Normalenvektor für $\mathcal{L}$.

[1] Aus der Logik: Es regnet (p). Dann ist die Straße naß (q), d.h. $p \to q$. Nun sei die Straße naß. Dann folgt *noch nicht*, dass es regnet (es kann jemand jemand die Strasse naß gespritzt haben). Aber wenn die Strasse *nicht* naß ist, kann es auch nicht regnen, d.h. man kann folgern, dass nicht-q die Aussage nicht-p impliziert. d.h. es gilt $\neg q \to \neg p$.

Basis, Teilbasis: Eine Menge $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ von linear unabhängigen Vektoren aus dem n-dimensionalen Vektorraum derart, dass $V_n = \mathcal{L}(\mathcal{B})$. $\mathcal{B}_r = \{\mathbf{b}_1, \ldots, \mathbf{b}_r\}$ mit $r < n$ ist eine Teilbasis, die einen r-dimensionalen Teilraum von V_n aufspannt. Gilt $\mathbf{b}_j'\mathbf{b}_k = 0$ für alle $j \neq k$, $\mathbf{b}_j'\mathbf{b}_j = 1$ für alle j, so ist die Basis eine Orthonormalbasis.

Fragen:

(a) Jede Menge von Vektoren ist ein Vektorraum, da ja eine Linearkombination von Vektoren wieder ein Vektor ist.

falsch

(b) Sei M eine Menge n-dimensionaler Vektoren. Die Maximalzahl linear unabhängiger Vektoren $\mathbf{b}_1, \ldots, \mathbf{b}_r$, die notwendig ist, um alle Vektoren der lineare Hülle $\mathcal{L}_M = \mathcal{L}(M)$ zu erzeugen, ist die Dimensionalität von $\mathcal{L}$.

korrekt

(c) Sei M eine Menge n-dimensionaler Vektoren. Die lineare Hülle $\mathcal{L}(M)$ ist die Menge aller Linearkombinationen der Vektoren aus M.

korrekt

(d) Die Vektoren einer Basis sind notwendig orthogonal zueinandner.

falsch

(e) Zwei kollineare Vektoren können nicht Elemente einer Basis sein.

korrekt

> **Kommentar:** Die Vektoren $\mathbf{b}_1, \ldots, \mathbf{b}_r$ einer Basis *müssen* linear unabhängig sein, denn eine Basis ist als Menge linear unabhängiger Vektoren definiert. Gilt etwa für $\mathbf{b}_1$ und $\mathbf{b}_2$, dass $\mathbf{b}_2 = a\mathbf{b}_1$, $a \in \mathbb{R}$, so sind $\mathbf{b}_1$ und $\mathbf{b}_2$ zueinander parallel und damit linear abhängig, können also *nicht beide* Basisvektoren sein, – höchstens einer von ihnen kann Element einer Basis sein.

(f) Die lineare Hülle von M ist ein Vektorraum (möglicherweise ein Teilraum eines Vektorraumes).

korrekt

(g) Ein Vektorraum hat beliebig viele Basen.

korrekt

4. **Matrizen:** entstehen durch Aneinander- oder Untereinanderschreiben von Vektoren gleicher Dimensionalität (gleicher Anzahl von Komponenten). Allgemein: es entsteht eine (m, n)-Matrix. Für $m = n$ ist

die Matrix quadratisch. Ist $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ eine Matrix mit den $\mathbf{x}_j$ als m-dimensionalen Spaltenvektoren, so ist X' die gestürzte oder transponierte Matrix (jetzt eine (n, m)-Matrix). Gilt $X' = X$, so ist X symmetrisch und damit notwendig auch quadratisch. Operationen: Multiplikationen mit einem Skalar: $aX = (ax_{ij})$, Addition: komponentenweise.

Multiplikation mit einem Vektor: A eine Matrix, $\mathbf{x}$ ein Vektor, $A\mathbf{x} = \mathbf{y}$ ist wieder ein Vektor; $\mathbf{y}$ ist Linearkombination der Spaltenvektoren von A. Voraussetzung: Anzahl der Salten von A gleich der Anzahl der Komponenten von $\mathbf{x}$.

$\mathbf{x}'A = \mathbf{y}'$; $\mathbf{y}'$ ist ein Zeilenvektor, der sich als Linearkombination der Zeilen von A ergibt. Voraussetzung: Anzahl der Komponenten von $\mathbf{x}$ ist gleich der Anzahl der Zeilen von A.

$A\mathbf{x} = \mathbf{y}$ bzw. $\mathbf{x}'A = \mathbf{y}'$ werden auch Transformationen von $\mathbf{x}$ genannt. A definiert eine Abbildung von Vektoren auf andere Vektoren.

