LEARNING MADE EASY

2nd Edition

Machine Learning

for dummies®

A Wiley Brand

- Perform in-depth analysis to extract interesting details
- Learn how machine learning algorithms are invaluable
- Implement algorithms in Python® and TensorFlow®

John Paul Mueller

Luca Massaron

Bestselling authors of the first edition

Machine Learning

2nd Edition

by John Paul Mueller and Luca Massaron

for dummies®
A Wiley Brand

Machine Learning For Dummies®, 2nd Edition

Published by: **John Wiley & Sons, Inc.,** 111 River Street, Hoboken, NJ 07030-5774, www.wiley.com

Copyright © 2021 by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permissions.

Trademarks: Wiley, For Dummies, the Dummies Man logo, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE

CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993, or fax 317-572-4002. For technical support, please visit https://hub.wiley.com/community/support/dummies.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this

material at http://booksupport.wiley.com. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2020952332

ISBN: 978-1-119-72401-8

ISBN 978-1-119-72406-3 (ebk); ISBN 978-1-119-72405-6 (ebk)

Machine Learning For Dummies®

To view this book's Cheat Sheet, simply go to www.dummies.com and search for “Machine Learning For Dummies Cheat Sheet” in the Search box.

Table of Contents

- Cover
- Title Page
- Copyright
- Introduction
  - About This Book
  - Foolish Assumptions
  - Icons Used in This Book
  - Beyond the Book
  - Where to Go from Here
- Part 1: Introducing How Machines Learn
  - Chapter 1: Getting the Real Story about AI
    - Moving beyond the Hype
    - Dreaming of Electric Sheep
    - Overcoming AI Fantasies
    - Considering the Relationship between AI and Machine Learning

Considering AI and Machine Learning Specifications
Defining the Divide between Art and Engineering
Predicting the Next AI Winter
Chapter 2: Learning in the Age of Big Data
Considering the Machine Learning Essentials
Defining Big Data
Considering the Sources of Big Data
Specifying the Role of Statistics in Machine Learning
Understanding the Role of Algorithms
Defining What Training Means
Chapter 3: Having a Glance at the Future
Creating Useful Technologies for the Future
Discovering the New Work Opportunities with Machine Learning
Avoiding the Potential Pitfalls of Future Technologies
Part 2: Preparing Your Learning Tools
Chapter 4: Installing a Python Distribution
Using Anaconda for Machine Learning
Installing Anaconda on Linux
Installing Anaconda on Mac OS X
Installing Anaconda on Windows
Downloading the Datasets and Example Code
Chapter 5: Beyond Basic Coding in Python
Defining the Basics You Should Know
Storing Data Using Sets, Lists, and Tuples
Defining Useful Iterators
Indexing Data Using Dictionaries
Chapter 6: Working with Google Colab
Defining Google Colab
Getting a Google Account
Working with Notebooks
Performing Common Tasks

Using Hardware Acceleration
Viewing Your Notebook
Executing the Code
Sharing Your Notebook
Getting Help
Part 3: Getting Started with the Math Basics
Chapter 7: Demystifying the Math Behind Machine Learning
Working with Data
Exploring the World of Probabilities
Describing the Use of Statistics
Chapter 8: Descending the Gradient
Acknowledging Different Kinds of Learning
The learning process
Optimizing with big data
Chapter 9: Validating Machine Learning
Considering the Use of Example Data
Checking Out-of-Sample Errors
Training, Validating, and Testing
Optimizing by Cross-Validation
Avoiding Sample Bias and Leakage Traps
Chapter 10: Starting with Simple Learners
Discovering the Incredible Perceptron
Growing Greedy Classification Trees
Taking a Probabilistic Turn
Part 4: Learning from Smart and Big Data
Chapter 11: Preprocessing Data
Gathering and Cleaning Data
Repairing Missing Data
Transforming Distributions
Creating Your Own Features
Delimiting Anomalous Data

