Vektoren und Matrizen

Übungen 4 + Antworten: SVD und PCA

1. **Aufgabe:** Gegeben sei eine $(m \times n)$-Datenmatrix: $m = 200$ Fälle, $n = 50$ Variablen. Was bedeutet es, wenn für diese Matrix gesagt werden kann, dass die 200 Zeilenvektoren in einem 5-dimensionalen Teilraum des $\mathbb{R}^{200}$ liegen? Ist es in diesem Fall möglich, dass die 50 200-dimensionalen Spaltenvektoren in einem 5-dimensionalen Teilraum liegen?

   Antwort: Es bedeutet, dass nur 5 linear unabhängige 200-dimensionale Vektoren als Basisvektoren benötigt werden, um die 50 200-dimensionalen Vektoren als Linearkombination dieser Teilbasis des 200-dimensionalen Raums darzustellen. Da der Rang einer Matrix die Anzahl der lin. unabhängigen Vektoren, die zur Darstellung sowohl der Spalten- wie der Zeilenvektoren benötigt werden angibt, können auch die 200 50-dimensionalen Zeilenvektoren durch nur 5, diesmal aber 50-dimensionale, linear unabhängige Vektoren als Linearkombinationen dargestellt werden.

2. **Aufgabe:** Es sei A eine (10×3)-Matrix. $\mathbf{x}$ und $\mathbf{y}$ seien Vektoren. Betrachten sie das Gleichungssystem $\mathbf{y} = A\mathbf{x}$.
   (a) Welche notwendige Bedingung muß erfüllt sein, damit eine Lösungsvektor $\mathbf{x}$ existiert?
   (b) Wieviele Dimensionen hat der Vektorraum, in dem $\mathbf{x}$ liegt, und wieviele Dimensionen hat der Vektorraum, in dem $\mathbf{y}$ liegt?
   (b) Die Matrix A und der Vektor $\mathbf{y}$ seien vorgegeben. Läßt sich $\mathbf{x}$ dann berechnen? Wenn ja, Welche Voraussetzung muß erfüllt sein?
   (c) Existiert eine zu A inverse Matrix A^{-1}?

   Antwort: (a) $\mathbf{y} = A\mathbf{x}$ ist die Behauptung, dass $\mathbf{y}$ eine Linearkombination der Spaltenvektgoren von A ist. Die notwendige Bedingung besteht darin, dass $\mathbf{y}$ tatsächlich eine Linearkombination dieser Spaltenvektoren ist. So kann A eine $(m \times n)$-Matrix sein, mit $m > n$; es gibt dann mehr Gleichungen als Unbekannte. $\mathbf{y}$ ist notwendig ein m-dimensionaler Vektor, und $\mathbf{x}$ ist n-dimensional. Da $\mathbf{y}$ eine Linearkombination der Spalten von A sein muß, diese Spalten aber nur einen n-dimensionalen Teilraum des $\mathbb{R}^m$ aufspannen können (dies ist die lineare Hülle $\mathcal{L}_s(A)$ der Spalten von A), kann es sein, dass $\mathbf{y}$ *kein* Element dieses Teilraums ist, $\mathbf{y} \notin \mathcal{L}_s(A)$, sondern Element des Komplementärraums von $\mathcal{L}_s(A)$ ist. Dann findet man im ganzen Universum keinen Vektor $\mathbf{x}$, der $\mathbf{y}$ als Linearkombination der Spalten von A darzustellen erlaubt.

Zu (b) $\mathbf{x} \in \mathbb{R}^n$ (= n-dimensionaler Vektorraum), und $\mathbf{y} \in \mathbb{R}^m$.

Zu (c) Damit A^{-1} existiert, muß (i) A quadratisch sein, und (ii) muß A volen Rang haben, d.h. die Zeilen- und Spaltenvektoren von A müssen linear unabhängig sein. Darüber ist aber in der Aufgabenstellung nichts ausgesagt worden, so dass man über die Existenz von A^{-1} nichts aussagen kann.

3. **Aufgabe:** Es seien A und B zwei nicht quadratische Matrizen.
   (a) Welche Voraussetzungen müssen erfüllt sein, damit Sie ein Produkt $C = AB$ berechnen können, und gilt dann $AB = BA$?
   (b) Angenommen, das Produkt C kann berechnet werden. Sind die Spaltenvektoren von C dann Linearkombinationen der Spalten von A oder der Spalten von B?
   (c) A habe den Rang $\text{rg}(A) = r_A$ und B habe den Rang $\text{rg}(B) = r_B < r_A$. Welche allgemeine Aussage läßt sich über den Rang von C machen, falls C berechenbar ist?
   (d) Welche Bedingung muß *notwendig* erfüllt sein, damit die Spaltenvektoren von C Linearkombinationen der Zeilenvektoren von B sind? Die Bedingung muß nicht hinreichend sein. (Hinweis: Die Anzahlen der Zeilen und Spalten von C hängen von den Anzahlen der Zeilen und Spalten von A und B ab!)