Rotationen: $\mathbf{y} = T\mathbf{x}$, $\|\mathbf{y}\| = \|\mathbf{x}\|$. Dann ist T orthonormal, $T'T = I$. Rotationen lassen die Skalarprodukte invariant: $\mathbf{x} = T\mathbf{u}$, $\mathbf{y} = T\mathbf{v}$, so ist $\mathbf{x}'\mathbf{y} = \mathbf{u}'T'T\mathbf{v} = \mathbf{u}'\mathbf{v}$.

Fragen:

(a) Es sei A eine (m, n)-Matrix und es gelte $\mathbf{y} = A\mathbf{x}'$. Macht diese Behauptung Sinn?

geht gar nicht, total falsch

(b) Es gelte $\mathbf{y} = \mathbf{x}'A$; ist $\mathbf{y}$ eine Zeilen- oder ein Spaltenvektor?

$\mathbf{x}'A$ muß ein Zeilenvektor sein, also muß $\mathbf{y}$ ein Zeilenvektor sein; dann aber müsste es $\mathbf{y}'$ heißen.

(c) $\mathbf{y} = A\mathbf{x}$.

$\mathbf{x}$ muß n-dimensional sein

korrekt

$\mathbf{x}$ muß m-dimensional sein

falsch

Multiplikation einer Matrix mit einer anderen: $C = AB$. Voraussetzung: Anzahl der Spalten von A gleich Anzahl der Zeilen von B. Die Spalten von C sind Linearkombinationen der Spaltenvektoren von A, die Zeilenvektoren sind Linearkombinationen der Zeilen von B.

Es sei A eine quadratische Matrix, λ eine Diagonalmatrix gleicher Dimensionalität. $A\Lambda = B$; die Spaltenvektoren von B sind die längenskalierten Spaltenvektoren von A. $\Lambda A = B$: die Zeilenvektoren von B sind die längenskalierten Zeilenvektoren von A.

Fragen:

(a) Anzahl der Zeilen von A muß gleich der Anzahl der Spalten von B sein sein.

falsch

(b) Die Spalten von C sind Linearkombinationen der Spalten von A.

korrekt

(c) Die Zeilen von C sind Linearkombinationen der Zeilen von A. falsch

Rang einer (m, n)-Matrix: Der Spaltenrang ist die Anzahl der m-dimensionalen, linear unabhängigen Vektoren $\mathbf{b}_j$, die benötigt werden, die Spaltenvektoren von A als Linearkombinationen der $\mathbf{b}_j$ darzustellen. Der Zeilenrang ist die Anzahl von linear unabhängigen, n-dimensionalen Vektoren $\mathbf{c}_k$, die benötigt werden, die Zeilenvektoren als Linearkombinatonen der $\mathbf{c}_k$ darzustellen. Es gilt: Zeilenrang = Spaltenrang = Rang.

Es gilt $\text{rg}(A) \leq \min(m, n)$. Sei $n = \min(m, n)$. Gilt $\text{rg}(A) = n$, so hat A "vollen Rang".

Es sei $C = AB$. Dann gilt $\text{rg}(C) \leq \min(\text{rg}(A), \text{rg}(B))$.

Der Rang der Matrix $\mathbf{xy}' = (x_i y_j)$ ist gleich 1. Denn

$$\begin{aligned} \mathbf{xy}' = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} (y_1, y_2, y_3) &= \begin{pmatrix} x_1y_1 & x_1y_2 & x_1y_3 \\ x_2y_1 & x_2y_2 & x_2y_3 \end{pmatrix} \\ &= \left(y_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, y_2 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, y_3 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) \end{aligned}$$

Das dyadische Produkt ist eine Matrix, deren Spaltenvektoren alle kollinear sind.

Fragen:

(a) Der Zeilenrang ist gleich dem Spalterang unter der Voraussetzung, dass die Matrix A sich in der Form $A = UV$ darstellen läßt, wobei U eine $(m, r)-$ und V eine (r, n)-Matrix ist.

falsch

(b) Der Rang kann höchstens gleich dem Spaltenrang sein, weil es stets weniger Variable als Fälle gibt.

falsch

(c) Der Rang r einer Matrix ist nur definiert, wenn zumindest einige der Zeilen- und Spaltenvektoren linear unabhängig sind.

falsch

(d) r bezieht sich nur auf die Spaltenvektoren von A, da die Spalten einer Matrix stets die Variablen repräsentieren.

falsch

(e) r ist stets größer als die kleinere der beiden Zahlen m und n.

falsch

(f) Zeilen- gleich Spaltenrang $\leq \min(m, n)$.

korrekt

(g) Die (m, n)-Matrix X habe den Rang $r \leq \min(m, n)$. Dann existiert eine (m, r)-Matrix U und eine (r, n)-Matrix V derart, dass $X = UV$.

korrekt

(h) Die(m, n) Matrix X habe den Rang r. Dann existiert eine (r, m)-Matrix U und eine (n, r)-Matrix V derart, dass $X = UV$.