Chapter 12: Leveraging Similarity

Measuring Similarity between Vectors

Using Distances to Locate Clusters

Tuning the K-Means Algorithm

Finding Similarity by K-Nearest Neighbors

Chapter 13: Working with Linear Models the Easy Way

Starting to Combine Features

Mixing Features of Different Types

Switching to Probabilities

Guessing the Right Features

Learning One Example at a Time

Chapter 14: Hitting Complexity with Neural Networks

Revising the Perceptron

Representing the Way of Learning of a Network

Introducing Deep Learning

Chapter 15: Going a Step Beyond Using Support Vector Machines

Revisiting the Separation Problem

Explaining the Algorithm

Classifying and Estimating with SVM

Chapter 16: Resorting to Ensembles of Learners

Leveraging Decision Trees

Working with Almost Random Guesses

Boosting Smart Predictors

Averaging Different Predictors

Part 5: Applying Learning to Real Problems

Chapter 17: Classifying Images

Working with a Set of Images

Revising the State of the Art in Computer Vision

Extracting Visual Features
Recognizing Faces Using Eigenfaces
Classifying Images
Chapter 18: Scoring Opinions and Sentiments
Introducing Natural Language Processing
Revising the State of the Art in NLP
Understanding How Machines Read
Using Scoring and Classification
Chapter 19: Recommending Products and Movies
Realizing the Revolution of E-Commerce
Downloading Rating Data
Catching the Limits of Behavioral Data
Integrating Text and Behaviors
Leveraging SVD
Part 6: The Part of Tens
Chapter 20: Ten Ways to Improve Your Machine Learning Models
Studying Learning Curves
Using Cross-Validation Correctly
Choosing the Right Error or Score Metric
Searching for the Best Hyper-Parameters
Testing Multiple Models
Averaging Models
Stacking Models
Applying Feature Engineering
Selecting Features and Examples
Looking for More Data
Chapter 21: Ten Guidelines for Ethical Data Usage
Obtaining Permission

Using Sanitization Techniques
Avoiding Data Inference
Using Generalizations Correctly
Shunning Discriminatory Practices
Detecting Black Swans in Code
Understanding the Process
Considering the Consequences of an Action
Balancing Decision Making
Verifying a Data Source

Chapter 22: Ten Machine Learning Packages to Master

Gensim
imbalanced-learn
OpenCV
SciPy
SHAP
Statsmodels
Modin
PyTorch
Poetry
Snorkel

Index
About the Authors
Advertisement Page
Connect with Dummies
End User License Agreement

List of Tables

Chapter 1

TABLE 1-1: Comparing Machine Learning to Statistics

Chapter 5

TABLE 5-1 Python Numeric Data Types

TABLE 5-2 Python Assignment Operators

TABLE 5-3 Python Arithmetic, Unary, and Bitwise Operators

TABLE 5-4 Python Relational and Logical Operators

TABLE 5-5 Python Membership and Identity Operators

TABLE 5-6 Python Operator Precedence

List of Illustrations

Chapter 2

FIGURE 2-1: The five tribes will combine their efforts toward the master algori...

Chapter 4

FIGURE 4-1: Tell the wizard how to install Anaconda on your system.

FIGURE 4-2: Configure the advanced installation options.

FIGURE 4-3: Anaconda Navigator provides centralized access to every development...

FIGURE 4-4: Jupyter Notebook provides an easy method to create machine learning...

FIGURE 4-5: New folders will appear with a name of Untitled Folder.

FIGURE 4-6: A notebook contains cells that you use to hold code.

FIGURE 4-7: Notebook uses cells to store your code.

FIGURE 4-8: The files that you want to add to the repository appear as part of ...

FIGURE 4-9: The `read_df` object contains the loaded dataset as a dataframe.

Chapter 6

FIGURE 6-1: Using Colab commands makes configuring your Notebook easy.

FIGURE 6-2: The Settings dialog box helps you configure the Colab IDE.

FIGURE 6-3: Customize shortcut keys for speed of access to commands.

FIGURE 6-4: Colab lets you compare two files to see how they differ.

FIGURE 6-5: Follow the prompts to create your Google account.

FIGURE 6-6: The sign-in page gives you access to all the general features, incl...

FIGURE 6-7: Create a new Python 3 Notebook using the same techniques as normal.

FIGURE 6-8: Use this dialog box to open existing notebooks.

FIGURE 6-9: When using GitHub, you must provide the location of the source code...

FIGURE 6-10: Your output may differ from the book's output when using Colab.

FIGURE 6-11: Colab maintains a history of the revisions for your project.

FIGURE 6-12: Using GitHub means storing your data in a repository.

FIGURE 6-13: Use Gists to store individual files or other resources.

FIGURE 6-14: Colab code cells contain a few extras not found in Notebook.

FIGURE 6-15: Use Cell panes to keep key cells easily available as needed.

FIGURE 6-16: Colab code cells contain a few extras not found in Notebook.

FIGURE 6-17: Use the GUI to make formatting your text easier.

