Zufallsmatrizen und hochdimensionale Wahrscheinlichkeitstheorie
Seminar im Wintersemester 2020/21

Markus Reiß
Humboldt-Universität zu Berlin

22. Januar 2021

Organisation

Termine: Mi 17.3. 9-13 Uhr, Do 18.3. 9-15 Uhr, Fr 9-13 Uhr

Das Seminar findet im Rahmen der digitalen Lehre statt. Jede/r Vortragende sollte eine 45-minütige Präsentation sowie eine schriftliche Ausarbeitung mit Resultaten, Beispielen und Beweistechniken von 5 bis 8 Seiten vorbereiten. Die Präsentation kann vorher aufgenommen oder per zoom vorgetragen werden. Ziel ist es, einen Überblick über die wichtigsten Resultate, Techniken und Beispiele zu geben sowie an ausgewählten Stellen bis in formale Details zu gehen. Erfahrungsgemäß sind Videovorträge länger als 45min schwer verdaubar, daher sollte die schriftliche Ausarbeitung ggf. auch zusätzliches Material aufnehmen und gut erklären.

Um die Diskussionskultur eines Seminars auch digital zu fördern, gibt es zu jedem Vortrag einen vorher bestimmten Diskutanten, der sich im Vorfeld mit dem Sprecher austauscht (ihm liegen Folien und Ausarbeitung vor!), auf etwa drei Folien seine Sicht des Vortrags zusammenfasst und auf dieser Grundlage in der Regel zwei Fragen stellt. Typischer Ablauf ist 45min Vortrag (gerne asynchron als Video), 5min Diskutant plus Sprecherantworten (zoom), 10min allgemeine Fragen und Diskussion (zoom).

Zeitplan für Vortragende: Folien und Ausarbeitung $\geqslant 6$ Tage vor dem Vortrag an die Diskutantin und an eine betreuende Mitarbeiterin; Austausch mit Diskutantin sowie Erstellung endgültiger Versionen und Hochladen von Ausarbeitung und Vortrag/Folien $\geqslant 1$ Tag vor dem Vortragstermin.
Zeitplan für Diskutanten/innen: $\geqslant 6$ Tage vor dem Vortrag Erhalt von Folien und Ausarbeitung, $\geqslant 3$ Tage vor dem Vortrag Senden eigener vorläufiger Folien an die Vortragende mit Bitte um Feedback, $\geqslant 1$ Tag vor dem Vortrag Hochladen der endgültigen Folien.

Details zu Hochladen, Videokonferenz, betreuenden Mitarbeitern und weitere Informationen erfolgen Anfang März.

Literatur

[T] Joel Tropp, An Introduction to Matrix Concentration Inequalities. Foundations and Trends in Machine Learning, 2015, `https://arxiv.org/abs/1501.01571`

[V] Roman Vershynin, High dimensional probability. An introduction with applications in Data Science, CUP, 2020, `https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.html#`

[W] Martin Wainwright, High-dimensional Statistics, CUP, 2019, `https://www.cambridge.org/core/books/highdimensional-statistics/8A91ECEEC38F46DAB53E9FF8757C7A4E` (Ebook bei HU-Zugang)

[BLM] Stéphane Boucheron, Gábor Lugosi, Pascal Massart. Concentration inequalities. OUP, 2013 `https://stephane-v-boucheron.fr/publication/zb-math-06586491`

Das Seminar orientiert sich an den einführenden Büchern [V] und [T]. Sehr viel weitergehende probabilistische Resultate finden sich in [BLM], während [W] sowohl die Theorie entwickelt als auch spannende Beispiele aus Statistik und Lerntheorie bietet.

Vortragsthemen

Angegeben sind jeweils die Abschnitte in den Büchern [V] und [T]. Für den Vortrag muss eine Auswahl des dort angegebenen Materials erfolgen. Gleichzeitig sollte weitere Literatur (s.o. und eigene Recherche) hinzugezogen werden für Diskussionen, Beispiele oder Ausblicke. Leitmotiv: der/die Vortragende weiß viel mehr als er/sie präsentiert.

1. Concentration for sums of random variables (V2.1–V2.6):
   short motivation, Hoeffding inequality, Chernoff inequality, subgaussianity, extended Hoeffding inequality, Khintchine inequality

2. Concentration for sums of random variables: (V2.7–V2.8):
   subexponentiality, Orlicz spaces, Bernstein inequality
   Stefanie Hesse (D: Thibaud Hadamczik)

---

3. High-dimensional random vectors (V3.1, V3.3, V3.4):
   concentration of norm, examples, subgaussianity in high dimensions
   Antigoni Todoulou (D: Nils Mattiss)

4. Bounding the matrix norm (V4.1, V4.2, V4.4):
   ε-net, covering number, norm of subgaussian matrices
   Lea Jacobs (D: Leon Ochmann)

5. Bounding the matrix norm (V4.5–V4.7):
   community detection, spectral clustering, two-sided bounds, covariance estimation, clustering
   Thibaud Hadamczik (D: Antigoni Todoulou)

---

6. Matrix Laplace Transform Method (parts of T3, T8):
   matrix moment generating function, tail bounds for eigenvalues, master bound, proof of Lieb's Theorem
   Nils Mattiss (D: Lea Jacobs)

7. Sums of bounded random matrices (parts of T6):
   matrix Bernstein inequality, proof, applications and examples
   Leon Ochmann (D: Stefanie Hesse)