



FOUNDATIONS AND FRONTIERS OF OPTIMIZATION IN MACHINE LEARNING

Dr. Sudipta Majumder

FOUNDATIONS AND FRONTIERS OF OPTIMIZATION IN MACHINE LEARNING

Dr. Sudipta Majumder

Senior Assistant Professor,
Department of Computer Science and Engineering,
Dibrugarh University Institute of Engineering and Technology
(DUIET), Dibrugarh University.

Kripa-Drishti Publications, Pune.

Book Title: **Foundations and Frontiers of Optimization in Machine Learning**

Author By: **Dr. Sudipta Majumder**

Price: ₹500

1st Edition

ISBN: **978-93-48091-25-3**



Published: **June 2024**

Publisher:



Kripa-Drishti Publications

A/ 503, Poorva Height, SNO 148/1A/1/1A,
Sus Road, Pashan- 411021, Pune, Maharashtra, India.

Mob: +91-8007068686

Email: editor@kdpublishations.in

Web: <https://www.kdpublishations.in>

© Copyright Dr. Sudipta Majumder

All Rights Reserved. No part of this publication can be stored in any retrieval system or reproduced in any form or by any means without the prior written permission of the publisher. Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages. [The responsibility for the facts stated, conclusions reached, etc., is entirely that of the author. The publisher is not responsible for them, whatsoever.]

PREFACE

Optimization serves as the lifeblood of machine learning, underpinning the success of models across diverse applications. As data grows in complexity and scale, so too does the challenge of training models effectively and efficiently. This book, *Foundations and Frontiers of Optimization in Machine Learning*, is an endeavour to navigate the vast landscape of optimization techniques, providing readers with a comprehensive understanding of the methods that propel machine learning forward.

From gradient-based approaches to derivative-free methods, and from classical second-order techniques to hyperparameter optimization strategies, this text spans the breadth of foundational concepts and cutting-edge innovations. Each chapter is crafted to balance theoretical insights with practical applications, ensuring that readers not only grasp the mathematics but also understand how to implement these methods in real-world scenarios.

This book is intended for researchers, practitioners, and students in machine learning, data science, and related fields who seek to deepen their understanding of optimization. Whether you are optimizing a simple regression model or tackling the challenges of distributed large-scale learning, the principles and techniques detailed in these pages will serve as a valuable resource.

As machine learning continues to evolve, so too must the tools and methods we use to optimize its potential. It is our hope that this book inspires innovation, fosters a deeper appreciation for the elegance of optimization, and equips readers with the skills to solve the complex problems of tomorrow.

Abbreviations

Automated Machine Learning (AutoML)

Batch Gradient Descent (BGD)

Bayesian Optimization (BO)

Broyden–Fletcher–Goldfarb–Shanno (BFGS)

Convolutional Neural Networks (CNNs)

Crossover Rate (CR)

Cyclical Learning Rate (CLR)

Deep Neural Networks (DNNs)

Derivative-Free Optimization (DFO)

Differentiable Architecture Search (DARTS)

Differential Evolution (DE)

Expected Improvement (EI)

Factor (F)

Federated Learning (FL)

Gaussian Process (GP)

Genetic Algorithms (GA)

Gradient Descent (GD)

Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS)

Linear Programming (LP)

Mean Squared Error (MSE)

Mini-Batch Gradient Descent (MBGD)

Nesterov Accelerated Gradient (NAG)

Neural Architecture Search (NAS)

Neural Architecture Search (NAS)

Particle Swarm Optimization (PSO)

Probability of Improvement (PI)

Projected Gradient Descent (PGD)

Quadratic Programming (QP)

Recurrent Neural Networks (RNNs)

Reinforcement Learning (RL)

Sequential Quadratic Programming (SQP)

Simulated Annealing (SA)

Stochastic Gradient Descent (SGD)

Successive Halving (SH)

Support Vector Machines (SVMs)

Upper Confidence Bound (UCB)

INDEX

Chapter 1: Optimization Techniques	1
1.1 Introduction:.....	1
1.2 Why Do We Need Optimization Techniques?.....	1
1.3 Aims of Optimization Techniques:	1
1.4 Categories of Optimization Techniques in Machine Learning:	2
Chapter 2: Gradient-Based Optimization	6
2.1 Introduction:.....	6
2.2 Categories of Gradient-Based Optimization:	7
2.2.1 Basic Gradient Descent Methods:.....	7
2.2.2 Gradient Descent Variants (Adaptive Methods):.....	12
2.2.3 Advanced Gradient-Based Methods:	20
Chapter 3: Second-Order Optimization.....	27
3.1 Introduction:.....	27
3.2 Types of Second-Order Optimization Techniques:	28
3.2.1 Newton’s Method:.....	28
3.2.2 Broyden–Fletcher–Goldfarb–Shanno (BFGS) Algorithm:	32
3.2.3 Limited-memory BFGS Algorithm:.....	36
Chapter 4: Derivative-Free Optimization	41
4.1 Introduction:.....	41
4.2 Why is Derivative-Free Optimization Required?.....	41
4.3 Applications of Derivative-Free Optimization:.....	42
4.4 How Derivative-Free Optimization Works:.....	43
4.5 Advantages of Derivative-Free Optimization:	43
4.6 Challenges and Limitations of DFO	44
4.7 Types of Derivative-Free Optimization:.....	45
4.7.1 Genetic Algorithms (GA):.....	45
4.7.2 Simulated Annealing (SA):	48
4.7.3 Particle Swarm Optimization (PSO):.....	50
4.7.4 Bayesian Optimization:	54
Chapter 5: Constrained Optimization.....	57
5.1 Introduction:	57

