90-minütige Klausur zur Vorlesung „Fortgeschrittene mathematische Methoden in der Statistik“ (Wintersemester 2024/25)

Prof. Dr. Thomas Nagler, Jana Gauss Klausur am 24.03.2025

Vorname:	
Nachname:	
Matrikelnummer:	
Studiengang + PO:	
Unterschrift:	

- **Überprüfen Sie, ob Ihre Angabe vollständig ist.** Sie sollte **4** Aufgaben auf **10** Seiten umfassen.
- Füllen Sie bitte das obenstehende Formular umgehend aus. Schreiben Sie bitte leserlich.
- Halten Sie für die Kontrolle bitte einen Lichtbildausweis und Ihren Studierendenausweis bereit.
- Die Bearbeitungszeit beträgt **90 Minuten**. Es ist keine vorzeitige Abgabe möglich.
- Verwenden Sie für die Bearbeitung ausschließlich die Ihnen zur Verfügung gestellten Papierbögen. Kennzeichnen Sie jedes zur Abgabe vorgesehene Blatt mit Namen und Matrikelnummer.
- Alle Lösungen müssen nachvollziehbar sein. Dazu gehört auch lesbare Schrift. Ergebnisse müssen klar erkennbar und zuzuordnen sein. Falls der Platz auf einem Blatt nicht ausreicht, muss auf dem betreffenden Blatt ein Verweis zur Lösung zu finden sein.
- **Machen Sie kenntlich, falls Sie mit einem Ersatzergebnis weiterrechnen!**
- Es dürfen nur **dokumentenechte Stifte** und keine roten Stifte verwendet werden.
- Zugelassene Hilfsmittel: 1 beidseitig (analog oder digital) beschriftetes DIN A4-Papier
- Bei Unterschleif erfolgt eine Meldung an das Prüfungsamt. Sie sind verpflichtet, durch Ihr Verhalten jegliche Missverständnisse diesbezüglich auszuschließen. Sorgen Sie insbesondere dafür, dass sich keinerlei Mobiltelefone und Uhren an Ihrem Arbeitsplatz befinden.
- Verlassen Sie den Prüfungsraum erst, nachdem Sie der Aufsicht die Klausur persönlich übergeben haben. Für den Eingang der kompletten Klausur bei der Aufsicht sind Sie selbst verantwortlich.

	mögliche Punkte	erreichte Punkte
1. Aufgabe	17	
2. Aufgabe	10	
3. Aufgabe	10	
4. Aufgabe	8	
Summe	45	

NAME:

Aufgabe 1

(a) Gib eine Funktion $f\colon \mathbb{R}^2 \to \mathbb{R}$ an, die konvex, aber nicht strikt konvex ist (ohne Begründung). (2 Punkte)

Lösung:
Zum Beispiel lineare Funktionen wie $f(\boldsymbol{x}) = 0$, $f((x_1, x_2)^T) = x_1$ etc.

(b) Beweise, dass die Menge aller stochastischen $n \times n$ Matrizen konvex ist. Eine stochastische Matrix ist eine Matrix $A \in \mathbb{R}^{n\times n}$ mit den Eigenschaften

(i) $a_{ij} \geq 0$ für alle $i, j = 1, \dots, n$ und

(ii) $\sum_{j=1}^n a_{ij} = 1$ für alle $i = 1, \dots, n$ (d.h. alle Zeilensummen sind 1). (4 Punkte)

Lösung:
Seien $A, B \in \mathbb{R}^{n\times n}$ stochastische Matrizen und $\lambda \in [0, 1]$ beliebig. Zu zeigen: die Matrix $C = \lambda A + (1-\lambda)B$ ist eine stochastische Matrix. Es gilt

$$c_{ij} = \lambda a_{ij} + (1-\lambda)b_{ij} \geq 0 \quad \text{für alle } i, j = 1, \dots, n$$

denn $a_{ij}, b_{ij}, \lambda, (1-\lambda) \geq 0$ per Definition, somit ist Eigenschaft (i) erfüllt.

$$\sum_{j=1}^n c_{ij} = \sum_{j=1}^n \lambda a_{ij} + (1-\lambda)b_{ij} = \lambda \sum_{j=1}^n a_{ij} + (1-\lambda)\sum_{j=1}^n b_{ij} = \lambda + 1 - \lambda = 1$$

für alle $i = 1, \dots, n$, somit ist auch Eigenschaft (ii) erfüllt und C ist eine stochastische Matrix.

