

# Einführung

---

Die Schlagwörter *Maschinelles Lernen* und *Künstliche Intelligenz* wecken bei uns zahlreiche Assoziationen. Dies liegt nicht zuletzt auch daran, dass sich Hollywood und andere Filmstudios weltweit immer öfter diesem Themenfeld widmen. Filme wie *Ex Machina* spielen mit den Vorstellungen und Ängsten der Kinogänger und dramatisieren das Konzept des maschinellen Lernens auf eine Weise, die nicht unbedingt der Realität entspricht. Denn im Alltag lassen sich mithilfe maschineller Lernverfahren zwar unglaublich vielseitige Aufgaben lösen – mit Androiden, die den Turing-Test bestehen und ihren Schöpfer von ihrer Menschlichkeit überzeugen, hat dies jedoch herzlich wenig zu tun. *Maschinelles Lernen mit Python und R für Dummies* bietet Ihnen anhand praktischer Anwendungsfälle einen Einblick in die faszinierende Welt des maschinellen Lernens. Die Aufgaben, die sich damit bewältigen lassen, mögen Ihnen auf den ersten Blick im Vergleich zu ihrer filmischen Umsetzung banal erscheinen. Beim Lesen dieses Buches werden Sie aber erkennen, dass eben genau diese scheinbar banalen Aufgaben das Leben aller Menschen auf der Welt entscheidend verbessern können. Kurz: Das maschinelle Lernen ist ein äußerst vielversprechender Ansatz – und die Realität ist dabei viel spannender als die Fiktion.

## Über dieses Buch


*Maschinelles Lernen mit Python und R für Dummies* soll Ihnen verdeutlichen, was mit maschinellen Lernverfahren heute bereits möglich ist und was vielleicht in Zukunft möglich sein wird. Obwohl dieses Buch zahlreiche Programmbeispiele enthält, sind zum Verständnis keine besonderen Informatikkenntnisse nötig. Mit der Mathematik sollten Sie jedoch nicht auf Kriegsfuß stehen, denn die mathematischen Grundlagen spielen bei unseren Erläuterungen eine wichtige Rolle. Wir zeigen Ihnen anhand konkreter Beispiele, wie Sie spezifische Algorithmen gezielt auf große Datenmengen anwenden, um ein zuverlässiges und nützliches Ergebnis zu erhalten. Die Betonung liegt dabei auf *nützlich*, denn durch maschinelles Lernen lässt sich ein breites Spektrum an Aufgaben effizienter lösen als je zuvor.

Wichtig ist auch die Auswahl geeigneter Programme und Hilfsmittel. In diesem Buch verwenden wir die Programmiersprachen R und Python, da sie sich aufgrund ihres Funktionsumfangs besonders für maschinelle Lernverfahren eignen. Während R vor allem durch seine hohe Benutzerfreundlichkeit überzeugt, verfügt Python über eine umfassende Standardbibliothek, mit der Ihnen quasi keine Grenzen gesetzt sind. *Maschinelles Lernen mit Python und R für Dummies* erläutert, wann Sie welche der beiden Sprachen jeweils nutzen sollten und was dabei zu beachten ist.

In diesem Buch lernen Sie außerdem einige interessante Techniken kennen. Wir zeigen Ihnen die Algorithmen, mit denen die jeweiligen Aufgaben gelöst werden, und erklären, wie die Algorithmen funktionieren. Denn auch ohne einen Doktor in Mathematik sollen Sie genau verstehen, was Sie da eigentlich tun. Nach dem Lesen dieses Buches sind Sie bestens gerüstet, um Aufgaben in Ihrem spezifischen Fachbereich mithilfe maschineller Lernverfahren optimal zu lösen.

