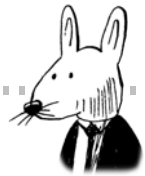
Grundlegende Konzepte: Wie lernen Maschinen? 2

In fünf Jahren wird Machine Learning für jeden erfolgreichen Börsengang verantwortlich sein.

– Eric Schmidt, Vorstandsvorsitzender bei Google, in einer Rede auf der Cloud Computing Platform Conference 2016



In diesem Kapitel:

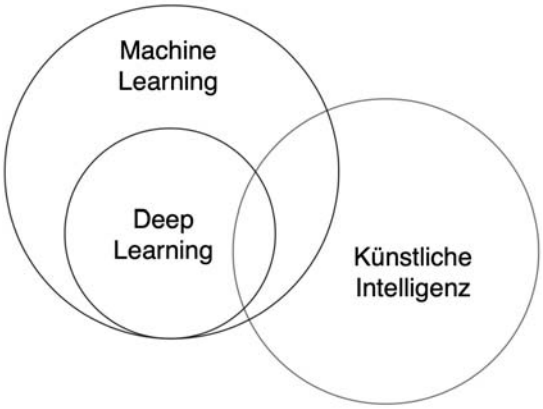
- Deep Learning, Machine Learning und künstliche Intelligenz
- Parametrische und nichtparametrische Modelle
- Überwachtes und unüberwachtes Lernen
- Wie können Maschinen lernen?

2.1 Was ist Deep Learning?

Deep Learning ist eine Teilmenge der Machine-Learning-Verfahren.

Deep Learning ist eine Teilmenge des Machine Learnings, das sich der Untersuchung und der Entwicklung von Maschinen widmet, die lernen können (manchmal auch mit dem Ziel, eine allgemeine künstliche Intelligenz zu erlangen).

In der Industrie wird Deep Learning in verschiedenen Bereichen zum Lösen praktischer Aufgaben eingesetzt, etwa bei Computer Vision (Bilder), bei der Verarbeitung natürlicher Sprache (Text) oder bei der automatischen Spracherkennung (Audio). Deep Learning ist also eine Teilmenge der *Verfahren*, die beim Machine Learning zum Einsatz kommen, wobei vor allem künstliche neuronale Netze verwendet werden, die zu einer Klasse von Algorithmen gehören, die mehr oder weniger durch das menschliche Gehirn inspiriert wurden.



Die Abbildung zeigt, dass es beim Deep Learning nicht nur um allgemeine künstliche Intelligenz geht (wie bei den empfindungsfähigen Maschinen in Spielfilmen). Viele Anwendungen dieser Technologie lösen die verschiedenartigsten Aufgaben. Dieses Buch soll sich auf die Grundlagen des Deep Learnings konzentrieren, das sowohl in der topaktuellen Forschung als auch in der Industrie zum Einsatz kommt, und dir diese Kenntnisse vermitteln.

2.2 Was ist Machine Learning?

> *Ein Fachgebiet, das Computern die Fähigkeit verleiht, zu lernen, ohne explizit programmiert zu werden.*
> – Arthur Samuel zugeschrieben

Wenn also Deep Learning eine Teilmenge des Machine Learnings ist, was ist dann eigentlich Machine Learning? Ganz allgemein ist es das, was der Name besagt. Machine Learning ist ein Teilgebiet der Informatik, das sich damit befasst, dass *Maschinen lernen*, Aufgaben zu erledigen, für die sie *nicht explizit programmiert* wurden. Oder kurz und bündig: Maschinen beobachten ein Muster und versuchen, es irgendwie zu imitieren, entweder direkt oder indirekt:

> Machine Learning ~= Affe sieht, Affe tut

Ich erwähne hier direktes und indirektes Imitieren, um die Analogie zu den beiden Haupttypen des Machine Learnings aufzuzeigen: *überwachtes* und *unüberwachtes* Lernen. Überwachtes Machine Learning ist das direkte Imitieren des Musters, das einem Datensatz zu eigen ist, bei einem anderen Datensatz. Es ist der Versuch, eine Eingabedatenmenge in eine Ausgabedatenmenge umzuwandeln. Das kann eine unglaublich mächtige und nützliche Fähigkeit sein. Sieh dir die folgenden Beispiele an (die Eingabedatenmengen sind fett gedruckt, die Ausgabedatenmengen kursiv):

