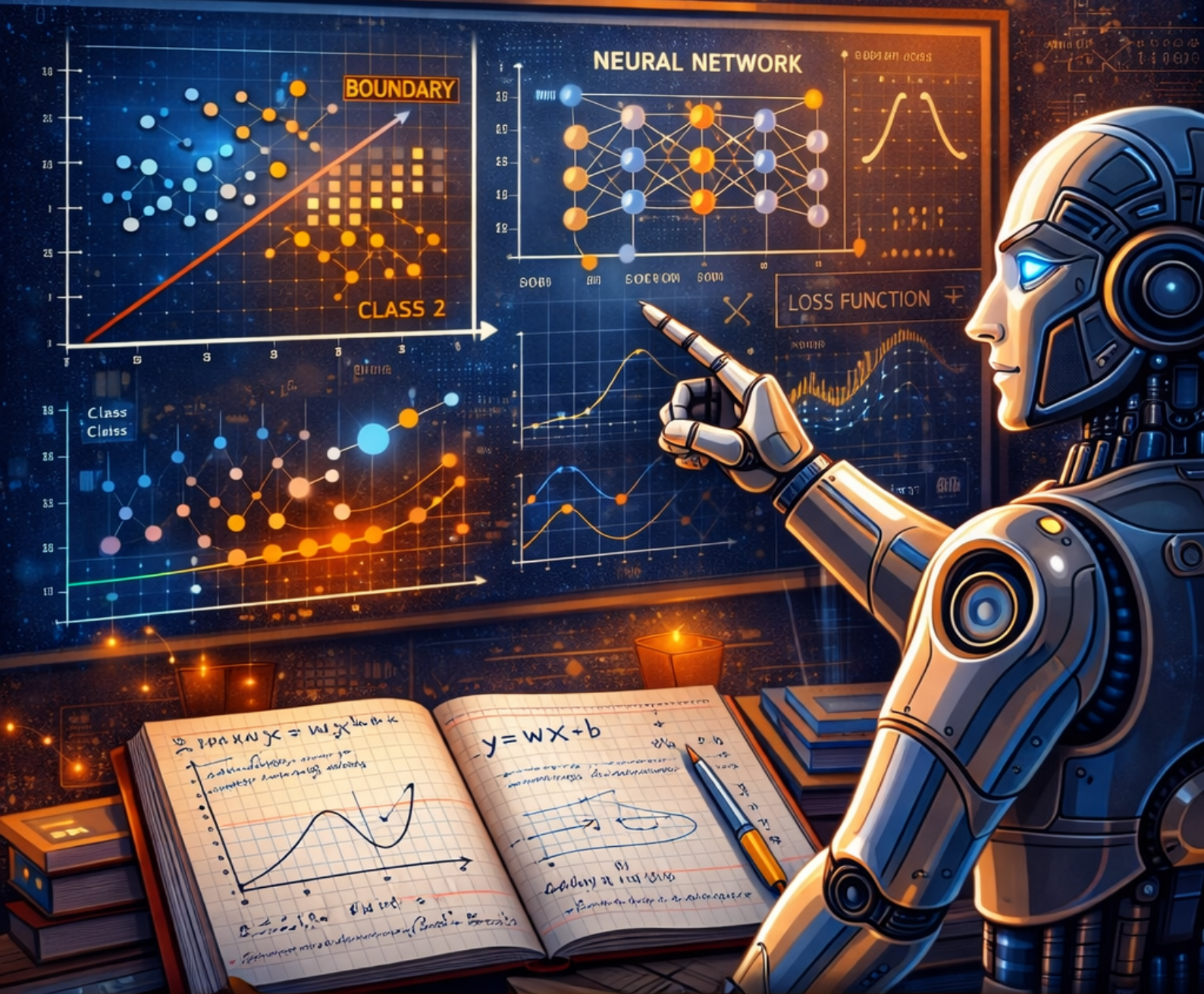


MACHINE LEARNING ALGORITHMS EXERCISE BOOK



Worked Problems and Practice Exercises

Machine Learning Algorithms Exercise Book

Worked Problems and Practice Exercises

Companion to Complete Machine Learning Algorithms Reference Guide

by Krzysztof Kołec

02-2026

Copyright © 2026 Krzysztof Kołek

All rights reserved.

No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the author, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law.

Disclaimer: The information provided in this book is for educational purposes only. While every effort has been made to ensure accuracy, the author makes no warranties or representations regarding the completeness or accuracy of the contents. The author shall not be liable for any loss or damage arising from the use of this book.

ISBN: 978-83-957596-3-5

First Edition: 2026

About This Book

This comprehensive reference guide was developed using AI tools to ensure thorough coverage of machine learning algorithms. As the author, I structured the outline, directed the content creation, verified technical accuracy, and edited all material for clarity and mathematical rigor. My goal was to create a practical, mathematically detailed resource for machine learning practitioners, students, and researchers.

