



Sebastian
Raschka

Vahid
Mirjalili

2.,
aktualisierte
und erweiterte
Auflage

Machine Learning mit **Python** und Scikit-learn und TensorFlow

Das umfassende **Praxis-Handbuch**
für Data Science, Deep Learning und Predictive Analytics

Vahid Mirjalili erlangte seinen Dokortitel als Maschinenbauingenieur mit einer Arbeit über neue Verfahren für Computersimulationen molekularer Strukturen. Derzeit erforscht er Anwendungen des Machine Learnings in verschiedenen Computer-Vision-Projekten (»maschinelles Sehen«) am Fachbereich für Informatik und Ingenieurwesen an der Michigan State University.

Vahid hat sich für Python als bevorzugte Programmiersprache entschieden und während seiner akademischen Laufbahn enorme Erfahrung in der Python-Programmierung gesammelt. Er hat an der Michigan State University Programmierkurse in Python für angehende Ingenieure geleitet, was es ihm ermöglichte, den Studenten ein besseres Verständnis verschiedener Datenstrukturen und die effiziente Entwicklung von Python-Code zu vermitteln.

Neben der Erforschung von Deep Learning und Anwendungen des maschinellen Sehens ist er besonders daran interessiert, Lernverfahren zu entwickeln, die biometrische Daten wie Bilder von Gesichtern besser schützen, sodass ein Benutzer nicht mehr Informationen preisgibt als beabsichtigt. Darüber hinaus arbeitet er mit einem Team von Ingenieuren zusammen, das selbstfahrende Autos entwickelt. Er entwirft dabei Modelle neuronaler Netzwerke, die anhand multispektraler Bilddaten Fußgänger erkennen.

Ich möchte meinem Doktorvater Dr. Arun Ross dafür danken, dass er mir die Möglichkeit bietet, in seinem Forschungslabor an neuartigen Aufgaben zu arbeiten. Außerdem möchte ich Dr. Vishnu Boddeti dafür danken, dass er mein Interesse an Deep Learning entfacht und mir die grundlegenden Konzepte vor Augen geführt hat.



Über die Korrektoren

Jared Huffmann ist Unternehmer, Spieler, Geschichtenerzähler, Machine-Learning-Fan und begeisterter Anhänger von Datenbanken. Die letzten zehn Jahre hat er mit der Entwicklung von Software und der Analyse von Daten verbracht. Seine früheren Tätigkeiten umfassen ein breites Spektrum, unter anderem Sicherheitsaspekte von Netzwerken, Finanzsysteme, systematische Unternehmensanalyse sowie Webdienste, Entwickler-Tools und Geschäftsstrategie. Erst kürzlich hat er das Data-Science-Team von Minecraft gegründet, das sich auf Big Data und Machine Learning konzentriert. Wenn er nicht arbeitet, ist er typischerweise beim Spielen oder genießt den wunderschönen pazifischen Nordwesten mit Freunden und der Familie.

Ich danke dem amerikanischen Originalverlag Packt für die Möglichkeit, an einem so tollen Buch mitwirken zu können, meiner Frau für die unermüdliche Ermutigung und meiner Tochter dafür, dass sie die meisten Abende durchgeschlafen hat, während ich mit der Durchsicht und dem Debuggen des Codes beschäftigt war.

Huai-En, Sun (Ryan Sun) verfügt über einen akademischen Grad der National Chiao Tung University in Statistik. Er ist derzeit als Data Scientist tätig und analysiert die Produktionsverfahren bei PEGATRON. Seine Forschungsschwerpunkte sind Machine Learning und Deep Learning.



Einleitung

Aus den Nachrichten und den sozialen Medien ist Ihnen vermutlich bekannt, dass das Machine Learning zu einer der spannendsten Technologien der heutigen Zeit geworden ist. Große Unternehmen wie Google, Facebook, Apple, Amazon, IBM und viele andere investieren aus gutem Grund kräftig in die Erforschung des Machine Learnings und dessen Anwendung. Auch wenn man manchmal den Eindruck bekommt, dass »Machine Learning« als leeres Schlagwort gebraucht wird, handelt es sich doch zweifellos nicht um eine Modeerscheinung. Dieses spannende Fachgebiet eröffnet viele neue Möglichkeiten und ist aus dem Alltag schon nicht mehr wegzudenken. Denken Sie an die virtuellen Assistenten von Smartphones, Produktempfehlungen für Kunden in Onlineshops, das Verhindern von Kreditkartenbetrug, Spamfilter in E-Mail-Programmen, die Erkennung und Diagnose von Krankheitssymptomen – die Liste ließe sich beliebig lang fortsetzen.

Wenn Sie zu einem Praktiker des Machine Learnings und einem besseren Problemlöser werden möchten oder vielleicht sogar eine Laufbahn in der Erforschung des Machine Learnings anstreben, dann ist dies das richtige Buch für Sie. Für einen Neuling können die dem Machine Learning zugrunde liegenden theoretischen Konzepte zunächst einmal erdrückend wirken. In den vergangenen Jahren sind aber viele praxisorientierte Bücher mit leistungsfähigen Lernalgorithmen erschienen, die Ihnen den Start erleichtern.