- Verwendung der **Pixel** eines Bilds, um das *Vorhandensein* oder das *Fehlen einer Katze* zu erkennen
- Verwendung der **Filme, die dir gefallen haben**, um *Filme* vorherzusagen, *die dir wahrscheinlich gefallen werden*
- Verwendung der **Wörter**, die jemand benutzt, um festzustellen, ob die Person *fröhlich* oder *traurig* ist
- Verwendung der **Daten einer Wetterstation**, um die *Regenwahrscheinlichkeit* vorherzusagen
- Verwendung von **Motorsensoren**, um die optimalen *Einstellungen* zu ermitteln
- Verwendung von **Nachrichtenmeldungen**, um die morgigen *Aktienkurse* vorherzusagen
- Verwendung eines **Eingabewerts** zur Berechnung des *doppelten Werts*
- Verwendung der Rohdaten einer **Audiodatei**, um eine *Transkription* des Inhalts zu erstellen

All diese Aufgaben gehören zum überwachten Lernen. Der Machine-Learning-Algorithmus versucht stets, das Muster, das die beiden Datenmengen verbindet, so zu imitieren, dass *die eine Datenmenge zur Vorhersage der anderen* verwendet werden kann. Stell dir vor, du könntest bei all diesen Beispielen die *Ausgabe*datenmenge nur anhand der *Eingabe*datenmenge vorhersagen – das wäre schon eine beachtliche Leistung.

2.3 Überwachtes Machine Learning

Überwachtes Lernen transformiert Datenmengen.

Überwachtes Lernen ist ein Verfahren, um eine Datenmenge in eine andere umzuwandeln. Wenn du beispielsweise eine Datenmenge namens *Aktienkurse am Montag* hättest, in der die Kurse aller Aktien an den Montagen der letzten zehn Jahre aufgezeichnet sind, und eine zweite, die die Aktienkurse an den Dienstagen desselben Zeitraums enthält, könnte ein überwachter Lernalgorithmus die eine Datenmenge verwenden, um die andere vorherzusagen.

Wenn du einen überwachten Machine-Learning-Algorithmus erfolgreich mit den Daten von zehn Jahren trainiert hast, bist du in der Lage, den Aktienkurs für einen

beliebigen Dienstag in der Zukunft vorherzusagen, wenn dir der Aktienkurs am direkt vorhergehenden Montag bekannt ist. Lege eine kurze Pause ein und denke einen Moment darüber nach.

Überwachtes Machine Learning gehört bei der angewandten künstlichen Intelligenz (der sogenannten schwachen KI) zum Alltag. Es erweist sich als nützlich, um das, *was du weißt*, als Eingabe zu verwenden und schnell in das umzuwandeln, *was du wissen möchtest*. Auf diese Weise können Machine-Learning-Algorithmen die Intelligenz und die Fähigkeiten des Menschen auf unzählige Arten erweitern.

Bei der Nutzung der Ergebnisse von Machine Learning besteht der größte Aufwand darin, einen überwachten Klassifikator zu trainieren. Bei der Entwicklung möglichst genauer überwachter Machine-Learning-Algorithmen kommt typischerweise sogar unüberwachtes Machine Learning (dazu gleich mehr) zum Einsatz.

Im folgenden Teil des Buchs wirst du Algorithmen entwickeln, die Eingabedaten mit bestimmten Eigenschaften entgegennehmen. Sie sind beobachtbar, lassen sich aufzeichnen, und sie sind *nachvollziehbar*. Außerdem lassen sie sich in wertvolle Ausgabedaten umwandeln, für die eine logische Analyse erforderlich ist. All das leistet überwachtes Machine Learning.

2.4 Unüberwachtes Machine Learning

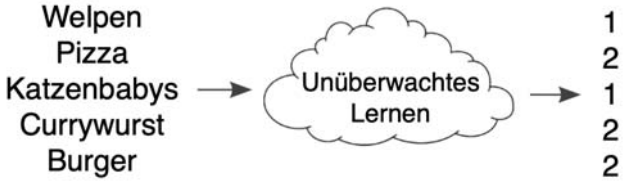
Unüberwachtes Lernen gruppiert deine Daten.