(c) Berechne die Ableitung $\frac{\partial f(A)}{\partial A} \in \mathbb{R}^{m\times n}$ der Abbildung

$$f\colon \mathbb{R}^{m\times n} \to \mathbb{R}, A \mapsto \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2.$$

(3 Punkte)

Lösung:

$$\frac{\partial f(A)}{\partial A} = \begin{pmatrix} \partial f(A)/\partial a_{11} & \cdots & \partial f(A)/\partial a_{m1} \\ \vdots & \ddots & \vdots \\ \partial f(A)/\partial a_{1n} & \cdots & \partial f(A)/\partial a_{mn} \end{pmatrix} = \begin{pmatrix} 2a_{11} & \cdots & 2a_{m1} \\ \vdots & \ddots & \vdots \\ 2a_{1n} & \cdots & 2a_{mn} \end{pmatrix} = 2A.$$

(d) Berechne die relative Konditionszahl von $f(x) = \frac{1}{x}$. (3 Punkte)

Lösung:
Es gilt

$$\frac{|f(\tilde{x}) - f(x)|}{|f(x)|} = \frac{\left|\frac{1}{\tilde{x}} - \frac{1}{x}\right|}{\left|\frac{1}{x}\right|} = \left|\frac{x - \tilde{x}}{\tilde{x}x}\right| |x| = \frac{|x|}{|\tilde{x}|} \frac{|\tilde{x} - x|}{|x|}.$$

Da $|x|/|\tilde{x}| \to 1$ für $\tilde{x} \to x$, ist $\kappa_{rel} = 1$.

(e) Finde für die gegebenen Folgen a_n jeweils eine möglichst einfache Folge b_n, sodass $a_n = O(b_n)$.

(i) $a_n = n^{10} + 10^n + \exp(n)$

(ii) $a_n = \frac{3}{n} + \frac{5}{\ln(n)}$

(iii) $a_n = c_n + d_n + 4$, wobei $c_n = o(1)$ und $d_n = O(1)$

(iv) Gilt die Aussage $a_n = c_n + \frac{1}{n} = O(1/n)$, wenn wir nur $c_n = o(1)$ wissen? Begründe deine Antwort. (5 Punkte)

Lösung:

(i) $n^{10} + 10^n + \exp(n) = O(10^n)$

(ii) $\frac{3}{n} + \frac{5}{\ln(n)} = O(1/\ln(n))$

(iii) $c_n + d_n + 4 = o(1) + O(1) + 4 = O(1)$

(iv) Nein, im Allgemeinen gilt die Aussage nicht, denn c_n könnte langsamer gegen 0 konvergieren als $1/n$, z.B. $c_n = 1/\sqrt{n}$. Dann gilt $a_n = O(1/\sqrt{n})$, aber nicht $a_n = O(1/n)$.

2+4+3+3+5=17 PUNKTE

NAME:

Aufgabe 2

(a) Zeige mithilfe der Jensen-Ungleichung, dass

$$\frac{1}{n}\sum_{i=1}^{n} x_i \leq \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

für alle $x_1, \dots x_n \in \mathbb{R}$ gilt.

Hinweis: Die Jensen-Ungleichung besagt, dass für jede konvexe Funktion $f : \mathcal{X} \to \mathbb{R}$

$$f\left(\sum_{i=1}^{n} w_i \boldsymbol{x}_i\right) \leq \sum_{i=1}^{n} w_i f(\boldsymbol{x}_i) \quad \forall n \in \mathbb{N}, \boldsymbol{x}_i \in \mathcal{X}, w_i \geq 0, \sum_{i=1}^{n} w_i = 1$$

gilt. (5 Punkte)

Lösung:

Da die Funktion x^2 monoton wachsend ist, kann man beide Seiten der Ungleichung quadrieren und die Aussage

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 \leq \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

beweisen. Anwenden der Jensen-Ungleichung mit $f(x) = x^2$ (konvex) und $w_i = 1/n$ liefert

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = f\left(\sum_{i=1}^{n} w_i x_i\right) \leq \sum_{i=1}^{n} w_i f(x_i) = \frac{1}{n}\sum_{i=1}^{n} x_i^2.$$

Alternative: Da für jede konkave Funktion g gilt, dass $-g$ konvex ist, gilt die Jensen-Ungleichung für konkave Funktionen g mit $\geq$, also

$$g\left(\sum_{i=1}^{n} w_i \boldsymbol{x}_i\right) \geq \sum_{i=1}^{n} w_i g(\boldsymbol{x}_i).$$

Die Funktion $g(x) = \sqrt{x}$ ist konkav, es gilt somit (mit $w_i = 1/n$)