falsch

(i) Die(m, n) Matrix X habe den Rang r. Dann existiert eine (m, r)-Matrix U und eine (r, n)-Matrix V derart, dass $X = UV$.

korrekt

5. **Quadratische Formen:** M sei eine symmetrische (n, n)-Matrix, $\mathbf{x}$ n-dimensionaler Vektor. $\mathbf{x}'M\mathbf{x} = k$ heißt 'Quadratische Form':

   $\mathbf{x}'M\mathbf{x} = \sum_{i=1} m_{ii}x_i^2 + 2\sum_{i<j} m_{ij}x_ix_j = k$

   $\mathcal{E}_x = \{\mathbf{x}|\mathbf{x}_M\mathbf{x} = k\}$ ist ein Ellipsoid.

   $\mathcal{E}_y = \{\mathbf{y}|\mathbf{y}'\Lambda\mathbf{y} = k\}$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ ist ein *achsenparalleles* Ellipsoid (keine Regression).

Es gelte $\mathbf{x} = T\mathbf{y}$, T eine Rotation. Dann folgt $\mathbf{x}'M\mathbf{x} = \mathbf{y}'T'MT\mathbf{y} = k$ und $T'MT = \Lambda$. Es folgt $TT'MT = MT = T\Lambda$. Für den Spaltenvektor $\mathbf{t}$ aus T folgt $M\mathbf{t} = \lambda\mathbf{t}$, d.h. die $\mathbf{t}$ sind *Eigenvektoren* von M, λ die zugehörigen Eigenwerte.

Es sei $\mathbf{t}$ ein Eigenvektor von M. M kann als Transformationsmatrix aufgefasst werden: $M\mathbf{t} = \lambda\mathbf{t}$ heißt, dass M die Orientierung von $\mathbf{t}$ invariant läßt.

Da T eine Rotationsmatrix ist, ist T orthonormal. Es folgt, dass die Eigenvektoren einer symmetrischen Matrix orthonormal sind.

Auch nicht-symmetrische Matrizen haben Eigenvektoren, die aber nicht notwendig auch orthogonal sind.

Es sei $\mathrm{rg}(M) = r$, $M' = M$. Dann hat M r von Null verschiedene Eigenwerte. Alle Eigenwerte sind größer/gleich Null, wenn M positiv semidefinit ist. M ist positiv semidefinit, wenn M von der Form $M = X'X$ ist, X eine (m, n)-Matrix (z.B. X ist eine Datenmatrix).

Kovarianz- und Korrelationsmatrizen sind positiv semidefinit, haben also nur Eigenwerte größer/gleich Null.

Satz von Courant-Fischer: Der Rayleigh-Quotient $\lambda = \frac{\mathbf{x}_A\mathbf{x}}{\mathbf{x}'\mathbf{x}}$, A symmetrisch, wird maximal, wenn $\lambda = \lambda_1$ der erste (= maximale) Eigenwert von A ist und $\mathbf{x} = \mathbf{t}_1$ der zugehörige Eigenvektor ist.

Fragen:

(a) $\mathbf{x}'M\mathbf{x} = k$ ist eine quadratische Form nur dann, wenn $M' = M$, d.h. wenn M symmetrisch ist.

korrekt

(b) Gilt $\mathbf{x}'M\mathbf{x} = k$, so ist $\mathbf{x}$ ein Eigenvektor von M.

falsch

(c) $\mathbf{x}$ ist nur dann ein Eigenvektor von M, wenn $M\mathbf{x} = \lambda\mathbf{x}$ gilt

korrekt

(d) Es sei $M' = M$ eine symmetrische (n, n)-Matrix. Unter einer quadratischen Form versteht man

i. Die Matrix $M^2 = m \cdot M$.

falsch

ii. Die Gleichung $M\mathbf{x} = \|\mathbf{x}\|^2$:

falsch

iii. Den Ausdruck $\mathbf{x}'M\mathbf{x} = k \in \mathbb{R}$, weil

iv. $\mathbf{x}'M\mathbf{x} = \sum_i m_{ii}x_i^2 + 2\sum_{i<j} m_{ij}x_i x_j = k$ und die Summe der Exponenten der Vektorkomponenten stets gleich 2 ist.

korrekt

v. stets ein Ellipsoid definiert.

falsch

(e) $M = C$ sei eine Kovarianzmatrix. Dann ist C nur dann eine Transformationsmatrix, wenn $M\mathbf{x} = \mathbf{y}$ eine Transformation des Vektors $\mathbf{x}$ in den Vektor $\mathbf{y}$ ist.

falsch

(f) T sei eine (n, n)-Matrix derart, dass $\mathbf{y} = T\mathbf{x}$ und $\|\mathbf{x}\| = \|\mathbf{y}\|$.

i. T ist orthonormal, so dass $T'T = I$.

