Deep Learning

A Practical Introduction

Manel Martínez-Ramón
Meenu Ajith
Aswathy Rajendra Kurup

WILEY

Deep Learning

Deep Learning

A Practical Introduction

Manel Martínez-Ramón
Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM, USA

Meenu Ajith
Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, and Emory University, Atlanta, GA, USA

Aswathy Rajendra Kurup
Machine learning Engineer, Intel Corporation, Hillsboro, OR, USA

WILEY

This edition first published 2024.

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data applied for:
Hardback ISBN: 9781119861867

Cover Design: Wiley
Cover Image: © Yuichiro Chino/Getty Images

Set in 9.5/12.5pt STIXTwoText by Straive, Chennai, India

To our families, who have been unwavering in their support and understanding throughout the long nights and weekends spent on this journey. Your love and encouragement have fueled our passion for the world of deep learning, and this book is dedicated to you with profound gratitude.

To all the dreamers, may this book inspire you to chase your passions and never give up. Keep reaching for the stars.

To our shared dreams, relentless passion, and enduring friendship – this book is a testament to our collective journey.

Contents

About the Authors

Dr. Manel Martínez-Ramón received his Telecommunication Engineering degree from Universitat Politècnica de Catalunya, Spain, in 1994 and his PhD in Telecommunication Engineering from Universidad Carlos III de Madrid, Spain, in 1999. He is currently a professor of Artificial Intelligence with the Department of Electrical and Computer Engineering of the University of New Mexico, NM, USA, where he holds the King Felipe VI Endowed Chair. His research interests are in the area of machine learning, where he has produced numerous contributions to kernel learning methods, Gaussian processes, and deep learning, with applications to electromagnetics and antenna array processing, smart grid, scientific particle accelerators, and others. As an instructor, he teaches graduate courses in statistical learning theory, Gaussian process learning, probabilistic machine learning, and deep learning both face-to-face and online.

Dr. Meenu Ajith earned her PhD in Electrical Engineering from the University of New Mexico, USA, in 2022. Presently, she serves as a postdoctoral research associate at the Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), a collaborative research institute supported by Georgia State, the Georgia Institute of Technology, and Emory University in Atlanta, GA, USA. Her research focuses on deep learning, image processing, time-series analysis, and neuroimaging. She obtained her MS in Electrical Engineering from the University of New Mexico in 2017 and her bachelor's degree in Electronics and Communication Engineering from Amrita School of Engineering in 2015. In her current role as a postdoctoral researcher, Dr. Ajith concentrates on implementing and applying various deep-learning models and neuroinformatic tools that leverage advanced brain imaging data. Her objective is to translate these approaches into biomarkers, addressing pertinent aspects of brain health and diseases.

Dr. Aswathy Rajendra Kurup earned her PhD in Electrical Engineering from the University of New Mexico USA in the year 2022, where her research focused on designing CNN-based deep-learning models for applications such as smart grids, medical diagnosis, and computer vision. She completed her MS in Electrical Engineering from the University of New Mexico, USA, in 2017. Currently, she serves as a Machine Learning Engineer at Intel Corporation, a leading force in the semiconductor chip manufacturing landscape.

In her current role, she applies her extensive knowledge to address real-world challenges, utilizing her expertise in handling diverse data types, including images, videos, and time-series data. As a part of the role, she applies data mining and statistical modeling techniques along with developing Machine learning/Deep learning solutions for enabling factory decision-making, improved equipment performance, and higher product yields.

Foreword

Deep Learning: A Practical Introduction, authored by Manel Martínez-Ramón, Meenu Ajith, and Aswathy Rajendra Kurup, stands as a pragmatic guide, which prepares the engaged student to digest and understand advanced deep learning concepts. Designed primarily as an educational resource for graduate-level courses in deep learning, this book is enriched with a valuable collection of exercises and practical Python tutorials, making it an ideal educational tool.

Deep learning, a cornerstone of modern artificial intelligence, has seen a meteoric rise in usage, powering the creation of text, images, and videos, from simple prompts, and enhancing our predictive capabilities in a diverse array of applications. This book offers a thorough exploration of deep learning fundamentals, an essential component for students in engineering or computer science.

The authors begin by tracing the intriguing history of deep learning, setting the stage for a deeper dive into the subject. They skillfully introduce various methods for training and optimizing algorithms, alongside an overview of essential programming tools and libraries which are prevalent today, including Python, NumPy, TensorFlow, and Pytorch.