   Antwort: Zu (a), Es sei A eine $(m \times n)$-Matrix und B sei eine $(r \times s)$-Matrix. $C = AB$ bedeutet, dass ein Spaltenvektor $\mathbf{c}_k$ von C eine Linearkombination $A\mathbf{b}_k$ der Spalten von A ist, wobei die Komponenten von $\mathbf{b}_k$ die Koeffizienten dieser Linearkombination sind. $\mathbf{b}_k$ muß dann soviele Komponenten haben, wie es Spalten von A gibt, – d.h. die Anzahl der Spalten von A und die Anzahl der Zeilen von B müssen übereinstimmen; in diesem Fall muß also $n = r$ gelten. $AB = BA$ kann aus analogen Gründen höchsten dann gelten, wenn die Anzahl der Spalten von B auch gleich der Anzahl der Zeilen von A ist, d.h. eine notwendige, aber keinesfalls hinreichende Bedingung für $AB = BA$ ist (i) $n = s$ und (ii) $s = m$. Zu (b): Es gelte $C = AB$. Dann sind die Spalten von C Linearkombinationen der Spalten von A, und die Zeilenvektoren von C sind Linearkombinationen der Zeilenvektoren von B.

   Zu (c): $\text{rg}(C) \leq \min(\text{rg}(A), \text{rg}(B))$, d.g. $\text{rg}(C) \leq \text{rg}(B)$.

   Zu (d): Es sei A eine $(m \times n)$-Matrix, B eine $(r \times s)$-Matrix. Ein Spaltenvektor $\mathbf{c}_k$ von C muß dann ein Element der linearen Hülle $\mathcal{L}(A)$ sein, d.h. $\mathbf{c}_k$ muß einer der möglichen Linearkombinationen der Spaltenvektoren von A sein, und dies muß für alle $\mathbf{c}_k$ von C gelten. (Zur Übung können Sie diese Bedingung elaborieren: den Rang von A in Abhängigkeit von der Zeilen- und Spaltenzahl von A betrachten etc)

4. **Aufgabe:** Die $(m \times n)$-Datenmatrix X habe den Rang r. Bekanntlich läßt sich X dann stets als Produkt zweier Matrizen U und V schreiben, d.h. $X = UV$.

(a) Welche Aussage läßt sich dann einerseits über die Anzahl der Zeilen von U und andererseits über die Zeilen und Spalten von V machen?
(b) Was läßt sich über die Ränge von U und V sagen?
(c) Müssen die Spaltenvektoren von U notwendig orthogonal sein?
(d) Welche Bezienung besteht zwischen der SVD von X und der Beziehung $X = UV$?

Antwort: Zu (A): U muß eine $(m \times r)$-Matrix sein, und V muß eine $(r \times n)$-Matrix sein.

Zu (b): $\text{rg}(U) = \text{rg}(V)$.

Zu (c): Nein, lineare Unabhängigkeit genügt.

Zu (d): $X = UV = Q\Lambda^{1/2}$. Da einerseits U eine $(m \times r)$-Matrix ist und Q ebenfalls eine $(m \times r)$-Matrix ist, wenn $\text{rg}(X) = r$, so kann man $U = Q$ oder $U = Q\Lambda^{1/2}$ annehmen ($\Lambda^{1/2}$ skaliert ja nur die Längen der Spaltenvektoren von Q, und diese Skalierung verändert den Rang von Q nicht). Je nach Wahl ist dann $V = T'$ oder $V = \Lambda^{1/2}T'$.

5. **Aufgabe:** Eine $(m \times n)$-Datenmatrix X habe den Rang r.
(a) Ist die Möglichkeit, die Spalten von X als Linearkommbinationen von Basisvektoren darzustellen, an Annahmen über die Wahrscheinlichkeitsverteilung der Elemente x_{ij} gekoppelt?
(b) Welche Dimensionalität haben die Basisvektoren, und welche Dimensionalität hat die lineare Hülle (i) der Spaltenvektoren von X, (ii) der Zeilenvektoren von X?
(c) Wieviele Möglichkeiten haben Sie, Basisvektoren für den Teilraum des $\mathbb{R}^m$ zu wählen, in dem die Spaltenvektoren von X liegen?

