Name, Vorname

Matrikelnummer

# Modulklausur 31821 – Multivariate Verfahren

| | |
|---|---|
| Datum | |
| Punkte | |
| Note | |

**Termin:** 22. März 2019, 11:30 - 13:30 Uhr
**Prüfer:** Univ.-Prof. Dr. H. Singer

## Hinweise zur Bearbeitung der Modulklausur 31821

1. Füllen Sie zunächst den **Kopf des Deckblatts** aus!
2. Es können insgesamt 80 Punkte erreicht werden. Bei Erreichen von 40 Punkten ist die Klausur bestanden. **Bitte kontrollieren Sie sofort, ob Sie ein vollständiges Klausurexemplar erhalten haben**.
3. Kursmaterialien des Kurses 00883 sind als Hilfsmittel zugelassen, sofern sie **keine handschriftlichen Ergänzungen** enthalten. Als Kursmaterialien gelten lediglich Lehrtexte, nicht jedoch alte Klausuren, Einsendearbeiten oder Musterlösungen.
4. Die Benutzung von Taschenrechnern ist nur gestattet, wenn das betreffende Modell
   - nicht programmierbar ist,
   - keine Differentiation, Integration und Matrixoperationen ermöglicht,
   - keine Texte oder Formeln speichern kann,
   - nicht drahtlos mit anderen Geräten kommunizieren kann,
   - über keine alphanumerische Tastatur verfügt,
   - kein grafisches Display (z.B. zur Darstellung von Funktionsgraphen) besitzt.
5. Bitte benutzen Sie für Ihre Rechnungen nur die beigefügten Lösungsbögen.
6. Wenn Sie die einzelnen Blätter der Klausur voneinander trennen, **vermerken Sie auf jedem Blatt Ihre Matrikelnummer**. Legen Sie bitte am Ende der Klausur die Blätter wieder zusammen.
7. Vergessen Sie nicht, die Klausur auf der letzten bearbeiteten Seite zu **unterschreiben**.
8. Falls Sie fertig sind, so geben Sie die Klausur vollständig ab (**Aufgaben** + Lösungsbogen!).

Wir wünschen Ihnen viel Erfolg!

## Aufgabe 1: Multiple Choice (15 Punkte)

Kreuzen Sie die richtige Antwort an!

a) Sei $\mathbf{x}$ eine $n$-dimensionale reelle Zufallsvariable. Ist $\mathbf{x}$ normalverteilt mit Erwartungswert $\boldsymbol{\mu}$ und positiv definiter Kovarianzmatrix $\boldsymbol{\Sigma}$, so hat die Wahrscheinlichkeitsverteilung von $\mathbf{x}$ die Dichtefunktion:

☐ $f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\}$

☐ $f(\mathbf{x}) = \begin{cases} \lambda \exp\{-\lambda x_1\} & \mathbf{x} \geq \mathbf{0} \\ 0 & \mathbf{x} < \mathbf{0} \end{cases}$ für $\lambda \in \mathbb{R}_{>0}$

☐ $f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n}} \log(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))$

(3 P.)

b) Das multiple lineare Regressionsmodell (MLR) hat die folgende Form:

☐ $Y_n = \alpha + \beta X_n - \epsilon_n \quad , n = 1, \ldots, N$

☐ $Y_n = \beta_0 + \beta_1 X_{n1} + \ldots \beta_q X_{nq} + \epsilon_n \quad , n = 1, \ldots, N$

☐ $Y_n = \beta_1 X_1 + \ldots \beta_q X_q \quad , n = 1, \ldots, N$

(3 P.)

c) Gegeben sei ein MLR für $\mathbf{y} = (Y_1, \ldots, Y_n)'$ mit Gewichtsvektor $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_q)'$. Um herauszufinden, ob Regressoren zur Erklärung der abhängigen Variable beitragen, so überprüft man die Hypothese $H_0 : \beta_1 = \ldots = \beta_q = 0$. Ein entsprechender F-Test für die Nullhypothese hat die Gestalt:

☐ $F = \frac{N-q-1}{q} \frac{R^2}{1-R^2}$

☐ $F = \frac{q}{N-q-1} \frac{1-R^2}{R^2}$

☐ $F = \frac{R^2}{q}(N - q - 1)$ (3 P.)

d) Die $L_q$-Distanz hat folgende Form:

☐ $d_q(x_n, x_m) = (\sum_{i=1}^{p} |x_{ni} - x_{mi}|^q)^{1/q}$

☐ $d_q(x_n, x_m) = (\sum_{i=1}^{p} |x_{ni} - x_{mi}|^{1/q})^q$

☐ $d_q(x_n, x_m) = (\sum_{i=1}^{q} |x_{ni} - x_{mi}|^q)^2$ (3 P.)

e) Die folgende Aussage ist richtig:

☐ $L_q$-Distanzen sind skaleninvariant, jedoch nicht translationsinvariant.

☐ $L_q$-Distanzen sind translationsinvariant, jedoch nicht skaleninvariant. (3 P.)

*Hinweis* : Für jede korrekte Antwort erhalten Sie 3 Punkte. Jede nicht oder falsch beantwortete Frage ergibt 0 Punkte.