Unüberwachtes und überwachtes Lernen haben etwas gemeinsam: Eine Datenmenge wird in eine andere Datenmenge umgewandelt. Beim unüberwachten Lernen ist die zu transformierende Datenmenge allerdings vorher *nicht bekannt*. Im Gegensatz zum überwachten Lernen gibt es keine »richtige Antwort«, die das Modell reproduzieren soll. Ein unüberwachter Lernalgorithmus sucht einfach nur nach Mustern in den Daten und zeigt an, was er gefunden hat.

Das Aufteilen einer Datenmenge in Gruppen, das sogenannte *Clustering*, ist eine Form des unüberwachten Lernens. Beim Clustering wird eine Menge von Datenpunkten in eine Sequenz von Clusterbezeichnungen transformiert. Wenn eine Aufteilung in zehn Cluster vorgenommen wird, verwendet man für die Bezeichnungen üblicherweise die Zahlen von 1 bis 10. Jedem Datenpunkt wird eine Zahl zugewiesen, die angibt, zu welchem Cluster er gehört. Die Menge aus Datenpunk-

ten wird also in eine Menge von Bezeichnungen umgewandelt. Und warum sind die Bezeichnungen Zahlen? Der Algorithmus sagt nichts darüber aus, was die Cluster eigentlich sind. Er stellt lediglich fest: »Hey, Forscher! Ich habe eine Struktur entdeckt. Sieht ganz danach aus, als ob es Gruppen in den Daten gibt – und zwar diese hier!«

Eine gute Nachricht: Dieses Prinzip des Clusterings kannst du dir als Definition des unüberwachten Lernens merken. Es gibt zwar eine Vielzahl verschiedener Arten, aber *alle Formen des unüberwachten Lernens können als eine Form des Clusterings betrachtet werden*. Mehr zu diesem Thema später im Buch.

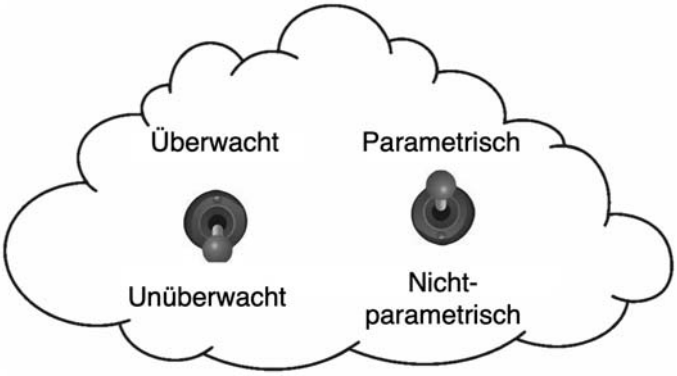


Sieh dir das Beispiel an. Der Algorithmus gibt zwar nicht an, wie die Cluster heißen, aber du kannst sicher erkennen, wonach er die Wörter gruppiert hat. (Antwort: 1 == niedlich, 2 == lecker.) Später werden wir uns damit befassen, dass andere Formen des unüberwachten Lernens auch nur eine Art Clustering sind und weshalb sich diese Cluster für das überwachte Lernen als nützlich erweisen.

2.5 Parametrisches und nichtparametrisches Lernen

Grob vereinfacht: Lernen durch Trial and Error vs. Zählen und Wahrscheinlichkeit

Auf den letzten Seiten wurden alle Machine-Learning-Algorithmen einer von zwei Gruppen zugeordnet: überwachtem und unüberwachtem Lernen. Jetzt werden wir eine andere Variante erörtern, die gleichen Machine-Learning-Algorithmen in zwei Gruppen zu unterteilen, nämlich in parametrische und nichtparametrische. Wenn wir unsere kleine Machine-Learning-Wolke genau betrachten, stellen wir fest, dass sie über zwei Einstellungen verfügt:



Wie du siehst, gibt es eigentlich vier verschiedene Arten von Algorithmen. Ein Algorithmus ist entweder unüberwacht oder überwacht und entweder parametrisch oder nichtparametrisch. Im vorangegangenen Abschnitt über überwachtes Lernen ging es um *die Art des erlernten Musters*, bei der Parametrisierung hingegen geht es darum, wie das Gelernte *gespeichert* wird, und im weiteren Sinne oft um die *Lernmethode*. Sehen wir uns zunächst einmal die formale Definition von parametrischen und nichtparametrischen Algorithmen an. Man hat sich übrigens noch nicht auf einen genauen Unterschied einigen können.

