
aracne

Classificazione Decimale Dewey

— 006.30151 (23.) INTELLIGENZA ARTIFICIALE. Matematica

Thema

— Soggetto: UYQ. Intelligenza artificiale

— Qualificatori: 4CT. Per l'istruzione superiore/terziaria/universitaria

GERARDO IOVANE

MATHEMATICAL METHODS FOR ARTIFICIAL INTELLIGENCE

EXERCISES AND SOLUTIONS

aracne

ISBN
979-12-218-2689-0

PRIMA EDIZIONE
ROMA 30 APRILE 2026

Table of Contents

INTRODUCTION

This volume, Mathematical Methods for Artificial Intelligence: Exercises and Solutions, serves as a natural practical extension of the theoretical text Mathematical Methods for Artificial Intelligence. While the latter provides a systematic and rigorous treatment of the mathematical foundations of artificial intelligence, this book aims to transform those foundations into practical skills through a structured series of exercises and solutions.

Modern artificial intelligence is inherently interdisciplinary, but its core remains deeply mathematical. Concepts such as vector spaces, probability distributions, loss functions, and optimization algorithms are not merely technical tools, but constitute the very language through which intelligent systems are designed, analyzed, and understood. This textbook emphasizes this theoretical dimension, demonstrating how every learning model can be traced back to a precise mathematical formulation.

However, a full understanding of these concepts requires a fundamental step: application. It is precisely in this context that this volume is situated. The exercises provided are not mere assessment tools, but represent an active extension of the learning process. They guide the reader in the direct exploration of the mathematical properties of the models, in the derivation of fundamental results, and in a deep understanding of the mechanisms governing the operation of the algorithms.

The structure of the book closely mirrors that of the main text. Each chapter corresponds to a specific thematic area—from mathematical foundations to advanced deep learning and general AI models—and is divided into two parts: one dedicated to exercises and one to detailed solutions. This organization allows the reader to progressively tackle problems of increasing difficulty, consolidating acquired knowledge and developing independent reasoning skills.

A distinctive feature of the volume is the emphasis on the completeness of the solutions. Each exercise is solved step by step, with clear explanations of the assumptions, logical steps, and techniques used. This approach not only facilitates understanding but also fosters the development of a rigorous problem-solving methodology, which is essential for those working in the field of artificial intelligence.

The book is aimed at a broad but demanding audience: undergraduate students, graduate students, researchers, and professionals. For the former, it serves as a tool for consolidating and testing knowledge; for the latter, it provides a practical reference for exploring specific aspects and testing new ideas. In both cases, the book aims to foster not a superficial but a structured and critical understanding of the mathematical methods of AI.

Ultimately, this exercise book serves as a bridge between theory and practice. It transforms abstract knowledge into practical skills, enabling the reader not only to understand the models but also to use, analyze, and, where possible, improve them. In a constantly evolving field such as artificial intelligence, this ability is essential for addressing current and future challenges.

1. THE MATHEMATICAL LANGUAGE OF ARTIFICIAL INTELLIGENCE

This chapter provides exercises and complete solutions for Chapter 1 of the main text. Part A contains the exercise statements. Part B provides detailed, step-by-step solutions with full mathematical justification.

PART A: EXERCISES

1.1. Exercise

Prove that the function $f(x) = \|x\|_2^2$ is convex on $\mathbb{R}^d$ by verifying the definition directly.

1.2. Exercise

Show that if $f: \mathbb{R}^d \to \mathbb{R}$ is convex and differentiable, then $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$ for all $x, y \in \mathbb{R}^d$.

1.3. Exercise

Let $f: \mathbb{R}^d \to \mathbb{R}$ be twice continuously differentiable. Prove that if $H_f(x)$ is positive definite for all x, then f is strictly convex.

1.4. Exercise

Consider the composition $h = g \circ f$ where $f: \mathbb{R} \to \mathbb{R}^2$ is defined by $f(t) = \big(\cos(t), \sin(t)\big)$ and $g: \mathbb{R}^2 \to \mathbb{R}$ by $g(x_1, x_2) = x_1^2 + x_2^2$. Compute $J_h(t)$ using the chain rule and verify by direct computation.

1.5. Exercise

Let $\sigma(z) = \frac{1}{1+e^{-z}}$ be the logistic sigmoid. Show that σ is Lipschitz continuous and compute its Lipschitz constant.

