

Einleitung

Sie verlassen sich Tag für Tag auf wissenschaftliche Daten, um eine Fülle von Aufgaben zu meistern, oder Sie haben die Aufgabe, bestimmte Dienstleistungen für andere zu erfüllen. Tatsächlich verlassen Sie sich aber auf wissenschaftliche Daten in einer Weise, die Ihnen sicher gar nicht klar ist. Haben Sie zum Beispiel heute Morgen für eine Suche Ihre bevorzugte Suchmaschine verwendet, haben Sie Vorschläge für alternative Suchbegriffe erhalten, die auf Basis wissenschaftlicher Daten geliefert werden. Sind Sie letzte Woche zu Ihrem Arzt gegangen und hörten, dass der Knoten, den Sie bei sich fanden, kein Krebs war, ist es wahrscheinlich, dass der Arzt diese Diagnose auf Grundlage wissenschaftlicher Daten gestellt hat. Tatsächlich könnten Sie jeden Tag mit wissenschaftlichen Daten arbeiten und würden es nicht einmal bemerken. *Data Science mit Python für Dummies* führt Sie nicht nur in den Gebrauch wissenschaftlicher Daten ein, um eine Fülle an praktischen Aufgaben lösen zu können, sondern zeigt Ihnen auch die Vielseitigkeit der Anwendung wissenschaftlicher Daten auf. Wenn Sie wissen, wie Probleme mit wissenschaftlichen Daten zu lösen sind und wo Sie sie anwenden können, erhöhen Sie Ihre Chancen etwa auf eine Beförderung oder Ihren Traumjob signifikant.

Über dieses Buch

Das große Ziel von *Data Science mit Python für Dummies* ist es, wissenschaftlichen Daten den Schrecken zu nehmen, indem gezeigt wird, dass die Arbeit mit diesen Daten und Python allgemein nicht nur interessant, sondern auch durchaus machbar ist. Vielleicht denken Sie, ein Computergenie sein zu müssen, um die komplexen Aufgaben, die normalerweise mit wissenschaftlichen Daten verbunden sind, zu lösen, jedoch verfügt Python über eine Vielzahl von nützlichen Bibliotheken, die all die schwierigen Aufgaben für Sie im Hintergrund erledigen. Sie werden gar nicht bemerken, was alles vor sich geht, und Sie müssen sich auch nicht damit beschäftigen. Sie müssen lediglich wissen, dass Sie spezifische Aufgaben lösen wollen und dass dieses Ziel durch Python ganz leicht erreicht werden kann.

Einer der Schwerpunkte dieses Buches liegt auf der Verwendung der richtigen Tools. Sie beginnen mit Anaconda, einem Paket, das IPython und IPython Notebook beinhaltet – zwei Tools, die die Arbeit mit Python erleichtern. Sie experimentieren mit IPython in einer komplett interaktiven Entwicklungsumgebung. Der Code, den Sie in IPython Notebook verwenden, hat Präsentationsqualität, und Sie können ihn mit zahlreichen Präsentationselementen in Ihrem Dokument kombinieren. Es ist also weit mehr als nur eine Entwicklungsumgebung.

Sie werden in diesem Buch viele interessante Techniken kennenlernen. Beispielsweise können Sie Grafiken aller Ihrer Daten mit Matplotlib erstellen, wofür Sie in diesem Buch alles finden, was Sie wissen müssen. Außerdem wird Ihnen in diesem Buch sehr umfassend gezeigt, welche Möglichkeiten es gibt und wie Sie diese nutzen können, um interessante Kalkulationen zu erstellen. Viele möchten gerne wissen, wie man die Handschrifterkennung realisieren kann – und wenn Sie auch dazugehören, können Sie dieses Buch nutzen, um einen Einblick darin zu erhalten.

Möglicherweise beunruhigt Sie die ganze Programmierumgebungsproblematik ein wenig, aber dieses Buch wird Sie ganz sicher nicht im Regen stehen lassen. Am Anfang finden Sie komplette Installationsanleitungen für Anaconda und ein Kurzlehrbuch (mit Referenzen) für die notwendigen Grundlagen der Python-Programmierung. Der Fokus liegt darauf, so schnell wie möglich beginnen zu können und einfache Beispiele zur Verfügung zu stellen, damit der Code keinen Stolperstein für das Lernen darstellt.

