

Proposal

Machine Intelligence: Applied Mathematics and Machine Learning

Authors

Ruye Wang is a Professor in the Engineering Department at Harvey Mudd College. Previously a Principal Investigator at the Jet Propulsion Laboratory, NASA, his research interests include image processing, computer vision, machine learning and remote sensing. Ruye's teaching spans machine learning, electrical engineering, and applied mathematics, and is the author of *Introduction to Orthogonal Transforms With Applications in Data Processing and Analysis*, (Cambridge University Press, 2012). He is the principal investigator of multiple research grants, including a NASA research grant of \$500,000. His collective works in image processing, pattern recognition, and signal processing have received over 1,500 citations.

340 E. Foothill Blvd., Claremont, CA 91711 · rwang@g.hmc.edu · (555) - 555 - 5555

Nick Richardson is a fourth-year undergraduate at Harvey Mudd College studying applied mathematics. His interests within machine learning include Bayesian nonparameterics, metric learning, applied information theory, and differential geometry. Nick has served as class president and student government chair at Harvey Mudd, as well as a tutor in the engineering, computer science, and mathematics departments. He was head teaching assistant and guest lecturer for engineering 190, an upper division machine learning course tailored for junior and senior engineering majors. Nick has worked as a machine learning engineer/applied mathematician, and currently works remotely from school as a machine learning consultant to a biometrics startup in Redwood City, CA.

340 E. Foothill Blvd. # 544, Claremont, CA 91711 · nrichardson@g.hmc.edu · (708) - 979 - 1500

Background

The pursuit of machine intelligence (AI or MI) in recent decades has been dominated by a largely *statistical* approach to achieve adaptive systems which can learn from data and reason under uncertainty. The aggregation of immense datasets, increased computational capacities, and modest theoretical advancements have cultivated a fascinating intersection of applied mathematics and computer science; the field of machine learning. It is difficult to overstate the current public excitement surrounding the development of commercially valuable MI systems. It is incontrovertible that we lie on a trajectory toward increasing the level of sophistication and breadth of application areas, which ideally will enhance our economic and scientific productivity. It is worth noting that we do not regard machine learning or MI as unique or special in this regard, but simply as one of several (e.g., genetic engineering, quantum computing) exciting technological developments beginning to mature.

Just several decades ago, enthusiasts in fields like statistics, optimization, statistical physics, signal processing, and operations research viewed machine learning research as reinventing the wheel¹. Likewise, the field of machine learning did not initially comprehend the important contributions made by these other fields as applied to their own work in prediction and pattern recognition. In the modern era, however, the sibling fields noted above are recognizing the novel contributions by machine learning, and machine learning researchers are utilizing the intellectual foundation of myriad disciplines (many previously thought irrelevant). It is in this sense that we say machine learning crystallized from both science and engineering. That is, there is room to view machine learning on the one hand as the application of rigorous tools from applied mathematics, and on the other as an 'art' dedicated to building systems at the cutting edge of empirical power, without mention of proofs or bounds regarding the behaviors of these systems. Subfields of machine learning like Bayesian inference and deep learning seemingly compete with one another for the notion of a 'correct' perspective on machine learning. All the while, the realization of commercial value in machine learning and related fields has drastically increased industrial/public attention. The consequence of this attention is an immense corpus of pedagogical resources: textbooks, so-called 'bootcamps', and courses (online or otherwise), to teach machine learning.

One thing is perhaps clear: empirically powerful engineered systems will continue to be applied to a growing variety of tasks, and many students of these methods will apply these methods without a sophisticated understanding of their boundaries. As we allude to above, we believe the historical dichotomy between rigor and empiricism in machine learning is superficial at best. Rigor should not be synonymous with theory, nor empiricism with ad-hocery. Our aim with this text is to provide a novel view on teaching machine learning; taking a student all the way from applied

¹Sentiment taken from [Wasserman, 2013]

mathematics to practical considerations in the construction of machine intelligence systems. We take ‘rigor’ to mean doing scientifically reproducible and systematic machine learning, and ‘empiricism’ to mean rationality in our engineering design decisions.

With this in mind, we’ve set out to capture the *fundamentals* of machine learning. Like any reasonably mature intellectual enterprise, machine learning cannot be ‘understood’ or ‘learned’ in a single book, any more than photography or economics. Instead, we set out to cover the central concepts of machine learning at depth, giving the reader the confidence and vocabulary to pursue further study from whichever department they reside.