Antworten: Zu (a): Nein. Alle Betrachtungen, die bisher in der Vorlesung angestellt worden sind, sind rein algebraischer Natur.

Zu (b): Die Spaltenvektoren sind m-dimensional, und die Basisvektoren für sie muüssen dann ebenfalls m-dimensional sein. Die Basisvektoren für die Zeilenvektoren sind notwendig n-dimensional. Ist der Rang von X gleich r, so hat die lineare Hülle der Spaltenvektoren die Dimensionalitätä r, d.h. sie besteht aus der Menge der Linearkombinationen von r m-dimensionalen Vektoren; eine analoge Aussage gilt für die lineare Hülle der Zeilenvektoren.

Zu (c): Unendlich viele (wenn man die Unterscheidung von abzählbar unendlich und überabzählbar unendlich vielen Möglichkeiten einmal außer Acht läßt).

6. **Aufgabe:** Für die Datenmatrix wird der Ansatz $X = LT'$ gemacht, wobei die Transformation von Vektoren $\mathbf{x}$ gemäß $\mathbf{y} = T\mathbf{x}$ eine Rotation der $\mathbf{x}$ bedeuten soll.
(a) Welche Beziehung besteht zwischen den Zeilenvektoren $\tilde{\mathbf{x}}_i$ von X und den Zeilenvektoren $\tilde{\mathbf{L}}_i$ $(i = 1, \ldots, m)$ von L?

(b) Die Punktekonfiguration der Fälle wird durch die Endpunkte der Vektoren $\tilde{\mathbf{x}}_i$ definiert. Warum liegt der Endpunkt jedes Vektors $\tilde{\mathbf{x}}_i$ auf einem Ellipsoid $\mathcal{E}_i$ und warum haben alle diese Ellipsoide dieselbe Orientierung?
(c) Impliziert die unter (b) genannte Beziehung zwischen Datenpunkten und Ellipsoiden, dass die Daten multivariat normalverteilt sind?
(d) Liegen die Variablen, die durch die Endpunkte der Spaltenvektoren $\mathbf{x}_j$ von X repräsentiert werden, ebenfalls auf Ellipsoiden?

Antworten: Zu (a): $\tilde{\mathbf{x}}_i = T\tilde{\mathbf{L}}_i$. $i = 1, \ldots, m$.
Zu (b): Die quadratische Form $\mathcal{E} = \{\mathbf{x} | M = M', \mathbf{x}'M\mathbf{x} = k\}$ definiert stets ein Ellipsoid, desses Orientierung relativ zu dem Koordinatensystem, in dem die $\mathbf{x}$ liegen, von der symmetrischen Matrix M abhängt. Die für jeden Vektor $\tilde{\mathbf{x}}_i$ läßt sich die quadratische Form $\tilde{\mathbf{x}}_i'(X'X)\tilde{\mathbf{x}}_i = k_i$ betrachten; für alle i werden sie durch dieselbe Matrix $X'X$ definiert, haben also dieselbe Orientierung.
Zu (c): Nein, die Beziehung ist rein algebraischer Natur. Überdies ist die multivariate Normalverteilung durch eine quadratische Form $\mathbf{x}'(X'X)^{-1}\mathbf{x}$ definiert, also durch die Inverse von $X'X$, wobei die $\mathbf{x}$ als zentriert vorausgesetzt werden.
Zu (d): Nein. Denn $X'X$ ist eine $(n \times n)$-Matrix, die $\tilde{\mathbf{x}}_i$ sind ebenfalls n-dimensionale Vektoren, die $\mathbf{x}_j$ (Spaltenvektoren von X) sind aber m-dimensional. Bei der PCA werden die Variablen durch die Ladungen a_{jk} (Ladung der j-ten Variablen auf der k-ten latenten Dimension) repräsentiert, und die Endpunkte der Vektoren, die die Variablen repräsentieren, liegen auf Hyperkugeln mit dem Radius 1.