1.6. Exercise

Prove that for any norm $\|\cdot\|$ on $\mathbb{R}^d$, the function $f(x) = \|x\|$ is convex.

1.7. Exercise

Consider the neural network layer $f(x) = \sigma(Wx + b)$ where σ is applied element-wise. Derive the formula for the Jacobian $J_f(x)$ in terms of W and the diagonal matrix $\mathrm{Diag}\big(\sigma'(Wx + b)\big)$.

1.8. Exercise

Show that if $f, g: \mathbb{R}^d \to \mathbb{R}$ are convex, then their pointwise maximum $h(x) = \max\big(f(x), g(x)\big)$ is also convex.

1.9. Exercise

Let $f: \mathbb{R}^d \to \mathbb{R}$ be differentiable and L-Lipschitz continuous. Prove that $\|\nabla f(x)\|_2 \leq L$ for all x.

1.10. Exercise

Verify that the empirical risk $R_{\text{emp}}(h; S)$ is an unbiased estimator of the true risk $R(h)$ when samples are drawn i.i.d. from D.

PART B: DETAILED SOLUTIONS

1.11. Convexity of squared Euclidean norm

Problem: Prove that $f(x) = \|x\|_2^2$ is convex on $\mathbb{R}^d$.

Strategy: We verify the definition of convexity directly: for all $x, y \in \mathbb{R}^d$ and $\lambda \in [0,1]$, we must show $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$.

Solution:

First, expand the left-hand side using the definition of the Euclidean norm squared:

$$f(\lambda x + (1-\lambda)y) = \|\lambda x + (1-\lambda)y\|_2^2$$

By definition, $\|v\|_2^2 = \langle v, v \rangle$, so:

$$= \langle \lambda x + (1-\lambda)y, \lambda x + (1-\lambda)y \rangle$$

Using bilinearity of the inner product:

$$= \langle \lambda x, \lambda x \rangle + 2\langle \lambda x, (1-\lambda)y \rangle + \langle (1-\lambda)y, (1-\lambda)y \rangle$$
$$= \lambda^2 \langle x, x \rangle + 2\lambda(1-\lambda)\langle x, y \rangle + (1-\lambda)^2 \langle y, y \rangle$$
$$= \lambda^2 \|x\|_2^2 + 2\lambda(1-\lambda)\langle x, y \rangle + (1-\lambda)^2 \|y\|_2^2$$

Now expand the right-hand side:

$$\lambda f(x) + (1-\lambda)f(y) = \lambda \|x\|_2^2 + (1-\lambda)\|y\|_2^2$$

To prove convexity, we need:

$$\lambda^2 \|x\|_2^2 + 2\lambda(1-\lambda)\langle x, y \rangle + (1-\lambda)^2 \|y\|_2^2 \leq \lambda \|x\|_2^2 + (1-\lambda)\|y\|_2^2$$

Rearranging:

$$\lambda^2 \|x\|_2^2 - \lambda \|x\|_2^2 + (1-\lambda)^2 \|y\|_2^2 - (1-\lambda)\|y\|_2^2 + 2\lambda(1-\lambda)\langle x, y \rangle \leq 0$$

Factor out common terms:

$$\lambda(\lambda - 1)\|x\|_2^2 + (1-\lambda)\big((1-\lambda) - 1\big)\|y\|_2^2 + 2\lambda(1-\lambda)\langle x, y \rangle \leq 0$$
$$-\lambda(1-\lambda)\|x\|_2^2 - \lambda(1-\lambda)\|y\|_2^2 + 2\lambda(1-\lambda)\langle x, y \rangle \leq 0$$

Factor out $\lambda(1-\lambda)$:

$$\lambda(1-\lambda)(-\|x\|_2^2 - \|y\|_2^2 + 2\langle x, y \rangle) \leq 0$$
$$-\lambda(1-\lambda)(\|x\|_2^2 + \|y\|_2^2 - 2\langle x, y \rangle) \leq 0$$
$$-\lambda(1-\lambda)\|x - y\|_2^2 \leq 0$$

Since $\lambda \in [0,1]$, we have $\lambda(1-\lambda) \geq 0$. Moreover, $\|x-y\|_2^2 \geq 0$ by definition of norms. Therefore:

$$-\lambda(1-\lambda)\|x - y\|_2^2 \leq 0$$

This inequality is always true, proving convexity.

