
Pattern Recognition and Machine Learning:

Concepts, Mathematics, and Intuitions

Pattern Recognition and Machine Learning:

Concepts, Mathematics, and Intuitions

Author

Partha Pratim Ray

Rp

RENU PUBLISHERS

New Delhi, Kolkata

This First Edition Published in 2026

© 2026 Renu Publishers, India

Title: Pattern Recognition and Machine Learning: Concepts, Mathematics and Intuitions

Author: Partha Pratim Ray

Description: First edition | Renu Publishers 2026 | Includes bibliographical references and index.

Identifiers: ISBN 9789392597985 (Print) | 9789392597312 (eBook)

Cover Design: New Delhi Publishers

All rights reserved. No part of this publication or the information contained herein may be reproduced, adapted, abridged, translated, stored in a retrieval system, computer system, photographic or other systems or transmitted in any form or by any means, electronic, mechanical, by photocopying, recording or otherwise, without written prior permission from the publisher.

Disclaimer: Whereas every effort has been made to avoid errors and omissions, this publication is being sold on the understanding that neither the editors (or authors) nor the publishers nor the printers would be liable in any manner to any person either for an error or for an omission in this publication, or for any action to be taken on the basis of this work. Any inadvertent discrepancy noted may be brought to the attention of the publisher, for rectifying it in future editions, if published.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.



RENU PUBLISHERS

Head Office: 90, Sainik Vihar, Mohan Garden, New Delhi, India

Corporate Office: 7/28, Room No. 208-209, Vardaan House, Mahavir Lane, Ansari Road, Daryaganj, New Delhi, India

Branch Office: 216, Flat-GC, Green Park, Narendrapur, Kolkata, India

Tel: 011-23256188, 011-45130562, 9971676330, 9582248909

Email: renupublishers@gmail.com

Website: www.renupublisher.in

*bhūh pādau yasya nābhir viyad asuranilaś candrasūryau ca netre,
karṇāv āśāḥ śiro dyauḥ mukham api dahano yasya vāso'yam abdhih;
antaḥsthaṃ yasya viśvaṃ sura-nara-khaga-go-bhogi-gandharva-daityaiḥ,
citraṃ ramyate taṃ tribhuvana-vapuṣaṃ viṣṇuṃ tīsaṃ namāmi.*

“I bow before that Lord Viṣṇu, the Sovereign of the three worlds —
whose feet are the Earth, whose navel is the Sky, whose breath is the Wind,
whose eyes are the Moon and the Sun, whose ears are the Directions,
whose head is Heaven, whose mouth is Fire, and whose garment is the Ocean.
Within Him resides the entire cosmos — gods, men, birds, beasts, serpents,
Gandharvas, and Daityas — all playing and rejoicing in His vast form.”

— Śrī Viṣṇu Sahasranāma Stotram 2

*oṃ śāntākāraṃ bhujagaśayanaṃ padmanābhaṃ sureśaṃ,
viśvādhāraṃ gaganasadr̥śaṃ meghavarṇaṃ śubhāṅgam;
lakṣmīkāntaṃ kamalanayanaṃ yogi-hṛd-dhyānagamyam,
vande viṣṇuṃ bhava-bhaya-haraṃ sarvalokaikanātham.*

“I bow before Lord Viṣṇu — embodiment of peace,
who reclines upon the serpent couch, whose navel bears a lotus,
who is the Lord of the gods and support of the universe,
who is vast as the sky, cloud-hued, and of auspicious form;
the beloved of Lakṣmī, lotus-eyed, perceived by sages in meditation,
the remover of worldly fear, and the sole sovereign of all the worlds.”

— Śrī Viṣṇu Sahasranāma Stotram 3

*meghaśyāmaṃ pīta-kaūśeya-vāsaṃ
śrīvatsāṅkaṃ kaustubhodbhāsītāṅgam;
punyopetaṃ puṇḍarikāyatākṣaṃ
viṣṇuṃ vande sarva-lokaika-nātham.*

“I bow before that Lord Viṣṇu, the sole ruler of all worlds —
cloud-hued, clad in yellow silk, marked with the Śrīvatsa,
radiant with the Kaustubha gem, lotus-eyed,
and ever surrounded by the blessed and the virtuous.”

— Śrī Viṣṇu Sahasranāma Stotram 4

*namaḥ samasta-bhūtānām ādibhūtāya bhūbhṛte,
anekarūpa-rūpāya viṣṇave prabhaviṣṇave.*

“Salutations to Lord Viṣṇu — the source of all beings,
the sustainer of the Earth, of manifold and infinite forms,
the primal origin and supreme cause of all creation.”

— Śrī Viṣṇu Sahasranāma Stotram 5

*viṣṇuṃ jiṣṇuṃ mahāviṣṇuṃ prabhaviṣṇuṃ maheśvaram,
anekarūpa-dāityāntaṃ namāmi puruṣottamam.*

“Om; Viṣṇu, the Conqueror, the Great Viṣṇu, the Mighty Creator, the Great Lord,
to Him of many forms, the Destroyer of demons, to the Supreme Person, I bow.”

— Śrī Viṣṇu Sahasranāma Stotram 22

Upaniṣadic Invocation

asato mā sad gamaya, tamaso mā jyotir gamaya, mṛtyor mā amṛtaṃ gamaya.

“Lead me from the unreal to the Real,
from darkness to Light,
from mortality to Immortality.”

— *Bṛhadāraṇyaka Upaniṣad 1.3.28*

Gītā Invocation

*karmaṇy evādhikāras te mā phaleṣu kadācana,
mā karma-phala-hetur bhūr mā te saṅgo 'stv akarmaṇi.*

“You have the right to action alone,
but never to its fruits.
Let not the fruits of action be your motive,
nor let your attachment be to inaction.”

— *Bhagavad Gītā 2.47*

*Dedicated to
Maa & Baba*

Acknowledgements

The completion of this book, entitled *Pattern Recognition and Machine Learning: Concepts, Mathematics, and Intuitions*, has been a journey shaped by the encouragement, guidance, and intellectual generosity of many individuals and institutions. To all those who have, in diverse ways, contributed to its realization, I express my sincere and heartfelt gratitude. The preparation of this manuscript required sustained reflection, study, and perseverance, and the support received from numerous sources made this endeavor both possible and meaningful.

First and foremost, I extend my deep appreciation to those who guided, inspired, and supported me throughout this intellectual undertaking. Their thoughtful discussions, constructive suggestions, and encouraging words helped refine both the conceptual structure and the pedagogical spirit of this work. Each exchange of ideas—whether in academic discussions or informal conversations—contributed to the evolution of the perspectives presented in this book.

I am profoundly grateful for the academic environment that nurtured this effort. It provided access to valuable scholarly resources as well as a vibrant community of teachers, colleagues, and students whose enthusiasm and curiosity continually stimulated deeper inquiry. Their intellectual engagement and encouragement helped transform complex theoretical ideas into explanations that strive for clarity, coherence, and accessibility.