Um die Konzepte einfacher auszuführen, nutzt dieses Buch die folgenden Grundsätze:

- ✓ Text, den Sie genau so eingeben sollen, wie er in diesem Buch steht, ist in Schreibmaschinenschrift gedruckt. Wenn Sie eine digitale Version dieses Buches auf einem Gerät lesen, das mit dem Internet verbunden ist, bedenken Sie, dass Sie die Webseiten anklicken können, um sie zu besuchen, beispielsweise diese hier: <http://www.wiley-vch.de/dummies>.
- ✓ Bei *kursiv* geschriebenen Wörtern als Teil eines Befehls müssen Sie diesen Wert durch Ihre eigenen Daten ersetzen. Wenn Sie zum Beispiel »Geben Sie *Ihren Namen* ein und drücken Sie Enter« sehen, müssen Sie *Ihren Namen* durch Ihren eigenen Namen ersetzen.
- ✓ Bedienelemente, wie Schaltflächen, Registerkarten und Menüs, stehen in **KAPITÄLCHEN**. Wenn Sie eine Befehlsfolge eingeben sollen, ist diese wie folgt dargestellt: **DATEI|NEUE DATEI**. In diesem Fall klicken Sie zuerst auf die Registerkarte **DATEI** und wählen dort **NEUE DATEI** aus. Als Ergebnis wird eine neue Datei angelegt.

Törichte Annahmen

Sie werden uns vielleicht nicht glauben, dass wir alle möglichen Annahmen über Sie getroffen haben – schließlich sind wir Ihnen noch nie begegnet! Obwohl die meisten Annahmen in der Tat töricht sind, brauchen wir sie, um einen Startpunkt für dieses Buch festzulegen. Es ist wichtig, dass Sie mit Ihrem Betriebssystem vertraut sind, denn dieses Buch bietet in dieser Hinsicht keinerlei Einführung. (Kapitel 3 stellt Installationsanweisungen für Anaconda zur Verfügung.) Um Ihnen ein Maximum an Informationen über Python und dessen Anwendbarkeit auf wissenschaftliche Daten bereitzustellen, werden in diesem Buch keine Betriebssystem-spezifischen Ausgaben diskutiert. Sie müssen wissen, wie Anwendungen installiert und verwendet werden, bevor Sie mit der Arbeit mit diesem Buch beginnen.

Dieses Buch ist kein Mathematiklehrbuch. Natürlich sehen Sie viele Beispiele komplexer Mathematik, aber der Schwerpunkt dieses Buchs liegt auf der Nutzung von Python und Data Science und nicht auf der mathematischen Theorie. Kapitel 1 und 2 vermitteln ein besseres Verständnis genau dessen, was Sie wissen müssen, damit Sie dieses Buch erfolgreich verwenden können.

Dieses Buch nimmt ebenso an, dass Sie auf Artikel im Internet zugreifen können. Es gibt durchgehend zahlreiche Verweise auf Onlinematerial, das Ihren Lernerfolg verbessern wird. Trotzdem sind diese zusätzlichen Quellen nur sinnvoll, wenn Sie sie finden und benutzen.

Im Buch verwendete Symbole

Wenn Sie dieses Buch lesen, werden Sie Symbole am Rand sehen, die auf wichtige Stellen hindeuten (oder möglicherweise auch nicht). Dieser Abschnitt beschreibt kurz jedes Symbol dieses Buches.



Tipps sind wichtig, denn sie helfen Ihnen, Zeit zu sparen oder eine Aufgabe ohne zusätzlichen Arbeitsaufwand zu lösen. Die Tipps in diesem Buch sind zeitsparende Techniken oder weisen auf Material hin, das Sie nutzen sollten, um maximal von Python profitieren zu können oder um mit Data Science verbundene Aufgaben zu bewältigen.



Wir wollen uns nicht wie zornige Eltern oder Verrückte anhören, aber Sie sollten es vermeiden, Dinge zu tun, die mit dem Warnungssymbol markiert sind. Andernfalls finden Sie vielleicht heraus, dass Ihre Anwendung nicht wie erwartet funktioniert, Sie bekommen eine falsche Ausgabe mit scheinbar sicheren Berechnungen oder Sie verlieren (im schlimmsten Fall) Daten.