Readership

This book is intended for undergraduates in mathematics, computing, engineering, and science. It can be used as a resource for a course in machine learning, pattern recognition, or inference, in the mathematics, computer science, or engineering departments. Several roadmaps are provided for possible course outlines. In addition, the software included with the book is meant to serve two communities. First, for the student. Working with these tools in a hands-on manner, at the correct level of abstraction, is invaluable as a pedagogical tool. Second, for the practitioner. The software we provide (as well as the software we point to) may serve as a reference for new or senior professionals working in related industries. It was with this second community in mind that we chose to write the software in both the Python and Matlab languages; Matlab for more senior professionals coming from an engineering background, and the corresponding Python code in an attempt to modernize the craft.

Brief Description of the Book

This book, in short, is concerned with the methods required to understand and build systems which learn their behaviors from data. That is, engineered systems which can autonomously recover structure in data, and utilize that structure to make inferences about the world. The book connects concepts from applied mathematics and mathematical inference with the practical engineering guidelines required to solve difficult problems in the real world.

The range of topics presented in the book was designed with the intent of introducing readers to the core ideas in machine learning; it is not meant to be encyclopaedic. As the book is intended for undergraduates, we do not assume a rich knowledge of linear algebra, probability theory, or signal processing. Instead, we introduce the important parts of these fields (important here meant with respect to machine learning) in the first section of the book: Foundations. Rather than introducing definitions outside of context, the core ideas from these foundational fields of applied math are introduced with sample applications, and hinting all the way at later machine learning applications to be introduced later in the text. Some examples of these hints are given in *Why another Machine Learning Textbook?*.

As we see it, there are two primary paths to gaining experience in machine learning as an undergraduate. The first is immersion in textbooks and courses on statistics, probability theory, optimization, and information theory. This student might indeed pick up a copy of a graduate level tome like [Murphy, 2012] or [Bishop, 2016]. Barring industry experience, however, this student often finds herself less confident in her ability to *apply* this theoretical foundation. The second path is to attend ‘data science’ coding camps, online courses, or online competitions (e.g., Kaggle). This student is often quickly able to attend to problems using premade packages, but often lacks a deep understanding of the methods being used, nor the effects of her design decisions. It is our opinion that neither of these paths is pedagogically sensible. On one hand, the genesis of some of the greatest insights and connections is when one attempts to *apply* theory, hence many of the wonderful discoveries through applied mathematics. On the other, one cannot hope to generalize to novel or serious engineering challenges without understanding the machinery of the fantastic packages available.

Our book attempts to ride the edge between providing a hearty theoretical foundation and genuinely useful advice on writing programs for these methods. We write our own software package to complement the mathematical presentation in the book, and illustrate specific conceptual details that might otherwise be lost by a purely theoretical explanation. As a concrete example, we discuss the theoretical discussion of bayesian versus frequentist methods, but more importantly, design software exercises and demos that illustrate the differences in an actual application domain. We write these programs in both Matlab and Python, with the intention that first learners are exposed to two tools (it’s always valuable to see an idea expressed in two languages), and so that senior researchers might see familiar Matlab implementations translated to the currently dominant language for data analysis.

In terms of content, the book is partitioned into the following sections: *Foundations*, *Representation*, *Inference*, *Application*, and *Frontiers*. The Foundations material covers probabilities and inference, linear algebra, and basic signal processing, all as applied to machine learning. The Representation section is concerned with constructing an effective basis for a set of data. This section is intended as the intellectual workhorse of the book, introducing basis function expansions, adaptive basis function methods, unsupervised learning methods, metric learning, and several examples of data representation. We mean ‘workhorse’ in the sense that we introduce topics like neural networks as adaptive basis function methods here, arguing that the central contribution of these methods is the construction of effective data representation, and that performing inference with these systems is simply extending their representational capacity with a linear regressor or classifier on that learned basis. In this sense, the Inference section describes the use of these basis representations to make predictions, typically with linear models like linear and logistic regression, which are discussed at length. Furthermore, the Application section engages a variety of practical questions concerning a machine learning engineer or researcher: How does one begin a machine learning project? How does one explore a dataset and understand the difficulty of a given problem? Once a problem has been specified, how can one choose a set of methods to engage with the problem, and then choose again among that set for the final system? How does one choose so called ‘hyperparameters’? None of these questions have *answers* in any precise sense, but practical advice and guidelines are given on these topics and others. Finally, Frontiers in machine learning and its applications are discussed. Sure to be out of date quickly, as any ‘Frontiers’ section ought to be in a growing scientific field, we discuss here important applications of machine learning to modern problems: autonomy in vehicles, medical robotics, and weapon systems, the interfacing of biological neural systems with computational ones, and ponderings on the quest for machine intelligence.

Why *another* Machine Learning Textbook?