Und keine Sorge wegen der Programmierumgebung! Am Anfang des Buches finden Sie eine komplette Installationsanleitung für die von uns genutzten integrierten Entwicklungsumgebungen, nämlich RStudio und Anaconda. Zusätzlich erleichtern Ihnen Kurzanleitungen (mit Links zu weiterführenden Informationen) das Verständnis der wesentlichen R- und Python-Programmierschritte, die Sie ausführen müssen. Wir möchten, dass Sie möglichst schnell loslegen und unseren Beispielen problemlos folgen können – unnötig komplexer Programmcode soll Sie dabei nicht aufhalten.

Zur übersichtlichen Darstellung nutzen wir in diesem Buch die folgenden Konventionen:

- ✓ Text, den Sie exakt übernehmen sollen, Programmcode und Internetadressen erscheinen in **Festbreitenschrift**. Falls Sie eine digitale Version dieses Buches auf einem Gerät mit bestehender Internetverbindung lesen, können Sie Links auch anklicken, zum Beispiel diesen hier: <http://www.fuer-dummies.de>.
- ✓ Wörter, die Sie eingeben sollen und die zusätzlich *kursiv* hervorgehoben sind, dienen als Platzhalter. Das heißt, Sie müssen sie jeweils mit für Sie passenden Wörtern ersetzen. Lautet die Anweisung beispielsweise: »Geben Sie *Ihren Namen* ein und bestätigen Sie mit , so müssen Sie *Ihren Namen* durch Ihren tatsächlichen Namen ersetzen.
- ✓ Für Definitionen nutzen wir die *Kursivschrift*. Sie brauchen Ihnen unbekannte Begriffe also nicht mehr in anderen Quellen nachzuschlagen.
- ✓ Menübefehle werden in **KAPITÄLCHEN** gesetzt. Wenn Sie mehrere Menübefehle hintereinander anklicken sollen, werden diese durch einen senkrechten Strich verknüpft, etwa so: **DATEI | NEUE DATEI**. In diesem Fall müssten Sie im Menü erst auf **DATEI** und dann auf **NEUE DATEI** klicken.

## Grundvoraussetzungen

Wir gehen davon aus, dass Sie mit Ihrem Computer und Betriebssystem vertraut sind. Dieses Buch enthält diesbezüglich keine Hinweise oder Anleitungen. (Kapitel 4 bietet jedoch Installationshinweise für RStudio und Kapitel 6 erklärt, wie Sie Anaconda installieren.) Wir möchten uns voll und ganz auf das maschinelle Lernen mit R und Python konzentrieren und gehen daher nicht auf plattformspezifische Fragen ein. Sie sollten also wissen, wie Sie Anwendungen installieren und aufrufen und wie Sie generell mit Ihrem gewählten Betriebssystem arbeiten.

Dieses Buch ist auch kein Auffrischkurs in Mathematik. Die zahlreichen komplexen mathematischen Beispiele sollen Ihnen die Nutzung von R, Python und maschinellen Lernverfahren für Ihre Analysen erleichtern und nicht die theoretische Mathematik beibringen. Viele der in diesem Buch genutzten Algorithmen werden aber genauer erklärt, damit Sie ihre Funktionsweise verstehen. Kapitel 1 und 2 vermitteln Ihnen gewisse Grundlagen, die Sie zur erfolgreichen Lektüre dieses Buches benötigen.

Und schließlich gehen wir auch davon aus, dass Sie wissen, wie man Inhalte im Internet abrufen. Das Buch enthält zahlreiche Verweise auf weiterführende Materialien, die natürlich nur dann nützlich sind, wenn Sie sie auch tatsächlich im Web finden und verwenden.

## Symbole, die in diesem Buch verwendet werden

Wichtige Hinweise oder zusätzliche Informationen in diesem Buch sind durch die folgenden Symbole am Seitenrand gekennzeichnet:



Tipps sparen Ihnen Zeit und unnötigen Arbeitsaufwand. Es handelt sich meist um Methoden oder nützliche Hilfsmittel, mit denen Sie R beziehungsweise Python optimal nutzen oder Aufgaben beim maschinellen Lernen effizienter ausführen.