The book then covers a broad range of fundamental models including recurrent neural networks, transformers, unsupervised learning, and deep Bayesian networks. Within each of these chapters, there is an accessible introduction and detailed explanation of each modeling framework, which allows the reader who is new to deep learning to gain a foothold in this extraordinarily important space, while also providing practical examples including code and data as well as references for further learning. Additionally, it offers references for extended learning, bridging the gap between fundamental concepts and recent advancements in the field.

The author's provides a clear and comprehensive introduction to deep learning, making it an essential addition to the field's literature. Whether you are an instructor designing a course or a student embarking on self-directed learning, this book is an invaluable resource for navigating the complexities and applications of deep learning.

In essence, *Deep Learning: A Practical Introduction* is not just a textbook; it is a gateway to understanding and applying one of the most influential technologies in the field of artificial intelligence today. It is a useful tool for (i) instructors who want to teach core deep learning topics to their students, (ii) researchers in a variety of fields, including my own field of neuroimaging, who want to develop domain-specific methods, and (iii) students who are interested in self-learning on this important topic.

Overall, I strongly endorse *Deep Learning: A Practical Introduction* as a valuable resource for both educators aiming to impart core deep learning concepts to their students and for learners pursuing self-study in this vital area. The book's blend of theoretical insights and practical applications, including code and data examples, makes it a standout choice for anyone looking to delve into the world of deep learning.

Vince Calhoun

Preface

The present book is intended to be a comprehensive introduction to deep learning that covers all major areas in this discipline. This document is designed to cover a full semester graduate class in deep learning, and it contains all the materials necessary to build the class. We structured our work in a classical way, starting from the fundamentals of neural networks, which are then used to describe the different elements of deep learning used in artificial intelligence, from the classic convolutional neural network and recurrent neural networks (RNNs) to the transformers, plus unsupervised learning structures and algorithms. In every chapter, we follow a schema where first the structures are described, and then the criteria and algorithms to optimize them are developed. In most cases, full mathematical developments are included in the description of the structure optimization.

Chapter 1 is a first contact with deep learning, where we introduce the most basic type of feedforward neural network (FFNN), which is called the multilayer perception (MLP). Here, we first introduce the low-level basic elements of most neural networks and then the structure and learning criteria.

Chapter 2 is complementary to Chapter 1, but its contents are valid for the rest of the book since it provides details about the practical training of deep learning structures, which we have omitted from the first chapter in order to make it more concise and compact.

These readers who do not have a knowledge of basic Python will benefit from using Chapter 3 in order to start experimenting with learning machines in this programming language. In this chapter, authors assume that the reader has reviewed Chapter 1, which implies that they have been introduced to the concepts of structure, criteria, and algorithms. If so, readers already had the opportunity to see some basic Python codes containing at least a class with methods and an instantiation of it to be used in the examples and exercises, without needing to understand their Python structure. In this chapter, we introduce the basic elements of Python to be used throughout the book, and we will revisit the code previously introduced in Chapter 3, among other examples.

The concepts and structure of convolutional neural structures are described in Chapter 4. It starts with the concept of convolution in two dimensions and its justification for its use in deep learning, after which the structure of a convolutional neural network is described. The training of such a structure is not commonly found in the literature, assuming that the students and practitioners understand and can apply the backpropagation to them. We offer in this chapter a full development of the backpropagation for convolutional neural networks and we summarize the algorithms, so the practitioner can program it. Still, most importantly, they will understand exactly how it works.

Chapter 5 covers the basics of the RNN. The chapter starts off with the architecture of the RNN and then explains how these networks are used for modeling sequential information. Further into the chapter, the training criterion is introduced, which describes the feed-forward training, loss functions, and backpropagation through time. Next, the different types of RNN and their application are discussed. The following section explains the shortcomings of RNNs and highlights the details on different types of gradient problems and the solutions to these problems. Then, the shortcomings of RNNs and highlights the details on different types of gradient problems and the solutions to these problems are explained. After that, the details on other RNN-derived structures which were introduced to mitigate the short-term memory problem associated with the traditional RNNs are discussed.

Chapter 6 provides a structured and comprehensive overview of the developments in attention-based networks. The first section summarizes the different types of attention mechanisms based on sequence, levels, positions, and representations. Finally, we review the network architectures that widely use attention and also discuss a few applications in which attention-based networks have shown a significant impact.