A special acknowledgement is due to open educational resources that continue to democratize learning. In particular, the NPTEL lecture series “*Electronics – Pattern Recognition and Application*” by Prof. P. K. Biswas, Electronics and Electrical Communication Engineering, IIT Kharagpur, has been an enduring source of conceptual clarity and inspiration. The accessibility of such scholarly resources reaffirms the collaborative nature of knowledge creation, and I gratefully acknowledge its influence in shaping the conceptual direction and pedagogical flow of this book. The clarity and rigor reflected in these lectures greatly contributed to strengthening the conceptual foundations underlying several discussions presented here.

I also express my sincere appreciation to my friends and peers whose intellectual companionship and encouragement sustained my motivation during the extended writing process. Their thoughtful suggestions, critical insights, and words of support served as a constant reminder of the collective nature of scholarly pursuit.

Above all, I remain deeply indebted to my family. Their unconditional love, patience, and quiet sacrifices provided the stability and encouragement necessary to sustain this work. In particular, I owe heartfelt gratitude to my wife, whose understanding, encouragement, and quiet strength stood beside me throughout the long and demanding process of writing this book. Her unwavering belief in the value of this effort provided constant motivation during moments of fatigue and uncertainty. I am equally grateful to my daughter, whose innocence, curiosity, and joyful presence brought balance and warmth to the many hours devoted to this manuscript. Their unwavering support offered both emotional strength and the sense of purpose that accompanies any meaningful intellectual endeavor.

Finally, I bow in gratitude and devotion to the **Bhagawan Sri Vishnu** for the strength, clarity, and perseverance granted to complete this journey. Without divine grace and guidance, none of this would have been possible.

Partha Pratim Ray

Preface

“Understanding patterns is understanding intelligence itself.”

Pattern recognition (PR) stands at the intellectual heart of artificial intelligence—the art and science of discerning order amid apparent chaos, of extracting meaning from complexity, and of translating data into structured knowledge. It is where mathematics meets perception, where geometry meets logic, and where computation begins to emulate cognition. At its essence, pattern recognition is the discipline that explains how intelligent systems perceive structure in their environment—how they identify, categorize, and generalize from experience.

Every modern intelligent system—whether it detects tumors in medical imagery, recognizes speech, translates text, or anticipates user intent—rests upon the foundations of pattern recognition. It is the silent mathematical engine beneath machine learning, computer vision, natural language processing, and signal understanding. To study PR is therefore to explore the mathematical fabric of intelligence itself, to understand how machines, like humans, learn to perceive, reason, and decide.

Pattern recognition transforms unstructured observations into structured insight. It teaches machines not merely to process information but to *comprehend* it—to discover relationships, abstractions, and meaning within data. In synergy with *machine learning* (ML), PR forms a dual narrative: PR defines the representations and theoretical framework, while ML provides the adaptive machinery that brings those representations to life. Together, they describe how perception evolves into understanding, and how intelligence emerges through structure, adaptation, and inference.

This book—*Pattern Recognition and Machine Learning: Concepts, Mathematics, and Intuitions*—is an attempt to tell that story with depth, coherence, and clarity. It unites rigorous mathematics with intuitive explanation, theory with geometry, and abstract reasoning with illustrative toy examples that make every idea tangible and alive.

Motivation and Philosophy

The motivation for this book arises from a deep conviction that the elegance of pattern recognition is too often hidden behind technical formality. Students are shown equations before intuition, algorithms before motivation, and derivations before discovery. This text seeks to restore balance—to present the subject not as a set of disjointed formulas, but as a coherent intellectual journey. It strives to make mathematics feel like reasoning rather than memorization, and to rekindle the sense of wonder that accompanies true understanding.

The guiding philosophy of this work is simple yet profound: *teach concepts as they are discovered, not merely as they are defined*. Each chapter begins with a question, unfolds its intuition, develops the mathematics, and concludes with a tangible example. This rhythm—curiosity, intuition, formalization, application—turns learning into an active dialogue between imagination and logic.

Mathematics here is not an obstacle but a language—a beautifully precise way to describe how intelligence operates. Equations are treated as sentences of thought, derivations as arguments of reason. Programming examples are intentionally omitted so that readers may internalize ideas before translating them into code, preserving focus on understanding rather than implementation.

Why Study Pattern Recognition?

Human beings are natural pattern recognizers. From the first moments of life, we discern shapes, rhythms, and relationships; we generalize, infer, and anticipate. This innate faculty, when expressed mathematically

and computationally, becomes the cornerstone of intelligent systems. To study pattern recognition is to study how perception becomes computation, and how cognition can be formalized.

It invites questions such as: *How can raw sensory data be transformed into meaningful representations? How can we distinguish between objects that appear similar? How can decisions be made reliably despite uncertainty?* Each of these leads to the triad of probability, geometry, and optimization—the pillars of this discipline. Probability captures uncertainty; geometry structures the space of representations; optimization drives learning and adaptation.

Pattern recognition is thus not only a tool for data analysis—it is a philosophy of reasoning under uncertainty. It teaches abstraction, generalization, and disciplined intuition—showing how meaning can emerge from data and how learning itself can be engineered.

From Observation to Representation

Every act of recognition begins with representation. An image, sound, or text in its raw form is incomprehensible to a machine. The first act of intelligence is to describe—to extract the essential and ignore the irrelevant. This transformation from observation to representation is achieved through *feature extraction*, which compresses high-dimensional reality into structured descriptors.

Once represented numerically, data acquire geometric identity—each object becomes a point in a multidimensional space. Recognition then becomes a geometric enterprise: identifying clusters, boundaries, or relationships among points. The process is both mathematical and visual, deterministic and probabilistic. Through careful visualization, examples, and intuitive commentary, the book builds geometric understanding alongside analytical skill.

Recognition as Intelligent Decision Making

Recognition is fundamentally a decision-making process. Given uncertain or overlapping evidence, a system must infer the most plausible category or hypothesis. Here, Bayes decision theory provides the foundation: priors encode belief, likelihoods express evidence, and posterior probabilities reconcile the two. This probabilistic reasoning mirrors the human approach to perception and judgment.

Bayesian thinking unites all classifiers, regressors, and modern neural models under a single interpretive framework. This text traces that lineage—from early discriminant functions to contemporary deep learning—revealing how the essence of Bayesian logic endures across generations of algorithms.

Pattern Recognition and Machine Learning: Two Perspectives, One Vision

Pattern recognition and machine learning are complementary facets of a unified field. PR emphasizes *representation and reasoning*—the modeling of how data arise and how they should be classified. ML emphasizes *adaptation and experience*—learning from data directly without presupposed structure. PR asks, “*What defines similarity?*” ML answers, “*Let me learn it from data.*”

Their convergence marks the true evolution of artificial intelligence—from rule-based design to data-driven inference. This book illustrates that evolution, showing how the statistical rigor of PR naturally flows into the optimization frameworks of ML—culminating in a unified science of intelligent behavior.