FIGURE 6-18: The table of contents makes Notebook information more accessible.

FIGURE 6-19: Hardware acceleration speeds code execution.

FIGURE 6-20: The notebook information includes both size and settings.

FIGURE 6-21: Colab tracks which code you execute and in what order.

FIGURE 6-22: Send a message or obtain a link to share your notebook.

Chapter 8

FIGURE 8-1: A lack of evidence makes it hard to map back to the target function...

FIGURE 8-2: Noise can cause mismatches in the data points.

FIGURE 8-3: A plotting of parameter data against the output of the cost functio...

FIGURE 8-4: Visualizing the effect of starting point on outcome.

Chapter 9

FIGURE 9-1: Example of a linear model struggling to map a curve function.

FIGURE 9-2: A K-Nearest Neighbor model correctly fitting the problem on the lef...

FIGURE 9-3: Learning curves affected by high bias (left) and high variance (rig...

FIGURE 9-4: A graphical representation of how cross-validation works.

FIGURE 9-5: Comparing grid-search to random-search.

Chapter 10

FIGURE 10-1: The separating line of a perceptron across two classes.

FIGURE 10-2: A visualization of the decision tree built from the play tennis da...

FIGURE 10-3: A visualization of the pruning alphas and their impurity cost.

FIGURE 10-4: A visualization of the pruned decision tree build from the Titanic...

Chapter 11

FIGURE 11-1: A boxplot of the LSTAT feature from the Boston dataset.

FIGURE 11-2: A scatterplot of the first two components of a PCA of the Boston d...

FIGURE 11-3: A scatterplot of the last two components of a PCA of the Boston da...

Chapter 12

FIGURE 12-1: Examples of values plotted as points on a chart.

FIGURE 12-2: Clusters of penguins plotted on a chart based on first PCA dimensi...

FIGURE 12-3: Plot of the Calinski and Harabasz score regarding different cluste...

FIGURE 12-4: Penguin species represented by five clusters.

FIGURE 12-5: The bull’s-eye dataset, a nonlinear cloud of points that is diffic...

Chapter 13

FIGURE 13-1: Adding random features increases in-sample performances but degrad...

FIGURE 13-2: Visualizing the different optimization paths on the same data prob...

FIGURE 13-3: How R^2 varies in training and test sets as iterations increase in ...

Chapter 14

FIGURE 14-1: Learning logical XOR using a single separating line isn’t possible...

FIGURE 14-2: Plots of different activation functions.

FIGURE 14-3: An example of the architecture of a neural network.

FIGURE 14-4: A detail of the feed-forward process in a neural network.

FIGURE 14-5: Overfitting occurs when there are too many learning iterations on ...

FIGURE 14-6: Be sure to use the Anaconda prompt for the installation and check ...

FIGURE 14-7: Use the TF_env channel for all TensorFlow examples in the book.

FIGURE 14-8: The bidimensional half-moon problem.

FIGURE 14-9: Dropout temporarily rules out a proportion of the connections from...

FIGURE 14-10: Decision boundaries display how a neural network solves the half-...

FIGURE 14-11: Some images from the fashion mnist dataset.

FIGURE 14-12: A folded and unfolded RNN cell processing a sequence input.

FIGURE 14-13: The Air Passenger Data.

FIGURE 14-14: Predictions on the last two years of the Air Passenger Data.

Chapter 15

FIGURE 15-1: Comparing some different approaches: perceptron, logistic regressi...

FIGURE 15-2: A case of nonlinearly separable points requiring feature transform...

FIGURE 15-3: An RBF kernel that uses diverse hyper-parameters to create unique ...

FIGURE 15-4: A polynomial (left) and an RBF kernel (right) applied to the same ...

Chapter 16

FIGURE 16-1: Comparing a single decision tree output (left) to an ensemble of d...

FIGURE 16-2: Seeing the accuracy of ensembles of different sizes.

FIGURE 16-3: Permutation importance of features computed on the test set.

Chapter 17

FIGURE 17-1: The image appears onscreen after you render and show it.

FIGURE 17-2: Different filters for different noise cleaning.

FIGURE 17-3: Cropping the image makes it smaller.

FIGURE 17-4: Considering the effects of filtering, cropping, and resizing the i...

FIGURE 17-5: Detection, localization and segmentation example from the Coco dat...

FIGURE 17-6: Finding the borders of an image.

FIGURE 17-7: The example application would like to find similar photos.

FIGURE 17-8: The output shows the results that resemble the test image.