Chapter 7 gives a comprehensive outline of deep unsupervised learning. The overview gives an introduction to the two main categories of deep unsupervised learning such as probabilistic and nonprobabilistic models. The chapter is mainly devoted to the autoencoder, which is one of the widely used nonprobabilistic deep unsupervised learning methods. First, the basic elements, training criteria, and the extensions of autoencoders are explained. Following this, an overview of the deep belief networks (DBNs) is given and it constitutes the basic blocks (restricted Boltzmann machines), training using contrastive divergence, and the variations of DBN. Finally, we also provide different applications of unsupervised deep learning.

Chapter 8 briefly covers the generative adversarial networks (GANs). Primarily, it introduces the two elements of GANs namely discriminator and generator. After this, the complete architecture of the GAN is illustrated to have a higher level of understanding of the network. Next, the training criteria are outlined which describes the alternate training process between the discriminator and the generator. The loss functions that model the probability distribution of the data is also added in this section. Finally, popular models derived from GAN are presented, and the chapter is concluded by summarizing the advantages and trade-offs of GAN.

Chapter 9 covers the main topics of deep Bayesian networks. Here, the authors do not intend to be exhaustive by covering the state of the art of deep Bayesian networks, Instead, we propose a chapter that gives the reader a general view of the characteristics and different philosophies of Bayesian networks with respect to previously introduced structures and algorithms. After introducing the general concepts of deep Bayesian networks, including structures and criteria (thus following the same format used in the rest of the book) we explain the main optimization algorithms used in the current literature, with several examples.

June, 2024
Albuquerque, New Mexico

Manel Martínez-Ramón
Meenu Ajith
Aswathy Rajendra Kurup

Acknowledgment

Manel Martínez-Ramón has been partially supported by the King Felipe VI Endowed Chair of the University of New Mexico, NM, USA.

About the Companion Website

A repository in GitHub with the URL

https://github.com/DeepLearning-book

contains all the additional materials of this book. In particular, readers will find:

- The Python code (in Jupyter Notebook format) of all the examples provided throughout the book, so that the student or the practitioner can run them immediately.
- A complete set of slides written in LaTex that summarize all chapters, intended to help instructors in the development of their lectures. The source files are also available so that instructors can modify the material and adapt it to each particular course design. All materials are available in the repository.

1

The Multilayer Perceptron

1.1 Introduction

The concept of artificial intelligence (AI) is relatively simple to explain, and it can be enunciated as a possible answer to the question of how to make a machine that is able to perform a given task without being explicitly programmed for it, but instead, extracting the necessary information from a set of data. Let us say, for example, that a machine is needed to classify green and red apples. The machine is provided with a camera, and all the mechanisms necessary to place one apple at a time in front of it and then throw it in one of two buckets. A machine wired to do this will relay in binary operators as "IF," "THEN." If the color is red, throw it in bucket A, otherwise, in bucket "B."

The limitations of this method are obvious. If a pear is mistakenly introduced in the process, it will be classified as a green apple. Also, how can we use the same or similar structure for a different or more complex task? As in the previous machine, an AI approach uses features found in the data in order to take the decision, but the algorithm is not explicitly programmed. Instead, the machine has a specific parametric structure capable of learning from data. The learning process involves the optimization of a certain measurable criterion with respect to the parameters. The deep learning (DL) structures for artificial intelligence are able to learn complex tasks from the available data, but they also have capabilities such as learning how to extract the useful features for the task at hand, provide probabilistic outputs (i.e. "the probability of apple is 97%"), and many others. The basic element of such a structure in DL is the so-called artificial neuron, a simple concept that provides the power and nonlinear properties.

This chapter is intended to be a first contact with DL, where we introduce the most basic type of feedforward neural network (FFNN), which is called the multilayer perceptron (MLP). Here, we first introduce the low-level basic elements of most neural network (NN)s, then the structure and learning criteria.

The elements introduced in this chapter will be used throughout the book. We start from the single perceptron, we construct a basic MLP, where the different activations are developed, and then the notation based on tensors is also justified as a generalized tool to be used throughout the book. After this, we present the maximum likelihood (ML) criterion as a general criterion, which is then particularized to the classic cases corresponding to the

Deep Learning: A Practical Introduction, First Edition.
Manel Martínez-Ramón, Meenu Ajith, and Aswathy Rajendra Kurup.

Companion website: https://github.com/DeepLearning-book

different output activations. Finally, the backpropagation (BP) is detailed and then summarized so that can be translated into a computer program.

In this chapter, examples and exercises are presented in a way that assumes that the student does not necessarily know about programming in Python. Examples will be focused on the behavior of the MLP, without focusing on the programming, and the exercises intended to modify, at a high-level data, parameters, and structures in order to answer questions to different practical cases. Chapter 3 explains, in particular, how the different examples have been coded, thus they will be reviewed in that chapter from the point of view of practical programming.