This book provides:

- Detailed mathematical foundations for each algorithm
- Step-by-step explanations of complex formulas
- Practical guidance on when to use each method
- Worked examples demonstrating real-world applications
- Implementation considerations and hyperparameter tuning advice

Whether you're a student learning machine learning fundamentals, a data scientist seeking a quick reference, or a researcher needing mathematical details, this guide is designed to serve as your comprehensive companion.

How to Use This Book

This reference is organized by algorithm type (classification, regression, clustering, etc.). Each section includes:

- **Core Concept:** Intuitive explanation of what the algorithm does
- **Mathematical Foundation:** Detailed formulas with variable explanations
- **Use Cases:** When to apply the algorithm
- **Practical Considerations:** Hyperparameters, pitfalls, and tips

You can read this book sequentially or jump directly to specific algorithms as needed.

Terms and Conditions

Permitted Use:

- Personal study and reference
- Educational use in academic settings with proper attribution
- Implementation of algorithms in your own projects

Prohibited Activities:

- Unauthorized reproduction or distribution of this book
- Commercial redistribution without written permission
- Creating derivative works without authorization

- Sharing digital copies with third parties
- Uploading to file-sharing websites or repositories

Violations of these terms may result in legal action under applicable copyright laws.

Contact

For permissions, corrections, or inquiries, please contact: krzysztof.kolek.praca@gmail.com

Krzysztof Kolek
September 2026

Contents

About This Book	3
Preface	9
1 Classification Algorithms	11
1.1 Logistic Regression	11
1.1.1 Worked Example 1: Binary Classification by Hand	11
1.1.2 Practice Exercises	12
1.2 Decision Trees	15
1.2.1 Worked Example 2: Computing Gini Impurity and Information Gain	15
1.2.2 Practice Exercises	16
1.3 Random Forest	20
1.3.1 Worked Example 3: Bootstrap Sampling and Prediction	21
1.3.2 Practice Exercises	22
2 Support Vector Machines	27
2.1 Linear SVM	27
2.1.1 Worked Example 4: Computing Margins and Support Vectors . .	27
2.1.2 Practice Exercises	28
2.2 Kernel SVM	33
2.2.1 Worked Example 5: RBF Kernel Computation	33
2.2.2 Practice Exercises	34
3 Clustering Algorithms	39
3.1 K-Means Clustering	39
3.1.1 Worked Example 6: K-Means by Hand	39
3.1.2 Practice Exercises	42
4 Regression Algorithms	47
4.1 Linear Regression	47
4.1.1 Worked Example 1: Simple Linear Regression by Hand	48
4.1.2 Practice Exercises	49
4.2 Ridge and Lasso Regularization	56
4.2.1 Worked Example 3: Understanding Regularization Effects	56
4.2.2 Practice Exercises	59
4.3 Polynomial Regression and Overfitting	66
4.3.1 Worked Example 4: Polynomial Overfitting	66
4.3.2 Practice Exercises	68
4.4 Evaluation Metrics	70
4.4.1 Worked Example 5: Computing All Metrics	70

4.4.2	Practice Exercises	72
5	Dimensionality Reduction	81
5.1	Principal Component Analysis (PCA)	81
5.1.1	Worked Example 1: PCA Step-by-Step (2D to 1D)	82
5.1.2	Worked Example 2: Choosing Number of Components	84
5.1.3	Practice Exercises	86
5.2	Choosing Number of Components	91
5.2.1	Worked Example 3: Comparing Selection Criteria	91
5.2.2	Practice Exercises	94
5.3	t-SNE vs UMAP	102
5.3.1	Worked Example 4: Comparing t-SNE and UMAP	102
5.3.2	Practice Exercises	110
5.4	Feature Selection Methods	131
5.4.1	Worked Example 5: Comparing Feature Selection Methods	131
5.4.2	Practice Exercises	141
6	Evaluation Metrics	163
6.1	Classification Metrics	163
6.1.1	Worked Example 1: Computing All Metrics from Confusion Matrix	164
6.1.2	Practice Exercises	165
6.2	ROC Curves and AUC	173
6.2.1	Worked Example 2: Constructing ROC Curve by Hand	173
6.2.2	Practice Exercises	176
6.3	Cost-Sensitive Evaluation	183
6.3.1	Worked Example 3: Finding Optimal Cost-Sensitive Threshold . .	184
6.3.2	Practice Exercises	187
7	Cross-Validation and Model Selection	193
7.1	K-Fold Cross-Validation	193
8	Neural Networks	195
8.1	Forward Propagation	195
8.1.1	Worked Example 1: Forward Pass Through 2-Layer Network . . .	196
8.1.2	Practice Exercises	198
8.2	Backpropagation	201
8.2.1	Worked Example 2: Complete Backpropagation Calculation . . .	202
8.2.2	Practice Exercises	207
8.3	Activation Functions	211
8.3.1	Worked Example 3: Comparing Activation Functions	212
9	Advanced Topics	217
9.1	Ensemble Methods	217
9.1.1	Voting Classifiers	217
9.1.2	Worked Example 1.1: Voting Classifier for Medical Diagnosis . . .	218
9.1.3	Practice Exercises	220
9.1.4	Stacking (Stacked Generalization)	222
9.2	Imbalanced Data Handling	228
9.2.1	Identifying Imbalance	228