Scope, Structure, and Pedagogical Approach

The book is structured as a gradual intellectual ascent—from foundational feature extraction and Bayes classification to discriminant analysis, neural computation, clustering, and temporal pattern modeling. It emphasizes clarity over brevity, intuition over abstraction. Each concept is introduced narratively, developed mathematically, visualized geometrically, and reinforced with a toy example.

Chapters progress logically—from linear separability to non-linear kernels, from perceptrons to deep networks, from static recognition to dynamic sequence modeling. Appendices provide self-contained mathematical supplements on optimization, loss functions, and evaluation metrics—allowing readers to explore without prerequisite dependencies.

Mathematics remains central, but always in service of comprehension. No equation appears without purpose, and no derivation stands without motivation. The goal is not rote mastery, but the cultivation of reasoning and insight.

Overview of Chapters

To guide the reader through the intellectual flow of the book, this section presents a concise overview of each chapter. The book comprises eleven principal chapters and four comprehensive appendices, each crafted to build conceptual, mathematical, and intuitive depth in a progressive manner. Every chapter begins with motivating questions, develops its theory with formal rigor, and concludes with illustrative toy examples and interpretive commentary. Together, they form a unified framework of pattern recognition and machine learning—from foundational ideas to advanced architectures.

- **Chapter 1 – Introduction to Pattern Recognition:** Lays the philosophical, mathematical, and historical foundations of the field. It discusses the nature of patterns, the distinction between supervised and unsupervised learning, and the concept of feature spaces. This chapter bridges classical philosophical thought—from Plato’s ideal forms to Aristotle’s inductive reasoning—with modern computational interpretations of learning from data.
- **Chapter 2 – Feature Extraction:** Explores how raw data are transformed into meaningful numerical representations. Topics include chain codes, boundary and region descriptors, spatial and frequency domain features, and polygonal approximations. It emphasizes the connection between perception and measurement, illustrating how features encode the essence of observed patterns.
- **Chapter 3 – Bayes Decision Theory:** Develops the probabilistic framework for decision-making under uncertainty. It introduces priors, likelihoods, posterior probabilities, and risk minimization, leading naturally to discriminant functions and optimal decision boundaries. The chapter demonstrates how Bayesian reasoning unifies inference and classification.
- **Chapter 4 – Probability Density Function Estimation:** Extends Bayes theory by examining how the underlying data distributions are modeled. Both parametric (e.g., Gaussian, Poisson) and non-parametric (e.g., histogram, Parzen window) estimation methods are presented. It also introduces dimensionality reduction techniques such as Principal Component Analysis (PCA) and Multiple Discriminant Analysis (MDA), preparing the ground for efficient pattern representation.
- **Chapter 5 – Discriminant Criteria Functions:** Unifies several classical classifiers under the framework of discriminant functions and optimization. It includes perceptrons, MSE and relaxation criteria, Widrow–Hoff learning, logistic regression, decision trees, and nearest-neighbour methods. This chapter establishes the conceptual transition from linear to nonlinear decision-making paradigms.

- **Chapter 6 – Neural Networks:** Presents the architecture, mathematics, and learning dynamics of artificial neural systems. Beginning with biological neurons and single-layer perceptrons, it extends to multilayer networks, backpropagation algorithms, Hopfield associative memories, and Radial Basis Function networks. The geometric and energy-based perspectives of learning are highlighted throughout.
- **Chapter 7 – Support Vector Machine:** Introduces the theory of maximum-margin classifiers and the geometry of separating hyperplanes. The Lagrangian and dual formulations are treated rigorously, with detailed discussion of kernel functions and nonlinear mappings. The chapter concludes with multi-class extensions and a reflective comparison between SVMs and neural networks.
- **Chapter 8 – Hyperbox Classifier:** Explains the min–max and fuzzy hyperbox models for pattern classification—an elegant blend of geometry and fuzzy reasoning. It details hyperbox construction, membership functions, and compensatory neuron architectures, emphasizing interpretability and adaptability in uncertain environments.
- **Chapter 9 – Clustering:** Focuses on unsupervised learning and the self-organization of data. Topics include hierarchical, partitioning, graph-based, and density-based methods such as K-means, DBSCAN, spectral clustering, and minimum spanning tree (MST) clustering. The chapter culminates with correlation clustering, which unifies similarity- and optimization-based perspectives.
- **Chapter 10 – Temporal Pattern Recognition:** Addresses time-dependent data where observations evolve sequentially. It covers Hidden Markov Models (HMMs), Dynamic Time Warping (DTW), Conditional Random Fields (CRFs), and Recurrent Neural Networks (RNNs) including LSTM architectures. This chapter connects the principles of memory, temporal dependence, and dynamic learning.
- **Chapter 11 – Convolutional Neural Networks:** Provides a detailed study of spatial pattern recognition and deep learning architectures. It explains convolution, padding, stride, pooling, and weight sharing using intuitive toy examples. Prominent models such as LeNet-5, AlexNet, VGG, ResNet, DenseNet, MobileNet, EfficientNet, and YOLO are analyzed with architectural and mathematical clarity.
- **Appendix A – Design and Performance Evaluation of Classifiers:** Discusses experimental design and model assessment using cross-validation, confusion matrices, ROC curves, AUC, Kappa coefficient, and bias–variance analysis. It highlights how evaluation metrics reflect model reliability and decision robustness.
- **Appendix B – Optimization Foundations:** Presents the mechanics of learning through gradient descent, momentum, learning rate scheduling, and regularization. A catalogue of loss functions is included, with derivations and practical insights into their comparative behavior.
- **Appendix C – Convolution and Pooling Operations:** Provides mathematical details of discrete and continuous convolution, padding, stride, dilation, and pooling—key to understanding CNN internals.
- **Appendix D – Explainable AI:** Explores interpretability and transparency in modern AI systems. Topics include SHAP, LIME, saliency maps, Grad-CAM, and quantitative measures for explanation quality. This appendix connects pattern recognition to ethical, interpretable, and responsible AI.

Objectives

1. To present pattern recognition and machine learning as a unified intellectual framework linking theory, geometry, and intuition.
2. To transform mathematical abstraction into conceptual clarity through illustrative examples.
3. To connect classical statistical foundations with modern learning paradigms.
4. To foster reasoning from first principles rather than formulaic recall.
5. To make the study of pattern recognition a pursuit of elegance, logic, and joy.

