

# Fundamental Mathematical Concepts for Machine Learning in Science

Umberto Michelucci

# Fundamental Mathematical Concepts for Machine Learning in Science

 Springer

Umberto Michelucci  
TOELT.AI  
Dübendorf, Switzerland

ISBN 978-3-031-56430-7      ISBN 978-3-031-56431-4 (eBook)  
<https://doi.org/10.1007/978-3-031-56431-4>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

*To Caterina and Francesca. To Francesca, my love, without whom nothing would be possible. To Caterina, my life, the person I am most proud of in my life.*

# Preface

In this book, I discuss all the mathematical and methodological aspects that are important when using machine learning techniques in natural sciences (but not only). This book was written with a clear purpose: to explain the complexities of **how to use** machine learning to students and professionals in the natural sciences who may not have a background in computer science. It is a bridge connecting two seemingly disparate worlds and offers a comprehensive guide to understanding and applying the important techniques when applying machine learning in the context of science. I will not discuss algorithms or neural networks, but I will explain all the methods that you need to know to use them properly (e.g. model validation, sampling theory, etc.). The book is structured to gradually build your understanding, starting from fundamental mathematical concepts and progressing to advanced machine learning methods. Each chapter is designed to be self-contained, allowing the reader to focus on specific topics of interest. The chapters on calculus, linear algebra, and statistics are particularly crucial as they lay the foundation for a comprehensive understanding of machine learning algorithms and approaches. Given the breadth and depth of many topics, adequately covering each one would fill many books. My aim has been to cover and explain the core concepts necessary for your journey. I highly recommend further exploring these topics through the additional readings and references that I provide throughout the text. As author, I have striven to present the material in an accessible, yet rigorous manner. The book relies heavily on mathematics as the language of machine learning, ensuring that concepts are conveyed with precision and clarity. Although the book does not delve into programming details, it points out its relevance in machine learning, especially considering Python. What sets this book apart is its focus on methods about **how** to use machine learning, and not on the algorithms themselves filling a gap in the existing machine learning literature. Whether you are a physicist, chemist, biologist, doctor, or involved in any scientific discipline, this book is your guide to harnessing the power of machine learning in your field (if you are a computer scientist, this book is for you too!). This book is a portal to the exciting world of machine learning, written to enrich your scientific research projects, and I hope spark a lifelong interest in this field. If you are an instructor, on the book webpage on Springerlink you will find an instructor guide that will help you to use this book for a course.

Dübendorf, Switzerland

*Umberto Michelucci*  
March, 2024

# Acknowledgements

This book would not have been possible without the help of many people. At the forefront is my wife and esteemed colleague, Prof. Dr. Francesca Venturini. Her role in this journey has been immeasurable. Francesca is a constant source of support (in life as in projects), always offering insightful critiques and invaluable suggestions that have significantly shaped the content and direction of this book. I extend my heartfelt gratitude to the numerous students who have patiently listened to my lectures. Their participation and feedback have been instrumental in identifying the most challenging concepts in the field, highlighting areas that are often under-represented in standard university curricula. Their curious minds and challenging questions have not only aided my understanding, but also enriched the content of this book, making it more accessible and relevant to learners at all levels.

A special acknowledgement goes to my editor, Paul Drougas. Paul's expertise, patience, and flexibility have been pivotal in bringing this book to fruition. His guidance through the more challenging phases of this project, combined with his unwavering support, has been invaluable. Paul's proficiency and dedication as an editor have greatly contributed to the quality and coherence of this manuscript.

I am also indebted to many others who have offered their insight, participated in thought-provoking discussions, and provided feedback in various forms. Among them, I am particularly grateful to Dr. Michela Sperti, Prof. Dr. Marco Deriu, and Prof. Dr. Els Ortibus. Each of them has contributed to the richness of the content presented in this book.

Although I have attempted to acknowledge as many contributors as possible, I apologise to those I may have inadvertently omitted. Your contributions have not gone unnoticed and I am deeply grateful for your input. Finally, I assume full responsibility for any errors that may remain in the text. The insights and support provided by all mentioned and those unmentioned have been invaluable, but the final product and any shortcomings it may possess are solely my responsibility. This book is not just a reflection of my efforts, but the result of the collective wisdom and support of many brilliant minds. To each and every one of you, I extend my deepest gratitude.

Dübendorf, February 2024

# Contents

<b>1</b>	<b>Introduction</b> .....	1
1.1	Introduction .....	1
1.2	Choice of Topics .....	2
1.3	Prerequisites .....	5
1.4	Book Structure .....	5
1.5	About This Book .....	6
1.6	Warnings, Info and Examples .....	6
1.7	★ Optional and Advanced Material .....	7
1.8	■ Further Exploration and Reading .....	7
1.9	References .....	7
1.10	Let us Start .....	8
	References .....	8
<b>2</b>	<b>Machine Learning: History and Terminology</b> .....	9
2.1	Brief History of Machine Learning .....	9
2.2	Machine Learning in Science .....	17
2.3	Types of Machine Learning .....	18
	References .....	19
<b>3</b>	<b>Calculus and Optimisation for Machine Learning</b> .....	21
3.1	Motivation .....	21
3.2	Concept of Limit .....	22
3.3	Derivative and its Properties .....	25
3.3.1	Derivatives' Properties .....	29
3.4	Partial Derivative .....	31
3.5	Gradient .....	31
3.6	Extrema of a Function .....	33
3.7	Optimisation for Machine Learning .....	37
3.8	Introduction to Optimisation for Neural Networks .....	39
3.8.1	First Definition of Learning .....	40
3.8.2	Constrained vs. Unconstrained Optimisation .....	42