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} w_i g(x_i^2) \leq g\left(\sum_{i=1}^{n} w_i x_i^2\right) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}.$$

NAME:

(b) Sei $\mathcal{F}$ ein Vektorraum von Funktionen $f\colon \mathcal{X} \to \mathbb{R}$ ($\mathcal{X}$ ist eine beliebige Menge) mit $\|f\| = \sup_{x\in\mathcal{X}} |f(x)|$ und sei $\tilde{x} \in \mathcal{X}$ fest. Wir betrachten die Abbildung

$$\Phi\colon \mathcal{F} \to \mathbb{R}, f \mapsto f(\tilde{x}).$$

(i) Zeige, dass die Abbildung $\Phi(f)$ konvex ist.

(ii) Was ist die Fréchet-Ableitung $D_f\Phi(h)$ der Abbildung $\Phi(f)$? Begründe deine Antwort. Zeige auch die Beschränktheit der Ableitung, d.h. zeige, dass es ein $M < \infty$ gibt, sodass $|D_f\Phi(h)| \le M\|h\| \quad \forall h \in \mathcal{F}$.

(5 Punkte)

Lösung:

(i) Seien $f_1, f_2 \in \mathcal{F}$ und $\lambda \in [0,1]$ beliebig. Es gilt

$$\begin{aligned}\Phi(\lambda f_1 + (1-\lambda)f_2) &= (\lambda f_1 + (1-\lambda)f_2)(\tilde{x}) = \lambda f_1(\tilde{x}) + (1-\lambda)f_2(\tilde{x}) \\ &= \lambda\Phi(f_1) + (1-\lambda)\Phi(f_2),\end{aligned}$$

somit ist die Abbildung konvex.

(ii) Aus der ersten Teilaufgabe ist ersichtlich, dass Φ eine lineare Funktion ist. Somit gilt $D_f\Phi(h) = \Phi(h)$. Die Beschränktheit der Abbildung folgt aus der Definition der Norm:

$$|D_f\Phi(h)| = |\Phi(h)| = |h(\tilde{x})| \le \sup_{x\in\mathcal{X}} |h(x)| = \|h\| \quad \text{(somit } M = 1\text{)}$$

5+5=10 PUNKTE

NAME:

Aufgabe 3

(a) Gesucht ist das Minimum der Funktion

$$f\colon \mathbb{R}^p \to \mathbb{R}, \boldsymbol{\beta} \mapsto (\boldsymbol{Y} - X\boldsymbol{\beta})^T(\boldsymbol{Y} - X\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}^T\boldsymbol{\beta}$$

mit $\boldsymbol{Y} \in \mathbb{R}^n, X \in \mathbb{R}^{n\times p}$ und einem Skalar $\lambda > 0$. Berechne die Lösung $\hat{\boldsymbol{\beta}}$ des Optimierungsproblem.

Hinweis: Du musst die Definitheit der Hessematrix nicht prüfen. Matrizen der Form $X^TX + \lambda I_p$ sind immer invertierbar. (5 Punkte)

Lösung:
Der Gradient ($\in \mathbb{R}^p$) von $(\boldsymbol{Y} - X\boldsymbol{\beta})^T(\boldsymbol{Y} - X\boldsymbol{\beta})$ ist $-2X^T(\boldsymbol{Y} - X\boldsymbol{\beta})$ (Produktregel oder Ausmultiplizieren und Kettenregel, siehe Übung Nr. 8), der Gradient von $\lambda\boldsymbol{\beta}^T\boldsymbol{\beta}$ ist $2\lambda\boldsymbol{\beta}$ (quadratische Form), somit

$$\nabla f(\boldsymbol{\beta}) = -2X^T\boldsymbol{Y} + 2X^TX\boldsymbol{\beta} + 2\lambda\boldsymbol{\beta}.$$

Nullsetzen liefert

$$\mathbf{0} \overset{!}{=} -X^T\boldsymbol{Y} + X^TX\boldsymbol{\beta} + \lambda\boldsymbol{\beta} = -X^T\boldsymbol{Y} + (X^TX + \lambda I_p)\boldsymbol{\beta}$$
$$\Leftrightarrow X^T\boldsymbol{Y} \overset{!}{=} (X^TX + \lambda I_p)\boldsymbol{\beta} \Leftrightarrow \hat{\boldsymbol{\beta}} = (X^TX + \lambda I_p)^{-1}X^T\boldsymbol{Y}.$$

NAME:

(b) Das Minimum $\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$ einer zweimal stetig differenzierbaren Funktion $f: \mathbb{R}^n \to \mathbb{R}$ soll durch ein iteratives Verfahren bestimmt werden. Dabei soll in der $k+1$-ten Iteration $f(\boldsymbol{x})$ an der Stelle $\boldsymbol{x}_k$ quadratisch approximiert werden und $\boldsymbol{x}_{k+1}$ ist das Minimum dieser quadratischen Approximation. Leite die Iterationsvorschrift her, d.h.