5.2 Methods for Constrained Optimization:	57
5.3 Types of Constrained Optimization:	58
5.3.1 Projected Gradient Descent (PGD):	58
5.3.2 Penalty Methods in Optimization:	60
5.3.3 Barrier Methods in Optimization:	64
Chapter 6: Convex Optimization.....	67
6.1 Introduction:.....	67
6.2 Importance of Convex Optimization:	67
6.3 Types of convex optimization:.....	68
6.3.1 Linear Programming:	68
6.3.2 Quadratic Programming:	74
Chapter 7: Distributed and Large-Scale Optimization	80
7.1 Introduction:.....	80
7.2 Why Distributed and Large-Scale Optimization is Needed:	81
7.3 Applications of Distributed and Large-Scale Optimization:	82
7.4 Types of Distributed and Large-Scale Optimization:.....	83
7.4.1 Parallel SGD:	83
7.4.2 Federated Learning:	86
7.4.3 Detailed Comparison: Parallel SGD vs. Federated Learning	89
Chapter 8: Hyperparameter Optimization	93
8.1 Introduction:.....	93
8.2 Types of Hyperparameters:.....	93
8.3 Key Methods for Hyperparameter Optimization:	94
8.4 Challenges in Hyperparameter Optimization:.....	95
8.5 Applications of Hyperparameter Optimization:.....	95
8.6 Types of Hyperparameter Optimization:	96
8.6.1 Grid Search Method of Hyperparameter Optimization:.....	96
8.6.2 Random Search: A Method of Hyperparameter Optimization:	99
8.6.3 Bayesian Optimization:	102
8.6.4 Gradient-Based Optimization for Hyperparameter Optimization: ..	106
8.6.5 Evolutionary Algorithms for Hyperparameter Optimization:	109
8.6.6 Hyperband and Successive Halving for Hyperparameter Optimization:	
.....	112
8.6.7 Neural Architecture Search (NAS): Hyperparameter Optimization	116
Chapter 9: Advanced Techniques.....	119
9.1 Introduction:.....	119
9.2 Region Methods:	119
9.3 Differential Evolution: Advanced Optimization Technique:.....	122

9.4 Learning Rate Schedulers:	126
9.5 Exponential Decay:.....	128
9.6 One Cycle Learning Rate:.....	130

List of Tables

Table Name	Page No.
Table 2.1: Comparisons	25
Table 5.1: Comparison of Penalty Methods	62
Table 7.1: Summary Table	92
Table 8.1: Comparison of Successive Halving and Hyperband	115

ABOUT THE AUTHOR



Dr. Sudipta Majumder

Dr. Sudipta Majumder embarked on his academic journey by completing his Bachelor of Technology (B.Tech) in Computer Science and Engineering from the North Eastern Regional Institute of Science and Technology (NERIST) in 2009. Eager to expand his knowledge, he pursued a Master of Technology (M.Tech) in Information Technology, followed by a Doctorate (Ph.D.) in Network Security, both from the prestigious North Eastern Regional Institute of Science and Technology. His doctoral research focused on advancing the field of network security, a domain he is deeply passionate about.

Dr. Majumder's primary research interests lie in the areas of internet of things, machine learning, Peer-to-Peer Networks, Wireless Networks, and Network Security. He is particularly fascinated by the challenges and opportunities presented by modern communication systems and strives to contribute innovative solutions to these fields.

Currently, Dr. Majumder serves as a Senior Assistant Professor in the Department of Computer Science and Engineering at the Dibrugarh University Institute of Engineering and Technology (DUIET), Dibrugarh University. His role allows him to guide the next generation of engineers, foster research, and collaborate with industry experts to bridge the gap between academia and real-world applications.

In addition to his academic responsibilities, Dr. Majumder is a dedicated life member of the Computer Society of India (CSI), where he actively engages with fellow professionals to promote knowledge-sharing and advancements in the field of computing.



Price: ₹ 500

Kripa-Drishti Publications

A-503 Poorva Heights, Pashan-Sus Road, Near Sai Chowk,
Pune – 411021, Maharashtra, India.

Mob: +91 8007068686

Email: editor@kdpublications.in

Web: <https://www.kdpublications.in>

ISBN: 978-93-48091-25-3



9 789348 091253