Key insight: The proof reduces to showing that $\lambda(1-\lambda)\|x-y\|_2^2 \geq 0$, which is obvious from the properties of norms and the fact that $\lambda(1-\lambda) \geq 0$ for $\lambda \in [0,1]$.

1.12. First-order characterization of convexity

Problem: If $f: \mathbb{R}^d \to \mathbb{R}$ is convex and differentiable, show $f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$.

Strategy: Use the definition of convexity with a point interpolating between x and y, then take a limit.

Solution:

By convexity, for any $\lambda \in [0,1]$:

$$f(\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda) f(x)$$

Rearranging:

$$f(\lambda y + (1-\lambda)x) - f(x) \leq \lambda f(y) + (1-\lambda) f(x) - f(x)$$

$$f(\lambda y + (1-\lambda)x) - f(x) \leq \lambda f(y) - \lambda f(x)$$

$$f(\lambda y + (1-\lambda)x) - f(x) \leq \lambda \big(f(y) - f(x)\big)$$

For $\lambda > 0$, divide both sides by λ:

$$\frac{f(\lambda y + (1-\lambda)x) - f(x)}{\lambda} \leq f(y) - f(x)$$

Now observe that $\lambda y + (1-\lambda)x = x + \lambda(y-x)$. The left-hand side becomes:

$$\frac{f\big(x + \lambda(y-x)\big) - f(x)}{\lambda}$$

This is a difference quotient along the direction $y-x$. Since f is differentiable, taking the limit as $\lambda \to 0^+$:

$$\lim_{\lambda \to 0^+} \frac{f\big(x + \lambda(y-x)\big) - f(x)}{\lambda} = \langle \nabla f(x), y-x \rangle$$

This is the directional derivative in direction $y-x$.

Since the inequality $\frac{f(x+\lambda(y-x))-f(x)}{\lambda} \leq f(y) - f(x)$ holds for all $\lambda \in (0,1]$, it holds in the limit:

$$\langle \nabla f(x), y - x \rangle \leq f(y) - f(x)$$

Rearranging:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

This is the first-order characterization of convexity.

Geometric interpretation: The inequality states that for convex functions, the linear approximation (tangent hyperplane) at any point x lies below the function everywhere. This is equivalent to the definition of convexity.

1.13. Positive definite Hessian implies strict convexity

Problem: If $H_f(x)$ is positive definite for all x, prove f is strictly convex.

Strategy: Use Taylor's theorem to relate function values at different points, then exploit positive definiteness.

Solution:

For f twice continuously differentiable, Taylor's theorem with Lagrange remainder gives:

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\langle y - x, H_f(\xi)(y - x) \rangle$$

for some ξ on the line segment between x and y.

Now consider $z = \lambda y + (1 - \lambda)x$ for $\lambda \in (0,1)$. We want to show:

$$f(z) < \lambda f(y) + (1 - \lambda) f(x)$$

for $x \neq y$ (strict inequality defines strict convexity).

Using Taylor expansion around x:

$$f(z) = f(x) + \langle \nabla f(x), z - x \rangle + \frac{1}{2}\langle z - x, H_f(\xi_1)(z - x) \rangle$$

where ξ_1 is between x and z. Since $z - x = \lambda(y - x)$:

$$f(z) = f(x) + \lambda \langle \nabla f(x), y - x \rangle + \frac{\lambda^2}{2}\langle y - x, H_f(\xi_1)(y - x) \rangle$$

Similarly, Taylor expansion for $f(y)$ around x:

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\langle y - x, H_f(\xi_2)(y - x) \rangle$$

for some ξ_2 between x and y.

Now compute:

$$\lambda f(y) + (1-\lambda)f(x) = \lambda\left[f(x) + \langle \nabla f(x), y-x\rangle + \frac{1}{2}\langle y-x, H_f(\xi_2)(y-x)\rangle\right] + (1-\lambda)f(x)$$

$$= f(x) + \lambda\langle \nabla f(x), y-x\rangle + \frac{\lambda}{2}\langle y-x, H_f(\xi_2)(y-x)\rangle$$

We need to show:

$$f(z) < \lambda f(y) + (1-\lambda)f(x)$$

This reduces to:

$$\frac{\lambda^2}{2}\langle y-x, H_f(\xi_1)(y-x)\rangle < \frac{\lambda}{2}\langle y-x, H_f(\xi_2)(y-x)\rangle$$

Since H_f is positive definite everywhere, for $x \neq y$:

$$\langle y-x, H_f(\xi)(y-x)\rangle > 0$$

for any ξ. Let $m = \min_{\xi\in[x,y]}\langle y-x, H_f(\xi)(y-x)\rangle / \|y-x\|^2 > 0$ (this exists by continuity and compactness).