korrekt

ii. T kann einen Rang kleiner als n haben.

falsch

iii. T läßt die Skalarprodukte für transformierte Vektoren invariant.

korrekt

(g) Es sei M symmetrisch und des gelte $\mathbf{x}'M\mathbf{x} = \mathbf{y}'\Lambda\mathbf{y} = k$, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Weiter gelte $\mathbf{x} = T\mathbf{y}$ mit $\|\mathbf{x}\| = \|\mathbf{y}\|$. Dann gilt

i. $T'MT = \Lambda$

korrekt

ii. $MT = T\Lambda$, T Matrix der Eigenvektoren, Λ Diagonalmatrix der Eigenwerte von M

korrekt

iii. $MT = T\Lambda$ nur dann, wenn $T = I$ die Einheitsmatrix.

falsch

iv. $MT = T\Lambda$ nur dann, wenn M den Rang ($\text{rg}(M) = n$) hat,

falsch

v. Hat M den Rang $r < n$, können alle Eigenwerte größer als Null sein.

falsch

6. **Bestimmung einer Basis für eine Matrix X:** Finde eine Matrix S derart, dass $XS = L$, L orthogonal, und es existiert eine Matrix T' mit $X = LT'$. Die Spaltenvektoren von L sind Basisvektoren für die Spaltenvektoren von X.

   Es gibt zwei Möglichkeiten:

(a) Definiere S als Rotationsmatrix. Dann folgt, dass $S = T$ die Matrix der Eigenvektoren von $X'X$ ist und $\lambda_1 = \mathbf{L}_1\mathbf{L}_1$ die maximale Varianz der Koordinaten der Fälle auf erster latenter Achse hat.

(b) Definiere die Spaltenvektoren $\mathbf{L}_k$ von L so, dass $\mathbf{L}_1'\mathbf{L}_1$ maximale Varianz hat. Dann folgt, dass $S = T'$ die Eigenvektoren von $X'X$ enthält und eine Rotation der Spaltenvektoren $\mathbf{x}_j$ erzeugen.

Der Fall (a): Ansatz: $XS = L$, L orthogonal, S eine Rotation.

$L'L = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) = S'(X'X)S$. S Rotation impliziert $S'S = I$ die Einheitsmatrix, und dann

$SS'(X'X)S = S\Lambda \Rightarrow S = T$, T Matrix der Eigenvektoren, Λ die Matrix der Eigenwerte.

Welche Orientierung der $L = [\mathbf{L}_1, \ldots, \mathbf{L}_n]$ liefert maximale Varianz der Koordinaten (Projektionen) der Fälle auf die Achsen?

Satz von Courant-Fischer:

$$\frac{\mathbf{s}'(X'X)\mathbf{s}}{\mathbf{s}'\mathbf{s}} = \frac{\mathbf{L}'\mathbf{L}}{\|\mathbf{s}\|^2} = \lambda = \max$$

wenn $\lambda = \lambda_1$ der erste Eigenwert und $\mathbf{s}/\|\mathbf{s}\| = \mathbf{t}_1$ der korrespondierende erste Eigenvektor ist, so dass $\|\mathbf{t}_1\|^2 = 1$.

Also $S = T$ die Matrix der orthonormalen Eigenvektoren, Λ Eigenwerte.

Der Fall (b): Ansatz: $XS = L$, $\mathbf{L}_1'\mathbf{L}_1$ soll maximal sein, $\mathbf{L}_2'\mathbf{L}_1 = 0$ (Orthogonalität). Dann wieder $S'(X'X)S = \Lambda$, und der Satz von Courant-Fischer liefert die Aussage, dass $S = T$ die Matrix der Eigenvektoren von $X'X$ und Λ die zugehörige Matrix der Eigenwerte; die Orthonormalität der Eigenvektoren impliziert, dass $S = T$ eine Rotationsmatrix ist.

Fragen:

(a) Der Ansatz $X = LT'$ mit orthogonaler Matrix L macht nur Sinn, wenn die Komponenten von X keine MEssfehler enthalten, da er eine rein algebraische Aussage ist.

falsch

(b) Die Annahme, dass L orthogonal sein soll, impliziert, dass T die Matrix der Eigenvektoren von $X'X$ ist.

korrekt

(c) Man muß annehmen, dass in der Forderung $XS = L$ die Matrix S eine Rotation repräsentiert.

falsch

(d) Der Ansatz $X = LT'$ ist doof, da trivial.

falsch

(e) Der Ansatz $X = LT'$ ist sinnlos, da das Psychische nicht algebraisch sein kann.

falsch

7. **Singularwertzerlegung (SVD):**

$\mathbf{L}_j'\mathbf{L}_j = \|\mathbf{L}_j\|^2 = \lambda_j \Rightarrow \|\mathbf{L}_j\|\sqrt{\lambda_j}$,

Normalisierung der $\mathbf{L}_j$: Multiplikation mit $1/\sqrt{\lambda_j} = \lambda_j^{-1/2} \Rightarrow Q = L\Lambda^{-1/2} \Rightarrow Q\Lambda^{1/2} = L$.