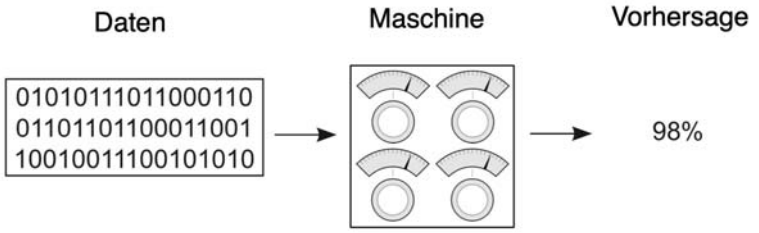
Ein parametrisches Modell ist dadurch gekennzeichnet, dass es eine feste Anzahl von Parametern besitzt, die Anzahl der Parameter eines nichtparametrischen Modells hingegen ist unbegrenzt (sie wird durch die Daten bestimmt).

Betrachten wir als Beispiel die Aufgabe, einen quadratischen Zapfen in eine passende (quadratische) Öffnung zu stecken. Manche Menschen (etwa Kleinkinder) stopfen den Zapfen einfach in alle Öffnungen, bis er irgendwo passt (parametrisch). Ein Teenager hingegen würde vielleicht die Anzahl der Seiten (vier) zählen und nach einer Öffnung suchen, die ebenso viele Seiten besitzt (nichtparametrisch). Parametrische Modelle verwenden tendenziell das Prinzip von Trial and Error, nichtparametrische dagegen versuchen eher, zu zählen. Das sehen wir uns genauer an.

2.6 Überwachtes parametrisches Lernen

Grob vereinfacht: Lernen durch Trial and Error mithilfe von Drehschaltern

Überwachte parametrisch lernende Maschinen besitzen eine festgelegte Anzahl von Drehschaltern (das ist der parametrische Teil), wobei das Lernen durch das Drehen der Schalter erfolgt. Vorliegende Eingabedaten werden entsprechend der Winkel, in dem die Drehschalter stehen, verarbeitet und in eine *Vorhersage* umgewandelt.



Das Lernen vollzieht sich in der Form, dass die Drehschalter in verschiedenen Winkeln positioniert werden. Wenn du versuchst, vorherzusagen, dass Bayern München die Champions League gewinnt, dann würde dieses Modell zuerst Daten entgegennehmen (wie Statistiken über Siege/Niederlagen oder die durchschnittliche Anzahl der Zehen pro Spieler) und eine Vorhersage treffen (etwa eine Chance von 98 Prozent). Anschließend würde das Modell beobachten, ob die Bayern tatsächlich gewinnen. Wenn feststeht, dass sie gewonnen haben, würde der Lernalgorithmus die *Position der Drehschalter aktualisieren*, um beim nächsten Mal eine genauere Vorhersage zu treffen, wenn die *gleichen oder ähnliche Eingabedaten* vorliegen.

Wenn sich der Schalter für »Siege/Niederlagen« bei der Vorhersage als nützlich erwiesen hat, würde das Modell ihn vielleicht etwas »aufdrehen«. Umgekehrt würde es den Schalter für die »durchschnittliche Anzahl der Zehen pro Spieler« »zurückdrehen«, wenn dieser Datenpunkt nicht besonders hilfreich war. Auf diese Weise lernen parametrische Modelle!

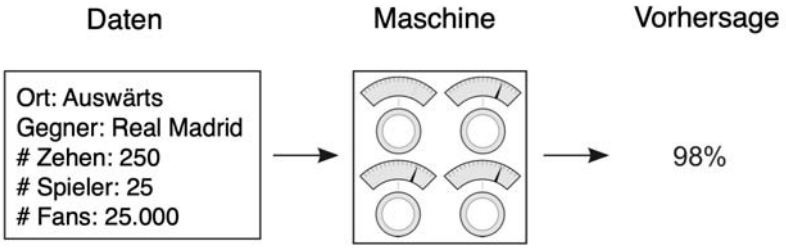
Hier ist zu beachten, dass alles, was das Modell gelernt hat, zu jedem Zeitpunkt in Form der Positionen der Drehschalter erfasst werden kann. Du kannst dir diese Art Lernmodell auch wie einen Suchalgorithmus vorstellen. Du »suchst« nach den geeigneten Stellungen der Drehschalter, indem du Konfigurationen ausprobierst, anpasst und wieder testest.