FIGURE 17-9: Examples from the training and test sets do differ in pose and exp...

Chapter 19

FIGURE 19-1: The output shows 1,000,209 cases and 10 features.

FIGURE 19-2: You can obtain a wealth of statistics about the movies.

Introduction

The term *machine learning* has all sorts of meanings attached to it today, especially after Hollywood (and other movie studios) have gotten into the picture. Films such as *Ex Machina* have tantalized the imaginations of moviegoers the world over and made machine learning into all sorts of things that it really isn't. Of course, most of us have to live in the real world, where machine learning actually does perform an incredible array of tasks that have nothing to do with androids that can pass the Turing Test (fooling their makers into believing they're human). *Machine Learning For Dummies,* 2nd Edition gives you a view of machine learning in the real world and exposes you to the amazing feats you really can perform using this technology.

Even though the tasks that you perform using machine learning may seem a bit mundane when compared to the movie version, by the time you finish this book, you realize that these mundane tasks have the power to impact the lives of everyone on the planet in nearly every aspect of their daily lives. In short, machine learning is an incredible technology — just not in the way that some people have imagined.

This second edition of the book contains a significant number of changes, not the least of which is that it's using pure Python code for the examples now upon request from our readers. You can still download R versions of every example, which is actually better than before when only some of the examples were available in R. In addition, the book contains new topics, including an entire chapter that discusses machine learning ethics.

About This Book

Machines and humans learn in entirely different ways, which is why the first part of this book is essential to your understanding of machine learning. Machines perform routine tasks at incredible speeds, but still require humans to do the actual thinking.

The second part of this book is about getting your system set up to use the various Python coding examples. The two setups work for desktop systems using Windows, Mac OS, or Linux, or mobile devices that have access to a Google Colab compatible browser.

TIP If you're using R, you'll find a README file in the R download file that contains instructions for configuring your R Anaconda environment.

The third part of the book discusses math basics with regard to machine learning requirements. It prepares you to perform math tasks associated with algorithms used in machine learning to make either predictions or classifications from your data.

The fourth part of the book helps you discover what to do about data that isn't quite up to par. This part is also where you start learning about similarity and working with linear models. The most advanced chapter tells you how to work with ensembles of learners to perform tasks that might not otherwise be reasonable to complete.

The fifth part of the book is about practical application of machine learning techniques. You see how to do things like classify images, work with opinions and sentiments, and recommend products and movies.

The last part of the book contains helpful information to enhance your machine learning experience. This part of the book also contains a chapter specifically oriented toward ethical data use.

To make absorbing the concepts easy, this book uses the following conventions:

- Text that you're meant to type just as it appears in the book is in **bold**. The exception is when you're working through a step list: Because each step is bold, the text to type is not bold.
- Web addresses and programming code appear in `monofont`. If you're reading a digital version of this book on a device connected to the Internet, you can click or tap the web address to visit that website, like this: https://www.dummies.com.
- When you need to type command sequences, you see them separated by a special arrow, like this: File ⇒ New File. In this example, you go to the File menu first and then select the New File entry on that menu.
- When you see words in *italics* as part of a typing sequence, you need to replace that value with something that works for you. For example, if you see "Type ***Your Name*** and press Enter," you need to replace *Your Name* with your actual name.

Foolish Assumptions

This book is designed for novice and professional alike. You can either read this book from cover to cover or look up topics and treat the book as a reference guide. However, we've made some assumptions about your level of knowledge when we put the book together. You should already know how to use your device and work with the

operating system that supports it. You also know how to perform tasks like downloading files and installing applications. You can interact with Internet well enough to locate the resources you need to work with the book. You know how to work with archives, such as the `.zip` file format. Finally, a basic knowledge of math is helpful.

Icons Used in This Book

As you read this book, you see icons in the margins that indicate material of interest. This section briefly describes each icon.

TIP The tips in this book are time-saving techniques or pointers to resources that you should try so that you can get the maximum benefit from machine learning.

WARNING You should avoid doing anything that's marked with a Warning icon. Otherwise, you might find that your application fails to work as expected, you get incorrect answers from seemingly bulletproof code, or (in the worst-case scenario) you lose data.

TECHNICAL STUFF Whenever you see this icon, think advanced tip or technique. Skip these bits of information whenever you like.

REMEMBER This text usually contains an essential process or a bit of information that you must know to perform machine learning tasks successfully.

