Brief Contents

Contents

Preface

In 2012, I published a 1200-page book called "Machine learning: a probabilistic perspective", which provided a fairly comprehensive coverage of the field of machine learning (ML) at that time, under the unifying lens of probabilistic modeling. The book was well received, and won the De Groot prize in 2013.

2012 was also the year that is generally considered the start of the "deep learning revolution". The term "deep learning" refers to a branch of ML that is based on neural networks with many layers (hence the term "deep"). Although this basic technology had been around for many years, it was not until 2012 that it started to significantly outperform other, more "classical" approaches to ML, on several challenging benchmarks. For example, [KSH12] used deep neural networks (DNNs) to win the ImageNet image classification challenge, [CMS12] used DNNs to win a different image classification challenge, and [DHK13] used DNNs to outperform existing methods for speech recognition by a large margin. These breakthroughs were enabled by advances in hardware technology (in particular, the repurposing of fast graphics processing units from video games to ML), data collection technology (in particular, the use of crowd sourcing to collect large labeled datasets such as ImageNet), as well as various new algorithmic ideas.

Since 2012, the field of deep learning has exploded, with new advances coming at an increasing pace. Interest in the field has also exploded, fueled by the commercial success of the technology, and the breadth of applications to which it can be applied. Therefore, in 2018, I decided to write a second edition of my book, to attempt to summarize some of this progress.

By Spring 2020, my draft of the second edition had swollen to about 1600 pages, and I was still not done. At this point, 3 major events happened. First, MIT Press told me they could not publish a 1600 page book, and that I would need to split it into two volumes. Second, the COVID-19 pandemic struck, so I decided to put the book on hold, so I could work 100% on various internal and external modeling efforts. Third, as a consequence of my "pivot" towards COVID-19 work, I realized that I would never finish the book unless I got help from others; fortunately I managed to recruit several colleagues to help me write the last $\sim 15\%$ of "missing content". (See acknowledgements below.)

The result is two new books, "Probabilistic Machine Learning: An Introduction", which you are currently reading, and "Probabilistic Machine Learning: Advanced Topics", which is the sequel to this book [Mur22]. Together these two books attempt to present a fairly broad coverage of the field of ML c. 2020, using the same unifying lens of probabilistic modeling and Bayesian decision theory that I used in the first book.

Most of the content from the first book has been reused, but it is now split fairly evenly between the two new books. In addition, each book has lots of new material, covering some topics from deep learning, but also advances in other parts of the field, such as generative models, variational inference and reinforcement learning. To make the book more self-contained and useful for students, I have also added some more background content, on topics such as optimization and linear algebra, that was omitted from the first book due to lack of space.

Another major change is that nearly all of the software now uses Python instead of Matlab. The new code leverages standard Python libraries, such as numpy, scipy, scikit-learn, etc. Some examples also rely on various deep learning libraries, such as TensorFlow, PyTorch and JAX. In addition to scripts to create some of the figures, there are Jupyter notebooks to accompany each chapter, which discuss practical aspects that we don't have space to cover in the main text. Details can be found at `http://mlbayes.ai`.

Acknowledgements

I would like to thank the following people for helping me to write various parts of this book:

- Frederik Kunstner, Si Yi Meng, Aaron Mishkin, Sharan Vaswani, and Mark Schmidt who helped write parts of Chapter 5 (Optimization algorithms).
- Lihong Lig, who helped write parts of Sec. 8.3 (Bandit problems).
- Justin Gilmer, who helped write Sec. 14.6 (Adversarial Examples).
- Krzysztof Choromanski, who helped write Sec. 15.6 (Efficient transformers).
- Andrew Wilson, who helped write Sec. 17.4.2 (Exploiting structure in the kernel matrix).
- Colin Raffel, who helped write Sec. 19.2 (Transfer learning) and Sec. 19.6 (Semi-supervised learning).
- Bryan Perozzi, who helped write Chapter 23 (Graph embeddings).
- Zico Kolter, who helped write parts of Chapter B (Linear algebra).

I would like to thank John Fearns for very carefully proofreading the entire book, Peter Cerno who also spotted many errors.

I would like to thank the following people for feedback on various sections: Sebastien Bratieres, Kai Brodersen, Daniel Galvez, Abhishek Kumar, Max Lepikhin, Aaron Michelony, Hal Varian.

I would like to thank the following people for help with Python coding: Andrew Carr, Duane Rich, Theodore Vasiloudis. I would also like to thank those who shared their open soure code (see credits in each file online).

I would like to thank all those who shared figures from their own papers (see credits in each figure caption), as well as Sandeep Choudhary for help making some of the figures, and Aurélien Géron for letting me use the Python code from his excellent book [Gér19] to make some of the figures.

Finally I would like to thank my manager at Google, Doug Eck, for letting me invest the (significant) time needed to make this book a reality. I hope my efforts to synthesize all this material together in one place will help to save you time in your journey of discovery into the "land of ML".

Kevin Patrick Murphy
Palo Alto, California
December 2020.