Außerdem musst du bedenken, dass das Prinzip von Trial and Error nicht der formalen Definition entspricht, es wird in parametrischen Modellen aber (mit Ausnahmen) häufig verwendet. Wenn es eine beliebige (aber feststehende) Anzahl von Drehschaltern gibt, muss eine Suche durchgeführt werden, um die optimale Konfiguration zu finden. Das ist beim nichtparametrischen Lernen anders, das oft auf Zählvorgängen beruht und (hin und wieder) neue Drehschalter hinzufügt, wenn etwas Neues gefunden wird, das gezählt werden kann. Wir unterteilen überwachtes parametrisches Lernen in drei Schritte.

Schritt 1: Vorhersage treffen

Zur Veranschaulichung des überwachten parametrischen Lernens bleiben wir bei der Analogie, vorherzusagen, ob Bayern München die Champions League gewinnt. Der erste Schritt besteht wie erwähnt darin, Statistiken zusammenzutragen, sie in

die Maschine einzuspeisen und eine Vorhersage über die Wahrscheinlichkeit zu treffen, dass die Bayern gewinnen.



Schritt 2: Vergleich mit Faktum

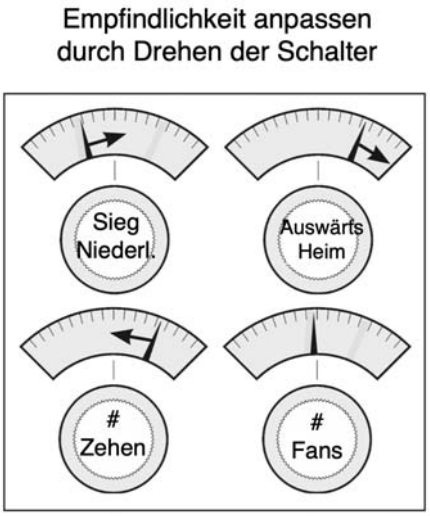
Der zweite Schritt besteht darin, die Vorhersage (98 Prozent) mit dem Faktum zu vergleichen, das von Bedeutung ist (ob Bayern München gewonnen hat). Leider haben sie verloren, somit ergibt sich:

Vorhersage: 98 % > Faktum: 0 %

Dieser Schritt erfasst, dass das Modell die bevorstehende Niederlage genau vorhergesehen hätte, wenn die Vorhersage 0 Prozent Gewinnchance gelautet hätte. Die Maschine soll möglichst genau arbeiten, was zu Schritt 3 führt.

Schritt 3: Erlernen des Musters

Dieser Schritt passt die Position der Drehschalter an und berücksichtigt dabei, wie sehr die Vorhersage daneben gelegen hat (98 Prozent) und was die Eingabedaten zum Zeitpunkt der Vorhersage waren (die Statistiken über Siege/Niederlagen). Auf diese Weise soll anhand der Eingabedaten eine genauere Vorhersage erfolgen.