Intended Audience

This book is designed for postgraduate and senior undergraduate students in Computer Science, Artificial Intelligence, Data Science, and related fields (B.Tech., M.Tech., MCA, M.Sc.), as well as for PhD/Post-Doc researchers seeking solid theoretical grounding. It also welcomes scholars from psychology, linguistics, cognitive science, and biomedical engineering who wish to understand how computational models parallel human reasoning.

Teachers will find its logical flow and stepwise derivations ideal for classroom exposition. Self-learners will appreciate its clarity and structure. Researchers and professionals will recognize in it a conceptual map that unifies the scattered algorithms of modern AI under one consistent philosophy.

Closing Reflections

Pattern recognition is more than an academic subject; it is a lens through which we understand the architecture of intelligence. It reveals that learning is not accidental but structured, not mysterious but principled. Behind every classifier lies a theory of uncertainty; behind every feature lies a theory of perception; behind every algorithm lies a theory of reasoning. The discipline teaches us that intelligence—whether biological or artificial—is ultimately the art of discovering order in a universe that offers no guarantees of clarity.

To recognize a pattern is to make a commitment: a commitment to the belief that meaning can be distilled from variation, that structure exists beneath noise, and that insight arises when the mind learns to differentiate, compare, and abstract. This commitment is both computational and philosophical. Computationally, it yields the models and decision rules we design; philosophically, it shapes the way we understand cognition, perception, and the nature of knowledge itself.

Pattern recognition encourages us to see learning not as memorization, but as transformation. A raw observation is not yet information; it becomes informative only when mapped into a representation that highlights what is essential and suppresses what is not. This delicate interplay—between measurement and meaning, evidence and belief, geometry and probability—forms the subconscious grammar of intelligent behavior. It is here that mathematics transcends symbolism and becomes a language of thought.

Machine learning extends this philosophy by showing that intelligence is plastic—that it adapts, improves, and internalizes structure through experience. Where pattern recognition provides the intellectual skeleton, machine learning provides the dynamic musculature. One offers the principles of representation, similarity, and inference; the other offers the mechanisms of optimization, generalization, and self-correction. Their union represents the modern vision of AI: models that do not simply compute, but evolve toward understanding.

As computational systems grow in scale—from shallow discriminant functions to deep hierarchical architectures, from Markovian temporal models to attention-driven transformers—the foundational questions endure.

- How should data be represented?
- What constitutes a good decision boundary?
- How do we quantify uncertainty?
- What does generalization truly mean?

These questions preserve the intellectual continuity between early pattern recognizers and contemporary generative models. They remind us that progress in AI is not just technological—it is conceptual.

For the curious reader, the study of pattern recognition becomes a form of intellectual introspection. It mirrors the cognitive mechanisms we intuitively employ in daily life: noticing regularities, predicting outcomes, revising beliefs, and reconciling conflicting evidence. Through this recognition, students and researchers develop not only technical skill but a more refined sense of reasoning—an ability to see patterns where others see complexity, to interpret phenomena where others see chaos.

To embrace this discipline is to cultivate a certain temperament:

- the patience to dissect abstractions,
- the imagination to visualize high-dimensional spaces,
- the humility to acknowledge uncertainty, and
- the confidence to craft models that balance elegance with utility.

Such qualities transcend software libraries and coding paradigms; they shape the mindset of a thoughtful scientist and a discerning engineer.

If this book achieves its purpose, it will not merely inform—it will transform the way readers think about data, models, and the very process of learning. It will help them see equations as distilled wisdom, derivations as narratives of logic, and algorithms as crystallized forms of human understanding. It will invite them to appreciate the subtle beauty of mathematical constructs that explain why recognition is possible at all.

In the larger arc of AI, pattern recognition stands as a reminder that intelligence is not magic—it is structure. And structure, when understood deeply, becomes insight; when encoded, becomes computation; and when refined through experience, becomes learning.

May the concepts in this book encourage readers to explore, question, and discover with renewed clarity. May they recognize that every dataset hides a geometry, every hypothesis hides a principle, and every model hides a story about how intelligence interprets the world.

Use pattern recognition to define what truly matters.

Use machine learning to refine how meaning is uncovered.

And let the joy of understanding remain a lifelong companion on the journey of scientific inquiry.

Contents

Dedication	ix
Acknowledgements	xi
Preface	xiii
1 Introduction to Pattern Recognition	1
1.1 Overview	1
1.2 What is a Pattern?	1
1.3 Origins of Pattern Recognition	2
1.4 Philosophical Roots: Plato and Aristotle	2
1.4.1 Plato’s View	3
1.4.2 Aristotle’s Contribution	3
1.5 Defining Pattern Recognition	4
1.6 Applications of Pattern Recognition	4
1.7 Supervised and Unsupervised Learning	4
1.7.1 Supervised Learning	4
1.7.2 Unsupervised Learning	4
1.8 Features	5
1.8.1 Boundary Features	6
1.8.2 Region Features	6
1.9 Feature Vectors	6
1.10 Why Scaling Helps	7
1.10.1 Standardization (z-score).	8
1.10.2 Min–max scaling (to $[0, 1]$).	8
1.10.3 Robust scaling (resistant to outliers).	8
1.11 Feature Space	9
1.12 Temporal, Spatial, and Spatio-Temporal Domains	10
1.13 Classification Approaches	11
2 Feature Extraction	15
2.1 How Do We Call Two Patterns “Similar”?	15
2.2 Warm-Up Example: Recognizing Circular Arcs	15
2.3 From Pattern to Feature Vector	16
2.4 What Features Should We Extract?	17
2.5 Boundary Representation by Chain Codes	17
2.5.1 Direction Coding: 4- vs. 8-Connectivity	17
2.5.2 How to Generate a Chain Code (Practical Steps)	18

