LaTeX Assignment - Book

Ritesh Dobhal

November 15, 2025

Contents

Preface

This book is a project for the **SEC course LaTeX TypeSetting For Beginners** designed to meet the requirements of the **final LaTeX assignment**. The primary goal is *not* to create new content, but to demonstrate **mastery over the LaTeX typesetting system**, specifically its ability to manage *large, multi-file documents* using the `book` class. The structure of this document directly mirrors the requirements of the assignment: to build a book complete with a title page, *front matter* (including this preface and acknowledgements), a *main body* containing at least three chapters, and *back matter* featuring a bibliography.

The content itself is sourced from two places. The foundational *"Notation"* chapter is a direct transcription from page 17 of the provided reference text, *Deep Learning* by Goodfellow, Bengio, and Courville. This exercise demonstrates the handling of **complex mathematical typesetting**.

The three main chapters—*"Introduction," "Linear Algebra,"* and *"Probability"*—are original abstracts written in the style of the reference book. Each abstract is over 15 lines, fulfilling the assignment's requirement to *"write about a book you like,"* using the provided *Deep Learning* text as the subject. This project showcases the use of `\include` and `\input` to manage files, automatic generation of a Table of Contents with `\tableofcontents`, and the inclusion of unnumbered chapters into the ToC using `\addcontentsline`.

Acknowledgements

In the spirit of the reference text, I would like to acknowledge the people who made this project possible. First, I must thank **our instructor for the SEC LaTeX Course** for providing the clear requirements and the reference materials necessary to complete this assignment. This project has been an *excellent practical exercise* in document preparation. I would also like to acknowledge the authors of the original *Deep Learning* book—**Ian Goodfellow, Yoshua Bengio, and Aaron Courville**. Their comprehensive text served as the *perfect model* for this typesetting project.

Finally, I am grateful for the **developers of the LaTeX system**, the **TeX Users Group (TUG)**, and the **countless individuals** who have created and maintained the packages used in this document, such as `amsmath`, `hyperref`, and `tocbibind`. Without this **powerful, free, and open-source ecosystem**, a project of this quality would not be possible.

Notation

This section provides a concise reference describing the notation used throughout this book.

Calculus

$\frac{dy}{dx}$ Derivative of y with respect to x

$\frac{\partial y}{\partial x}$ Partial derivative of y with respect to x

$\nabla_x y$ Gradient of y with respect to x

$\nabla_X y$ Matrix derivatives of y with respect to X

$\nabla_{\mathbf{X}} y$ Tensor containing derivatives of y with respect to **X**

$\frac{\partial f}{\partial x}$ Jacobian matrix $J \in \mathbb{R}^{m \times n}$ of $f : \mathbb{R}^n \to \mathbb{R}^m$

$H(f)(x)$ The Hessian matrix of f at input point x (also $\nabla_x^2 f(x)$)

$\int f(x)dx$ Definite integral over the entire domain of x

$\int_S f(x)dx$ Definite integral with respect to x over the set S

Probability and Information Theory

$a \perp b$ The random variables a and b are independent

$a \perp b | c$ They are conditionally independent given c

$P(a)$ A probability distribution over a discrete variable

$p(a)$ A probability distribution over a continuous variable, or over a variable whose type has not been specified

$a \sim P$ Random variable a has distribution P

Expectation of $f(x)$ with respect to $P(x)$

$Var(f(x))$ Variance of $f(x)$ under $P(x)$

$Cov(f(x), g(x))$ Covariance of $f(x)$ and $g(x)$ under $P(x)$

$H(x)$ Shannon entropy of the random variable x

$D_{KL}(P||Q)$ Kullback-Leibler divergence of P and Q

$\mathcal{N}(x; \mu, \Sigma)$ Gaussian distribution over x with mean μ and covariance Σ

Functions

$f : \mathbb{A} \rightarrow \mathbb{B}$ The function f with domain $\mathbb{A}$ and range $\mathbb{B}$

