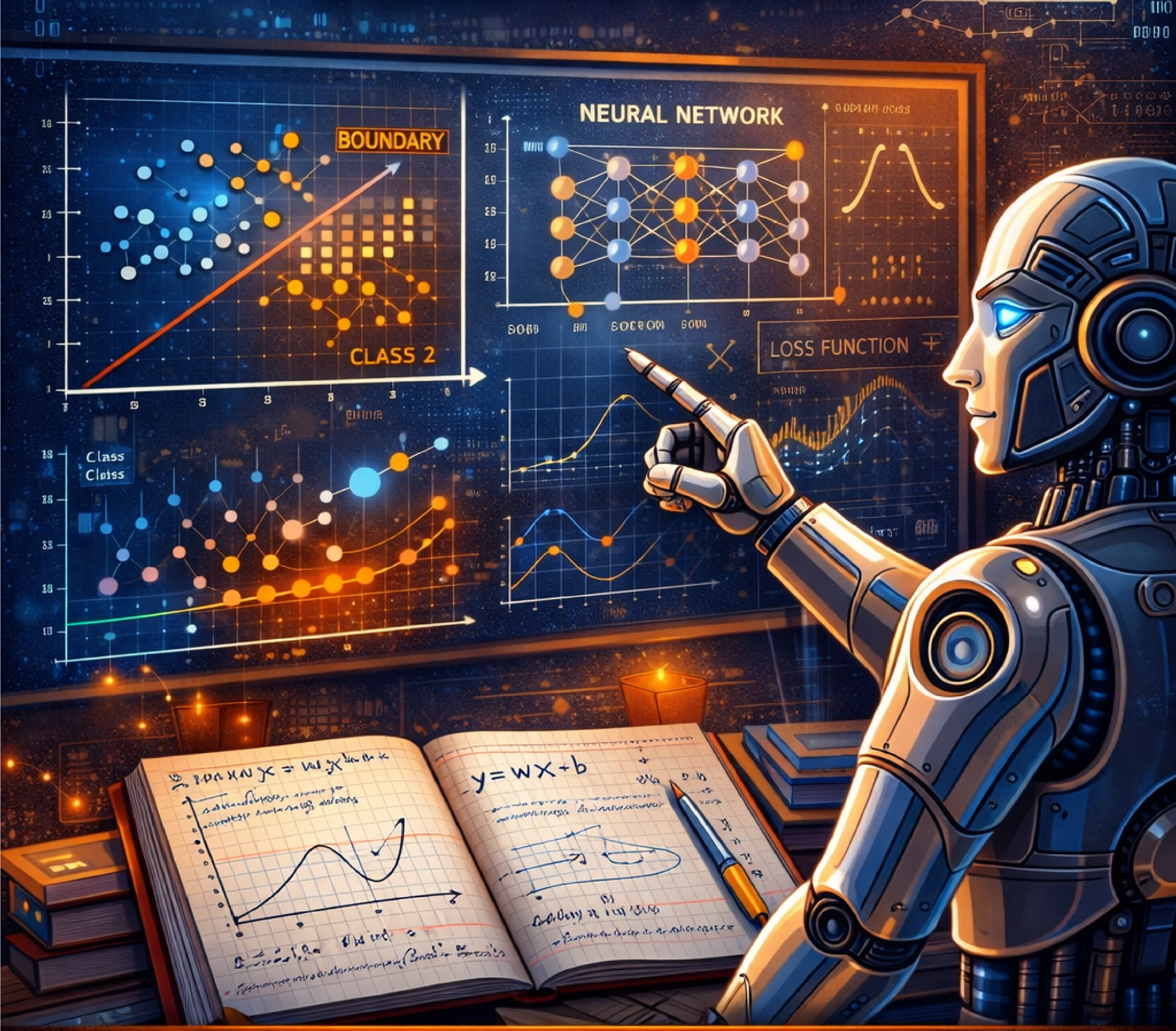
MACHINE LEARNING ALGORITHMS

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 Kołek

02-2026

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

- 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 Kołek
September 2026

Contents

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

- $\star$ **Basic:** Tests fundamental understanding
- $\star\star$ **Intermediate:** Requires synthesis of concepts
- $\star\star\star$ **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

[$\star$] 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

[$\star\star$] 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

[$\star\star\star$] 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