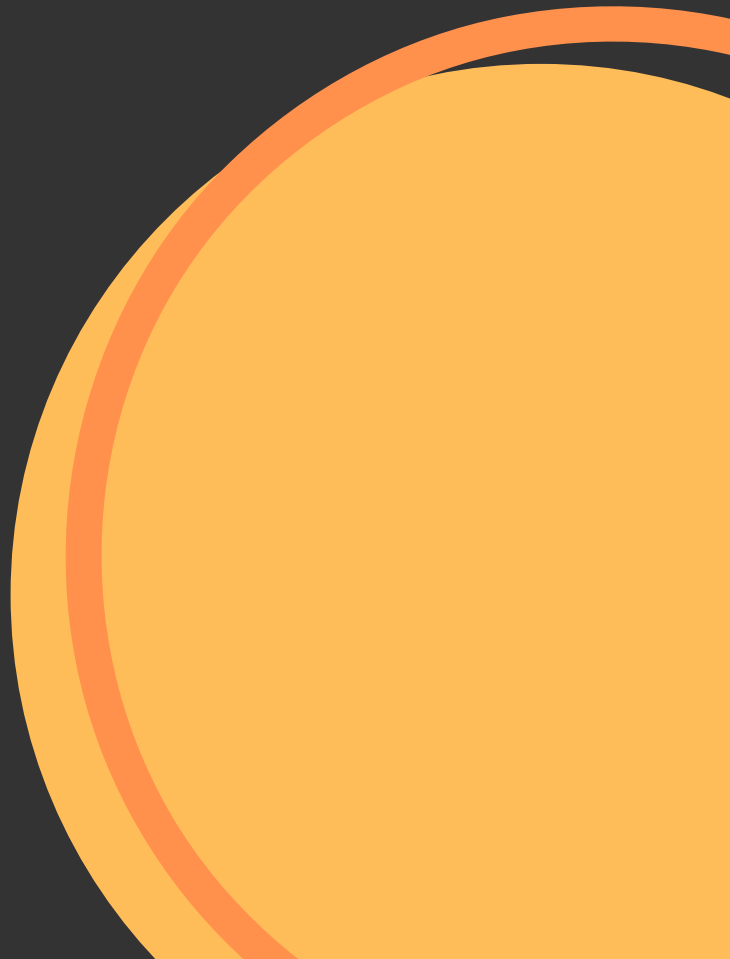Andriy Burkov's

THE HUNDRED-PAGE MACHINE LEARNING BOOK

"All models are wrong, but some are useful."
— George Box

The book is distributed on the "read first, buy later" principle.

Preface

Let's start by telling the truth: machines don't learn. What a typical "learning machine" does, is finding a mathematical formula, which, when applied to a collection of inputs (called "training data"), produces the desired outputs. This mathematical formula also generates the correct outputs for most other inputs (distinct from the training data) on the condition that those inputs come from the same or a similar statistical distribution as the one the training data was drawn from.

Why isn't that learning? Because if you slightly distort the inputs, the output is very likely to become completely wrong. It's not how learning in animals works. If you learned to play a video game by looking straight at the screen, you would still be a good player if someone rotates the screen slightly. A machine learning algorithm, if it was trained by "looking" straight at the screen, unless it was also trained to recognize rotation, will fail to play the game on a rotated screen.

So why the name "machine learning" then? The reason, as is often the case, is marketing: Arthur Samuel, an American pioneer in the field of computer gaming and artificial intelligence, coined the term in 1959 while at IBM. Similarly to how in the 2010s IBM tried to market the term "cognitive computing" to stand out from competition, in the 1960s, IBM used the new cool term "machine learning" to attract both clients and talented employees.

As you can see, just like artificial intelligence is not intelligence, machine learning is not learning. However, machine learning is a universally recognized term that usually refers to the science and engineering of building machines capable of doing various useful things without being explicitly programmed to do so. So, the word "learning" in the term is used by analogy with the learning in animals rather than literally.
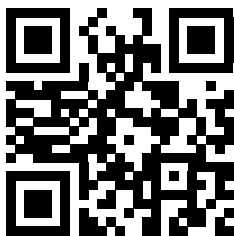
Who This Book is For

This book contains only those parts of the vast body of material on machine learning developed since the 1960s that have proven to have a significant practical value. A beginner in machine learning will find in this book just enough details to get a comfortable level of understanding of the field and start asking the right questions.

Practitioners with experience can use this book as a collection of directions for further self-improvement. The book also comes in handy when brainstorming at the beginning of a project, when you try to answer the question whether a given technical or business problem is "machine-learnable" and, if yes, which techniques you should try to solve it.

How to Use This Book

If you are about to start learning machine learning, you should read this book from the beginning to the end. (It's just a hundred pages, not a big deal.) If you are interested in a

specific topic covered in the book and want to know more, most sections have a QR code.

By scanning one of those QR codes with your phone, you will get a link to a page on the book's companion wiki theMLbook.com with additional materials: recommended reads, videos, Q&As, code snippets, tutorials, and other bonuses. The book's wiki is continuously updated with contributions from the book's author himself as well as volunteers from all over the world. So this book, like a good wine, keeps getting better after you buy it.

Scan the QR code on the left with your phone to get to the book's wiki.

QR Code

Some sections don't have a QR code, but they still most likely have a wiki page. You can find it by submitting the section's title to the wiki's search engine.

Should You Buy This Book?

This book is distributed on the "read first, buy later" principle. I firmly believe that paying for the content before consuming it is buying a pig in a poke. You can see and try a car in a dealership before you buy it. You can try on a shirt or a dress in a department store. You have to be able to read a book before paying for it.

The *read first, buy later* principle implies that you can freely download the book, read it and share it with your friends and colleagues. Only if you read and liked the book, or found it helpful or useful in any way, you have to buy it.

Now you are all set. Enjoy your reading!