Beyond the Book

If you want to email us, please do! Make sure you send your book-specific requests to: John@JohnMuellerBooks.com. We want to ensure that your book experience is the best one possible. The blog entries at http://blog.johnmuellerbooks.com/ contain a wealth of additional information about this book. You can check out John's website at http://www.johnmuellerbooks.com/. You can also access other cool materials:

- **Cheat Sheet:** A cheat sheet provides you with some special notes on things you can do with machine learning that not every other scientist knows. You can find the Cheat Sheet for this book at www.dummies.com. Type **Machine Learning For Dummies** in the Search box and click the Cheat Sheets option that appears.
- **Errata:** You can find errata by entering this book's title in the Search box at www.dummies.com, which takes you to this book's page. In addition to errata, check out the blog posts with answers to reader questions and demonstrations of useful book-related techniques at http://blog.johnmuellerbooks.com/.
- **Companion files:** The source code is available for download. All the book examples tell you precisely which example project to use. You can find these files at this book's page at www.dummies.com/go/machinelearningfd2e.

We've also had trouble with the datasets used in the previous edition of this book. Sometimes the datasets change or might become unavailable. Given that you likely don't want to download a large dataset unless you're interested in that example, we've made the non-toy datasets (those available with a package) available at https://github.com/lmassaron/datasets. You don't actually need to download them, though; the example code will perform that task for you automatically when you run it.

Where to Go from Here

Most people will want to start this book from the beginning, because it contains a good deal of information about how the real world view of machine learning differs from what movies might tell you. However, if you already have a first grounding in the reality of machine learning, you can always skip to the next part of the book.

Chapter 4 is where you want to go if you want to use a desktop setup, while Chapter 6 is helpful when you want to use a mobile device. Your preexisting setup may not work with the book's examples because you might have different versions of the various products. It's essential that you use the correct product versions to ensure success. Even if you choose to go with your own setup, consider reviewing Chapter 5 unless you're an expert Python coder already.

If you're already an expert with Python and know how machine learning works, you could always skip to Chapter 7. Starting at Chapter 7 will help you get into the examples quickly so that you spend less time with basics and more time with intermediate machine

learning tasks. You can always go back and review the previous materials as needed.

Part 1

Introducing How Machines Learn

IN THIS PART ...

- Discovering how AI really works and what it can do for you
- Considering what the term *big data* means
- Understanding the role of statistics in machine learning
- Defining where machine learning will take society in the future

Chapter 1

Getting the Real Story about AI

IN THIS CHAPTER

- **» Seeing the dream; getting beyond the hype of artificial intelligence (AI)**
- **» Comparing AI to machine learning**
- **» Understanding the engineering portion of AI and machine learning**
- **» Delineating where engineering ends and art begins**

A*rtificial Intelligence (AI)*, the appearance of intelligence in machines, is a huge topic today, and it's getting bigger all the time thanks to the success of new technologies (see some current examples at https://thinkml.ai/top-5-ai-achievements-of-2019/). However, most people are looking for everyday applications, such as talking to their smartphone. Talking to your smartphone is both fun and helpful to find out things like the location of the best sushi restaurant in town or to discover how to get to the concert hall. As you talk to your smartphone, it learns more about the way you talk and makes fewer mistakes in understanding your requests. The capability of your smartphone to learn and interpret your particular way of speaking is an example of an AI, and part of the technology used to make it happen is *machine learning*, the use of various techniques to allow algorithms to work better based on experience.

You likely make limited use of machine learning and AI all over the place today without really thinking about it. For example, the capability to speak to devices and have them actually do what you intend is an example of machine learning at work. Likewise, recommender systems, such as those found on Amazon, help you make purchases based on criteria such as previous product purchases or products that complement a current choice. The use of both AI and machine learning will only increase with time.

In this chapter, you delve into AI and discover what it means from several perspectives, including how it affects you as a consumer and as a scientist or engineer. You also discover that AI doesn't equal machine learning, even though the media often confuse the two. Machine learning is definitely different from AI, even though the two are related.

Moving beyond the Hype

As any technology becomes bigger, so does the hype, and AI certainly has a lot of hype surrounding it. For one thing, some people have decided to engage in fear mongering rather than science. Killer robots, such as those found in the film *The Terminator*, really aren't going to be the next big thing. Your first real experience with an android AI is more likely to be in the form a health care assistant (https://www.robotics.org/blog-article.cfm/The-Future-of-Elder-Care-is-Service-Robots/262) or possibly as a coworker (https://www.computerworld.com/article/2990849/meet-the-virtual-woman-who-may-take-your-job.html). The reality is that you interact with AI and machine learning in far more mundane ways already. Part of the reason you need

to read this chapter is to get past the hype and discover what AI can do for you today.