2.5.3	What Changes the Chain Code?	18
2.5.4	Differential Chain Codes: Making Rotation Invariant	19
2.5.5	Handling the Starting-Point Sensitivity	20
2.5.6	Handling the Scale Sensitivity	20
2.6	Polygonal Approximation of Boundary	21
2.6.1	Minimum Perimeter Length Polygonization	22
2.6.2	Splitting	24
2.6.3	From Polygonization to Feature Vector Formation	27
2.6.4	Signature	29
2.7	Region Features	34
2.7.1	Spatial Domain Features	35
2.7.2	Features Derived from the Co-occurrence Matrix	37
2.7.3	Transformed Domain Features	39
3	Bayes Decision Theory	49
3.1	From Features to a Feature Space	49
3.1.1	Linear Separability	49
3.1.2	Beyond Linear Separability	49
3.2	Bayes Decision Theory from First Principles	54
3.2.1	A Naive Decision Rule	54
3.2.2	Combining Priors with Observed Features	54
3.2.3	Probability of Error in Bayes Classification	58
3.3	Generalization of Bayes Theorem	62
3.3.1	Generalization Purpose	62
3.3.2	Minimum Risk Classifier	64
3.3.3	Minimum Error Rate Classifier	67
3.3.4	Discriminant Function	69
3.4	Normal Density and Discriminant Function	76
3.4.1	Normal (Gaussian) Probability Density Function	76
3.4.2	Multivariate Normal Probability Density Function	78
3.4.3	Bivariate Normal Distribution	80
3.4.4	Higher-Dimensional Normal Distribution	85
3.4.5	From Probability Densities to Decision Surfaces	88
3.5	Discriminant Function and Decision Surfaces: Special Case	91
3.5.1	Case 1: Covariance Matrix $\Sigma_i = \sigma^2 I$	91
3.5.2	Case 2: Arbitrary but Common Covariance Matrix $\Sigma_i = \Sigma$	99
3.5.3	Case 3: Arbitrary Covariance Matrices ($\Sigma_i \neq \Sigma_j$)	105
3.6	Bayes Decision Theory for Discrete Binary Feature Vectors	109
3.6.1	Binary Feature Vectors	109
3.6.2	Conditional Independence Assumption	110
3.6.3	Probabilities for the Two Classes	110
3.6.4	Class-Conditional Probability for ω_1	111
3.6.5	Class-Conditional Probability for ω_2	111
3.6.6	Likelihood Ratio	111
3.6.7	Decision Function Using Likelihood Ratio	113
3.6.8	Linear Discriminant Function	114
3.6.9	Role of the Weights	115
3.7	Maximum Likelihood Estimation	119
3.7.1	Problem Formulation for Parameter Estimation	120

3.7.2	Validation with Gaussian Distribution (Known Variance, Unknown Mean)	125
3.7.3	Validation with Gaussian Distribution (Unknown Mean and Unknown Variance)	130
3.7.4	Validation with Exponential Distribution (Unknown Parameter θ)	134
4	Probability Density Function Estimation	139
4.1	Recap of Bayes Decision Theory	139
4.1.1	Posterior Probabilities and Bayes Rule	139
4.1.2	Discriminant Functions	140
4.1.3	Role of Class-Conditional Probability Densities	140
4.2	Parametric Probability Density Function	141
4.2.1	Gaussian Distribution	141
4.2.2	Effect of Mean and Variance on Shape	143
4.2.3	Multivariate Gaussian PDF	143
4.2.4	Log-Normal Distribution	145
4.2.5	Exponential Distribution	146
4.2.6	Poisson Distribution	147
4.3	Non-Parametric Probability Density Function Estimation	150
4.3.1	Histogram Based Estimation	151
4.3.2	Kernel Based Estimation (Window Functions)	169
4.4	The Dimensionality Problem	179
4.4.1	Motivation from Bayes Decision Theory	179
4.4.2	Adding More Features	179
4.4.3	Computational Complexity	180
4.4.4	Effect of Dimensionality	180
4.4.5	Principal Component Analysis	183
4.4.6	Multiple Discriminant Analysis	193
4.4.7	Fisher's Linear Discriminant: The Two-Class Case	195
5	Discriminant Criteria Functions	215
5.1	Linear Discriminant Function without Density Assumptions	215
5.1.1	Definition of Linear Discriminant Function	215
5.1.2	Geometric Nature of \mathbf{w}	216
5.1.3	Interpretation of $g(\mathbf{x})$ as Distance	216
5.1.4	Algebraic Distance and Sign	217
5.1.5	Homogeneous Form of the Linear Discriminant Function	218
5.1.6	Decision Rule in Homogeneous Form	219
5.1.7	Designing the Weight Vector for a Linear Classifier	220
5.2	Batch Perceptron Learning	222
5.2.1	Why Do We Need a Criterion Function?	222
5.2.2	The Batch Perceptron Criterion Function	222
5.2.3	Gradient Descent Minimization	223
5.2.4	Key Challenges of the Batch Perceptron Approach	232
5.2.5	Online Sequential Perceptron Approach	232
5.2.6	Need for a Margin in Perceptron Learning	240
5.3	Relaxation Criteria Procedure and Restricted Solution Region	241
5.3.1	Derivation of the Relaxation Update Rule	243
5.4	Minimum Squared Error Criteria	247
5.4.1	Formulation in Matrix Form	248
5.4.2	Error Vector and Criterion Function	248
5.4.3	Gradient Descent Derivation	248

5.4.4	Closed-Form Solution	249
5.4.5	Pseudo-Inverse Formulation	249
5.4.6	Sequential Update Rule	250
5.5	MSE with Adaptive Margin Vector	254
5.5.1	The Problem with Plain MSE	255
5.5.2	Incorporating a Margin Vector	255
5.5.3	Redefining the Criterion Function	256
5.5.4	Gradients with Respect to \mathbf{a} and \mathbf{b}	256
5.5.5	Positivity Constraint on \mathbf{b}	257
5.5.6	Updating Weight Vector \mathbf{a}	258
5.6	Widrow–Hoff / Least Mean Square Criteria	262
5.7	Extension of the Two-Class Linear Discriminant Criteria	266
5.7.1	Nonparametric Case and Multi-Class Inequalities	266
5.7.2	Discriminant Functions in the Multi-Class Case	267
5.7.3	Unified Approach	267
5.7.4	Constructing Equivalent Feature Vectors	268
5.8	Nearest Neighbour	280
5.8.1	The Nearest Neighbour Rule (1-NN)	280
5.8.2	Limitations of the NN Rule	282
5.8.3	The k -Nearest Neighbours Rule (kNN)	282
5.8.4	Distance-weighted kNN	284
5.9	Fuzzy k -Nearest Neighbour Classifier	287
5.9.1	Mathematical Formulation	287
5.9.2	Membership Assignment Rule	287
5.9.3	Classification Decision	288
5.9.4	Understanding in Any Dimension	288
5.9.5	Algorithmic Steps	288
5.9.6	Illustrative Perspective	289
5.9.7	Practical Considerations	289
5.10	Parzen Window Classifier	293
5.10.1	Mathematical Formulation for d -Dimensional Data	293
5.10.2	Understanding in Any Dimension	294
5.10.3	Classification Using Estimated Densities	294
5.10.4	Choice of Kernel Functions	294
5.10.5	Influence of the Bandwidth h	295
5.10.6	Practical Considerations	295
5.11	Linear Regression Model	299
5.11.1	Simple Linear Regression	299
5.11.2	Multiple Linear Regression	308
5.12	Logistic Regression Classifier	313
5.12.1	Explanatory and Categorical Variables	314
5.12.2	Mathematical Model and Parameters	314
5.12.3	Logit Function, Odds, and Odds Ratio	319
5.12.4	Parameter Estimation and Logistic Loss	323
5.13	Decision Tree	337
5.13.1	Types and Extensions of Decision Trees	338
5.13.2	Algorithmic Variants of Decision Tree	341
5.13.3	The CART Algorithm	342