7. **Aufgabe:** Es gelte wieder $X = LT'$. Die Annahmen, dass (i) die Spaltenvektoren von L sind orthogonal, und (ii) T repräsentiert eine Rotation implizieren, dass die Hauptachsen der unter (c) der vorangegangenen Aufgabe genannten Ellipsoide als neues Koordinatensystem betrachtet werden, dessen Achsen unkorrelierte latente Variablen repräsentieren.
(a) Wie werden die Koordinaten der Datenpunkte auf diesen Achsen berechnet?
(b) Durch welche Eigenschaft sind die Eigenvektoren von $X'X$ charakterisiert?
(c) in welcher Beziehung stehen die Varianzen der Koordinaten der Datenpunkte (Fälle) auf den Hauptachsen der Ellipsoide zu den Eigenwerten von $X'X$?
(d) Für die zentrierten oder standardisierten Messwerte x_{ij} gilt gemäß dem Ansatz $X = LT'$

$$x_{ij} = \begin{cases} q_{i1}a_{j1} + q_{i2}a_{j2} + \cdots + q_{in}a_{jn}, \text{ oder} \\ L_{i1}t_{j1} + L_{i2}t_{j2} + \cdots + L_{in}t_{jn}. \end{cases} \tag{1}$$

Von welcher Umformulierung von $X = LT'$ wurde hier Gebrauch gemacht und welche Zielsetzungen liegt Ihre Entscheidung für eine der beiden Mög-

lichkeiten zugrunde?
(e) Ein Sozialpsychologe berichtet, er habe bei der Analyse einer Datenmatrix X gefunden, dass die Ränge von X und L gleich r seien, T aber einen Rang $s < r$ habe. Ein pädagogischer Psychologe stimmt dem Befund zu, weil er bei einer Datenanalyse gefunden habe, X und T hätten denselben Rang r gehabt, L aber habe einen Rang $s > r$ gehabt; die Anzahl der Dimensionen für die Fälle einerseits und die Variablen andererseits könnten also verschieden sein. Können Sie Bedingungen angeben, unter denen die beiden Forscher recht haben?

Antworten: Zu (a): $X = LT'$; T eine Rotation impliziert $T'TR = I$, so dass $XT = LT'T = L$. Hat man T identifiziert, so läßt sich L und damit die Koordinaten der Fälle auf den "neuen" Koordinaten berechnen. Weiter: $X'X = TL'LT' = T\Lambda T'$, Λ eine Diagonalmatrix wegen der vorausgesetzten Orthogonalität der Spalten von L, und dann $(X'X)T = T\Lambda T'T = T\Lambda$, woraus folgt, dass T die Matrix der Eigenvektoren von $X'X$ sein muß. T läßt sich berechnen (Verfahren der numerischen Mathematik).
Zu (b): Die Eigenvektoren sind orthogonal und werden als normalisiert betrachtet, d.h. $\|\mathbf{t}_k\| = 1$ für alle $k = 1, \ldots, n$. Diese Normalisierung ergibt sich im Rahmen der hier betrachteten Herleitung aus der Annahme, dass T eine Rotation repräsentiernen soll, – dann muß T orthonormal sein. Andererseits: es sei $\mathbf{t}$ ein Eigenvektor von $X'X$. dann ist auch $\mathbf{s} = \mu\mathbf{t}$, $\mu \in \mathbb{R}$ ein Eigenvektor, denn $(X'X)\mathbf{s} = \lambda\mathbf{s}$, λ der zu $\mathbf{t}$ gehörende Eigenwert. Das Wesentliche am Eigenwert ist, dass $(X'X)\mathbf{s}$ und $\}lambda\mathbf{s}$ dieselbe Orientierung haben. Natürlich kürzt sich μ sofort aus der Gleichung $(X'X)\mathbf{s} = \lambda\mathbf{s}$ heraus, denn

$$(X'X)\mathbf{s} = (X'X)\mu\mathbf{t} = \mu(X'X)\mathbf{t} = \mu\lambda\mathbf{t} \Rightarrow (X'X)\mathbf{t} = \lambda\mathbf{t},$$

d.h. die Setzung $\|\mathbf{t}\| = 1$ ist keine Einschränkung der Allgemeinheit.
Zu (c): $\lambda_k = \mathbf{L}_k'\mathbf{L}_k = \|\mathbf{L}_k\|^2$ ist proportional (= gleich bis auf den Faktor $1/m$) zur Varianz der L_{ik} ($\sum_i L_{ik}^2$), weil die $\mathbf{L}_k$ zentriert sind (die Komponenten haben den Mittelwert Null), wenn die $\mathbf{x}_j$ zentriert sind.
Zu (d): Von der SVD $X = Q\Lambda^{1/2}T'$, $L = Q\Lambda^{1/2}$ Im Fall $x_{ij} = \sum_k q_{ik}a_{jk}$ wird $A = T\Lambda^{1/2}$ gesetzt, $A = (a_{jk})$, im Fall $x_{ij} = \sum_k L_{ik}t_{jk}$ wird $L = Q\Lambda^{1/2}$ gesetzt. IM ersten Fall fokussiert man auf die Struktur der Variablen, im zweiten Fall auf die der Fälle ("Typen").
Zu (e): Angenommen, der Sozialpsychologe habe recht. da $X = LT'$ vorausgesetzt wird, folgt, dass $\mathrm{rg}(X) = \mathrm{rg}(LT')$ ist. Allgemein gilt aber, dass $\mathrm{rg}(LT') \leq \min(\mathrm{rg}(L), \mathrm{rg}(T'))$ ist. Wäre also $\mathrm{rg}(T') = s < r$, so würde $\mathrm{rg}(LT') = s < r$ sein, im Widerspruch zur Aussage $\mathrm{rg}(X) = \mathrm{rg}(LT')$, die gelten *muß*, weil die Matrizen X und LT' ja identisch sind. Also kann die Behauptung des Sozialpsychologen nicht korrekt sein, – er hat sich verrechnet oder den Ausdruck des von ihm benutzten Programms falsch gelesen. Für den pädagogischen Psychologen gilt ein analoges Argument: Er sagt,