9.2.2	Techniques for Handling Imbalanced Data	231
9.2.3	Practice Exercises	235
9.3	Cross-Validation Strategies	237
9.3.1	K-Fold Cross-Validation	238
9.3.2	Stratified K-Fold for Imbalanced Data	240
9.3.3	Time Series Cross-Validation	243
9.4	Model Deployment Considerations	247
9.4.1	Practice Exercises	251
9.5	Chapter Summary	257
9.5.1	Ensemble Methods	257
9.5.2	Imbalanced Data	257
9.5.3	Cross-Validation	257
9.5.4	Model Deployment	257
9.5.5	Key Principles	258

Preface

This exercise book is designed as a companion to the *Complete Machine Learning Algorithms Reference Guide*. It provides:

- **Worked Examples:** Step-by-step solutions to help you understand algorithm mechanics
- **Practice Exercises:** Problems for self-study with varying difficulty levels
- **Conceptual Questions:** Test your understanding of when and why to use each algorithm
- **Implementation Exercises:** Hands-on coding problems
- **Real-World Applications:** Applied problems from various domains

How to Use This Book

1. **Start with Worked Examples:** Each section begins with fully solved problems
2. **Attempt Practice Exercises:** Try solving on your own before checking solutions
3. **Use Hints Strategically:** Read hints if stuck, but try to complete independently
4. **Check Your Understanding:** Solutions include detailed explanations
5. **Progress Gradually:** Exercises are ordered by increasing difficulty

Difficulty Levels

- ★ **Basic:** Tests fundamental understanding
- ★★ **Intermediate:** Requires synthesis of concepts
- ★★★ **Advanced:** Challenging problems requiring deep understanding

Chapter 1

Classification Algorithms

1.1 Logistic Regression

Key Concept

Logistic Regression models the probability that an instance belongs to a particular class using the sigmoid function:

$$P(y = 1|x) = \sigma(w^T x + b) = \frac{1}{1 + e^{-(w^T x + b)}}$$

The model outputs probabilities in the range $[0,1]$, and we typically classify as positive if $P(y = 1|x) > 0.5$.

1.1.1 Worked Example 1: Binary Classification by Hand

Exercise

Given a trained logistic regression model with weights $w_1 = 0.5$, $w_2 = 0.8$, and bias $b = -1.2$, classify the following points:

- Point A: $(x_1, x_2) = (1, 2)$
- Point B: $(x_1, x_2) = (0, 0)$
- Point C: $(x_1, x_2) = (3, 1)$

Step 1: Recall the sigmoid function

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Step 2: Calculate for Point A

Linear combination:

$$\begin{aligned} z_A &= w_1 x_1 + w_2 x_2 + b \\ &= 0.5(1) + 0.8(2) + (-1.2) \\ &= 0.5 + 1.6 - 1.2 \\ &= 0.9 \end{aligned}$$

Apply sigmoid:

$$\begin{aligned}
 P(y = 1|x_A) &= \frac{1}{1 + e^{-0.9}} \\
 &= \frac{1}{1 + e^{-0.9}} \\
 &= \frac{1}{1 + 0.4066} \\
 &= \frac{1}{1.4066} \\
 &= 0.711
 \end{aligned}$$

Since $0.711 > 0.5$, classify as **Positive (Class 1)**.

Step 3: Calculate for Point B

Linear combination:

$$z_B = 0.5(0) + 0.8(0) + (-1.2) = -1.2$$

Apply sigmoid:

$$P(y = 1|x_B) = \frac{1}{1 + e^{1.2}} = \frac{1}{1 + 3.320} = 0.232$$

Since $0.232 < 0.5$, classify as **Negative (Class 0)**.

Step 4: Calculate for Point C

Linear combination:

$$z_C = 0.5(3) + 0.8(1) + (-1.2) = 1.5 + 0.8 - 1.2 = 1.1$$

Apply sigmoid:

$$P(y = 1|x_C) = \frac{1}{1 + e^{-1.1}} = \frac{1}{1.333} = 0.750$$

Since $0.750 > 0.5$, classify as **Positive (Class 1)**.