$X = LT' = Q\Lambda^{1/2}T' = \text{SVD}$.

Fragen:

(a) Die PCA ist eine Anwendung der SVD

korrekt

(b) Die SVD gilt nur, wenn $X = LT'$ postuliert werden kann.

falsch

(c) Die SVD gilt nur bei intervallskalierten Messwerten.

falsch

8. **Approximation Faktorenanalyse (PCA):** L enthält die Koordinaten der Fälle, – aber man ist im Allgemeinen an den Variablen interessiert. Die SVD erlaubt zwei Interpretationen:

$$X = Q\Lambda^{1/2}T' = \begin{cases} LT', & \text{Fokus auf Fälle} \\ QA', & A = T\Lambda^{1/2} \text{ "Ladungen", Fokus auf Variablen} \end{cases}$$

A: Zeilen = Variablen, Spalten = latente Dimensionen, a_{jk} Ladung der j-ten Variablen auf der k-ten Dimension.

Warum nicht anders herum? Q ist (m,n) – m Fälle, n latente Variablen. $X = QA' \Rightarrow x_{ij} = \tilde{\mathbf{q}}_i'\tilde{\mathbf{a}}_j$, $\tilde{\mathbf{q}}_i$ i-ter Zeilenvektor von Q, $\tilde{\mathbf{a}}_j$ j-ter Zeilenvektor von A = j-ter Spaltenvektor von A', und j steht für die j-te Variable. $\tilde{\mathbf{a}}_j = (a_{j1}, a_{j2}, \ldots, a_{jn})$, und die Komponenten a_{jk} stehen für die k-te Dimension bei der j-ten Variablen.

Es sei $X = Z$ (spaltenstandardisiert). Dann ist $R = \frac{1}{m} Z'Z$ die Matrix der Korrelationen, $Z = QA' \Rightarrow R = \frac{1}{m} AQ'QA' = \frac{1}{m} AA'$. Korrelation zwischen den Variablen u und v: $r_{uv} = \tilde{\mathbf{a}}_u \tilde{\mathbf{a}}_v = \sum_{k=1}^{n} a_{uk} a_{vk}$

$r_{jj} = 1 = \sum_{k=1}^{n} a_{jk}^2 \Rightarrow$ Endpunkte der Ladungsvektoren für die Variablen liegen auf n-dimensionaler Hyperkugel, z.B. auf einem Kreis, wenn es nur 2 latente Dimensionen gibt.

Aus $Z = QA'$ folgt $Z' = AQ' \Rightarrow ZQ = A$, d.h. $a_{jk} = \mathbf{z}_j' \mathbf{q}_k$, d.h. die Ladung a_{jk} der j-ten Variablen auf der k-ten latenten Dimension entspricht einer Korrelation zwischen der j-ten Variablen und der k-ten Dimension.

Fragen:

(a) Die Komponenten des Spaltenvektors $\mathbf{a}_k$ sind die Ladungen der Variablen auf der k-ten latenten Dimension.

korrekt

(b) Der Zeilenvektor $\tilde{\mathbf{a}}_j$ enthält die Ladungen der j-ten Variablen auf den latenten Dimensionen.

korrekt

(c) Die Summe der Quadrate der Komponenten von $\tilde{\mathbf{a}}_j$ (d.h. $\|\tilde{\mathbf{a}}_j\|^2$, ist stets gleich 1.

korrekt

(d) $\|\mathbf{a}_k\|^2 = \lambda_k =$ der k-te Eigenwert von $Z'Z$.

korrekt

(e) $\|\mathbf{L}_k\|^2 = \lambda_k$

korrekt

Abschätzung der Anzahl latenter Dimensionen: Üblicherweise der Scree-Test. Aber mit Vorsicht anwenden: Auch wenn die Variablen alle statistisch unabhängig sind, sind die empirischen Korrelationen zufällig (um Null) verteilt es kann mehrere Eigenwerte deutlich größer als 1 geben (Kaiser-Kriterium). Deswegen erst Sphärizitäts-Test anwenden!!! (Der nimmt die multivariate Normalverteilung der Daten an, aber diese Annahme ist i.A. approximativ gerechtfertigt.)

Fragen:

(a) Das Kaiser-Kriterium erlaubt stets eine korrekte Abschätzung der bedeutungsvollen Eigenwerte.

falsch

(b) Der Sphärizitätstest ist vernachlässigbar, da nur bei multivariat normalverteilten Variablen anwendbar.

falsch, zumindest nicht ganz korrekt

(c) Exponentiell abfallende Eigenwerte im Scree-Plot legen stochastische Unabängigkeit der Variablen nahe.