1.2 The Concept of Neuron

The idea of the artificial neural network (ANN) is obviously inspired by the structure of the nervous system. The first attempt to understand how neural tissue works from a logical perspective was published in 1943 by Warren S. McCulloch and Walter Pitts (1943) (Fig. 1.1). They proposed the first mathematical model for a biological neuron in his paper. In this model, the neuron has two possible states, defined as 0 or 1 depending on whether the neuron is resting or it has been activated or *fired*. This represents the *axon* of the neuron. The input of this neuron model consists of a number of *dendrites* whose excitation is also binary. This elemental structure is completed with an inhibitory input. If this input

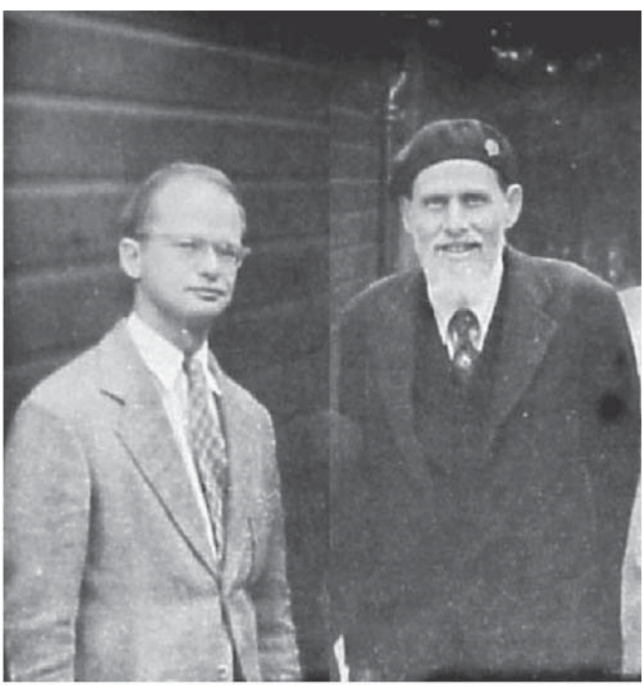

Figure 1.1 Warren S. McCulloch (left) and Walter Pitts in 1949. Source: R. Moreno-Díaz and A. Moreno-Díaz (2007)/with permission from Elsevier.

is activated, the neuron cannot fire. If the inhibitory input is deactivated, the neuron can be activated if the combination of inputs is larger than a given threshold. This model is fully binary and, since it includes mathematical functions that cannot be differentiated, it cannot be treated mathematically in an easy way. Certain modifications that will be seen further give rise to what is known as the artificial neuron in use today.

Section 1.2.1 contains an introduction to the concept of artificial perceptron from an algebraic point of view. A possible way to train a single perceptron is introduced in Sections 1.2.2 and 1.2.3, as well as the limitations of this structure as a linear classifier.

The concept of artificial NN was introduced by the psychologist Frank Rosenblatt (Fig. 1.2) in 1958 Rosenblatt (1957, 1958). In this paper, he proposed a structure of the visual cortex *perceptron* (Fig. 1.3). The structure presented in Rosenblatt (1958) contained the fundamental idea that is used in any artificial learning structure. In the first stage

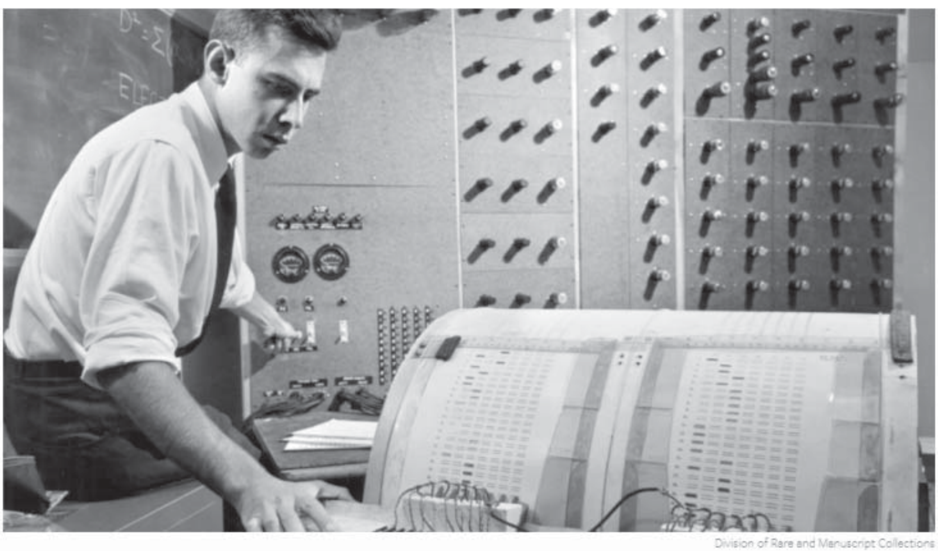

Figure 1.2 Frank Rosenblatt. Source: https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon/ last accessed November 30, 2023.