Wir möchten uns ja ungern wie strenge Eltern aufspielen und Ihnen missbilligend auf die Finger klopfen, aber die Ratschläge neben diesem Warnsymbol sollten Sie unbedingt befolgen! Ansonsten kann es sein, dass Ihr Programm nicht wie gewünscht funktioniert oder dass eigentlich narrensichere Gleichungen doch die falschen Ergebnisse liefern oder dass (schlimmstenfalls) Ihre Daten verloren gehen.



Dieses Symbol weist auf vertiefende Tipps oder Methoden hin. Manche dieser Informationshappen finden Sie vielleicht langweilig – oder aber sie liefern Ihnen den entscheidenden Hinweis, wie Sie eines Ihrer Programme zum Laufen bringen. Auf jeden Fall können Sie diese Abschnitte bei Bedarf einfach überspringen.



Den Text neben diesem Symbol sollten Sie unbedingt lesen und verinnerlichen. Hier werden wesentliche Prozesse oder Zusammenhänge erläutert, die Sie zur Arbeit mit R oder Python oder zur erfolgreichen Umsetzung von maschinellen Lernverfahren benötigen.



RStudio und Anaconda können standardmäßig zahlreiche allgemeine Aufgaben erledigen. Beim maschinellen Lernen sind jedoch manchmal sehr spezifische Aufgaben erforderlich, für die Sie zusätzliche Komponenten aus dem Internet herunterladen müssen. Dieses Symbol weist auf entsprechende Websites hin, die Sie beachten sollten, damit Sie alles korrekt installieren und unsere Beispiele nachvollziehen können.

## Weitere Ressourcen

Bestimmt haben Sie keine Lust, den Programmcode aus unserem Buch mühsam abzutippen. Viel lieber würden Sie wahrscheinlich praktische Aufgaben lösen und testen, was mit R, Python und maschinellen Lernverfahren tatsächlich alles möglich ist. Da haben wir gute Nachrichten: Alle Beispiele aus den verschiedenen Kapiteln stehen als Download bereit. Sie finden sie unter [www.wiley-vch.de/publish/dt/books/ISBN3-527-71363-8](http://www.wiley-vch.de/publish/dt/books/ISBN3-527-71363-8).

## Und nun?

Nun kann es losgehen! Wenn Sie sich bisher noch nie mit dem maschinellen Lernen befasst haben, beginnen Sie am besten mit Kapitel 1. Lesen Sie sich das Buch in Ihrem eigenen Rhythmus durch, sodass Sie alle Informationen gut verarbeiten können. Machen Sie sich vor allem auch mit R und Python vertraut, denn beide Programmiersprachen kommen in unseren Beispielen zum Einsatz.

Sollten Sie es auch als Neuling ganz eilig haben und sofort mit dem maschinellen Lernen loslegen wollen, können Sie direkt zu Kapitel 4 springen. Beschweren Sie sich aber nicht, wenn Ihnen einige Themen später etwas verwirrend vorkommen. Falls Sie RStudio bereits installiert haben, brauchen Sie Kapitel 4 nur zu überfliegen. Gleiches gilt für Anaconda: Bei schon vorhandener Installation ist Kapitel 6 nebensächlich. Für die Beispiele aus diesem Buch benötigen Sie R in der Version 3.2.3 und Python in der Version 2.7.11. Mit Version 3.x von Python funktionieren die Beispiele leider nicht, da diese Version einige der von uns genutzten Bibliotheken nicht unterstützt.

Leser, die mit R und Python vertraut sind und die korrekten Versionen installiert haben, können direkt mit Kapitel 8 fortfahren. Bei Bedarf können Sie natürlich jederzeit zu früheren Kapiteln zurückblättern. Überspringen sollten Sie Abschnitte aber nur, wenn Sie sicher sind, dass Sie die enthaltenen Informationen wirklich beherrschen. Jede Methode, jedes Beispiel und jedes Verfahren vermittelt wichtige Lektionen, die Ihnen das Verständnis der nächsten Kapitel erleichtern.