Preface

This book is a sequel to [Mur22]. That book mostly focused on techniques for learning functions $f : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ is the set of possible inputs (typically $\mathcal{X} = \mathbb{R}^D$), $\mathcal{Y}$ represents the set of labels (for classification problems) or real values (for regression problems), and f is some nonlinear model, such as a deep neural network. We assumed that the training data consists of iid labeled samples, $\mathcal{D} = \{(\boldsymbol{x}_n, \boldsymbol{y}_n) \sim p(\boldsymbol{x}, \boldsymbol{y}) : n = 1 : N\}$, 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 learning and testing under multiple different distributions; we consider generation of high dimensional outputs, such as images, text and graphs; we discuss methods for discovering "insights" about data, based on latent varibale models; and we discuss how to use probabilistic models and inference for decision making and control tasks.

We assume the reader has some prior exposure to (supervised) 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 the current book is self-contained, and does not require that you read [Mur22] first.

Since this book cover so many topics, it was not possible to fit all of the content into these pages. Some of the extra material can be found in an online supplement at probml.ai. This site also contains Python code for reproducing most of the figures in the book. In addition, because of the broad scope of the book, about one third of the chapters are written, or co-written, with guest authors, who are domain experts (see the full list of contributors below). I hope that by collecting all this material in one place, new ML researchers will find it easier to "see the wood for the trees", so that we can collectively advance the field using a larger step size.

Contributing authors

I would like to thank the following people who co-wrote various parts of this book:

- Alex Alemi (Google), who wrote Section 5.1 (KL divergence).
- Marco Cuturi (Apple, work done at Google), who wrote Section 6.9 (Optimal Transport).
- Jeff Bilmes (U. Washington), who wrote Section 6.10 (Submodular optimization).
- Justin Gilmer (Google), who wrote Section 20.8 (Adversarial examples).

- Roy Frostig (Google), who wrote Section 16.4 (Automatic differentiation).
- 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 24 (Normalizing Flows).
- Yang Song (Stanford) and Durk Kingma (Google), who helped write Chapter 25 (Energy-based models).
- Mihaela Rosca (Deepmind / UCL), Shakir Mohamed (Deepmind) and Balaji Lakshminarayanan (Google), who wrote Chapter 27 (Generative adversarial networks).
- Viayak Rao (Purdue), who wrote Chapter 33 (Non-parametric Bayesian models).
- Ben Poole (Google) and Simon Kornblith (Google), who wrote Chapter 34 (Representation learning **(Unfinished)**).
- Been Kim (Google) and Finale Doshi-Velez (Harvard), who wrote Chapter 35 (Interpretability).
- Lihong Li (Amazon, work done at Google), who helped write Section 36.4 (Contextual bandits) and Chapter 37 (Reinforcement learning).
- Victor Veitch (Google / U. Chicago) and Alexander D'Amour (Google), who wrote Chapter 38 (Causality).

Acknowledgements

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

- Mahmoud Soliman, for help with many issues related to latex, python, GCP, TPUs, etc.
- Aleyna Kara, for help with many figures and software examples.
- Gerardo Duran-Martin for help with many software examples.
- Participants in the Google Summer of Code for 2021 and 2022.
- Numerous people who proofread parts of the book, including: Kay Brodersen (Section 19.3.5), Krzysztof Choromanski (Chapter 6, Chapter 11), John Fearns (an earlier version of the the whole book), Lehman Pavasovic Krunoslav (Chapter 10, Chapter 12), Amir Globerson (Chapter 4), Ravin Kumar (Chapter 2) Scott Linderman (Section 31.4.1), Hal Varian (Section 19.3.5), Chris Williams (Part IV), Raymond Yeh (Chapter 16, Chapter 23), et al.
- Numerous people who contributed figures (acknowledged in the captions).
- Numerous people who made their open source code available (acknowledged in the online code).

About the cover

The cover illustrates a variational autoencoder (Chapter 22) being used to map from a 2d Gaussian to image space.

Kevin Patrick Murphy
Palo Alto, California
March 2022.