6	Neural Networks	365
6.0.1	Biological Perspective	365
6.1	Artificial Neuron Model	366
6.1.1	Structure and Function	366
6.1.2	Threshold and Classification	366
6.2	Artificial Neuron with Bias	367
6.2.1	Incorporating the Bias Term	367
6.2.2	Decision Rule with Bias	368
6.3	Single-Layer Perceptron with Single Output	369
6.3.1	Single-Layer Perceptron with Single Output: Error Minimization and Learning	371
6.3.2	Single-Layer Perceptron with Multiple Outputs	383
6.4	Multilayer Perceptrons (MLPs)	396
6.4.1	Motivation: Beyond Single-Layer Networks	396
6.4.2	Network Architecture	397
6.4.3	Various Activation Functions	398
6.4.4	Forward Propagation of Signals	402
6.4.5	Error Function	405
6.4.6	Final Weight Update Rule	410
6.4.7	Backpropagation Learning Algorithm	411
6.5	Hopfield Network	417
6.5.1	Associative Memory Concept	417
6.5.2	Binary Feature Vectors	418
6.5.3	Network Structure	418
6.5.4	Storing Patterns in Weights	419
6.5.5	Network Dynamics and Recall	420
6.5.6	Energy Function	420
6.6	Radial Basis Function Neural Network	433
6.6.1	Motivation for RBF Networks	434
6.6.2	Dimensionality Expansion	435
6.6.3	Transformation to Higher Dimensions	435
6.6.4	Different Choices of Radial Basis Functions	436
6.6.5	Distance Measure in RBFs	437
6.6.6	Architecture of Radial Basis Function Neural Networks	439
6.6.7	Training in RBF Neural Networks	440
6.6.8	XOR Classification: Low-to-High Dimensional Mapping	445
6.6.9	Training the Output Layer in RBF Networks	448
7	Support Vector Machine	459
7.1	Classification Rule in SVM	459
7.1.1	Training process	460
7.1.2	Decision boundary	460
7.1.3	Classification rule	460
7.2	Why Arbitrary Hyperplanes May Fail?	461
7.3	Maximizing the Margin	461
7.3.1	Basic Linear Separation	462
7.3.2	Class Encoding with Labels	462
7.3.3	From Separation to Margin Maximization	463
7.3.4	Normalization and Scaling	463
7.3.5	Support Vectors	463

7.3.6	Formulating the Optimization Problem	465
7.4	The Lagrangian Formulation	466
7.4.1	Constrained Optimization Setup	466
7.4.2	Introducing the Lagrangian	467
7.4.3	Expand the Lagrangian	467
7.4.4	Stationarity Conditions	467
7.4.5	Substitution into the Lagrangian	468
7.4.6	The Dual Optimization Problem	470
7.5	Decision Function	470
7.5.1	Express \mathbf{w} in terms of support vectors	470
7.5.2	Compute the discriminant value	470
7.5.3	Apply the decision rule	471
7.6	Kernel Functions	473
7.6.1	Nonlinear Separability and the Kernel Trick	473
7.6.2	Mapping to Feature Space	474
7.6.3	The Kernel Trick	475
7.6.4	Common Kernels	476
7.6.5	Computation of the Bias Term b	479
7.6.6	Decision Rule and Decision Boundary in Kernel SVMs	480
7.7	Extension to Multi-Class Problems	495
7.7.1	One-vs-Rest (OvR)	495
7.7.2	One-vs-One (OvO)	500
7.7.3	Direct Methods	503
8	Hyperbox Classifier	507
8.0.1	Motivation for Hyperbox Classifiers	507
8.1	Principle of Hyperbox Classifiers	508
8.1.1	Hyperboxes in the Feature Space	508
8.1.2	Assigning Classes to Hyperboxes	509
8.1.3	Classification of New Samples	509
8.2	Min–Max Hyperboxes	510
8.2.1	Generalization to Higher Dimensions	511
8.2.2	Membership Condition for a Point	511
8.2.3	Constructing a Min–Max Hyperbox from Training Data	511
8.3	Fuzzy Hyperboxes and Membership Functions	512
8.3.1	Handling Unknown Samples	512
8.4	From Crisp to Fuzzy Min–Max Hyperboxes	513
8.4.1	Fuzzy Classification Decision Rules	513
8.4.2	Mathematical Formulation of Fuzzy Membership	514
8.4.3	Fuzzy Membership Across Multiple Hyperboxes	516
8.4.4	Building and Expanding Hyperboxes During Training	518
8.4.5	The Risk of Over Expansion	519
8.5	Fuzzy Min–Max Neural Network	523
8.5.1	Network Architecture	523
8.5.2	Training and Hyperbox Creation	525
8.5.3	Classification Strategy	526
8.5.4	Containment and Overlap of Hyperboxes	529
8.6	Fuzzy Min–Max Neural Network with Compensatory Neurons	537
8.6.1	Concept of Compensatory Neurons and Graded Membership	538

8.6.2	FMCN Architecture Overview	538
8.6.3	Compensation Mechanism in FMCN	541
8.6.4	Implications of FMCN	542
8.6.5	Comparison of FMCN, FMNN, and Traditional Neural Networks	542
8.6.6	Computation of Compensation Neurons in FMCN	549
9	Clustering	559
9.1	Hierarchical Clustering Approaches	560
9.1.1	Agglomerative Nesting (Bottom-Up) Clustering	561
9.1.2	Divisive (Top-Down) Clustering	562
9.2	Similarity Measures Between Clusters Approaches	564
9.2.1	Single Linkage (Nearest Neighbor) Algorithm	564
9.2.2	Complete Linkage (Farthest Neighbor) Algorithm	567
9.2.3	Average Linkage Algorithm	571
9.2.4	Ward's Method	573
9.3	Batchelor and Wilkins Algorithm	581
9.3.1	Comparison with Hierarchical Methods	582
9.3.2	Batchelor and Wilkins Algorithm	582
9.4	Graph-Based Clustering Algorithm	587
9.4.1	Theoretical Background	587
9.4.2	Clustering Algorithm	588
9.4.3	Comparison with Sequential Methods	588
9.5	Minimal Spanning Tree Based Clustering	591
9.5.1	Concept and Idea	591
9.5.2	Minimal Spanning Tree Based Clustering Algorithm	591
9.5.3	Stopping Criteria for Clustering	592
9.5.4	Properties and Remarks	592
9.6	K-Means Clustering	597
9.6.1	Criterion Function Based on Clustering	598
9.6.2	Net Effect of Transfer Between Clusters	603
9.7	Guard Zone Clustering Algorithm	611
9.7.1	Conceptual Foundation	612
9.7.2	Guard Zone Algorithm	612
9.7.3	Properties and Discussion	613
9.8	DBSCAN Clustering	616
9.8.1	Conceptual Foundation	616
9.8.2	DBSCAN Algorithm	616
9.8.3	Properties and Insights	617
9.9	Spectral Clustering	620
9.9.1	Motivation	620
9.9.2	Similarity Graph	621
9.9.3	Graph Laplacians	622
9.9.4	Spectral Properties	623
9.9.5	Spectral Clustering Algorithm	623
9.9.6	Theoretical Basis	624
9.9.7	Applications, Advantages, and Limitations	625
9.10	Divisive Analysis Clustering Algorithm (DIANA)	634
9.10.1	Intuitive Idea	635
9.10.2	Algorithmic Foundation of the DIANA Method	635