So einfach ist das nicht! Aber manchmal ist es so.

Biplot: Die latenten Dimensionen sind dieselben für Fälle und Variablen. Dieser Sachverhalt legt nahe, sowohl die Fälle wie die Variablen in ein und demselben Plot zu repräsentieren. Schließlich ist x_{ij} bzw-z_{ij} (Messwert des i-ten Falls bei der j-ten Variablen) durch das Skalarprodukt

$z_{ij} = q_{i1}a_{j1} + q_{i2}a_{j2} + \cdots + q_{in}a_{jn}$

gegeben. Dabei sind aber die q_{ik} standardisierte Variablen, die a_{jk} dagegen skalierte Variablen: t_{jk} mit $\sqrt{\lambda_k}$ multipliziert. Möglichkeit:

$Z = Q\Lambda^{1/2}T' = (Q\Lambda^{1/4})(\Lambda^{1/4}T') = LA'$ mit

$L = Q\Lambda^{1/4}$, $A = T\Lambda^{1/4}$. Man erhält eine Repräsentation der Beziehungen zwischen Fällen und Variablen.

Fragen:

(a) Ein Biplot ist die Darstellung von Fällen und Variablen bzw. von Zeilen- und Spaltenkategorien in einem einzigen Plot.

korrekt

(b) Der Biplot ist nur möglich, wenn die Daten auf einer Absolutskala (Häufigkeiten, vergl. Korrespondenzanalyse) gemessen werden.

falsch

(c) Der Biplot ist eine mentale Verirrung der Statistiker, da im realen Leben jeder fall ganzheitlich erfasst werden muß.

falsch

Rotationen: $Z = Q\Lambda^{1/2}T' = QSS'\Lambda^{1/2}T' = Q\Lambda^{1/2}SS'T'$.

Fragen:

(a) Bei der PCA machen Rotationen der Achsen keinen Sinn, weil dann die Unabhängigkeit der latenten Variabeln verloren geht (PCA: die orientierte Konfiguration der Fälle wird in ein achsenparalleles Koordinatensystem transformiert, – Nullregression!)

falsch

(b) Man kann entweder die Fälle, repräsentiert in der Matrix L, rotieren, verliert dann aber die Beziehung zu den Variablen oder umgekehrt: man rotiert die Variablen, hat aber keinen Bezug mehr zu den Fällen. In beiden Fällen macht der Biplot keinen Sinn mehr.

korrekt

9. **Faktorenanalyse (FA):** Modell $\frac{1}{\sqrt{m}} Z = FA' + \mathbf{e}$. Dann

   $R = \frac{1}{m} Z'Z = AA' + V, \ V = \mathrm{diag}(v_1^2, \ldots, v_n^2)$. $h_j^2 = 1 - v_j^2$ ist die Kommunalität für die j-te Variable, das ist die durch die gemeinsamen Faktoren (common factors) erklärte Varianz der j-ten Variablen.

   $AA' = T\Lambda T' = R_h,\ R = AA' + V$.

   $F = ZT\Lambda^{-1/2} = ZR_h^{-1}A = [\mathbf{F}_1, \ldots, \mathbf{F}_p]$

   $\mathbf{F}_k$ Hauptfaktoren.

   $\lambda_1 + \cdots + \lambda_p = h_1^2 + \cdots + h_p^2$.

   Fragen:

   (a) Vorteil der FA ist die explizite Schätzung der Varianz der Variablen (entspricht der Reliabilität im Sinne der Theorie psychometrischer Tests, bei der PCA wird $\sigma_j^2 = 1$ gesetzt für alle j).

   korrekt

   (b) Ein Vorteil der FA ist die begriffliche Trennung von *common factors* und *unique factors*.

   korrekt

   (c) Der Nachteil der FA ist die Schwierigkeit, die Kommunalität zu schätzen.

   korrekt

   Rolle der Skalen: Insbesondere $(0,1)$-Daten, $r_{uv} = \phi_{uv}$ der Phi-Koeffizient. Variiert nur im vollen Bereich $[-1, +1]$, wenn $p_u = p_v$.

   Fragen:

   (a) $p_u = p_v = .5$ impliziert $\phi_{uv} = .5$

   falsch

   (b) $p_u = p_v = .5 \Rightarrow -1 \leq \phi \leq 1$

   korrekt

   (c) $p_u \neq p_v$ unvermeidlich, wenn die Grundquoten für die 1-Antworten unterschiedlich sind (unterschiedlich schwierige Items).

   korrekt

   (d) $p_u \neq p_v$ impliziert "triviale" Schwierigkeitsfaktoren.

   korrekt

10. **Korrespondenzanalyse:** (Correspondence Anbalysis (CA)) Man geht von 2-dimensionalen Häufigkeitstabellen aus. Sind die gegebenen Tabellen höherdimensional, so kann man sie u. U. durch Neben- oder Untereinanderschreiben in 2-dimensionale Tabellen verwandeln (vergl. die Selbstmorddaten).