Summary:

Point	z	$P(y = 1 x)$	Classification
A	0.9	0.711	Positive
B	-1.2	0.232	Negative
C	1.1	0.750	Positive

1.1.2 Practice Exercises

Exercise

[★] A logistic regression model has parameters $w_1 = 1.5$, $w_2 = -0.5$, $b = 0$.

(a) Calculate $P(y = 1|x)$ for point $(2, 3)$.

(b) What is the classification if threshold is 0.5?

(c) If we lower the threshold to 0.3, does the classification change?

Hint

First compute $z = w^T x + b$, then apply $\sigma(z) = \frac{1}{1+e^{-z}}$. Remember that lowering the threshold makes the model more likely to predict positive.

(a) Calculate the probability:

$$z = 1.5(2) + (-0.5)(3) + 0 = 3 - 1.5 = 1.5$$

$$P(y = 1|x) = \frac{1}{1 + e^{-1.5}} = \frac{1}{1 + 0.223} = 0.818$$

(b) Since $0.818 > 0.5$, classify as **Positive**.

(c) Since $0.818 > 0.3$ as well, classification remains **Positive**. The lower threshold doesn't change this particular classification because the probability is already quite high.

Exercise

[**] You're building a medical diagnosis system where:

- True disease prevalence is 5%
- Your logistic regression model predicts $P(disease) = 0.45$ for a patient

(a) Using threshold 0.5, what would you predict?

(b) The cost of missing a disease (false negative) is \$100,000, while the cost of a false alarm (false positive) is \$1,000. Should you adjust your threshold? If so, in which direction?

(c) What probability threshold would make the expected costs equal for this patient?

Hint

Consider the asymmetric costs. When false negatives are much more expensive than false positives, you should lower the threshold to catch more diseases, accepting more false alarms as a tradeoff.

(a) With threshold 0.5 and $P(disease) = 0.45 < 0.5$, predict **healthy**.

(b) **Yes, lower the threshold.** Since missing a disease costs 100× more than a false alarm (\$100k vs \$1k), we should be more conservative and predict disease even at lower probabilities.

(c) Let p be the probability threshold. Expected costs:

- Predict disease (when $P \geq p$): Cost if wrong = $(1 - P) \times \$1,000$
- Predict healthy (when $P < p$): Cost if wrong = $P \times \$100,000$

At optimal threshold, these should be equal:

$$(1 - p) \times 1,000 = p \times 100,000$$

$$1,000 - 1,000p = 100,000p$$

$$1,000 = 101,000p$$

$$p = \frac{1,000}{101,000} = 0.0099 \approx 0.01$$

The optimal threshold is approximately **1%**. This makes sense: given the 100:1 cost ratio, we should predict disease if there's even a 1% chance.

For our patient with $P = 0.45$, we would definitely predict disease using this threshold.

Exercise

[★ ★ ★] Derive the gradient of the logistic regression loss function with respect to weights.

Given:

- Loss: $L(w, b) = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$
- Prediction: $\hat{y}_i = \sigma(w^T x_i + b)$
- Sigmoid: $\sigma(z) = \frac{1}{1+e^{-z}}$

Show that $\frac{\partial L}{\partial w} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) x_i$.

Hint

Use the chain rule: $\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial w}$.
Remember that $\frac{d}{dz} \sigma(z) = \sigma(z)(1 - \sigma(z))$.

Step 1: Gradient with respect to prediction

For a single example:

$$\frac{\partial}{\partial \hat{y}} [-y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})] = -\frac{y}{\hat{y}} + \frac{1 - y}{1 - \hat{y}}$$

Step 2: Gradient of sigmoid

Let $z = w^T x + b$, then $\hat{y} = \sigma(z)$:

$$\frac{\partial \sigma}{\partial z} = \sigma(z)(1 - \sigma(z)) = \hat{y}(1 - \hat{y})$$

Step 3: Gradient with respect to z

Chain rule:

$$\begin{aligned} \frac{\partial L}{\partial z} &= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \\ &= \left(-\frac{y}{\hat{y}} + \frac{1 - y}{1 - \hat{y}} \right) \cdot \hat{y}(1 - \hat{y}) \\ &= -y(1 - \hat{y}) + (1 - y)\hat{y} \\ &= -y + y\hat{y} + \hat{y} - y\hat{y} \\ &= \hat{y} - y \end{aligned}$$

Step 4: Gradient with respect to w

Since $z = w^T x + b$ and $\frac{\partial z}{\partial w} = x$:

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w} = (\hat{y} - y) \cdot x$$

Step 5: Average over all samples