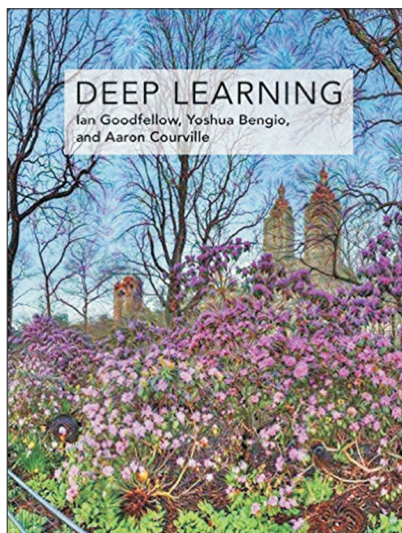# HIR
Healthcare Informatics Research

Deep Learning

Kwang Gi Kim, PhD

Biomedical Engineering Branch, Division of Precision Medicine and Cancer Informatics, National Cancer Center, Goyang, Korea
E-mail: kimkg@ncc.re.kr

Author: Ian Goodfellow, Yoshua Bengio, and Aaron
    Courville
Year: 2016
Name and location of publisher: The MIT Press,
    Cambridge, MA, USA
Number of pages: 800
Language: English
ISBN: 978-0262035613

Deep learning is a form of machine learning that enables computers to learn from experience and understand the world in terms of a hierarchy of concepts. Because the computer gathers knowledge from experience, there is no need for a human computer operator formally to specify all of the knowledge needed by the computer. The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones; a graph of these hierarchies would be many layers deep. This book introduces a broad range of topics in relation to deep learning.

The text offers a mathematical and conceptual background covering relevant concepts in linear algebra, probability theory and information theory, numerical computation, and machine learning. It describes deep learning techniques which are used by practitioners in industry, including deep feedforward networks, regularization, optimization algorithms, convolutional networks, sequence modeling, and practical methodology; and it surveys such applications as natural language processing, speech recognition, computer vision, online recommendation systems, bioinformatics, and videogames. Finally, the book offers research perspectives covering such theoretical topics as linear factor models, autoencoders, representation learning, structured probabilistic models, Monte Carlo methods, the partition function, approximate inference, and deep generative models.

Deep learning can be used by undergraduate or graduate students who are planning careers in either industry or research, and by software engineers who want to begin using deep learning in their products or platforms. A website offers supplementary material for both readers and instructors.

This book can be useful for a variety of readers, but the author wrote it with two main target audiences in mind. One of these target audiences is university students (undergraduate or graduate) who study machine learning, includ-

ing those who are beginning their careers in deep learning and artificial intelligence research. The other target audience consists of software engineers who may not have a background in machine learning or statistics but who nonetheless want to acquire this knowledge rapidly and begin using deep learning in their fields. Deep learning has already proven useful in many software disciplines, including computer vision, speech and audio processing, natural language processing, robotics, bioinformatics and chemistry, video games, search engines, online advertising and finance.

   This book has been organized into three parts so as to best accommodate a variety of readers. In Part I, the author introduces basic mathematical tools and machine learning concepts. Part II describes the most established deep learning algorithms that are essentially solved technologies. Part III describes more speculative ideas that are widely believed to be important for future research in deep learning.

   In this book, certain areas assume that all readers have a computer science background. The authors assume familiarity with programming and a basic understanding of computational performance issues, complexity theory, introductory-level calculus and some of the terminology of graph theory.

Chapter 1. Introduction
This book offers a solution to more intuitive problems in these areas. These solutions allow computers to learn from experience and understand the world in terms of a hierarchy of concepts, with each concept defined in terms of its relationship to simpler concepts. By gathering knowledge from experience, this approach avoids the need for human operators to specify formally all of the knowledge needed by the computer. The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones. If the authors draw a graph to show how these concepts have been built on top of each other, the graph will be deep, with many layers. For this reason, the authors call this approach "AI Deep Learning."

Chapter 2. Linear Algebra
Linear algebra is a branch of mathematics that is widely used throughout science and engineering. However, because linear algebra is a form of continuous rather than discrete mathematics, many computer scientists have little experience with it. This chapter will completely omit many important linear algebra topics that are not essential for understanding deep learning.

Chapter 3. Probability and Information Theory
This chapter describes probability and information theory. Probability theory is a mathematical framework for representing uncertain statements. It provides a means of quantifying uncertainties and axioms to derive new uncertainty statements. In addition, probability theory is a fundamental tool of many disciplines of science and engineering. The authors mention this chapter to ensure that readers whose fields are primarily in software engineering with limited exposure to probability theory can understand the material in this book.

Chapter 4. Numerical Computation
This chapter includes a brief overview of numerical optimization in general. Machine learning algorithms usually require a large amount of numerical computation. This typically refers to algorithms that solve mathematical problems by methods that update estimates of the solution via an iterative process rather than analytically deriving a formula and thus providing a symbolic expression for the correct solution.

Chapter 5. Machine Learning Basics
This chapter introduces the basic concepts of generalization, underfitting, overfitting, bias, variance and regularization. Deep learning is a specific type of machine learning. In order to understand deep learning well, one must have a solid understanding of the basic principles of machine learning. This chapter provides a brief course in the most important general principles, which will be applied throughout the rest of the book. Novice readers or those who want to gain a broad perspective are recommended to consider machine learning textbooks with more comprehensive coverage of the fundamentals, such as those by Murphy [1] or Bishop [2].

