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.

The year 2012 is also 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 in 2012 when [KSH12] used deep neural networks (DNNs) to win the ImageNet image classification challenge by such a large margin that it caught the attention of the wider community. Related advances on other hard problems, such as speech recognition, appeared around the same time (see e.g., [Cir+10; Cir+11; Hin+12]). These breakthroughs were enabled by advances in hardware technology (in particular, the repurposing of fast graphics processing units (GPUs) from video games to ML), data collection technology (in particular, the use of crowd sourcing tools, such as Amazon's Mechanical Turk platform, to collect large labeled datasets, such as ImageNet), as well as various new algorithmic ideas. some of which we cover in this book.

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 March 2020, my draft of the second edition had swollen to about 1600 pages, and I still had many topics left to cover. As a result, MIT Press told me I would need to split the book into two volumes. Then the COVID-19 pandemic struck. I decided to pivot away from book writing, and to help develop the risk score algorithm for Google's exposure notification app [MKS21] as well as to assist with various forecasting projects [Wah+21]. However, by the Fall of 2020, I decided to return to working on the book.

To make up for lost time, I asked several colleagues to help me finish by writing various sections (see acknowledgements below). The result of all this 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. 2021, using the same unifying lens of probabilistic modeling and Bayesian decision theory that I used in the 2012 book.

Nearly all of the content from the 2012 book has been retained, but it is now split fairly evenly between the two new books. In addition, each new book has lots of fresh material, covering topics from deep learning, as well as advances in other parts of the field, such as generative models, variational inference and reinforcement learning.

To make this introductory book more self-contained and useful for students, I have added some background material, on topics such as optimization and linear algebra, that was omitted from the 2012 book due to lack of space. Advanced material, that can be skipped during an introductory level course, is denoted by an asterisk * in the section or chapter title. Exercises can be found at the end of some chapters. Solutions to exercises marked with an asterisk * are available to qualified instructors by contacting MIT Press; solutions to all other exercises can be found online. See the book web site at probml.ai for additional teaching material (e.g., figures and slides).

Another major change is that all of the software now uses Python instead of Matlab. (In the future, we may create a Julia version of the code.) The new code leverages standard Python libraries, such as NumPy, Scikit-learn, JAX, PyTorch, TensorFlow, PyMC3, etc. Details on how to use the code can be found at probml.ai.

Acknowledgements

I would like to thank the following people for helping me with the book:

- Zico Kolter, who helped write parts of Chapter 7 (Linear Algebra).
- Frederik Kunstner, Si Yi Meng, Aaron Mishkin, Sharan Vaswani, and Mark Schmidt who helped write parts of Chapter 8 (Optimization).
- Mathieu Blondel, who helped write Section 13.3 (Backpropagation).
- Krzysztof Choromanski, who helped write Section 15.6 (Efficient transformers *).
- Colin Raffel, who helped write Section 19.2 (Transfer learning) and Section 19.3 (Semi-supervised learning).
- Bryan Perozzi, Sami Abu-El-Haija and Ines Chami, who helped write Chapter 23 (Graph Embeddings *).
- John Fearns and Peter Cerno for carefully proofreading the book.
- Many members of the github community for finding typos, etc (see `https://github.com/probml/pml-book/issues?q=is:issue` for a list of issues).
- The 4 anonymous reviewers solicited by MIT Press.
- Mahmoud Soliman for writing all the magic plumbing code that connects latex, colab, github, etc, and for teaching me about GCP and TPUs.
- The 2021 cohort of Google Summer of Code students who worked on code for the book: Aleyna Kara, Srikar Jilugu, Drishti Patel, Ming Liang Ang, Gerardo Durán-Martín. (See `https://probml.github.io/pml-book/gsoc2021.html` for a summary of their contributions.)
- Many members of the github community for their code contributions (see `https://github.com/probml/pyprobml#acknowledgements`).
- The authors of [Zha+20], [Gér17] and [Mar18] for letting me reuse or modify some of their open source code from their own excellent books.
- My manager at Google, Doug Eck, for letting me spend company time on this book.
- My wife Margaret for letting me spend family time on this book.

About the cover

The cover illustrates a neural network (Chapter 13) being used to classify a hand-written digit $\boldsymbol{x}$ into one of 10 class labels $y \in \{0, 1, \ldots, 9\}$. The histogram on the right is the output of the model, and corresponds to the conditional probability distribution $p(y|\boldsymbol{x})$.

Kevin Patrick Murphy
Palo Alto, California
August 2021.