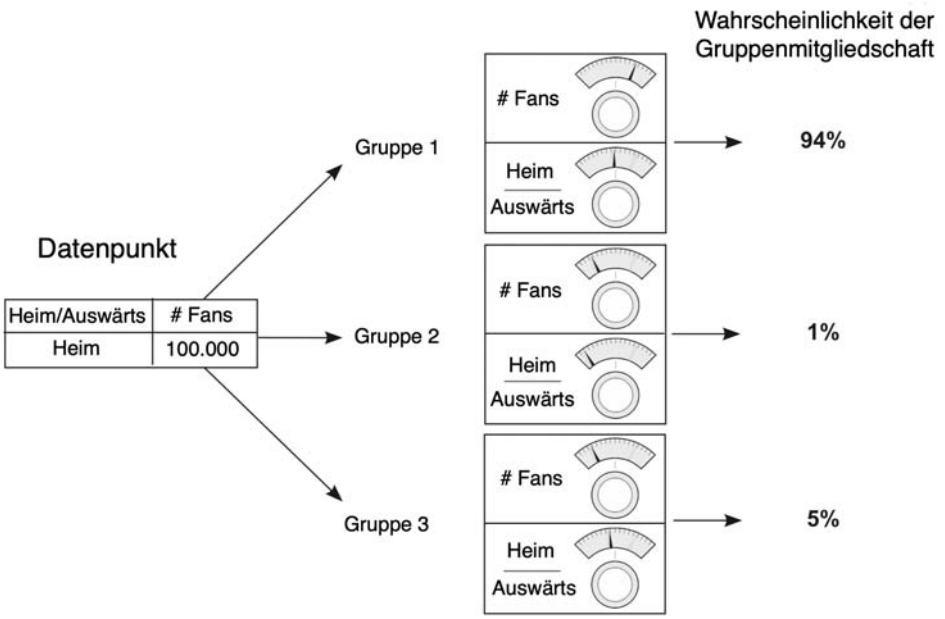
Theoretisch sollte dieser Schritt beim nächsten Vorliegen der gleichen Statistiken über Siege/Niederlagen weniger als 98 Prozent vorhersagen. Beachte auch, dass jeder Drehschalter die *Empfindlichkeit der Vorhersage für verschiedene Typen von Eingabedaten* repräsentiert. Das ist das, was du änderst, wenn die Maschine »lernt«.

2.7 Unüberwachtes parametrisches Lernen

Heim/Auswärts	# Fans
Heim	**100.000**
Auswärts	50.000
Heim	**100.000**
Heim	**99.000**
Auswärts	50.000
Auswärts	*10.000*
Auswärts	*11.000*

Das unüberwachte parametrische Lernen verwendet einen sehr ähnlichen Ansatz. Wir gehen die Schritte einmal der Reihe nach durch. Wie du weißt, geht es beim unüberwachten Lernen immer um das Gruppieren von Daten. Das unüberwachte *parametrische* Lernen verwendet die Drehschalter, um Daten zu gruppieren. Aber in diesem Fall gibt es für jede Gruppe mehrere Drehschalter, die jeweils die Stärke der Zugehörigkeit der Eingabedaten zu einer bestimmten Gruppe angeben (mit Ausnahmen – das ist eine allgemeine Beschreibung). Betrachten wir ein Beispiel, bei dem du die Daten in drei Gruppen unterteilen möchtest.

Die Datenmenge enthält drei Cluster, die das parametrische Modell entdecken soll. Sie sind durch Textformatierung als **Gruppe 1**, Gruppe 2 und *Gruppe 3* gekennzeichnet. Nun übergeben wir einem trainierten unüberwachten Modell den ersten Datenpunkt. Die Zuordnung (auch *Mapping* genannt) zur **Gruppe 1** ist am stärksten.



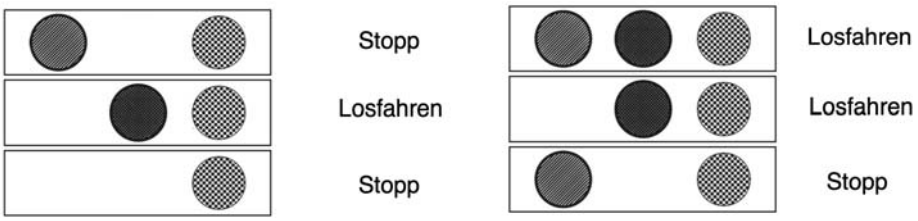
Die Maschine jeder Gruppe versucht, die Eingabedaten in eine Zahl zwischen 0 und 1 umzuwandeln und somit die Wahrscheinlichkeit anzugeben, dass die Einga-

bedaten zu dieser Gruppe gehören. Das Training für die Modelle unterscheidet sich deutlich und führt zu unterschiedlichen Eigenschaften, aber ganz allgemein werden die Parameter angepasst, um die Eingabedaten auf die zugehörige/n Gruppe/n abzubilden.