$\text{rg}(X) = \text{rg}(T)$, aber $\text{rg}(L) = s > r$. Es muß aber $\text{rg}(X) = \text{rg}(L'T)$ bzw $\text{rg}(XT) = \text{rg}(L)$ gelten. Es ist aber $\text{rg}(XT) \leq \min(\text{rg}(X), \text{rg}(T)) = r$, wegen $\text{rg}(X)X = \text{rg}(T)$, so dass $\text{rg}(L) = s > r$ unmöglich ist. Auch der pädagogische Psychologe ht sich geirrt. Es existieren keine Bedingungen, unter denen zumindest einer der beiden Forscher recht haben kann.

8. **Aufgabe:** Die rechte Seite der Gleichungen (1) enthält lauter unbekannte Grössen.
   (a) Sie werden mit der Methode der Kleinsten Quadrate geschätzt.
   (b) Sie lassen sich direkt aus den Daten ausrechnen.
   (c) Sie lassen sich aus Hypothesen über die Beziehungen zwischen den Variablen herleiten.

   Antworten: Zu (a): Nein.
   Zu (b): Ja
   Zu (c): Nein.

9. **Aufgabe:** Für eine gegebene Datenmatrix X berechnen Sie die Ladungen der Variablen.
   (a) In welcher Beziehung stehen die Vektoren $\mathbf{a}_j$ und $\mathbf{a}_k$ der Ladungen für die Variablen j und k zu den Korrelationen r_{jk}?
   (b) Welche allgemeine Interpretation für die Ladung a_{jk} (j-te Variable, k-te latente Dimension) kennen Sie?
   (c) Welche Implikation hat die Tatsache, dass $r_{jj} = 1$ für alle Variablen $j = 1, \ldots, n$, für die Position der Endpunkte der Ladungsvektoren $\mathbf{a}_j$?
   (d) Der Rang von X sei $r < n$; welchem geometrischen Gebilde entsprechen die Endpunkte der $\mathbf{a}_j$ in diesem Fall?

   Antworten: Zu (a): $r_{jk} = \sum_k a_{js} a_{ks}$:
   Zu (b): a_{jk} ist die Korrelation zwischen der j-ten Variablen und der k-ten latenten Dimension.
   Zu (c): Die Endpunkte der Ladungsvektoren für die Variablen liegen auf einer Hyperkugel mit dem Radius 1.
   Zu (d): Einer r-dimensionalen Hyperkugel mit dem Radius 1.

10. **Aufgabe:** X sei wieder eine $(m \times n)$-Datenmatrix. Welche der folgenden Aussagen ist korrekt, und wenn ja, warum?, und wenn nicht, warum nicht?
    (a) Die SVD $X = Q\Lambda^{1/2}T'$ läßt sich nur berechnen, wenn $m > n$.
    (b) Die Matrix A der Ladungen ist stets quadratisch.
    (c) Welcher Matrix entspricht das Produkt AA', – können Sie Ihre Antwort herleiten?
    (d) Welcher Matrix entspricht das Produkt $A'A$, – können Sie Ihre Antwort herleiten?

    Antworten: Zu (a): falsch
    Zu (b): Im fehlerfreien Fall kann $r < n$ gelten, A ist dann eine $(n \times r)$-Matrix: Numerisch findet man aber $r = n$, so dass im Allgemeinen eine

quadratische Matrix ausgegebebn wird.
Zu (c): $X = QA' \Rightarrow X'X = AQ'QA' = AA'$, da $Q'Q = I$. $X'X$ isgt die Matrix der Kovarianzen oder der Korrelationen.
Zu (d): $A = T\Lambda^{1/2} \Rightarrow A'A = \Lambda^{1/2}T'T\Lambda^{1/2} = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.