9.10.3	Mathematical Criterion and Theoretical Basis	637
9.11	Correlation Clustering	647
9.11.1	Concept and Motivation	647
9.11.2	Relational Representation	648
9.11.3	Why It Matters?	649
9.11.4	Graph-Theoretic Intuition	649
9.11.5	Problem Setup	650
9.11.6	Formal Definitions	651
9.11.7	Optimization Formulations	651
9.11.8	Alternative Cost Representation	652
9.11.9	Unified Optimization Expression	652
9.11.10	Computation of Negative Inter-Cluster Edges	653
10	Temporal Pattern Recognition	667
10.0.1	Examples of Temporal Patterns	667
10.0.2	Sequential Machines and Finite State Models	668
10.0.3	An Illustrative Finite State Machine	668
10.1	Hidden Markov Model	670
10.1.1	Concept of State Transitions	670
10.1.2	Transition Probabilities	671
10.1.3	Probability of a State Sequence in an HMM	673
10.1.4	Absorbing State and Observable Symbols in Hidden Markov Models	675
10.1.5	Structure of a Hidden Markov Model	676
10.1.6	Hidden Markov Model with an Accepting (Absorbing) State	678
10.1.7	Central Problems in HMMs (Evaluation, Decoding, and Learning)	680
10.1.8	Evaluation Problem in HMM	681
10.1.9	Forward Algorithm for Hidden Markov Model	689
10.1.10	Decoding Problem in HMM	696
10.1.11	Viterbi Algorithm for Decoding Optimal State Sequence	698
10.1.12	Classification of a Sequence of Visible Symbols	703
10.1.13	Learning or Training of HMMs	705
10.2	Dynamic Time Warping	718
10.2.1	Concept	719
10.2.2	Local Distance Measure	719
10.2.3	Recursive Formulation	720
10.2.4	Warping Path	720
10.2.5	Algorithmic Steps	721
10.2.6	Locality Constraint (Sakoe–Chiba Band)	721
10.2.7	Complexity and Optimizations	722
10.2.8	Relation to Hidden Markov Models	722
10.3	Conditional Random Fields	725
10.3.1	Why Structured Prediction Matters	725
10.3.2	Core Idea in Simple Words	726
10.3.3	Mathematical Formulation of Conditional Random Fields	727
10.3.4	Representing Dependencies Using Graphs	727
10.3.5	Linear-Chain CRF	727
10.3.6	Conditional Markov Property	728
10.3.7	Meaning of the Equation	728
10.3.8	Analogy for Understanding	729

10.3.9	Mathematical Definition of the Model	730
10.3.10	Linear-Chain Conditional Random Fields	732
10.3.11	Parameter Learning	735
10.4	Vanilla Recurrent Neural Networks	754
10.4.1	Intuition	755
10.4.2	Architecture of the Vanilla RNN	755
10.4.3	Computational Graph and Unfolding in Time	757
10.4.4	Forward Computation	757
10.4.5	Sequence Modeling and Loss Function	763
10.4.6	Backpropagation Through Time	768
10.4.7	Vanishing and Exploding Gradients	806
10.5	Long Short-Term Memory Networks	810
10.5.1	Motivation	810
10.5.2	Intuitive Analogy	810
10.5.3	Why the Name “Long Short-Term Memory”?.	811
10.5.4	Architecture of an LSTM Cell	811
10.5.5	Gate-Level Interpretation	814
10.5.6	Forward Computation	816
10.5.7	Backpropagation Through Time in LSTM	824
10.5.8	How LSTM Solves the Vanishing and Exploding Gradient Problems?	853
10.5.9	Constant Error Carousel (CEC)	855
10.5.10	Variants of LSTM	856
11	Convolutional Neural Networks	863
11.1	Basics of CNN	863
11.1.1	Biological Inspiration	863
11.1.2	Illustration: Limitations of Basic ANN in Image Recognition	864
11.1.3	Convolutional Simplification and Weight Sharing	866
11.1.4	Weight Initialization in Convolutional Neural Networks	870
11.1.5	Basic Architecture of a Convolutional Neural Network	872
11.1.6	Mathematical Foundation of Convolutional Neural Network	886
11.1.7	Complete Workflow of a Single-Channel CNN	903
11.1.8	Evolution of Dimensions from Input to Classification	905
11.1.9	Forward Propagation in Convolutional Neural Networks	911
11.1.10	Backward Propagation in Convolutional Neural Networks	914
11.1.11	Toy Example: CNN Training on 3×3 Pixel Size “x” Symbol Image	931
11.2	LeNet-5	957
11.2.1	Historical Context and Significance	957
11.2.2	Motivation and Design Philosophy	959
11.2.3	Architecture of LeNet-5	959
11.2.4	Mathematical Insight	960
11.3	AlexNet	961
11.3.1	Historical Context and Impact	961
11.3.2	Architectural Overview	961
11.3.3	Architectural Design and Layer Functions	962
11.3.4	Mathematical Viewpoint	963
11.3.5	VGGNet-16	963
11.3.6	Introduction and Motivation	963
11.3.7	Architectural Philosophy	964