Then:

$$\frac{\lambda^2}{2}\langle y-x, H_f(\xi_1)(y-x)\rangle \geq \frac{\lambda^2 m}{2}\|y-x\|^2$$

$$\frac{\lambda}{2}\langle y-x, H_f(\xi_2)(y-x)\rangle \geq \frac{\lambda m}{2}\|y-x\|^2$$

Since $0 < \lambda < 1$, we have $\lambda^2 < \lambda$. Therefore:

$$\frac{\lambda^2 m}{2}\left\|y-x\right\|^2 < \frac{\lambda m}{2}\left\|y-x\right\|^2$$

This implies:

$$\frac{\lambda^2}{2}\langle y-x, H_f(\xi_1)(y-x)\rangle < \frac{\lambda}{2}\langle y-x, H_f(\xi_2)(y-x)\rangle$$

proving strict convexity.

1.14. Chain rule for composition

Problem: For $h = g \circ f$ with $f(t) = (\cos t, \sin t)$ and $g(x_1, x_2) = x_1^2 + x_2^2$, compute $J_h(t)$.

Strategy: Apply the multivariate chain rule:

$$J_h = J_g\big(f(t)\big) \cdot J_f(t)$$

Solution:

First compute $J_f(t)$. Since $f: \mathbb{R} \to \mathbb{R}^2$:

$$f(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

The Jacobian is:

$$J_f(t) = \begin{pmatrix} \frac{d}{dt}[\cos t] \\ \frac{d}{dt}[\sin t] \end{pmatrix} = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}$$

This is a 2×1 matrix (column vector).

Next compute J_g at point (x_1, x_2). Since $g: \mathbb{R}^2 \to \mathbb{R}$:

$$g(x_1, x_2) = x_1^2 + x_2^2$$

The Jacobian (gradient transposed) is:

$$J_g(x_1, x_2) = \begin{pmatrix} \frac{\partial g}{\partial x_1} & \frac{\partial g}{\partial x_2} \end{pmatrix} = (2x_1 \quad 2x_2)$$

This is a 1×2 matrix (row vector).

At point $f(t) = (\cos t, \sin t)$:

$$J_g\big(f(t)\big) = (2\cos t \quad 2\sin t)$$

By the chain rule:

$$J_h(t) = J_g\big(f(t)\big) \cdot J_f(t)$$

$$= (2\cos t \quad 2\sin t) \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}$$

$$= 2\cos t \cdot (-\sin t) + 2\sin t \cdot \cos t$$

$$= -2\cos t \sin t + 2\sin t \cos t = 0$$

Verification by direct computation:

$$h(t) = g\big(f(t)\big) = g(\cos t, \sin t) = \cos^2 t + \sin^2 t = 1$$

Since $h(t) = 1$ is constant, $h'(t) = 0$, confirming our result.

Note: This exercise illustrates that the chain rule correctly captures the fact that h is constant along the unit circle parametrized by f.

1.15. Lipschitz continuity of sigmoid

Problem: Show $\sigma(z) = \frac{1}{1+e^{-z}}$ is Lipschitz continuous and find its constant.

Strategy: For differentiable functions, Lipschitz continuity with constant L is equivalent to $|\sigma'(z)| \leq L$ for all z.

Solution:

First compute $\sigma'(z)$. Using the chain rule:

$$\sigma'(z) = \frac{d}{dz}\left[\frac{1}{1+e^{-z}}\right] = \frac{d}{dz}[(1+e^{-z})^{-1}]$$

$$= -(1+e^{-z})^{-2} \cdot \frac{d}{dz}[1+e^{-z}]$$

$$= -(1+e^{-z})^{-2} \cdot (-e^{-z}) = \frac{e^{-z}}{(1+e^{-z})^2}$$

We can rewrite this using $\sigma(z)$. Note that:

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

Therefore:

$$\sigma'(z) = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}} = \sigma(z)\big(1-\sigma(z)\big)$$

Now find the maximum of $|\sigma'(z)|$. Since $\sigma(z) \in (0,1)$ for all $z \in \mathbb{R}$ and $\sigma'(z) = \sigma(z)\big(1-\sigma(z)\big) > 0$, we have:

$$|\sigma'(z)| = \sigma(z)\big(1-\sigma(z)\big)$$

This is a product that we want to maximize. Consider the function $g(s) = s(1-s)$ for $s \in [0,1]$:

$$g'(s) = 1 - 2s$$

Setting $g'(s) = 0$ gives $s = 1/2$. The second derivative is $g''(s) = -2 < 0$, confirming this is a maximum.

