

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 community. Related work appeared around the same time in several other papers, including [\[Cir+10; Cir+11; Hin+12\]](#). 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 March 2020, my draft of the second edition had swollen to about 1600 pages, and I was still not done. Then the COVID-19 pandemic struck. I decided to put the book writing on hold, and to “pivot” towards helping with various COVID-19 projects (see e.g., [\[MKS21; Wah+21\]](#)). However, in the Fall, when these projects were taking less of my cycles, I decided to try to finish the book. To make up for lost time, I asked several colleagues to help me finish the last $\sim 10\%$ of “missing content”. (See acknowledgements below.)

In the meantime, MIT Press told me they could not publish a 1600 page book, and that I would need to split it into two volumes. 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 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. Advanced material, that can be skipped during an introductory level course, is denoted by an asterisk * in the section or chapter title. In the future, we hope to post sample syllabuses and slides online.

Another major change is that nearly all of the software now uses Python instead of Matlab. (In the future, we hope to have a Julia version of the code.) 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 the code to create all the figures, there are supplementary 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 [problm.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 8 \(Optimization\)](#).
- Lihong Li, who helped write [Sec. 5.3 \(Bandit problems *\)](#).
- Mathieu Blondel, who helped write [Sec. 13.3 \(Backpropagation\)](#).
- Roy Frostig, who wrote [Sec. 7.8.8 \(Functional derivative notation *\)](#) and [Sec. 13.3.5 \(Automatic differentiation in functional form *\)](#).
- Justin Gilmer, who helped write [Sec. 14.7 \(Adversarial Examples *\)](#).
- Krzysztof Choromanski, who helped write [Sec. 15.6 \(Efficient transformers *\)](#).
- Colin Raffel, who helped write [Sec. 19.2 \(Transfer learning\)](#) and [Sec. 19.3 \(Semi-supervised learning\)](#).
- Bryan Perozzi, who helped write [Chapter 23 \(Graph embeddings *\)](#).
- Zico Kolter, who helped write parts of [Chapter 7 \(Linear algebra\)](#).

I would like to thank John Fearnas and Peter Cerno for carefully proofreading the book, as well as feedback from many other people, including 4 anonymous reviewers solicited by MIT Press.

I would like to thank Mahmoud Soliman for writing all the “back-office” code, that connects latex, colab, github and GCP. I would like to thank 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. I would also like to thank many other members of the github community for their code contributions (see the [full list of names](#)).

Finally I would like to thank my manager at Google, Doug Eck, for letting me spend company time on this book, and my wife Margaret for letting me spend family time on it, too. 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
April 2021.