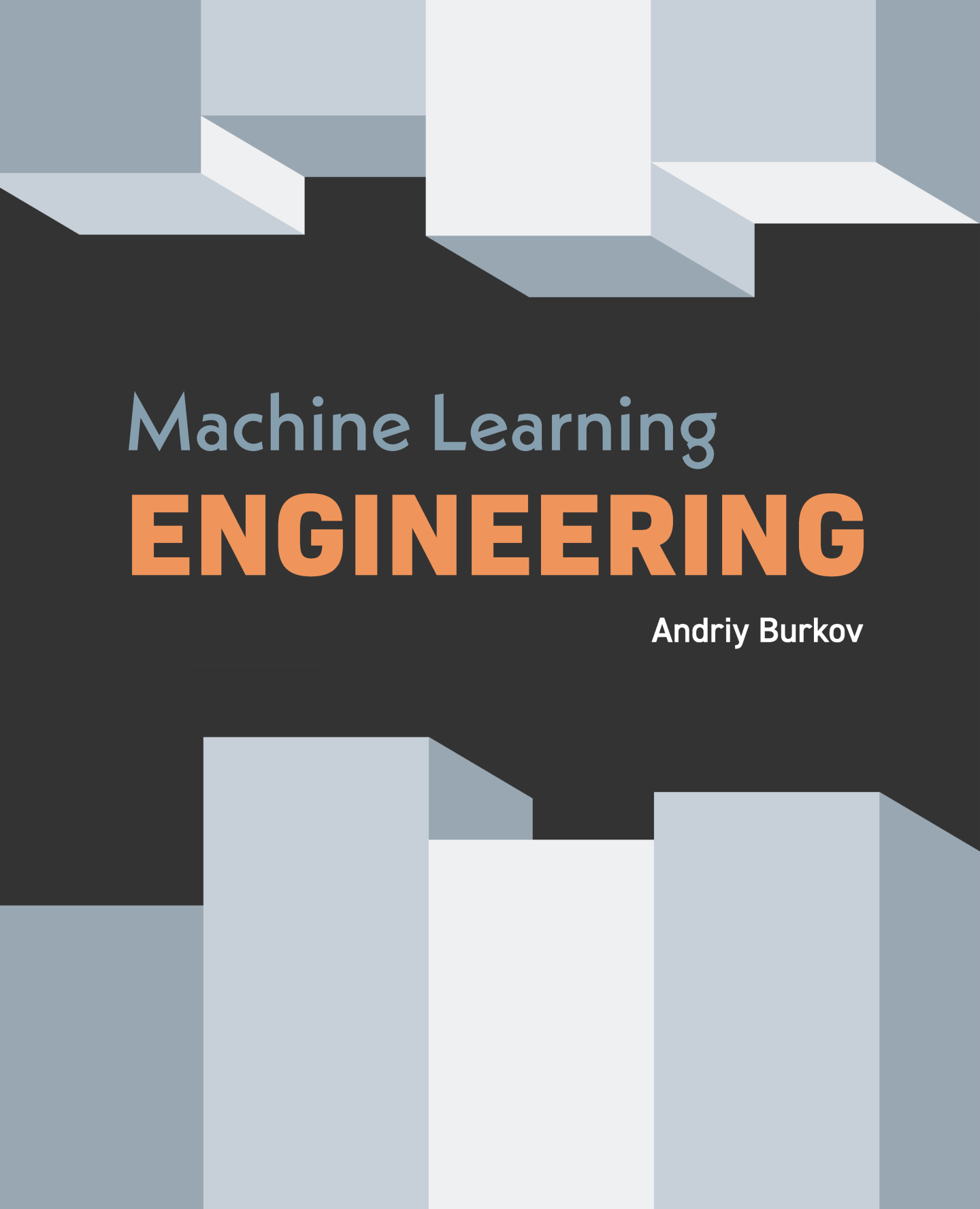
Machine Learning
ENGINEERING
Andriy Burkov

“In theory, there is no difference between theory and practice. But in practice, there is.”

— Benjamin Brewster

“The perfect project plan is possible if one first documents a list of all the unknowns.”

— Bill Langley

“When you’re fundraising, it’s AI. When you’re hiring, it’s ML. When you’re implementing, it’s linear regression. When you’re debugging, it’s printf().”

