

1. INTRODUCTION

Deep Learning is the heart of Artificial Intelligence and will become a most important field in Data Science in the near future. For general introduction to Deep Learning see for example [27]. The monographs [1], [2], [12], [18], [20], [21], are standard textbooks.

Deep Learning is one of Machine Learning methods and it has attracted much attention recently. Learning in general is divided into the supervised learning and unsupervised learning. In this book we discuss only supervised learning.

Deep Learning is usually carried out by the gradient descent method, which is not always easy to understand for beginners (or non-experts). When one starts studying Deep Learning first hurdles are

- 1) how to choose the learning rate ϵ
- 2) how to avoid being trapped by local minima
- 3) what is a deep meaning of the minibatch

In this book I plan to offer intuitive answers to these questions within my understandings.

As a matter of course, when beginners study Deep Learning some mathematical knowledge from Calculus, Linear Algebra, Statistics and Information are required. For example:

- a) Calculus \Rightarrow Gradient Descent Method
- b) Linear Algebra \Rightarrow Linear Independence
- c) Statistics \Rightarrow Least Squares Method

d) Information Retrieval \Rightarrow Vector Space Model

In this book I give minimum knowledge required for understanding Deep Learning.

Let me explain the contents (main results) of this book briefly.

The least squares method in Statistics plays an important role in almost all disciplines, from Natural Science to Social Science. When we want to find properties, tendencies or correlations hidden in huge and complicated data we usually employ the method. See for example [30].

Deep Learning may be stated as a successive learning method based on the least squares method. Therefore, to reconsider it from the view point of Deep Learning is very natural. The calculation of the successive approximation called gradient descent sequence will be carried out thoroughly.

Then it is emphasized that choosing the learning rate is important. When the learning rate changes a difference in methods between Statistics and Deep Learning gives different results.

Theorem I and II in Section 6.2 are the first group of the main results and a related problem is presented for readers. These results may give a new insight to both Statistics and Deep Learning.

Deep Learning is based on big data, so the supervised learning will give a heavy load to computers. In order to alleviate the burden a practical method called the minibatch (a collection of randomly selected small data from big data, see Section 7.3) is usually applied. Although the method is commonly used it is not widely understood why it is so effective from the mathematical point of view. In Theorem III in Section 7.1 and IV in Section 7.3 it is pointed out that a certain condition, which is practically satisfied in ordinary applications, is essential for the effectiveness of the minibatch method. Theorem III and IV are the second group of the main results and some related problems are presented.

This book is based on three papers on Deep Learning and one paper on Information Retrieval:

- 1) K. Fujii: Introduction to Artificial Intelligence (in Japanese), Lecture Note at

Yokohama City University, 2014.

2) K. Fujii: Mathematical reinforcement to the minibatch of Deep Learning, Advances in Pure mathematics, 8, 307-320, 2018.

3) K. Fujii: Least squares method from the view point of Deep Learning, Advances in Pure mathematics, 485-493, 2018.

4) K. Fujii and H. Oike: How to understand Google from the mathematical point of view, preprint, Yokohama City University, 2014.

The target of this book is undergraduates in the world who like Mathematics, Computer and Artificial Intelligence. Some elementary knowledge of Calculus, Linear Algebra, Statistic and Information is assumed. A lot of figures are provided to make the contents of the book easier to understand. Many exercises are included for self study. I believe that the book will be enjoyable.

After reading the book, readers are encouraged to challenge advanced books of Deep Learning (or Artificial Intelligence).

Acknowledgments

I wish to thank **Hiroshi Oike** and **Ryu Sasaki** and **Tatsuo Suzuki** for useful suggestions and comments.