Machine Learning

IBM Limited Edition

by Judith Hurwitz and Daniel Kirsch

for **dummies**®

A Wiley Brand

Machine Learning For Dummies®, IBM Limited Edition

Copyright © 2018 by John Wiley & Sons, Inc.

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, The Dummies Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. IBM and the IBM logo are registered trademarks of International Business Machines Corporation. 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.

Publisher's Acknowledgments

Table of Contents

Introduction

Machine learning is having a dramatic impact on the way software is designed so that it can keep pace with business change. Machine learning is so dramatic because it helps you use data to drive business rules and logic. How is this different? With traditional software development models, programmers wrote logic based on the current state of the business and then added relevant data. However, business change has become the norm. It is virtually impossible to anticipate what changes will transform a market.

The value of machine learning is that it allows you to continually learn from data and predict the future. This powerful set of algorithms and models are being used across industries to improve processes and gain insights into patterns and anomalies within data.

But machine learning isn't a solitary endeavor; it's a team process that requires data scientists, data engineers, business analysts, and business leaders to collaborate. The power of machine learning requires a collaboration so the focus is on solving business problems.

About This Book

Machine Learning For Dummies, IBM Limited Edition, gives you insights into what machine learning is all about and how it can impact the way you can weaponize data to gain unimaginable insights. Your data is only as good as what you do with it and how you manage it. In this book, you discover types of machine learning techniques, models, and algorithms that can help achieve results for your company. This information helps both business and technical leaders learn how to apply machine learning to anticipate and predict the future.

Chapter **1**

Understanding Machine Learning

achine learning, artificial intelligence (AI), and cognitive computing are dominating conversations about how emerging advanced analytics can provide businesses with a competitive advantage to the business. There is no debate that existing business leaders are facing new and unanticipated competitors. These businesses are looking at new strategies that can prepare them for the future. While a business can try different strategies, they all come back to a fundamental truth — you have to follow the data. In this chapter, we delve into what the value of machine learning can be to your business strategy. How should you think about machine learning? What can you offer the business based on advanced analytics technique that can be a game-changer?

What Is Machine Learning?

Machine learning has become one of the most important topics within development organizations that are looking for innovative ways to leverage data assets to help the business gain a new level of understanding. Why add machine learning into the mix? With the appropriate machine learning models, organizations have the ability to continually predict changes in the business so that they are best able to predict what's next. As data is constantly added, the machine learning models ensure that the solution is constantly updated. The value is straightforward: If you use the most appropriate and constantly changing data sources in the context of machine learning, you have the opportunity to predict the future.

Machine learning is a form of AI that enables a system to learn from data rather than through explicit programming. However, machine learning is not a simple process.

Machine learning uses a variety of algorithms that iteratively learn from data to improve, describe data, and predict outcomes. As the algorithms ingest training data, it is then possible to produce more precise models based on that data. A machine learning model is the output generated when you train your machine learning algorithm with data. After training, when you provide a model with an input, you will be given an output. For example, a predictive algorithm will create a predictive model. Then, when you provide the predictive model with data, you will receive a prediction based on the data that trained the model. Machine learning is now essential for creating analytics models.

You likely interact with machine learning applications without realizing. For example, when you visit an e-commerce site and start viewing products and reading reviews, you're likely presented with other, similar products that you may find interesting. These recommendations aren't hard coded by an army of developers. The suggestions are served to the site via a machine learning model. The model ingests your browsing history along with other shoppers' browsing and purchasing data in order to present other similar products that you may want to purchase.

are seeing major advancements and commercial solutions. Why has the market become real? There are six key enablers:

» Modern processors have become increasingly powerful and increasingly dense. The density to performance ratio has improved dramatically.

» The cost of storing and managing large amounts of data has been dramatically lowered. In addition, new storage innovations have led to faster performance and the ability to analyze vastly larger data sets.

» The ability to distribute compute processing across clusters of computers has dramatically improved the ability to analyze complex data in record time.

» There are more commercial data sets available to support analytics, including weather data, social media data, and medical data sets. Many of these are available as cloud services and well-defined Application Programming Interfaces (APIs).

» Machine learning algorithms have been made available through open-source communities with large user bases. Therefore, there are more resources, frameworks, and libraries that have made development easier.

» Visualization has gotten more consumable. You don't need to be a data scientist to interpret results, making use of machine learning broader within many industries.

Defining Big Data

Big data is any kind of data source that has at least one of four shared characteristics, called the four Vs:

» Extremely large *Volumes* of data

» The ability to move that data at a high *Velocity* of speed

» An ever-expanding *Variety* of data sources

» *Veracity* so that data sources truly represent truth

The accuracy of a machine learning model can increase substantially if it's trained on big data. Without enough data, you are

broadest way of describing systems that can "think." For example, thermostats that learn your preference or applications that can identify people and what they are doing in photographs can be thought of as AI systems.