— Baron Schwartz

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

Foreword

Foreword by **Cassie Kozyrkov**, Chief Decision Scientist at Google, author of the course ***Making Friends with Machine Learning*** on Google Cloud Platform

I'd like to let you in on a secret: when people say "machine learning" it sounds like there's only one discipline here. Surprise! There are actually two machine learnings, and they are as different as innovating in food recipes and inventing new kitchen appliances. Both are noble callings, as long as you don't get them confused; imagine hiring a pastry chef to build you an oven or an electrical engineer to bake bread for you!

The bad news is that almost everyone does mix these two machine learnings up. No wonder so many businesses fail at machine learning as a result. What no one seems to tell beginners is that most machine learning courses and textbooks are about Machine Learning Research — how to build ovens (and microwaves, blenders, toasters, kettles... the kitchen sink!) from scratch, not how to cook things and innovate with recipes at enormous scale. In other words, if you're looking for opportunities to create innovative ML-based solutions to business problems, you want the discipline called Applied Machine Learning, not Machine Learning Research, so most books won't suit your needs.

And now for the good news! You're looking at one of the few true Applied Machine Learning books out there. That's right, you found one! A real applied needle in the haystack of research-oriented stuff. Excellent job, dear reader... unless what you were actually looking for is a book to help you learn the skills to design general purpose algorithms, in which case I hope the author won't be too upset with me for telling you to flee now and go pick up pretty much any other machine learning book. This one is different.

When I created Making Friends with Machine Learning in 2016, Google's Applied Machine Learning course loved by more than ten thousand of our engineers and leaders, I gave it a very similar structure to the one in this book. That's because doing things in the right order is crucial in the applied space. As you use your newfound data powers, tackling certain steps before you've completed others can lead to anything from wasted effort to a project-demolishing kablooie. In fact, the similarity in table of contents between this book and my course is what originally convinced me to give this book a read. In a clear case of convergent evolution, I saw in the author a fellow thinker kept up at night by the lack of available resources on Applied Machine Learning, one of the most potentially-useful yet horribly-misunderstood areas of engineering, enough to want to do something about it. So, if you're about to close this book, how about you do me a quick favor and at least ponder why the Table of Contents is arranged the way it is. You'll learn something good just from that, I promise.

So, what's in the rest of the book? The machine learning equivalent of a bumper guide to innovating in recipes to make food at scale. Since you haven't read the book yet, I'll put

it in culinary terms: you'll need to figure out what's worth cooking / what the objectives are (*decision-making and product management*), understand the suppliers and the customers (*domain expertise and business acumen*), how to process ingredients at scale (*data engineering and analysis*), how to try many different ingredient-appliance combinations quickly to generate potential recipes (*prototype phase ML engineering*), how to check that the quality of the recipe is good enough to serve (*statistics*), how to turn a potential recipe into millions of dishes served efficiently (*production phase ML engineering*), and how to ensure that your dishes stay top notch even if the delivery truck brings you a ton of potatoes instead of the rice you ordered (*reliability engineering*). This book is one of the few to offer perspectives on each step of the end-to-end process.

Now would be a good moment for me to be blunt with you, dear reader. This book *is* pretty good. It is. Really. But it's not perfect. It cuts corners on occasion — just like a professional machine learning engineer is wont to do — though on the whole it gets its message right. And, since it covers an area with rapidly-evolving best practices, it doesn't pretend to offer the last word on the subject. But even if it were terribly sloppy, it would still be worth reading. Given how few comprehensive guides to Applied Machine Learning are out there, a coherent introduction to these topics is worth its weight in gold. I'm so glad this one is here!

One of my favorite things about this book is how fully it embraces the most important thing you need to know about machine learning: mistakes are possible... and sometimes they hurt. As my colleagues in site reliability engineering love to say, "Hope is not a strategy." Hoping that there will be no mistakes is the worst approach you can take. This book does so much better. It promptly shatters any false sense of security you were tempted to have about building an AI system that is more "intelligent" than you are. (Um, no. Just no.) Then it diligently takes you through a survey of all kinds of things that can go wrong in practice and how to prevent/detect/handle them. This book does a great job of outlining the importance of monitoring, how to approach model maintenance, what to do when things go wrong, how to think about fallback strategies for the kinds of mistakes you can't anticipate, how to deal with adversaries who try to exploit your system, and how to manage the expectations of your human users (there's also a section on what to do when your, er, users are machines). These are hugely important topics in practical machine learning, but they're so often neglected in other books. Not here.

If you intend to use machine learning to solve business problems at scale, I'm delighted you got your hands on this book. Enjoy!

Cassie Kozyrkov

September 2020