REMEMBER You may also have heard machine learning and AI used interchangeably. AI includes machine learning, but machine learning doesn't fully define AI. This chapter helps you understand the relationship between machine learning and AI so that you can better understand how this book helps you move into a technology that used to appear only within the confines of science fiction novels.

Machine learning and AI both have strong engineering components. That is, you can quantify both technologies precisely based on *theory* (substantiated and tested explanations) rather than simply *hypothesis* (a suggested explanation for a phenomenon). In addition, both have strong science components, through which people test concepts and create new ideas of how expressing the thought process might be possible. Finally, machine learning also has an artistic component, and this is where a talented scientist can excel. In some cases, AI and machine learning both seemingly defy logic, and only the true artist can make them work as expected.

YES, FULLY AUTONOMOUS WEAPONS EXIST

Before people send us their latest dissertations about fully autonomous weapons, yes, some benighted souls are working on such technologies. You'll find some discussions of the ethics of AI in this book, but for the most part, the book focuses on positive, helpful uses of AI to aid humans, rather than kill them, because most AI research reflects these uses. You can find articles on the pros and cons of AI online, such as the Towards Data Science article at https://towardsdatascience.com/advantages-and-disadvantages-of-artificial-

intelligence-182a5ef6588c and the *Emerj* article at https://emerj.com/ai-sector-overviews/autonomous-weapons-in-the-military/.

If you really must scare yourself, you can find all sorts of sites, such as https://www.reachingcriticalwill.org/resources/fact-sheets/critical-issues/7972-fully-autonomous-weapons, that discuss the issue of fully autonomous weapons in some depth. Sites such as Campaign to Stop Killer Robots (https://www.stopkillerrobots.org/) can also fill in some details for you. We do encourage you to sign the letter banning autonomous weapons at https://futureoflife.org/open-letter-autonomous-weapons/ — there truly is no need for them.

However, it's important to remember that bans against space-based, chemical, and certain laser weapons all exist. Countries recognize that these weapons don't solve anything. Countries will also likely ban fully autonomous weapons simply because the citizenry won't stand for killer robots. The bottom line is that the focus of this book is on helping you understand machine learning in a positive light.

Dreaming of Electric Sheep

Androids (a specialized kind of robot that looks and acts like a human, such as Data in *Star Trek: The Next Generation*) and some types of *humanoid robots* (a kind of robot that has human characteristics but is easily distinguished from a human, such as C-3PO in *Star Wars*) have become the poster children for AI (see the dancing robots at https://www.youtube.com/watch?v=lTckiTBaWkw). They present computers in a form that people can *anthropomorphize* (give human characteristics to, even though they aren't human). In fact, it's entirely possible that one day you won't be able to distinguish between human and artificial life with ease. Science fiction authors, such as Philip K. Dick, have long predicted such an occurrence, and it seems all too possible today. The story "Do Androids Dream of Electric Sheep?" discusses the whole concept of more real than real. The idea appears as part of the plot in the movie *Blade Runner*

(https://www.warnerbros.com/movies/blade-runner). However, some uses of robots today are just plain fun, as in the Robot Restaurant show at https://www.youtube.com/watch?v=l1vvTtz8hpg. The sections that follow help you understand how close technology currently gets to the ideals presented by science fiction authors and the movies.

The current state of the art is lifelike, but you can easily tell that you're talking to an android. Viewing videos online can help you understand that androids that are indistinguishable from humans are nowhere near any sort of reality today. Check out the Japanese robots at https://www.youtube.com/watch?v=LyyytwT-BMk and https://www.cnbc.com/2019/10/31/human-like-androids-have-entered-the-workplace-and-may-take-your-job.html. One of the more lifelike examples is Erica (https://www.youtube.com/watch?v=oRlwvLubFxg), who is set to appear in a science fiction film. Her story appears on *HuffPost* at https://www.huffpost.com/entry/erica-japanese-robot-science-fiction-film:n_5ef6523dc5b6acab284181c3. The point is, technology is just starting to get to the point where people may eventually be able to create lifelike robots and androids, but they don't exist today.

Understanding the history of AI and machine learning

There is a reason, other than anthropomorphization, that humans see the ultimate AI as one that is contained within some type of android. Ever since the ancient Greeks, humans have discussed the possibility of placing a mind inside a mechanical body. One such myth is that