## Aufgabe 2: Rechenregeln für Zufallsvektoren (15 Punkte)

Es seien $\mathbf{x}$, $\mathbf{y}$ und $\mathbf{z}$ stetige Zufallsvektoren und $\mathbf{A}, \mathbf{B}, \mathbf{c}$ Konstanten mit passenden Dimensionen. Beweisen Sie:

a) $E[\mathbf{A}\mathbf{x} + \mathbf{c}] = \mathbf{A}\, E[\mathbf{x}] + \mathbf{c}$ (3 P.)

b) $\mathrm{Var}(\mathbf{x} + \mathbf{y}) = \mathrm{Var}(\mathbf{x}) + \mathrm{Cov}(\mathbf{x}, \mathbf{y}) + \mathrm{Cov}(\mathbf{y}, \mathbf{x}) + \mathrm{Var}(\mathbf{y})$ (3 P.)

c) $\mathrm{Var}(\mathbf{A}\mathbf{x} + \mathbf{c}) = \mathbf{A}\, \mathrm{Var}(\mathbf{x})\mathbf{A}'$ (3 P.)

d) $\mathrm{Cov}(\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}) = \mathbf{A}\, \mathrm{Cov}(\mathbf{x}, \mathbf{y})\mathbf{B}'$ (3 P.)

e) $\mathrm{Cov}(\mathbf{x} + \mathbf{y}, \mathbf{z}) = \mathrm{Cov}(\mathbf{x}, \mathbf{z}) + \mathrm{Cov}(\mathbf{y}, \mathbf{z})$ (3 P.)

*Hinweis* : Es ist $\mathrm{Var}(\mathbf{x}) := E[\mathbf{x}\mathbf{x}'] - E[\mathbf{x}]E[\mathbf{x}']$ und $\mathrm{Cov}(\mathbf{x}, \mathbf{y}) := E[\mathbf{x}\mathbf{y}'] - E[\mathbf{x}]E[\mathbf{y}']$.

## Aufgabe 3: Agglomerative Verfahren (20 Punkte)

a) Definieren Sie die euklidische Distanz und beschreiben Sie die Single-Linkage-Methode. Nach welchem Kriterium werden die Klassen miteinander fusioniert? (5 P.)

b) Gegeben seien die Objekte **A**, **B**, **C**, **D** und **E**, die jeweils durch vier Merkmale beschrieben werden. Die entsprechende Datenmatrix **X** hat dabei folgende Gestalt:

$$\mathbf{X} = \begin{bmatrix} 2 & 2 & 3 & 1 \\ 4 & 1 & 2 & 5 \\ 1 & 1 & 1 & 3 \\ 3 & 5 & 4 & 2 \\ 5 & 4 & 3 & 1 \end{bmatrix}$$

*Hinweis* : Zeile 1 enthält die Merkmale von Objekt A, Zeile 2 die von Objekt B usw.

Konstruieren Sie agglomerativ eine Hierachie. Verwenden Sie dabei die Single-Linkage-Methode, indem Sie zunächst die euklidische Distanzmatrix bilden. Geben Sie stets die Klassen und den Indexwert der Fusion, sowie die Distanzmatrix an. Zeichnen Sie zudem das entsprechende Dendogram. (15 P.)

## Aufgabe 4: Maximum-Likelihood-Schätzer (30 Punkte)

a) Es sei $X$ poisson-verteilt mit Parameter $\lambda$. Leiten Sie den Maximum-Likelihood-Schätzer (MLS) $\hat{\lambda}$ für $\lambda$ her. Beweisen oder widerlegen Sie zudem: $\hat{\lambda}$ ist Erwartungstreu. Berechnen Sie den Schätzwert für die Stichprobe $x_1 = 2$, $x_2 = 3$, $x_3 = 8$ und $x_4 = 13$.

*Hinweis*: Die Dichtefunktion der Poissonverteilung ist gegeben durch $f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$.

Weiter ist $x! = 1 \cdot 2 \cdot \ldots \cdot x$ (Bsp: $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$). (10 P.)

b) Es sei $\mathbf{X}' = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ eine Datenmatrix der Größe $n \times p$, wobei $\mathbf{x}_i$ ein $p-$dimensionaler reeller Zufallsvektor ist. Zudem seien die $\mathbf{x}_i$ unabhängig multivariat normalverteilt mit Parameter $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$, also $\mathbf{x}_i \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Leiten Sie die MLS von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ her.

*Hinweis*: Verwenden Sie den Ansatz aus dem Skript und nutzen Sie $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathrm{tr}(\mathbf{A}\mathbf{x}\mathbf{x}')$. (20 P.)

## LÖSUNGSBOGEN

Punkte

## LÖSUNGSBOGEN

Punkte

## LÖSUNGSBOGEN

Punkte

# LÖSUNGSBOGEN

Punkte

# LÖSUNGSBOGEN

Punkte

Punkte

## LÖSUNGSBOGEN

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | |

Punkte

Punkte

# LÖSUNGSBOGEN

Punkte

# LÖSUNGSBOGEN

Punkte

## LÖSUNGSBOGEN

Punkte

# LÖSUNGSBOGEN

Punkte

# LÖSUNGSBOGEN

Punkte

## LÖSUNGSBOGEN

Punkte

# LÖSUNGSBOGEN

Punkte

Punkte

# LÖSUNGSBOGEN

Punkte

# LÖSUNGSBOGEN

Punkte

# LÖSUNGSBOGEN

Punkte

# LÖSUNGSBOGEN

Punkte