As illustrated in Figure 1-1, there are four main subsets of AI. In this book, we focus on machine learning. However, in order to understand machine learning, it is important to put it in perspective.
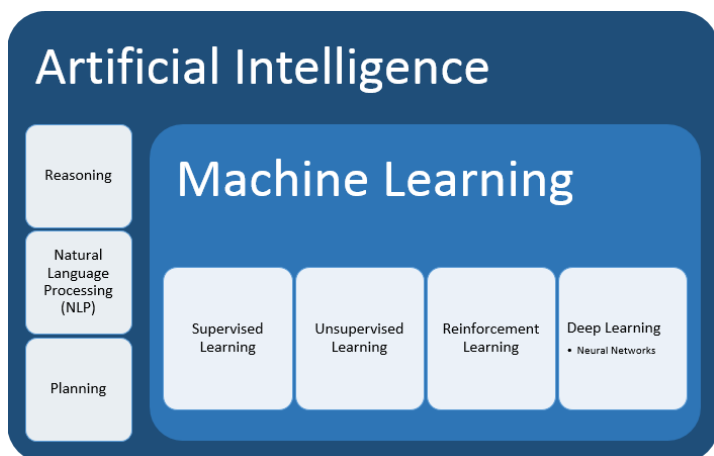
FIGURE 1-1: AI is the overall category that includes machine learning and natural language processing.

When we explore machine learning, we focus on the ability to learn and adapt a model based on the data rather than explicit programming. In Chapter 6, we focus on applying machine learning to solving business problems.

Before we delve into the types of machine learning, it is important to understand the other subsets of AI:

» **Reasoning:** Machine reasoning allows a system to make inferences based on data. In essence, reasoning helps fill in the blanks when there is incomplete data. Machine reasoning helps make sense of connected data. For example, if a system has enough data and is asked "What is a safe internal temperature for eating a drumstick?" the system would be capable of telling you that the answer is 165 degrees. The

Chapter **2**

Applying Machine Learning

With machine learning, you have the opportunity to use the data generated by your business to anticipate business change and plan for the future. While it is clear that machine learning is a sophisticated set of technologies, it is only valuable when you find ways to tie technology to outcomes. Your business is not static; therefore, as you learn more and more from your data, you can be prepared for business change.

Getting Started with a Strategy

Before you can define the strategy, you have to understand the problem that you're trying to solve. As businesses go through major strategy transitions, certain challenges present themselves. What is the status of existing business and existing customer engagement? What does the future hold for what customers will buy and expect from you in the future? The obvious answer is to ask customers if they are happy and what they will purchase in the future. While this is a sound starting point, it is not enough. Customers that are happy one minute become unhappy when something transformational comes along. If you do traditional

assessing your future. All of this data has the potential to help you look beyond the obvious and anticipate the future.

>> **How can you prepare to get your data in order?**

The challenge is to make sure that you have your data ready to perform the type of analytical analysis so that you can learn from the data that you have. Are you using the right data sources that are the most up to date? Have you put the data into a form that is usable? Are you protecting the identity of your customers' private data? Are you selecting the best third-party data sources that will put your own data in context with your industry?

While machine learning has captured the attention of the technology and business market, you want to make sure that you select the approach and tools that best match the problem you need to solve. There will be different approaches depending on your industry and the type of data you're dealing with and the type of results you're looking to achieve.

For many organizations, being able to understand the hidden patterns within their data offers a huge potential advantage. Most companies have important data that is stored in silos across different business units. Some of the important data may be found in social media sources. Data may also be found in unstructured data sources such as documents related to new research findings. Data is also found in semi-structured sources such as sensor and IoT-based systems.

Your first task is to determine what data sources and types of data are best suited to solving your problem. After you understand this, you will be in a good position to determine which algorithms will be used to create the most appropriate models. While there are hundreds of use cases to illustrate how to use machine learning algorithms to solve specific problems, we give you three examples in this section.

Bringing data silos together

You are in a competitive market with a lot of emerging companies determined to disrupt the market. Therefore, you have to figure out a way to understand the subtle changes to customer preferences and requirements. While you are diligent about conducting customer surveys and responding to customer complaints, this

Chapter **7**

Ten Predictions on the Future of Machine Learning

Machine learning is emerging as one of the most important developments in the software industry. While this advanced technology has been around for decades, it is now becoming commercially viable. We're moving into an era where machine learning techniques are essential tools to create value for businesses that want to understand the hidden value of their data. What does the future hold for machine learning? In this chapter, you explore our top ten predictions.

adding automation, technical users will be able to focus on more challenging work rather than simply automating repetitive tasks. There are many tedious details involved with machine learning that are important but ripe for automation (for example, data cleaning). Data visualization is another area where automation is helping to streamline the machine learning process. Systems can be designed to select the most appropriate visualization for a given data set, making it easy to understand the relationship between data points.

Specialized Hardware Will Improve the Performance of Machine Learning

We are approaching an era where sophisticated hardware is now affordable. Therefore, many organizations can procure hardware that is powerful enough to quickly process machine learning algorithms. In addition, this powerful hardware removes the processing bottleneck of machine learning, thus allowing machine learning to be embedded in more applications.

Traditionally, CPUs have been used to support the deep learning training process with mixed results. These CPUs are problematic because of the cumbersome way that they process steps in a neural network. In contrast, GPUs have hundreds of simpler cores that allow thousands of concurrent hardware threads. Because of the importance of GPUs in deep learning applications, there has been considerable research going into the technology in order to offer more powerful chips. Cloud computing vendors also recognize the value of GPUs, and more of them are offering GPU environments on the cloud.

In addition to GPUs, researchers are using Field-Programmable Gate Arrays (FPGAs) to successfully run machine learning workloads. Sometimes FPGAs outperform GPUs when running neural network and deep learning operations.

Automate Algorithm Selection and Testing Algorithms

Data scientists typically need to understand how to use dozens of specific machine learning algorithms. In Chapter 3, we discuss