Mathematische Semesterberichte (2025) 72:73–76
https://doi.org/10.1007/s00591-025-00386-9

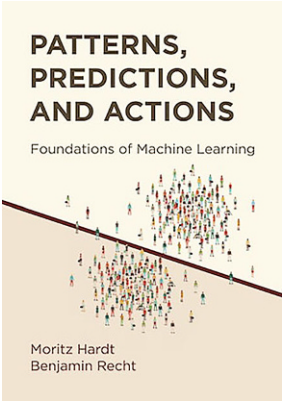
BUCHBESPRECHUNG – BOOK REVIEW

Moritz Hardt, Benjamin Recht: Patterns, Predictions, and Actions

Princeton University Press 2022, 320 Seiten, ISBN 9780691233734/pbk, 9780691233727/ebbok

Felix Voigtlaender

Angenommen: 20. Januar 2025 / Online publiziert: 19. Februar 2025



Das Buch *Patterns, Predictions, and Actions* befasst sich mit der mathematischen Beschreibung und Analyse von Verfahren des *Machine Learning* und speziell des *Deep Learning*, geschrieben von zwei Koryphäen auf diesem Gebiet. Dabei wird im Vergleich zu anderen Büchern besonderes Augenmerk auf die aus Sicht der klassischen Theorie teilweise überraschenden empirischen Beobachtungen der letzten ca. 10 Jahre gelegt und es wird diskutiert, inwieweit diese Beobachtungen mittlerweile mathematisch verstanden sind.

Die Literatur im Bereich des Machine Learning reicht von rein angewandten „Programmierbüchern“ [1, 2] wie *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*, die sich mit der praktischen Umsetzung mittels der einschlägigen Softwarebibliotheken befassen, über Bücher, die die Verfahren mathematisch sauber beschreiben [3, 4], aber keine mathematische Theorie (in Form von Sätzen und Beweisen) beinhalten, bis hin zu theoretischen, mathematisch anspruchsvollen Büchern [5, 6] wie *Foundations of Machine Learning*, in denen es um rigoros beweisbare Aussagen über klassische Algorithmen und Techniken des Machine Learning (z. B. logistische Regression, SGD, SVMs, Regularisierung, PAC learning, VC dimension) geht.

In diesem Kontext fällt *Patterns, Predictions, and Actions* zwischen die zweite und dritte Kategorie: Das Buch ist mathematisch anspruchsvoll (Kenntnisse min-

✉ Felix Voigtlaender
Mathematisches Institut für Maschinelles Lernen und Data Science (MIDS), Katholische Universität Eichstätt-Ingolstadt, Auf der Schanz 49, 85049 Ingolstadt, Deutschland
E-Mail: Felix.Voigtlaender@ku.de

destens im Umfang der ersten drei Semester eines Mathematikstudiums an einer deutschen Universität sollten vorhanden sein), die besprochenen Verfahren werden mathematisch exakt beschrieben und es werden mathematische Sätze formuliert und bewiesen. Im Gegensatz zu rein theoretischen Büchern geht es aber nicht um „Theorie der Theorie willen", sondern es werden in den allermeisten Fällen zuerst wichtige empirische Beobachtungen erläutert, bevor besprochen wird, inwieweit sich diese empirischen Resultate mathematisch erklären lassen. Die behandelten mathematischen Resultate stammen somit des Öfteren aus der neueren Forschungsliteratur, wobei die Beweise in der Literatur oftmals lang und technisch sind; deshalb beschränkt sich das Buch in diesen Fällen darauf, Beweisskizzen zu geben und für die vollständigen Beweise auf die Originalliteratur zu verweisen. Dies wird auch bereits im Vorwort des Buches erwähnt:

> In contemporary learning theory important results often have short sketches, yet making these arguments rigorous and precise may require dozens of pages of technical calculations. [...] We aim to strike a balance, [...] frequently referring readers to the relevant literature for full details.