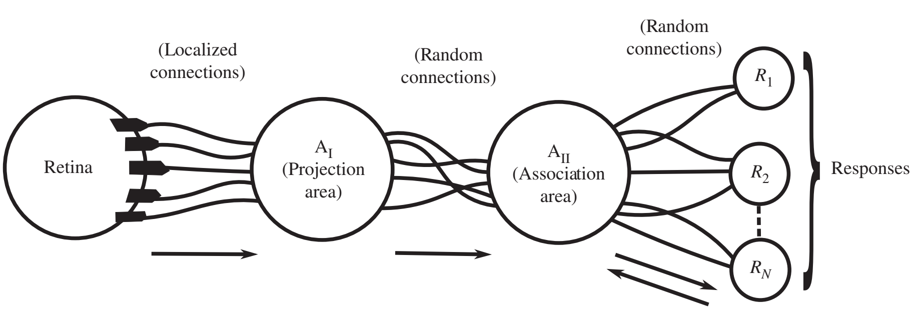

Figure 1.3 The perceptron as described in Rosenblatt (1958)/American Psychological Association.

(Retina), the device collects the available observation or input pattern intended to be processed in order to extract knowledge of it. The second stage (Projection area) is in charge of processing this observation to extract the information needed for the task at hand. This information is commonly called the set of features of the input pattern. The third stage (Association area) is intended to process these features to map them into a given response. For example, the response may be to recognize some given object classes present in the scene. Rosenblatt is the father of the artificial perceptron. He proved that by modifying the McCulloch–Pitts neuron model, the neuron could actually learn tasks from the data. In particular, his model had weights that multiplied each of the inputs to the neuron as well as the input bias or threshold that could be adjusted for the neuron to perform a given task. He developed the Mark 1 perceptron machine, which was the first implementation of his perceptron algorithm. This device was not a computer but an electromechanical *learning machine*. The machine consisted of a camera constructed with an array of 400 photocells, the output of each one connected randomly to the dendrites of a set of neurons. The weights, or attenuations applied to these inputs, were controlled with potentiometers whose axes were connected to electric motors. During the learning procedure, the motors adjusted the input weights. This machine was able to distinguish *linearly separable* patterns, or patterns that were at one or another side of a hyperplane in the space of 400 dimensions spanned by the camera inputs depending on its binary class. The invention was then limited in its capabilities until it was proven that a perceptron constructed with more than one layer of neurons MLP had nonlinear capabilities, that is, the ability to separate patterns that could not be separated by a hyperplane. Nevertheless, the MLP could not be trained using the techniques introduced by Rosenblatt for his perceptron. It was in 1971 that Paul Werbos, in his PhD thesis (P. J. Werbos 1974) introduced the BP algorithm, which made it possible to adjust the weights of a multilayer perceptron.

1.2.1 The Perceptron

From a conceptual point of view, a perceptron is a function made to perform a binary classification. In order to describe this function, let us first introduce the necessary notation and concepts associated with it. Assume a given observation that consists of a collection of D magnitudes observed from a physical phenomenon. These magnitudes are stored in a column vector, which will be called $\mathbf{x} \in \mathbb{R}^D$, which lies in a space of D dimensions. For illustrative purposes, let us construct a set of artificial data in a space of $D = 2$ dimensions as in Fig. 1.4.

The figure shows a set of points with coordinates $\mathbf{x} = (x_1, x_2)^\top$, where operator $\top$ denotes the transpose operation, meaning that the vector is a column one even if it is written as a row vector. In this toy example, the data belongs to one of two classes (black or white) that we will label arbitrarily with the labels $1, -1$, though in some cases, labels $1, 0$ are more convenient. It can be seen that the data is linearly separable, that is, both classes can be separated by placing a line between the black and white clusters of data. That is, roughly speaking, the idea of the perceptron. It must be trained to place a *separating hyperplane* between both classes. We define the hyperplane (particularized to a line in the two-dimensional example) as

$$\mathbf{w}^\top \mathbf{x} + b = 0 \tag{1.1}$$