- 3.9 Optimization Algorithms ..... 43
  - 3.9.1 Line Search and Trust Region Approaches ..... 44
  - 3.9.2 Steepest Descent ..... 44
  - 3.9.3 ■ Additional Directions for the Line Search Approach ..... 47
  - 3.9.4 The Gradient Descent Algorithm ..... 48
  - 3.9.5 Choosing the Right Learning Rate ..... 51
  - 3.9.6 Variations of Gradient Descent ..... 51
  - 3.9.7 How to Choose the Right Mini-batch Size ..... 55
  - 3.9.8 ★ Stochastic Gradient Descent and Fractals ..... 56
- 3.10 Conclusions ..... 61
- References ..... 62
  
- 4 Linear Algebra ..... 63**
  - 4.1 Motivation ..... 63
  - 4.2 Vectors ..... 65
    - 4.2.1 Geometrical Interpretation of Vectors ..... 66
    - 4.2.2 Norm of Vectors ..... 68
    - 4.2.3 Dot Product ..... 69
    - 4.2.4 ★ Cross Product ..... 71
  - 4.3 Matrices ..... 73
    - 4.3.1 Sum, Subtraction and Transpose ..... 74
    - 4.3.2 Multiplication of Matrices and Vectors ..... 75
    - 4.3.3 Inverse and Trace ..... 77
    - 4.3.4 ★ Determinant ..... 78
    - 4.3.5 ★ Matrix Calculus and Linear Regression ..... 82
  - 4.4 Relevance for Machine Learning ..... 85
  - 4.5 Eigenvectors and Eigenvalues ..... 87
  - 4.6 ★ Principal Component Analysis ..... 88
    - 4.6.1 Basis of a Vector Space ..... 89
    - 4.6.2 ★ Definition of a Vector Space ..... 90
    - 4.6.3 ★ Linear Transformations (maps) ..... 90
    - 4.6.4 PCA Formalisation ..... 93
    - 4.6.5 Covariance Matrix ..... 94
    - 4.6.6 Overview of Assumptions ..... 95
    - 4.6.7 PCA with Eigenvectors and Eigenvalues ..... 96
    - 4.6.8 One Implementation Limitation ..... 97
- References ..... 97
  
- 5 Statistics and Probability for Machine Learning ..... 99**
  - 5.1 Motivation ..... 99
  - 5.2 Random Experiments and Variables ..... 100
  - 5.3 Algebra of Sets ..... 102
  - 5.4 Probability ..... 103
    - 5.4.1 Relative Frequency Interpretation of Probability ..... 103
    - 5.4.2 Probability as a Set Function ..... 104



- 5.4.3 ★ Axiomatic Definition of Probability ..... 104
- 5.4.4 Properties of Probability Functions ..... 106
- 5.5 The Softmax Function ..... 107
  - 5.5.1 ■ Softmax Range of Applications ..... 108
- 5.6 ★ Some Theorems about Probability Functions ..... 109
- 5.7 Conditional Probability ..... 110
- 5.8 Bayes Theorem ..... 111
- 5.9 ★ Bayes Error ..... 111
- 5.10 ★ Naïve Bayes Classifier ..... 112
- 5.11 Distribution Functions ..... 114
  - 5.11.1 Cumulative Distribution Function (CDF) ..... 115
  - 5.11.2 Probability Density (PDF) and Mass Functions (PMF) ..... 116
- 5.12 Expected Values and its Properties ..... 118
- 5.13 Variance and its Properties ..... 118
  - 5.13.1 Properties ..... 119
- 5.14 Normal Distribution ..... 121
- 5.15 ■ Other Distributions ..... 125
- 5.16 The MSE and its Distribution ..... 128
  - 5.16.1 ★ Moment Generating Functions ..... 131
  - 5.16.2 ★ Central Limit Theorem ..... 132
- 5.17 Central Limit Theorem without Mathematics ..... 134
- References ..... 135
- 6 Sampling Theory (a.k.a. Creating a Dataset Properly) ..... 137**
  - 6.1 Introduction ..... 137
  - 6.2 Research Questions and Hypotheses ..... 138
    - 6.2.1 Research Questions ..... 139
    - 6.2.2 Hypothesis ..... 139
    - 6.2.3 Relevance of Hypothesis and Research Questions in  
Machine Learning ..... 141
  - 6.3 Survey Populations ..... 141
  - 6.4 Survey Samples ..... 143
    - 6.4.1 Non-probability Sampling ..... 143
    - 6.4.2 Probability Sampling ..... 145
  - 6.5 Stratification and Clustering ..... 146
  - 6.6 Random Sampling without Replacement ..... 147
  - 6.7 Random Sampling with Replacement ..... 147
  - 6.8 ■ Bootstrapping ..... 148
  - 6.9 Random Stratified Sampling ..... 149
  - 6.10 Sampling in Machine Learning ..... 150
  - References ..... 151

