Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618

1 author:

Jeffrey Theodore Heaton
Washington University in St. Louis
47 PUBLICATIONS **2,630** CITATIONS

CrossMark

B O O K   R E V I E W

Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning

The MIT Press, 2016, 800 pp, ISBN: 0262035618

Jeff Heaton[1] ⓘ

Deep Learning provides a truly comprehensive look at the state of the art in deep learning and some developing areas of research. The authors are Ian Goodfellow, along with his Ph.D. advisor Yoshua Bengio, and Aaron Courville. All three are widely published experts in the field of artificial intelligence (AI). In addition to being available in both hard cover and Kindle the authors also make the individual chapter PDFs available for free on the Internet.[1] The book is aimed at an academic research audience with prior knowledge of calculus, linear algebra, probability, and some programming capabilities. A non-mathematical reader will find this book difficult. A comprehensive, well cited coverage of the field makes this book a valuable reference for any researcher. The book provides a mathematical description of a comprehensive set of deep learning algorithms, but could benefit from more pseudocode examples. The authors provide an adequate explanation for the many mathematical formulas that are used to communicate the ideas expressed in this book. The lack of both exercises and examples in any of the major machine learning software packages makes this book difficult as a primary undergraduate textbook.

While a review of a book focused entirely on deep learning might not be the usual topic for *Genetic Programming and Evolvable Machines*, there are many areas of interest for the genetic programming (GP) and evolutionary algorithm research communities. The effect of deep learning upon the field of AI has been profound. Deep learning's application to diverse cases ranging from self-driving cars to the game of Go have been widely reported. This book provides a solid deep learning

---

[1] http://www.deeplearningbook.org/.

✉ Jeff Heaton
  jtheaton@wustl.edu

[1] School of Engineering and Applied Science, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130-4899, USA

foundation for any AI researcher. Neural networks are the primary algorithm of deep learning, Neural networks and evolutionary algorithms have seen a great deal of combined research. Evolutionary algorithms are often used to evolve the complex structure of neural networks, an example of this is Kenneth Stanley's Neuroevolution of Augmenting Topologies (NEAT). The foundation of deep learning implementation are software packages, such as TensorFlow, MXNet and Theano that implement highly efficient computation engines that can be executed over distributed grids of GPUs and CPUs.

Deep Learning comprises 20 chapters which are divided into three distinct parts: prerequisite knowledge, current mainstream deep learning, and emerging future areas of deep learning research. The first part, which spans the first five chapters, provides an overview of the prerequisite mathematical concepts that the rest of the book is built upon. Chapters 1–5 only present a mathematical overview, the reader is expected to have previously studied each of these topics. The review contained in these first five chapters covers areas of mathematics and computer science that are valuable to any machine learning researcher, including GP and evolutionary algorithms. Specific areas of coverage are machine learning basics, linear algebra, and numerical computation. Of particular interest to GP researchers is the section on numeric computation that describes the mathematical and computational underpinnings of the graph evaluation that is performed by packages such as Google TensorFlow and Apache MXNet. These computational engines are not specifically tied to deep learning and can speed the calculation of any mathematically intense application through distributed computing and running on parallel graphics hardware (known as GPGPU).

Chapters six through twelve comprise the second part of the book, which provides a comprehensive review of the current mainstream deep learning technologies. The second part begins with the classic feedforward neural network and introduces regularization. Regularization is essentially a means of controlling the complexity of neural networks as they are trained. These neural network regularization techniques often parallel GP techniques that simplify genetic programs to avoid overfitting. Convolution is demonstrated as an effective means of recognizing images. Usually neural network layers are feed forward, in the sense that they connect to later layers. Recurrent neural networks contain connections to previous layers and maintain a state that allows their application to time series problems. Recurrent neural networks, such as the Long Short-Term Memory (LSTM) are introduced for signal processing and other time series tasks. The second section ends with several examples of the application of deep neural networks.

The final part of the book explores newer and more speculative directions in which deep learning may be headed. For many subfields of machine learning research, feature representation is important. The third part of the book covers feature representation with chapters devoted to dimension reduction and representation learning. Partition functions can be used to segment and prioritize the search space. Partitioning and sampling of the search space is discussed through chapters on Monte Carlo methods, partition functions, and approximate inference. Many of the techniques described in the final part of the book are general machine

learning principles that are not directly tied to neural networks and can be used in evolutionary algorithms.

Deep Learning provides a solid comprehensive foundation to any researcher interested in the current and future directions of deep learning research. It covers aspects from network design, training, evaluation and tuning. In addition to a theoretical background, the authors present practical advice from their own research. The explanations make the mathematical presentation of deep learning approachable to researchers from other subfields of AI. The bibliography is extensive and provides a great starting point for additional information. Because of *Deep Learning*'s comprehensive, yet relatively approachable treatment of deep learning and related technologies, it is very good value, and I highly recommend it for any AI researcher interested in neural networks.