$f \circ g$ Composition of the functions f and g

$f(x; \theta)$ A function of x parametrized by θ. (Sometimes we write $f(x)$ and omit the argument to lighten notation)

$\log x$ Natural logarithm of x

$\sigma(x)$ Logistic sigmoid, $\frac{1}{1+exp(-x)}$

$\zeta(x)$ Softplus, $log(1 + exp(x))$

$||x||_p$ L^p norm of x

$||x||$ L^2 norm of x

x^+ Positive part of x, i.e., $max(0, x)$

$\mathbf{1_{condition}}$ 1 if the condition is true, 0 otherwise

Chapter 1

Introduction

Abstract

This introductory chapter serves as the primary gateway into the world of deep learning, arguably the most capable and fastest-growing approach to modern artificial intelligence. We begin by defining what artificial intelligence (AI) actually is, placing machine learning as a subfield of AI that enables systems to learn from data without being explicitly programmed. Deep learning is then presented as a specific, more powerful type of machine learning that uses "deep" neural networks. These networks are inspired by the human brain's structure. The chapter provides a high-level overview of the book's three main parts: foundations, modern practical models, and research. We also explore the historical trends that have led to the current "AI boom," citing the availability of "big data" and the massive computational power of modern GPUs. This historical context is crucial for understanding why deep learning is succeeding now, when similar ideas in the 1980s and 1990s failed to gain traction. This book is intended for both students and software engineers. This chapter sets the stage, providing the necessary motivation and context to understand the deep concepts that follow, from basic linear algebra to advanced generative models.

Chapter 2

Linear Algebra

Abstract

Linear algebra is a crucial branch of mathematics for understanding and working with machine learning algorithms, especially deep learning. This chapter provides the essential mathematical foundation for the rest of the book. We will see that the core data objects we manipulate are not single numbers, but tensors (multidimensional arrays). These objects include scalars (0D tensors), vectors (1D tensors), and matrices (2D tensors). A batch of images, for example, can be represented as a 4D tensor. The fundamental operations within a neural network, such as the transformation in a fully-connected layer, are expressed using matrix multiplication. The parameters, or 'weights', of a network are themselves stored as matrices. Understanding how these objects interact is therefore essential. We will review key operations like the transpose and the dot product. Furthermore, we will cover foundational concepts like eigendecomposition and Singular Value Decomposition (SVD). These concepts are not just theoretical; they are the basis for powerful algorithms like Principal Components Analysis (PCA). PCA itself is. a simple example of a representation learning algorithm, a core theme of this book. This chapter builds the vocabulary we need to describe deep models. Without a solid grasp of these linear algebra concepts, it is nearly impossible to understand how deep learning models are structured, how they process data, or how they are trained.

Chapter 3

Probability and Information Theory

Abstract

Probability theory is a fundamental tool for machine learning, as it allows us to quantify and manage uncertainty. This chapter introduces the core probabilistic concepts needed to understand deep learning models. We begin by defining random variables and probability distributions, for both discrete and continuous cases. Understanding these concepts is vital, as many models are designed to output a probability distribution. We also cover key ideas like marginal and conditional probability, the chain rule of probability, and Bayes' rule, which is the foundation for many generative models. The chapter then introduces information theory, a branch of mathematics built on probability. We will define self-information, Shannon entropy, and the Kullback-Leibler (KL) divergence. These concepts are not just theoretical; they are used to define the ”loss functions” that train neural networks. For example, the cross-entropy function, which is widely used in classification tasks, is derived directly from these principles. This chapter provides the language for reasoning about uncertainty and information, which is essential for all modern AI.

Bibliography

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[2] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998b). Gradient based learning applied to document recognition. *Proc. IEEE*.

[3] Krizhevsky, A., Sutsever, I., and Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. In *NIPS'2012*.

[4] Hinton, G. E., Osindero, S., and Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18, 1527–1554.