<b>7</b>	<b>Model Validation and Selection</b> .....	153
7.1	Introduction .....	153
7.2	Bias-Variance Tradeoff .....	154
7.3	★ Bias-Variance Tradeoff - a Mathematical Discussion .....	155
7.4	High-Variance Low-Bias regime .....	160
7.5	Low-Variance High-Bias regime .....	163
7.6	Overfitting and Underfitting .....	164
7.7	The Simple Split Approach (a.k.a. Hold-out Approach) .....	165
7.8	Data Leakage .....	166
7.8.1	Data Leakage with Correlated Observations .....	168
7.8.2	Stratified Sampling .....	169
7.9	Monte Carlo Cross-Validation .....	171
7.9.1	How to use the Monte Carlo Cross-Validation Approach .....	173
7.10	Monte-Carlo Cross Validation with Bootstrapping .....	175
7.11	$k$ -Fold Cross Validation .....	176
7.12	The Leave-One-Out Approach .....	178
7.13	Choosing the Cross-Validation Approach .....	178
7.14	Model Selection .....	180
7.14.1	Model Selection with Supervised Learning .....	181
7.14.2	■ Model Selection with Unsupervised Learning .....	182
7.15	■ Qualitative Criteria for Model Selection .....	183
	References .....	184
<b>8</b>	<b>Unbalanced Datasets and Machine Learning Metrics</b> .....	185
8.1	Introduction .....	185
8.2	A Simple Example .....	186
8.3	Approaches to Deal with Unbalanced Datasets .....	190
8.3.1	Oversampling .....	191
8.3.2	(Random) Undersampling .....	193
8.4	Synthetic Minority Oversampling TEchnique (SMOTE) .....	194
8.5	■ Summary of Methods for Dealing with Unbalanced Datasets .....	195
8.6	Important Metrics .....	196
8.6.1	★ The Notion of Metric .....	196
8.6.2	Confusion Matrix .....	200
8.6.3	Sensitivity and Specificity .....	202
8.6.4	Precision .....	203
8.6.5	$F_\beta$ -score .....	203
8.6.6	Balanced Accuracy .....	205
8.6.7	Receiving Operating Characteristic (ROC) Curve .....	205
8.6.8	★ Probability Interpretation of the AUC .....	208
	References .....	209

- 9 Hyper-parameter Tuning** ..... 213
  - 9.1 Introduction ..... 213
  - 9.2 Black-box Optimisation ..... 214
    - 9.2.1 Notes on Black-box Functions ..... 215
  - 9.3 The Problem of Hyper-parameter Tuning ..... 216
  - 9.4 Sample Black-box Problem ..... 217
    - 9.4.1 Grid Search ..... 217
    - 9.4.2 Random Search ..... 221
    - 9.4.3 Coarse to Fine Optimisation ..... 223
    - 9.4.4 Sampling on a Logarithmic Scale ..... 225
  - 9.5 ■ Overview of Approaches for Hyper-parameter Tuning ..... 225
  - References ..... 226
  
- 10 Feature Importance and Selection** ..... 229
  - 10.1 Introduction ..... 229
  - 10.2 Feature Importance Taxonomy ..... 230
    - 10.2.1 ■ Filter Methods ..... 231
    - 10.2.2 ■ Wrapper Methods ..... 232
    - 10.2.3 ■ Embedded Methods ..... 233
  - 10.3 Forward Feature Selection ..... 234
    - 10.3.1 Forward Feature Selection Practical Example ..... 236
  - 10.4 Backward Feature Elimination ..... 238
  - 10.5 Permutation Feature Importance ..... 240
  - 10.6 Information Content Elimination ..... 241
  - 10.7 Summary ..... 241
  - 10.8 ■ SHapley Additive exPlanations (SHAP) ..... 241
  - References ..... 242
  
- Index** ..... 243

# Acronyms

## Acronyms

AE	Autoencoder
AUC	Area Under the Curve
BFE	Backward Feature Elimination
CART	Classification and Regression Trees
CNN	Convolutional Neural Network
DL	Deep Learning
FFNN	Feed Forward Neural Network
FFS	Forward Feature Selection
GAN	Generative Adversarial Networks
GD	Gradient Descent
GBM	Gradient Boosting Machines
kNN	k-Nearest Neighbour
LLM	Large Language Model
MAE	Mean Absolute Error
ML	Machine Learning
MNIST	Modified National Institute of Standards and Technology
MSE	Mean Square Error
NLP	Natural Language Processing
PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
RFE	Recursive Feature elimination
RL	Reinforcement Learning
ROC	Receiving Operating Characteristic
t-SNE	t-Distributed Stochastic Neighbour Embeddings
SMOTE	Synthetic Minority Oversampling Techniques
SVM	Support Vector Machine
VAE	Variational Autoencoder