Das Verständnis dieser Beweisskizzen setzt ein gewisses Maß an mathematischer Reife, bzw. eine gut entwickelte Intuition voraus.

Im Folgenden möchte ich die Besonderheiten des Buchs *Patterns, Predictions, and Actions* ausführlicher beschreiben:

- Andere mathematische Bücher im Bereich des Machine Learning diskutieren zumeist die *klassische Theorie des Machine Learning*, wie z. B. die Theorie der *VC Dimension* und Resultate bezüglich *(S)GD* für Verfahren, bei denen der Loss eine konvexe Funktion der Modell-Parameter ist. Dann wird üblicherweise nur kurz erwähnt, dass die modernen Verfahren des *Deep Learning* überraschend erfolgreich sind, obwohl sie die Annahmen der klassischen Theorie keineswegs erfüllen, oder obwohl die klassische Theorie nur triviale Aussagen für sie liefern würde (weil z. B. in der modernen Praxis die Anzahl der Parameter und/oder die VC Dimension größer ist als die Anzahl der Trainingssample).
  Im Gegensatz zu diesen Büchern liegt der Fokus im Buch *Patterns, Predictions, and Actions* darauf, die modernen Verfahren des Machine Learning (insbesondere Deep Learning) und ihre empirisch beobachteten Eigenschaften zu verstehen. Hierzu werden oftmals Resultate aus der jüngeren Forschung diskutiert, welche die klassischen Techniken geschickt auf die modernen Verfahren anwenden, aber teilweise auch neue Techniken entwickeln. Insbesondere wird ein spezielles Augenmerk auf sogenannte *überparametrisierte Modelle* (also Modelle mit mehr Parametern als Trainingsdaten) gelegt; siehe zum Beispiel Seite 87ff und Seite 116. Dies ist zentral, da im Wesentlichen alle modernen Deep Learning Modelle in diese Kategorie fallen. In diesem Kontext wird auch die sogenannte *Neural Tangent Kernel* Theorie diskutiert (siehe Seite 140). Weiterhin wird erwähnt, dass die Größe (Norm) der Gewichte eines neuronalen Netzes relevanter für Fragen der Generalisierung ist als die reine Anzahl der Gewichte (siehe Seite 138). Außerdem wird auch das Phänomen der *impliziten Regularisierung durch die Wahl des Optimierungsalgorithmus* angesprochen; siehe Seite 91f und Seite 120.

- Es wird ausführlich diskutiert, wie das Paradigma, Datensätze wie ImageNet als „Benchmark“ zu verwenden, zu den Fortschritten der letzten 10–15 Jahre im Machine Learning beigetragen hat.
  In diesem Kontext wird mathematisch analysiert, inwiefern sich die Praxis bezüglich der Handhabung von Datensätzen im Machine Learning rechtfertigen lässt. So hat es sich im Machine Learning durchgesetzt, sorgfältig erstellte Datensätze als Grundlage für Wettbewerbe (Competition/Challenge) zu verwenden. Im besten Fall ist der Datensatz hierbei in einen öffentlich verfügbaren Trainingsdatensatz und zwei nicht veröffentlichte Testdatensätze aufgeteilt; der erste Testdatensatz wird während der „Laufzeit“ des Wettbewerbs benutzt, um eine Rangliste der Teilnehmer zu erstellen; der endgültige Gewinner nach Ablauf des Wettbewerbs wird dann mittels des zweiten Testdatensatzes ermittelt.
  Dies kann dazu führen, dass die Teilnehmer des Wettbewerbs implizit „auf den Testdaten trainieren“, indem sie die Struktur ihres Modells so lange variieren, bis sie eine gute Platzierung in der Rangliste erreichen. Selbst der zweite Testdatensatz wird bei großen Wettbewerben zur Evaluierung von einer großen Zahl von Modellen benutzt, so dass nicht ohne Weiteres klar ist, ob das Ergebnis des Gewinners des Wettbewerbs repräsentativ für das Verhalten auf neuen Daten ist.
  Bei vielen beliebten Datensätzen ist sogar der Testdatensatz öffentlich verfügbar, so dass das Problem des „Trainierens auf den Testdaten“ mit der Zeit immer schwerwiegender wird, da die gleichen Daten zur Evaluierung von mehr und mehr Modellen verwendet werden.
  Die sich ergebenden Fragestellungen werden in Kapitel 8 des Buches diskutiert. Insbesondere wird gezeigt, dass das Problem deutlich weniger schwerwiegend ist, wenn man nur daran interessiert ist, dass die Leistungsfähigkeit *des bisher besten Modells* akkurat gemessen wird („*Ranglisten-Prinzip*“).
- Das Buch beinhaltet zwei Kapitel die sich mit „Causality“ und „Causal Inference“ beschäftigen. Diese Themen werden in den meisten anderen Büchern zum Thema Machine Learning ausgelassen.
- Das Buch spricht die potenziellen negativen Aspekte von großen Datensätzen und von Machine Learning an, wie zum Beispiel die Verewigung von Stereotypen, potenzielle Probleme bezüglich Datenschutz, Fragen des Copyrights in Bezug auf generative künstliche Intelligenz, sowie die Nutzung von Machine Learning für Anwendungen wie Targeted Advertising mit potenziell gesellschaftlich fragwürdigen Auswirkungen. In den Worten der Autoren:

  > Commercially [the] most successful use cases [of machine learning] to date are targeted advertising and digital content recommendation, both of questionable value to society.

In der vorliegenden ersten Auflage enthält das Buch noch zahlreiche kleine Tippfehler und manche andere Ungenauigkeiten. So sind zum Beispiel sowohl die Aussage als auch der Beweis einer Version des Neyman-Pearson Lemmas (siehe Lemma 2

in Kapitel 2) nicht korrekt[1]. Auf der Webseite zum Buch (https://mlstory.org/) finden sich Übungsaufgaben (welche im Buch selber leider nicht enthalten sind), eine Liste von Errata, sowie eine kostenlos erhältliche Preprint-Version des Buchs.

Insgesamt ist das Buch eine exzellente, hochaktuelle und sehr lesbare Quelle für Leser mit dem nötigen mathematischen Hintergrund, die sich eine Übersicht über die wichtigsten modernen Fragestellungen und den Stand der Forschung im Bereich der mathematischen Analyse von Machine Learning Methoden verschaffen wollen. So kann ich das Buch allen angehenden Doktorandinnen und Doktoranden in diesem Bereich empfehlen. Bezüglich der mathematischen Details der Resultate und ihrer Beweise wird man aber in den meisten Fällen nicht umhin kommen, die Originalliteratur zu konsultieren. Dies wird durch das gelungene Literaturverzeichnis und die „Chapter Notes“ am Ende jedes Kapitels erleichtert.

Funding Open Access funding enabled and organized by Projekt DEAL.

Literatur

1. Chollet, F.: Deep Learning with Python. Manning Publications, second edition (2021)
2. Géron, A.: Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc., second edition (2022)
3. Prince, S.J.D.: Understanding Deep Learning. MIT Press (2023)
4. Bishop, C.M., Bishop, H.: Deep learning. Foundations and concepts. Springer, Cham (2024)
5. Shalev-Shwartz, S., Ben-David, S.: Understanding machine learning. From theory to algorithms. Cambridge University Press, Cambridge (2014)
6. Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT Press, Cambridge, MA, second edition (2018)

[1] Ein Gegenbeispiel ist gegeben durch $X \sim \mathcal{N}(0,1)$ stochastisch unabhängig von $Y \sim \mathrm{Ber}(1/2)$; allgemeiner ist jeder Fall problematisch, bei dem ein $a > 0$ existiert für das $p(x \mid y = 1) = a \cdot p(x \mid y = 0) > 0$ auf einer Menge von positivem Lebesgue-Maß gilt.