Preface

> I am writing a longer [book] than usual because there is not enough time to write a short one. (Blaise Pascal, paraphrased.)

This book is a sequel to [Mur22]. That book mostly focused on techniques for learning functions $f : \mathcal{X} \to \mathcal{Y}$, where f is some nonlinear model, such as a deep neural network, $\mathcal{X}$ is the set of possible inputs (typically $\mathcal{X} = \mathbb{R}^D$), and $\mathcal{Y} = \{1, \ldots, C\}$ represents the set of labels for classification problems or $\mathcal{Y} = \mathbb{R}$ for regression problems. In that book, as in most of current ML, we assumed that the training data consists of iid labeled samples, and that the test distribution is the same as the training distribution. Judea Pearl, a well known AI researcher, has called this kind of ML a form of "glorified curve fitting" (quoted in [Har18]).

In this book, we expand the scope of ML to encompass more challenging problems. For example, we consider training and testing under different distributions; we consider generation of high dimensional outputs, such as images, text and graphs, so the output space is, say, $\mathcal{Y} = \mathbb{R}^{256x256}$ or $\mathcal{Y} = \{1, \ldots, C\}^T$; we discuss methods for discovering "insights" about data, based on latent variable models; and we discuss how to use probabilistic models and inference for causal inference and decision making.

We assume the reader has some prior exposure to ML and other relevant mathematical topics (e.g., probability, statistics, linear algebra, optimization). This background material is covered in the prequel to this book, [Mur22], although it is not strictly necessary to read [Mur22] first if you already have acquired such background knowledge from other sources.

Python code (mostly in JAX) to reproduce nearly all of the figures can be found online. In particular, if a figure caption says "Generated by foo", then you can find the corresponding notebook at `probml.github.io/notebooks.html#foo`. Clicking on the figure link in the pdf version of the book will take you to this list of notebooks. Clicking on the notebook link will open it inside Google Colab, which will let you easily reproduce the figure for yourself, and modify the underlying source code to gain a deeper understanding of the methods. (Colab gives you access to a free GPU, which is useful for some of the more computationally heavy demos.)

In addition to the online code, there is some additional supplementary online content at `probml.github.io/supp`. this contains additional material which was excluded from the main book for space reasons. For exercises (and solutions) related to the topics in this book, see [Gut22].

Contributing authors

This book is the result of a lot of effort from a lot of people. I would especially like to thank the following people who wrote or co-wrote various sections or chapters:

- Alex Alemi (Google), who wrote Section 5.1 (KL divergence).
- Roy Frostig (Google), who wrote Section 6.2 (Automatic differentiation).
- Marco Cuturi (Apple, work done at Google), who wrote Section 6.10 (Optimal Transport).
- Jeff Bilmes (U. Washington), who wrote Section 6.11 (Submodular optimization).
- Justin Gilmer (Google), who wrote Section 19.8 (Adversarial examples).
- Andrew Wilson (NYU), who helped write Chapter 17 (Bayesian neural networks) and Chapter 18 (Gaussian processes).
- George Papamakarios (Deepmind) and Balaji Lakshminarayanan (Google), who wrote Chapter 23 (Normalizing Flows).
- Yang Song (Stanford) and Durk Kingma (Google), who helped write Chapter 24 (Energy-based models).
- Mihaela Rosca (Deepmind / UCL), Shakir Mohamed (Deepmind) and Balaji Lakshminarayanan (Google), who wrote Chapter 26 (Generative adversarial networks).
- Vinayak Rao (Purdue), who wrote Chapter 31 (Non-parametric Bayesian models).
- Ben Poole (Google) and Simon Kornblith (Google), who wrote Chapter 32 (Representation learning **(Unfinished)**).
- Been Kim (Google) and Finale Doshi-Velez (Harvard), who wrote Chapter 33 (Interpretability).
- Lihong Li (Amazon, work done at Google), who helped write Section 34.4 (Contextual bandits) and Chapter 35 (Reinforcement learning).
- Victor Veitch (Google / U. Chicago) and Alexander D'Amour (Google), who wrote Chapter 36 (Causality).

Other contributors

I would like to thank the following people who helped in various other ways:

- Participants in the Google Summer of Code (GSOC) for 2021, including Aleyna Kara, Gerardo Duran-Martin, Ming Liang Ang, Srikar Reddy Jilugu, Drishti Patel, and Mahmoud Soliman.
- Participants in the Google Summer of Code (GSOC) for 2022, including Zeel Patel, Karm Patel, Nipun Batra, Peter Chang, Giles Harper-Donnelly, Qingyao Sun, Xinglong Li, and Scott Linderman.
- Mahmoud Soliman, Zeel Patel and Karm Patel, who helped significanty with latex, python, JAX, GCP, TPUs, etc, used to make many of the figures in this book.
- Many people who proofread parts of the book, including: Aalto Seminar students (numerous chapters), Kay Brodersen (Section 29.12.7), Peter Chang (Chapter 2, Chapter 3), Krzysztof Choromanski (Chapter 6, Chapter 11), Gerardo Duran-Martin (Chapter 19), John Fearns (almost the entire book!), Lehman Pavasovic Krunoslav (Chapter 10, Chapter 12), Amir Globerson (Chapter 4), Giles Harper-Donnelly (Chapter 24), Ravin Kumar (Chapter 2), Junpeng Lao (Chapter 13, Section 29.2, Chapter 29), Stephen Mandt (Chapter 21), Simon Prince (numerous chapters), Erik Sudderth (various chapters), Hal Varian (Section 29.12.7), Chris

Williams (numerous chapters), Raymond Yeh (numerous chapters), and others listed on github at https://github.com/probml/pml2-book/issues?q=is:issue.

- Many people who contributed figures (acknowledged in the captions).
- Many people who made their open source code available (acknowledged in the online code).

About the cover

The cover illustrates a variational autoencoder (Chapter 21) being used to map from a 2d Gaussian to image space.

Kevin Patrick Murphy
Palo Alto, California
July 2022.