1. Schreibe $f(\boldsymbol{x}_k + \boldsymbol{h})$ als Taylorapproximation zweiter Ordnung.
2. Bestimme das $\boldsymbol{h}$, das die Approximation minimiert, und setze $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{h}$.

Hinweis: Du darfst annehmen, dass die Hessematrix $H_f(\boldsymbol{x})$ immer positiv definit ist. (5 Punkte)

Lösung:

1.
$$f(\boldsymbol{x}_k + \boldsymbol{h}) \approx f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{h} + \frac{1}{2}\boldsymbol{h}^\top H_f(\boldsymbol{x}_k)\boldsymbol{h} \quad (\text{mit } \nabla f(\boldsymbol{x}_k) \in \mathbb{R}^n)$$

2. Ableiten und Nullsetzen liefert (wobei genutzt wird, dass die Hessematrix symmetrisch ist)
$$\boldsymbol{0} \stackrel{!}{=} \nabla f(\boldsymbol{x}_k) + H_f(\boldsymbol{x}_k)\boldsymbol{h} \quad \Leftrightarrow \quad \boldsymbol{h} = -H_f(\boldsymbol{x}_k)^{-1}\nabla f(\boldsymbol{x}_k),$$
wobei $H_f(\boldsymbol{x}_k)$ positiv definit $\implies$ $H_f(\boldsymbol{x}_k)$ invertierbar genutzt wird. Es ist ein Minimum, da $H_f(\boldsymbol{x}_k)$ positiv definit ist. Somit
$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - H_f(\boldsymbol{x}_k)^{-1}\nabla f(\boldsymbol{x}_k).$$

5+5=10 PUNKTE

NAME:

Aufgabe 4

Seien $A \in \mathbb{R}^{n\times n}$ und $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$ und $k, n \in \mathbb{N}$ mit $1 \leq k \leq n$. Angenommen wir wissen, dass A besonders strukturiert ist, so dass $a_{ij} = 0$ für alle $i < j$ und alle $i \geq j + k$. Eine solche Matrix A könnte mit $n = 6, k = 3$ zum Beispiel so aussehen:

$$\begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 1 & 1 & 0 & \dots & \dots & 0 \\ 1 & 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

Gib einen möglichst effizienten Algorithmus zum Lösen des LGS $A\boldsymbol{x} = \boldsymbol{b}$ an und leite dessen Laufzeitkomplexität in Abhängigkeit von n und k her. Wie verhält sich die Komplexität im Vergleich zur Situation, wenn

(i) über die Struktur von A nichts bekannt ist,

(ii) A eine untere Dreiecksmatrix ist.

(8 Punkte)

Lösung:

Das LGS kann durch Vorwärtssubstitution gelöst werden, wobei die besondere Struktur von A zu einer schnelleren Laufzeit führt:
Da jede Zeile maximal k Einträge ungleich 0 hat, sind bei der Rückwärtssubstitution in jeder der n Zeilen maximal $O(k)$ Elementaroperationen (und nicht $O(n)$) nötig, somit ist die Laufzeitkomplexität $O(nk)$.

Alternative: Der obere $k \times k$ Block von A ist eine obere Dreiecksmatrix, das Lösen durch Vorwärtssubstitution benötigt $O(k^2)$ Operationen. In den letzten $n - k$ Zeilen sind jeweils $O(k)$ Operationen nötig, die Laufzeit ist somit insgesamt $O(k^2 + (n - k)k) = O(nk)$.

Wenn die Struktur von A nicht bekannt ist, beträgt die Laufzeit $O(n^3)$, z.B. mit der LU-Zerlegung, und ist somit auf jeden Fall schlechter als im vorliegenden Fall.

Wenn A eine untere Dreiecksmatrix ist, beträgt die Laufzeit $O(n^2)$ für die Vorwärtssubstitution. Diesen Fall erhält man für $k = n$. Für $k < n$ ist die Komplexität im vorliegenden Fall geringer.

8=8 PUNKTE

NAME:

Zusatzblatt

NAME:

Zusatzblatt