Wann immer Sie dieses Symbol sehen, deutet dies auf einen Tipp oder eine Technik für Fortgeschrittene hin. Vielleicht finden Sie diese Leckerbissen nützlicher Informationen einfach langweilig, aber bedenken Sie, dass sie entscheidend sein könnten, um Ihr Programm zum Laufen zu bringen. Sie können diesen Teil überspringen, wenn Sie möchten.



Wenn Sie irgendetwas aus einem bestimmten Kapitel oder Abschnitt nicht bekommen, erinnern Sie sich an das Material, das mit diesem Symbol gekennzeichnet ist. Dieser Teil beinhaltet für gewöhnlich einen wichtigen Prozess oder Informationen, die Sie für die Arbeit mit Python oder für die erfolgreiche Bewältigung mit Data Science verbundener Aufgaben benötigen.

Über das Buch hinaus

Dieses Buch ist nicht das Ende unserer Arbeit mit Data Science – es ist nur der Anfang. Wir stellen Onlinematerial zur Verfügung, um dieses Buch flexibler zu gestalten und an Ihre Bedürfnisse anzupassen. Wenn wir E-Mails von Ihnen erhalten, können wir auf Fragen eingehen und Ihnen mitteilen, wie Sie entweder Python aktualisieren oder mit verknüpften Add-Ons umgehen, die den Inhalt des Buches beeinflussen. Tatsächlich haben Sie Zugang zu all diesen tollen Ergänzungen:

- ✓ **Onlineartikel:** Auf der Seite <http://www.dummies.com/extras/pythonfordata-science> finden Sie – in englischer Sprache – zusätzliches Onlinematerial. Es handelt sich dabei um Ergänzungen, die nicht mehr ins Buch gepasst haben.

Vielleicht sind auch die Blog-Posts mit Antworten zu Leserfragen sowie die Demonstration nützlicher und für dieses Buch relevanter Techniken auf <http://blog.johnmuellerbooks.com/> für Sie interessant.

- ✓ **Begleitdateien:** Hey! Wer will schon den gesamten Code des Buches abtippen und die Grafiken manuell rekonstruieren? Die meisten Leser bevorzugen es, ihre Zeit mit der Arbeit mit Python zu verbringen, Aufgaben im Data-Science-Umfeld zu lösen und die interessantesten Dinge auszuprobieren. Glücklicherweise sind die Beispiele, die in diesem Buch verwendet werden, als Download verfügbar, und Sie müssen nichts weiter tun, als dieses Buch zu lesen und die Techniken von Python für Data Science zu erlernen. Sie können die Dateien unter <http://www.wiley-vch.de/publish/dt/books/ISBN3-527-71208-9> herunterladen.

Wie es weitergeht

Es ist Zeit, mit dem Abenteuer Data Science mit Python zu beginnen! Wenn Python und der Umgang mit Data Science vollkommen neu für Sie sind, sollten Sie mit Kapitel 1 beginnen und sich in Ihrem ureigenen Tempo so effektiv wie möglich durch das Buch hindurcharbeiten.

Wenn Sie ein Anfänger sind, der es kaum erwarten kann, mit Python für Data Science so schnell es geht voranzukommen, können Sie zu Kapitel 3 springen, sollten dabei aber einkalkulieren, dass Sie so manche Dinge etwas verwirrend finden werden. Zu Kapitel 4 zu springen, ist möglich, wenn Sie Anaconda (die Programmierumgebung dieses Buches) bereits installiert haben, jedoch sollten Sie Kapitel 3 zumindest überfliegen, damit Sie wissen, welche Voraussetzungen wir geschaffen haben, um dieses Buch zu schreiben. Sie müssen Anaconda mit Python der Version 2.7.9 installiert haben, um die besten Ergebnisse mit dem Code dieses Buches zu erzielen.

Leser, die schon etwas Erfahrung mit Python mitbringen und Anaconda bereits installiert haben, können Zeit sparen, indem sie direkt mit Kapitel 5 beginnen. Sie können, falls erforderlich, immer zu früheren Kapiteln zurückgehen, wenn Ihnen etwas unklar ist. Trotzdem ist es wichtig für Sie, zu verstehen, wie eine Technik funktioniert, bevor Sie zur nächsten gehen. Jede Technik, jedes Codebeispiel und jede Methode ist wichtig für Sie, und Sie könnten entscheidende Inhalte übersehen, wenn Sie zu viele Informationen überspringen.