2.8 Nichtparametrisches Lernen

Grob vereinfacht: Auf Zählvorgängen beruhende Verfahren

Nichtparametrisches Lernen ist eine Algorithmenklasse, bei der die Anzahl der Parameter von den Daten abhängt (und nicht vorab festgelegt ist). Daher werden im Allgemeinen Verfahren verwendet, die auf die eine oder andere Weise zählen und so die Anzahl der Parameter anhand der in den Daten vorhandenen Objekte erhöhen. Bei einem überwachten nichtparametrischen Modell könnte beispielsweise gezählt werden, wie oft das Auftreten einer bestimmten Farbe einer speziellen Ampel dafür sorgt, dass Autos »losfahren«. Schon nach dem Zählen einiger weniger Beispiele wäre das Modell in der Lage vorherzusagen, dass das Aufleuchten der *mittleren* Lampe immer (in 100 Prozent der Fälle) dafür sorgt, dass Autos losfahren und das Aufleuchten der *rechten* Lampe nur manchmal (in 50 Prozent der Fälle).



Dieses Modell hätte drei Parameter: drei Zähler, die angeben, wie oft eine der Lampen aufleuchtet und Autos losfahren (eventuell geteilt durch die Anzahl der Beobachtungen). Bei fünf Lampen gäbe es fünf Zähler (fünf Parameter). Dieses einfache Modell ist *nichtparametrisch*, weil es die Eigenschaft hat, dass die Anzahl der Parameter sich in Abhängigkeit von den Daten (hier die Anzahl der Lampen) ändert. Das ist bei parametrischen Modellen anders, bei denen die Anzahl der Parameter anfangs zwar festgelegt ist, diese aber nach Ermessen des Forschers, der das Modell trainiert, erhöht oder verringert werden kann (unabhängig von den Daten).

Ein kritischer Beobachter könnte das infrage stellen. Beim vorangegangenen parametrischen Modell gab es offenbar für jeden Eingabedatenpunkt einen Drehschalter. Die meisten parametrischen Modelle benötigen irgendeine Art *Eingabe*, die von der Anzahl der Klassen in den Daten abhängt. Es gibt also eine *Grauzone* zwischen parametrischen und nichtparametrischen Algorithmen. Auch parametrische Algo-

rithmen werden in gewisser Weise von der Anzahl der Klassen in den Daten beeinflusst, sogar wenn es sich überhaupt nicht um ein Zählmuster handelt.

Hier wird auch deutlich, dass *Parameter* ein allgemeiner Begriff ist, der sich lediglich auf die Menge der Zahlen bezieht, die zum Modellieren eines Musters verwendet werden (ohne Berücksichtigung, wie diese Zahlen verwendet werden). Zähler sind Parameter. Gewichte sind Parameter. Normalisierte Varianten von Zählern und Gewichten sind ebenfalls Parameter. Korrelationskoeffizienten können auch Parameter sein. Der Begriff bezeichnet die Menge der Zahlen, die zur Modellierung eines Musters verwendet werden. Und Deep Learning ist eine Klasse von parametrischen Modellen. Ich werde in diesem Buch nicht näher auf nichtparametrische Modelle eingehen, aber es handelt sich um eine hochinteressante und leistungsstarke Algorithmenklasse.

2.9 Zusammenfassung

In diesem Kapitel haben wir die verschiedenen Ausprägungen des Machine Learnings etwas genauer betrachtet. Du hast erfahren, dass ein Machine-Learning-Algorithmus entweder überwacht oder unüberwacht und entweder parametrisch oder nichtparametrisch ist. Darüber hinaus haben wir erkundet, was genau diese vier Gruppen von Algorithmen unterscheidet. Du hast gelernt, dass überwachtes Machine Learning zu einer Algorithmenklasse gehört, bei der eine Datenmenge anhand einer anderen vorhergesagt wird, und dass unüberwachtes Lernen im Allgemeinen eine einzelne Datenmenge zu verschiedenen Clustern gruppiert. Zudem hast du erfahren, dass parametrische Algorithmen eine festgelegte Anzahl von Parametern besitzen und dass nichtparametrische Algorithmen die Anzahl der Parameter den Daten entsprechend anpassen.

Deep Learning verwendet sowohl für überwachte als auch für unüberwachte Vorhersagen neuronale Netze. Bislang sind wir auf theoretischer Ebene verblieben, damit du die Bedeutung für das Fachgebiet als Ganzes erkennst. Im nächsten Kapitel wirst du dein erstes neuronales Netz entwickeln, und alle nachfolgenden Kapitel sind *projektbasiert*. Also starte dein Jupyter Notebook und los geht's!