Unlike many of the more popular machine learning books available ([Murphy, 2012], [Bishop, 2016], [Hastie et al., 2013]), this book does not assume or require a great deal of mathematical maturity. The primary feature of our text will be its well-paced and clear bridging of the applied mathematics and the practical engineering. To be precise, we introduce all of the basic mathematics required to understand later sections of the book in the preliminary Foundations portion, and feel we have done so in a unique and effective manner. Rather than parroting the typical theorems and definitions given in a book on mathematics, each section in Foundations provides hints and nods to later content, giving the reader several visits to the same idea, within the same textbook. Some examples of these hints are given below:

- In the section on probability, we introduce point estimation (maximum likelihood and maximum a posteriori), markov chain monte carlo, and the language of Bayesian inference.
- In the section on linear algebra, we introduce k -means clustering as a taste of unsupervised learning, compositions of affine functions as a preview to neural networks (which simply place a nonlinearity between each affine function), the singular value decomposition as background for principal components analysis, and the least squares problem as a noise-model-agnostic interpretation of linear regression.
- In the section on signal processing, we introduce the Kalman filter and the Hidden Markov model, later discussed in the context of latent variable models. We also introduce many basic data preprocessing and exploratory data analysis steps like rescaling, centering, histograms, viewing the decay of the data spectrum, filtering, etc.

In addition to our approach in presenting machine learning fundamentals in a palatable and comprehensive way, we feel our commitment toward practical guidelines and advice is an even more powerful departure from current texts. In particular, we have written an entire section (one of five in the text) on machine learning applications, providing a complete beginner with a small guidebook to her machine learning problem. In that section we engage with a variety of common questions (see *Brief Description of the Book*), as well as providing several sample machine learning projects with the accompanying code (in the form of documented Jupyter notebooks) to follow along and replicate the results.

This book is in one sense a reaction to and a unification of machine learning as a field of science and engineering for the next generation in the field. The book does not represent a contribution to the debate, per se, but an understanding of the value in both perspectives. To the scientist, we write so that she is able to be rigorous and powerful in the *application* of her theory. To the engineer, we write so that she may gain a rich *understanding* of the powers, the weaknesses, and the boundaries of the rapidly expanding set of tools at her disposal.

Competition

Books Marketed Towards Senior Undergraduates and Graduate Students

§ *Machine Learning: A Probabilistic Perspective*, Kevin Murphy

§ *Pattern Recognition and Machine Learning*, Christopher Bishop

§ *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Hastie et al.

Of these texts, our book differs strongly in its expectations of mathematical maturity in the reader, and our emphasis on depth rather than breadth. Unlike [Murphy, 2012], we do not attempt to cover anywhere near the entirety of the field. We also do not assume a full course in statistics, or an understanding of linear regression, as does [Hastie et al., 2013]. Our book covers a smaller number of methods, with an emphasis on a thorough explanation of these methods and the accompanying applications. In general, the texts noted above have a much stronger emphasis on theory as opposed to implementation.

Books Marketed Towards Junior Undergraduates

§ *Foundations of Machine Learning*, Mehryar Mohri et al.

[Mohri et al., 2018] takes a particular learning theoretic flavor that we do not. For example, entire sections of Mohri’s book are dedicated to PAC learning, VC-dimensions, and Learning Automata and Languages. Mohri’s text also aims to briefly cover wide ground, moving quickly from learning theory, to ranking problems, then to online learning and eventually reinforcement learning. Though we find all of these topics important, our aim is to write a book that presents a thorough vision of a few parts of the field, rather than a bit of everything.

§ *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, Aurelien Geron

While [Géron, 2017] does an excellent job introducing the reader to the practical tools in machine learning, it is intended more as a handbook for the interested practitioner, treating most methods as black boxes, and generally avoiding theoretic discourse about the methods.

§ *Machine Learning: An Algorithmic Perspective*, Stephen Marsland

Marsland’s text is written primarily for the software developer and engineer, as opposed to the applied mathematician. This is to say that Marsland generally avoids theory in favor of providing useful and practical application guidelines. For example, he writes that “When I started to look for a textbook that was suitable for classes of undergraduate computer science and engineering students, I discovered that the level of mathematical knowledge required was (unfortunately) rather in excess of that of the majority of the students” [Marsland, 2015]. While we broadly agree with his assertion, our solution is to revisit the mathematical foundations rather than to avoid them, and the synthesis of mathematics and application is a major theme of our text.

References

- [Bishop, 2016] Bishop, C. (2016). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer New York.
- [Géron, 2017] Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly Media.
- [Hastie et al., 2013] Hastie, T., Tibshirani, R., and Friedman, J. (2013). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer New York.
- [Marsland, 2015] Marsland, S. (2015). *Machine Learning: An Algorithmic Perspective, Second Edition*. CRC Press.
- [Mohri et al., 2018] Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2018). *Foundations of Machine Learning*. Adaptive Computation and Machine Learning series. MIT Press.
- [Murphy, 2012] Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning series. MIT Press.
- [Wasserman, 2013] Wasserman, L. (2013). *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer New York.