Chapter 6. Deep Feedforward Networks
Deep feedforward networks, also often called neural networks or multilayer perceptrons (MLPs), are the quintessential deep-learning models. Feedforward networks are of extreme importance to machine learning practitioners. They form the basis of many important commercial applications. For example, the convolutional networks used for object recognition from photos are a specialized type of feedforward network. Feedforward networks are a conceptual stepping stone on the path to recurrent networks, which power many natural-language applications.

Chapter 7. Regularization for Deep Learning
In this chapter, the authors describe regularization in more

detail, focusing on regularization strategies for deep models or models that may be used as building blocks to form deep models. Some sections of this chapter deal with standard concepts in machine learning. If readers are already familiar with these concepts, they may want to skip the relevant sections. However, most of this chapter is concerned with extensions of these basic concepts to the particular case of neural networks.

Chapter 8. Optimization for Training Deep Models

This chapter focuses on one particular case of optimization: finding the parameters θ of a neural network that significantly reduce a cost function $J(\theta)$, which typically includes a performance measure evaluated on the entire training set as well as additional regularization terms.

Chapter 9. Convolutional Networks

Convolutional networks [3], also known as neural networks or CNNs, are a specialized type of neural network for processing data that has a known, grid-like topology. In this chapter, the authors initially describe what convolution is. Next, they explain the motivation behind the use of convolution in a neural network, after which they describe an operation, called pooling, employed by nearly all convolutional networks.

Chapter 10. Sequence Modeling: Recurrent and Recursive Nets

Recurrent neural networks or RNNs are a family of neural networks for the processing of sequential data [4].

This chapter extends the idea of a computational graph to include cycles. These cycles represent the influence of the present value of a variable on its own value at a future time step. Such computational graphs allow one to define recurrent neural networks. Also in this chapter, the authors describe several different ways to construct, train, and use recurrent neural networks.

Chapter 11. Practical Methodology

Successfully applying deep learning techniques requires more than merely knowing what algorithms exist and understanding the principles by which they work.

Correct application of an algorithm depends on mastering some fairly simple methodology. Many of the recommendations in this chapter are adapted from Ng [5].

Chapter 12. Applications

In this chapter, the authors describe how to use deep learn-

ing to solve applications in computer vision, speech recognition, natural language processing, and other application areas of commercial interest. The authors begin by discussing the large-scale neural network implementations that are required for most serious AI applications.

Chapter 13. Linear Factor Models

In this chapter, the authors describe some of the simplest probabilistic models with latent variables, i.e., linear factor models. These models are occasionally used as building blocks of mixture models [6-8] or larger, deep probabilistic models. Additionally, they show many of the basic approaches that are necessary to build generative models, which are more advanced deep models.

Chapter 14. Autoencoders

An autoencoder is a neural network that is trained to attempt to copy its input to its output. Internally, it has a hidden layer 'h' that describes the code used to represent the input.

Chapter 15. Representation Learning

This chapter initially discusses what it means to learn representations and how the notion of representation can be useful to design deep architectures. Secondly, it discusses how learning algorithms share statistical strength across different tasks, including the use of information from unsupervised tasks to perform supervised tasks.

Chapter 16. Structured Probabilistic Models for Deep Learning

Deep learning draws upon many modeling formalisms that researchers can use to guide their design efforts and describe their algorithms. One of these formalisms is the idea of structured probabilistic models.

Chapter 17. Monte Carlo Methods

Randomized algorithms fall into two rough categories: Las Vegas algorithms and Monte Carlo algorithms. Las Vegas algorithms always return precisely the correct answer (or report their failure). These algorithms consume a random amount of resources, usually in the form of memory or time. In contrast, Monte Carlo algorithms return answers with a random amount of error.

Chapter 18. Confronting the Partition Function

In this chapter, the authors describe techniques used for training and evaluating models that have intractable parti-

tion functions.

Chapter 19. Approximate Inference

This chapter introduces several of the techniques used to confront these intractable inference problems.

Chapter 20. Deep Generative Models

This describes how to use these techniques to train probabilistic models that would otherwise be intractable, such as deep belief networks and deep Boltzmann machines.

This book provides the reader with a good overview of deep learning, and in the future this knowledge can serve as good material when researching content related to the study of artificial intelligence.

In particular, in medical imaging data, there are a number of images which require some preparation processes. These images can create a higher detection rate with artificial intelligence of tumors and diseases to help medical staff. At present, much effort is required to create the proper data values. In the future, it will be possible to generate useful images automatically, with more input image data then utilized. This book describes a wide range of different methods that make use of deep learning for object or landmark detection tasks in 2D and 3D medical imaging; it also examines a varied selection of techniques for semantic segmentation or detection using deep learning principles in medical imaging.

References

1. Murphy KP. Machine learning: a probabilistic perspective. Cambridge (MA): MIT Press; 2012.
2. Bishop CM. Pattern recognition and machine learning. New York (NY): Springer; 2006. p. 98-108.
3. Le Cun Y, Jackel LD, Boser B, Denker JS, Graf HP, Guyon I, et al. Handwritten digit recognition: applications of neural network chips and automatic learning. IEEE Commun Mag 1989;27(11):41-6.
4. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323(6088):533-6.
5. Ng A. Advice for applying machine learning [Internet]. Stanford (CA): Stanford University; 2015 [cited at 2016 Oct 22]. Available from: http://cs229.stanford.edu/materials/ML-advice.pdf.
6. Hinton GE, Dayan P, Frey BJ, Neal RM. The "wake-sleep" algorithm for unsupervised neural networks. Science 1995;268(5214):1158-61.
7. Ghahramani Z, Hinton GE. The EM algorithm for mixtures of factor analyzers. Toronto, Canada: University of Toronto; 1996.
8. Roweis ST, Saul LK, Hinton GE. Global coordination of local linear models. Adv Neural Inf Process Syst 2002;2:889-96.