11.3.8	Layer-by-Layer Overview of VGG-16	964
11.3.9	Mathematical Intuition	965
11.4	GoogLeNet / Inception v1	967
11.4.1	Historical Context and Motivation	967
11.4.2	Architectural Insight: The Inception Module	967
11.4.3	Dimensionality Reduction Example	967
11.4.4	Architectural Composition	967
11.4.5	Mathematical Representation of Inception	968
11.5	ResNet-34	970
11.5.1	Historical Motivation and Context	970
11.5.2	Core Concept: Residual Learning	970
11.5.3	Architecture of ResNet-34	972
11.5.4	Mathematical Analysis of Residual Mapping	972
11.5.5	Residual Block Variants	973
11.5.6	Residual Signal Flow and Shortcut Projection	973
11.5.7	ResNet Family and Successive Variants	974
11.6	DenseNet-121	976
11.6.1	Concept and Motivation	976
11.6.2	High-Level Architectural Flow	976
11.6.3	Dense Block and Transition Layer Mechanics	977
11.6.4	DenseNet-121 Structure	978
11.6.5	Growth Rate and Feature Evolution	980
11.6.6	Transition Layers	980
11.6.7	Growth Rate (k)	980
11.6.8	DenseNet Variants	981
11.7	MobileNet V1 Architecture	983
11.7.1	Overview	983
11.7.2	Depthwise Separable Convolution	983
11.7.3	Network Structure	984
11.7.4	Width and Resolution Multipliers	986
11.7.5	Applications and Evolution of MobileNet Series	988
11.7.6	Evolution of MobileNet Series	989
11.8	EfficientNet-B0	990
11.8.1	Concept and Motivation	990
11.8.2	Baseline: EfficientNet-B0	990
11.8.3	EfficientNet Architecture Overview	990
11.8.4	Depthwise Separable Convolution	992
11.8.5	Residual vs. Inverted Residual	992
11.8.6	Squeeze-and-Excitation (SE) Mechanism	993
11.8.7	Swish Activation Function	994
11.8.8	Illustration of Compound Scaling	995
11.8.9	Variants of EfficientNet	997
11.9	YOLO v1	998
11.9.1	Flow of Object Detection Process	998
11.9.2	YOLO v1: Basic Architecture and Concept	1001
11.9.3	YOLOv1 in Action	1005
11.9.4	Output Tensor Representation in YOLO v1	1007
11.9.5	Visualization of YOLO v1 Detection Process	1008
11.9.6	Feature Vector Representation in YOLO v1	1009

11.9.7	Training of YOLOv1	1009
11.9.8	Comparison with Two-Stage Architectures	1011
A	Design and Performance Evaluation of Classifiers	1015
A.1	Overview of Classifier Types	1015
A.2	Designing a Robust Classifier: The Leave-One-Out Technique	1016
A.2.1	Basic Idea	1017
A.2.2	Estimation of Error Rate	1017
A.2.3	Practical Interpretation	1019
A.2.4	Computational Considerations	1020
A.3	Generalized Cross-Validation	1020
A.3.1	Concept	1020
A.3.2	Procedure	1020
A.3.3	Choice of k	1021
A.3.4	Variants of Cross-Validation	1023
A.4	Classifier Performance Metrics	1024
A.4.1	Binary Classification Framework	1024
A.4.2	Sensitivity and Specificity	1024
A.4.3	Additional Derived Metrics	1025
A.4.4	Confusion Matrix for Multi-Class Problems	1026
A.5	ROC and AUC	1027
A.5.1	Mathematical Construction	1027
A.5.2	Area Under the Curve (AUC)	1028
A.6	Matthews Correlation Coefficient	1030
A.6.1	Definition	1030
A.6.2	Interpretation	1031
A.7	Cohen's Kappa Coefficient (κ)	1031
A.7.1	Definition	1032
A.7.2	Computation from Confusion Matrix	1032
A.7.3	Interpretation of κ	1032
A.7.4	Advantages and Limitations of Cohen's Kappa	1033
A.8	Bias-Variance Trade-Off	1034
A.8.1	Conceptual Background	1034
A.8.2	Mathematical Formulation	1034
A.8.3	Training vs. Test Error Behavior	1037
A.8.4	Practical Insight	1037
A.9	Experimental Evaluation and Reproducibility Framework	1040
A.9.1	Dataset Preparation and Partitioning	1040
A.9.2	Performance Logging and Statistical Validation	1041
A.9.3	Computational Efficiency Metrics	1041
A.9.4	Reproducibility Protocol	1041
B	Optimization Foundations	1043
B.1	Gradient Descent	1044
B.1.1	Concept	1044
B.1.2	Geometric Interpretation	1044
B.1.3	Dataset and Learning Dynamics	1044
B.1.4	The Learning Curve and Convergence Behavior	1045
B.1.5	Batch Gradient Descent	1046
B.1.6	Stochastic Gradient Descent	1046

B.1.7	Mini-Batch Gradient Descent	1047
B.1.8	Epochs, Iterations, and Generalization	1048
B.1.9	Visualization of the Optimization Process	1050
B.2	Momentum and Learning Rate Scheduling	1061
B.2.1	Motivation	1061
B.2.2	Momentum	1061
B.2.3	Learning Rate Scheduling	1062
B.3	Regularization: Controlling Overfitting	1071
B.3.1	L1 Regularization (Lasso)	1072
B.3.2	L2 Regularization (Ridge)	1072
B.3.3	Dropout Regularization	1073
B.3.4	Comparative Insight	1075
B.4	Overview of Common Loss Functions	1091
B.4.1	Mean Squared Error	1092
B.4.2	Mean Absolute Error	1093
B.4.3	Huber Loss	1094
B.4.4	Log-Cosh Loss	1095
B.4.5	Cross-Entropy Loss	1096
B.4.6	Focal Loss	1097
B.4.7	Hinge Loss	1098
B.4.8	Kullback–Leibler (KL) Divergence	1099
B.4.9	Dice Loss	1100
B.4.10	Cosine Similarity Loss	1102
B.4.11	Triplet Loss	1103
B.4.12	Contrastive Loss	1104
C	Convolution and Pooling Operations	1107
C.1	Discrete Convolution	1107
C.1.1	Definition: Discrete Convolution	1108
C.1.2	Polynomial and Sequence Interpretation	1114
C.1.3	Fast Fourier Convolution Algorithm	1116
C.2	Continuous Convolution	1120
C.3	Stride, Padding, and Dilation	1126
C.3.1	Stride: Controlling the Step of Convolution	1127
C.3.2	Padding: Preserving Boundary Information	1127
C.3.3	Dilation: Expanding the Receptive Field	1133
C.4	Pooling Operations	1135
C.4.1	Mathematical Description	1137
C.4.2	Max Pooling	1139
C.4.3	Average Pooling	1140
C.4.4	Generalized and Hybrid Pooling Methods	1143
C.4.5	Hierarchical and Structured Pooling	1145
D	Explainable AI	1147
D.1	Motivation and Rationale	1148
D.2	Explainability	1148
D.3	Principles and Theoretical Foundations	1149
D.4	Explainability versus Interpretability	1151
D.5	Relation to Responsible and Trustworthy AI	1151
D.6	Relevance to Pattern Recognition	1152

D.7	Feature Attribution Methods	1153
D.7.1	Gradient-Based Attribution	1153
D.7.2	SHapley Additive exPlanations	1159
D.7.3	Local Interpretable Model-agnostic Systems (LIMS)	1167
D.8	Visualization-Based Explanations	1176
D.8.1	Saliency Maps	1176
D.8.2	Gradient-weighted Class Activation Mapping (Grad-CAM)	1180
D.9	Quantifying Explanation Quality	1186
D.9.1	Faithfulness (or Fidelity)	1186
D.9.2	Stability (or Robustness)	1187
D.9.3	Sparsity (or Compactness)	1187
D.9.4	Complementary Metrics	1188