    $\chi^2_{(m,n)} = \sum_{i=1}^m \sum_{j=1}^n \frac{(n_{ij}-n_{i+}n_{+j})^2}{n_{i+}n_{+j}}$

    Definiere

    $$x_{ij} = \frac{n_{ij} - n_{i+}n_{+j}}{\sqrt{n_{i+}n_{+j}}}$$

    $X = (x_{ij})$, SVD $X = U_0 \Lambda^{1/2} V_0'$.

    Skalierung: $f_{ik} = u_{ik}\sqrt{\lambda_k / r_i}$, $g_{jk} = v_{jk}\sqrt{\lambda_k / c_j}$. $r_i = \sum_j n_{ij}/N$, $c_j = \sum_i n_{ij}/N$.

    χ^2-Distanzen:

    $d_{ii'} = \sqrt{\sum_{k=1}^p (f_{ik} f_{i'k})^2}$
    $d_{jj'} = \sqrt{\sum_{k=1}^p (g_{jk} - g_{j'k})^2}$

    Relation j-te Zeilenkategorie, k-te Spaltenkategorie:

    $\mathbf{f}_i'\mathbf{g}_j = \|\mathbf{f}_i\| \|\mathbf{g}_j\| \cos\theta_{jk}$

    Fragen:

    (a) Die CA ist eine PCA der Häufigkeiten in einer Kontingenztabelle.
    falsch

    (b) Die CA beruht auf einer Anwendung der SVD auf die Matrix der standardisierten Residuen x_{ij} (wie oben definiert).
    korrekt

    (c) Die Reskalierung der Matrizen U_0 und V_0 liefert den Bezug zu unabhängigen χ^2-Anteilen.
    korrekt

    (d) Zwischen den Fällen und den Variablen kann für die Interpretation im Biplot eine Distanz berechnet werden.
    falsch

11. **PCA-Regression:**

    $\mathbf{y} = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \cdots + b_p\mathbf{x}_p + \mathbf{e} = X\mathbf{b} + \mathbf{e}$

Schätzung von $\mathbf{b}$: $\hat{\mathbf{b}} = (X'X)^{-1}X'\mathbf{y}$.

$(X'X)^{-1} = V\Lambda^{-1}V'$

Eigenschaften der Schätzungen:

$\mathbb{E}(\hat{\mathbf{b}}) = \mathbf{b}$, $\mathrm{Var}(\hat{\mathbf{b}}) = \sigma^2(X'X)^{-1} = \sigma^2 \sum_{k=1}^{p} \frac{\mathbf{v}_k\mathbf{v}'_k}{\lambda_k}$

$\mathrm{Var}(\hat{\mathbf{b}}) = \sum_{j=1}^{p} \frac{v_{kj}^2}{\lambda_j}$

Korrelationen zwischen den Prädiktoren
$\Rightarrow$ kleine Eigenwerte
$\Rightarrow$ große Varianz der Schätzungen!

PCA-Regression:
$X = Q\Lambda^{1/2}T' \Rightarrow \mathbf{y} = Q\Lambda^{1/2}T'\mathbf{b} + \mathbf{e}$

$\mathbf{u} = \Lambda^{1/2}T'\mathbf{b} \Rightarrow \mathbf{y} = Q\mathbf{u} + \mathbf{e}$

$\hat{\mathbf{u}} = (Q'Q)^{-1}Q'\mathbf{u} + Q'\mathbf{e} = Q'\mathbf{u} + Q'\mathbf{e}$.

Fragen:

(a) Substantielle (dh nicht nur zufällig von Null abweichende) Korrelationen zwischen den Prädiktoren erhöhen die Varianz der Schätzungen der Regressionsgewichte.

korrekt

(b) Die Spaltenvektoren von Q (neue Prädiktoren) ensprechen nicht notwendig Variablen in der Realität.

korrekt

Zur letzten Frage: die Spaltenvektoren von Q sind Linearkombinationen der gemessenen Variablen; insofern sind die Q-Variablen Mischungen von realen Variablen. Ob diese Mischungen unabhängig von den gemessenen Variablen gefunden werden können, ist letztlich eine empirische Frage. Man mache sich klar: Es gibt unendlich viele Linearkombinationen der $\mathbf{x}_1, \ldots, \mathbf{x}_n$, – warum sollen sie alle in der Natur vorkommen?

12. **Klassifikation:** $\mathbf{y} = u_1\mathbf{x}_1 + \cdots + u_p\mathbf{x}_p$.

Gesucht: $\mathbf{u} = (u_1, u_2, \ldots, u_p)'$ derart, dass Fälle, die zu verschiedenen Klassen gehören, maximal getrennt werden.