Die Verwendung praxisorientierter Codebeispiele dient einem wichtigen Zweck: Konkrete Beispiele verdeutlichen die allgemeinen Konzepte, indem das Erlernete unmittelbar in die Tat umgesetzt wird. Allerdings darf man dabei nicht vergessen, dass mit großer Macht auch immer große Verantwortung einhergeht! Neben der unmittelbaren Erfahrung, Machine Learning mithilfe der Programmiersprache Python und auf Python beruhenden Lernbibliotheken in die Tat umzusetzen, stellt das Buch auch die den Machine-Learning-Algorithmen zugrunde liegenden mathematischen Konzepte vor, die für den erfolgreichen Einsatz von Machine Learning unverzichtbar sind. Das Buch ist also kein rein praktisch orientiertes Werk, sondern ein Buch, das die erforderlichen Details der Konzepte des Machine Learnings erörtert, die Funktionsweise von Lernalgorithmen und ihre Verwendung verständlich, aber dennoch informativ erklärt und – was noch wichtiger ist – das zeigt, wie man die häufigsten Fehler vermeidet.

Wenn Sie bei Google Scholar den Suchbegriff *machine learning* eingeben, erhalten Sie als Resultat eine riesige Zahl (ca. 1.800.000) von Treffern. Nun können wir in

diesem Buch natürlich nicht sämtliche Einzelheiten der in den letzten 60 Jahren entwickelten Algorithmen und Anwendungen erörtern. Wir werden uns jedoch auf eine spannende Tour begeben, die alle wichtigen Themen und Konzepte umfasst, damit Sie eine gründliche Einführung erhalten. Sollte Ihr Wissensdurst auch nach der Lektüre noch nicht gestillt sein, steht Ihnen eine Vielzahl weiterer hilfreicher Ressourcen zur Verfügung, die Sie nutzen können, um die entscheidenden Fortschritte auf diesem Fachgebiet zu verfolgen.

Falls Sie sich schon ausführlich mit der Theorie des Machine Learnings beschäftigt haben, zeigt Ihnen dieses Buch, wie Sie Ihre Kenntnisse in die Praxis umsetzen können. Wenn Sie bereits entsprechende Techniken eingesetzt haben, aber deren Funktionsweise besser verstehen möchten, kommen Sie hier ebenfalls auf Ihre Kosten. Und wenn Ihnen das Thema Machine Learning noch völlig neu ist, haben Sie umso mehr Grund, sich zu freuen, denn ich kann Ihnen versprechen, dass dieses Verfahren Ihre Denkweise über Ihre in Zukunft zu lösenden Aufgaben verändern wird – und ich möchte Ihnen zeigen, wie Sie Problemstellungen in Angriff nehmen können, indem Sie die den Daten innewohnende Kraft freisetzen.

Bevor wir uns eingehender mit dem Machine Learning befassen, soll aber zunächst noch Ihre vermutlich vordringlichste Frage beantwortet werden: Warum Python? Die Antwort ist einfach: Es ist leistungsfähig und doch leicht zu erlernen. Python ist zur beliebtesten Programmiersprache im Bereich Data Science geworden, weil man sich die lästigen Teile bei der Programmierung erspart und eine Umgebung bereitsteht, in der sich Ideen und Konzepte sofort umsetzen lassen.

Wir, die Autoren, können aus eigener Erfahrung sagen, dass wir durch die Beschäftigung mit dem Machine Learning zu besseren Wissenschaftlern, Denkern und Problemlösern geworden sind. In diesem Buch möchten wir unsere diesbezüglichen Erkenntnisse mit Ihnen teilen. Wissen wird durch Lernen erworben, was wiederum einen gewissen Eifer erfordert, und erst Übung macht den sprichwörtlichen Meister. Der vor Ihnen liegende Weg ist manchmal nicht ganz einfach, und einige der Themenbereiche sind deutlich schwieriger als andere, aber wir hoffen dennoch, dass Sie die Gelegenheit nutzen und sich auf den Lohn der Mühe konzentrieren. Im weiteren Verlauf des Buches werden Sie Ihrem Repertoire eine ganze Reihe leistungsfähiger Techniken hinzufügen können, die dabei helfen, auch die schwierigsten Aufgaben auf datengesteuerte Weise zu bewältigen.

Zum Inhalt des Buches

Kapitel 1, Wie Computer aus Daten lernen können, führt Sie in die wichtigsten Teilbereiche des Machine Learnings ein, mit denen sich verschiedene Probleme in Angriff nehmen lassen. Darüber hinaus werden die grundlegenden Schritte beim Entwurf eines typischen Machine-Learning-Modells erörtert, auf die wir in den nachfolgenden Kapiteln zurückgreifen.