Komponenten des unbekannten Vektors $\mathbf{y}$ sind Koordinaten der Fälle.

Komponenten sollen so angeordnet werden, dass die Variation der Koordinaten *innerhalb* der Gruppen bzw. Kategorien klein ist im Vergleich zur Variation der Mittelwerte der Komponenten für die einzelnen Gruppen bzw. Katgegorien:

$$\lambda = \frac{QS_{zw}(\mathbf{u})}{QS_{inn}(\mathbf{u})} = \max!$$

$QS_{zw}(\mathbf{u}) = \mathbf{u}'B\mathbf{u}$, $QS_{inn} = \mathbf{u}'W\mathbf{u}$

$$\lambda(\mathbf{u}) = \frac{\mathbf{u}'B\mathbf{u}}{\mathbf{u}'W\mathbf{u}} \Rightarrow W^{-1}B\mathbf{u} = \lambda\mathbf{u}$$

Der Rang von $W^{-1}B$ ist gleich der Anzahl der Eigenvektoren $\mathbf{u}_k$ mit zugehörigen Eigenwerten λ_k. Die $\mathbf{y}_k = X\mathbf{u}_k$ sind orthogonal.

Das Verfahren funktioniert insbesondere dann gut, wenn die Konfigurationen der Fälle in den einzelnen Kategorien bzw. Gruppen Ellipsoide sind; sie sollten sich möglichst wenig überlappen. Dann können die Gruppen linear getrennt werden.

Bei korrelierten Prädiktoren ergeben sich ähnliche Probleme wie bei der multiplen Korrelation. Ausweg: Regularisierte Diskriminanzanalyse und verwandte Verfahren (nicht in dieser Vorlesung behandelt).

Fragen:

(a) Lineare Trennbarkeit der Kategorien bedeutet, dass die Konfigurationen der Fälle in den einzelnen Kategorien Ellipsoide sind, – die Daten also multivariat normalverteilt sind.

falsch

(b) Man benötigt minimal so viele Fälle, wie es Prädiktoren gibt, um zu stabilen Schätzungen der $\mathbf{u}$ zu kommen.

falsch[2]

(c) Im Ansatz $\mathbf{y} = X\mathbf{u}$ gibt es keinen Fehlerterm $\mathbf{e}$. Da stets Messfehler auftreten, ist der Mangel eines Fehlerterms ein Konstruktionsfehler der Diskriminanzanalyse.

falsch

(d) Es sei $\bar{\mathbf{y}}_k$ der mittlere Vektor für die k-te Klasse. $|\mathbf{y} - \bar{\mathbf{y}}_k|$ entspricht einer Mahalanobis-Distanz.

korrekt

(e) Die Diskriminanzanalyse entspricht nicht dem Bayesschen Ansatz, nach dem die Wahrscheinlichkeit $P(K|D)$, einen Fall auf der Basis D von Prädiktorwerten der Kategorie zuzuordnen, durch $P(K|D) = P(D|K)P((K)/P(D)$ ggeben ist, so dass a priorie Wahrscheinlichkeiten $P(K)$ für die Kategorien eingehen.
Das macht nichts, da die $P(K)$ sowieso subjektive Größen sind.

falsch

13. **Logistische Regression:** Man kann nach Maßgabe des Bayesschen Theorems klassifizieren:

$$P(K|D) = \frac{P(K \cap D)}{P(D)} = \frac{P(D|H)P(H)}{P(D|H)P(H) + P(D|\neg H)P(\neg H)}$$

K Kategorie, D Daten (Prädiktoren). Division der rechten Seite durch $P(D|H)P(H)$:

$$P(H|D) = \frac{1}{1 + \frac{P(D|\neg H)P(\neg H)}{P(D|H)P(H)}} = \frac{1}{1+q}, \quad q = \frac{P(D|\neg H)P(\neg H)}{P(D|H)P(H)}$$

Es ist $q = e^{\log q}$. Nimmt man an, dass die Prädiktoren multivariat normalverteilt sind, so ergibt sich nach ein wenig Rechnerei

$$\log q = b_1\mathbf{x}_1 + \cdots + b_p\mathbf{x}_p$$

[2] mindestens 4- bis 5-mal so viele!

und man erhält die

$$P(H|D) = \frac{1}{1 + e^{b_1 \mathbf{x}_1 + \cdots + b_p \mathbf{x}_p}}$$

Nach ein wenig Rechnen erhält man

$$\frac{P(H|D)}{1 - P(H|D)} = e^{b_1 \mathbf{x}_1 + \cdots + b_p \mathbf{x}_p}$$

und die Logits

$$\log \frac{P(H|D)}{1 - P(H|D)} = b_1 \mathbf{x}_1 + \cdots + b_p \mathbf{x}_p$$

Schätzung der b_j: Maximum-Likelihood-Methode.