At $s = 1/2$:

$$g(1/2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Now find where $\sigma(z) = 1/2$:

$$\frac{1}{1+e^{-z}} = \frac{1}{2} 1 + e^{-z} = 2e^{-z} = 1 z = 0$$

At $z = 0$:

$$\sigma'(0) = \sigma(0)\big(1 - \sigma(0)\big) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

.

As $z \to \pm\infty$, $\sigma'(z) \to 0$. Therefore, the maximum of $|\sigma'(z)|$ over all $z \in \mathbb{R}$ is $1/4$.

By the mean value theorem, for any z_1, z_2:

$$|\sigma(z_1) - \sigma(z_2)| = |\sigma'(\xi)||z_1 - z_2|$$

for some ξ between z_1 and z_2. Since $|\sigma'(\xi)| \leq 1/4$:

$$|\sigma(z_1) - \sigma(z_2)| \leq \frac{1}{4}|z_1 - z_2|$$

Therefore, σ is Lipschitz continuous with Lipschitz constant $L = 1/4$.

1.16. Convexity of norms

Problem: Prove $f(x) = \|x\|$ is convex for any norm.

Strategy: Use the defining properties of norms, particularly the triangle inequality and positive homogeneity.

Solution:

Let $\|\cdot\|$ be any norm on $\mathbb{R}^d$. We must show that for $x, y \in \mathbb{R}^d$ and $\lambda \in [0,1]$:

$$\|\lambda x + (1 - \lambda)y\| \leq \lambda\|x\| + (1 - \lambda)\|y\|$$

By the triangle inequality:

$$\|\lambda x + (1 - \lambda)y\| \leq \|\lambda x\| + \|(1 - \lambda)y\|$$

By positive homogeneity of norms ($\|\alpha v\| = |\alpha|\|v\|$ for all $\alpha \in \mathbb{R}$, $v \in \mathbb{R}^d$):

$$\|\lambda x\| = |\lambda|\|x\|\|(1 - \lambda)y\| = |1 - \lambda|\|y\|$$

Since $\lambda \in [0,1]$, we have $\lambda \geq 0$ and $1 - \lambda \geq 0$, so $|\lambda| = \lambda$ and $|1 - \lambda| = 1 - \lambda$.

Therefore:

$$\|\lambda x\| = \lambda\|x\|\|(1 - \lambda)y\| = (1 - \lambda)\|y\|$$

Substituting back:

$$\|\lambda x + (1 - \lambda)y\| \leq \lambda\|x\| + (1 - \lambda)\|y\|$$

This proves convexity.

Key insight: This proof works for any norm because it only uses the triangle inequality and positive homogeneity, which are axioms satisfied by all norms. Thus, ℓ^1, ℓ^2, ℓ^∞, and any other norm yield a convex function.

1.17. Jacobian of neural network layer

Problem: For $f(x) = \sigma(Wx + b)$ with element-wise σ, derive $J_f(x)$.

Strategy: Use the chain rule, treating $z = Wx + b$ as an intermediate variable.

Solution:

Let $z = Wx + b \in \mathbb{R}^m$ where $W \in \mathbb{R}^{m \times d}$, $x \in \mathbb{R}^d$, $b \in \mathbb{R}^m$.

The function f can be decomposed as:

$$f(x) = \sigma\big(z(x)\big)$$

where $z(x) = Wx + b$ and σ is applied element-wise.

The Jacobian $J_f(x) \in \mathbb{R}^{m \times d}$ has entries:

$$[J_f(x)]_{ij} = \frac{\partial f_i(x)}{\partial x_j}$$

By the chain rule:

$$\frac{\partial f_i(x)}{\partial x_j} = \frac{\partial \sigma(z_i)}{\partial z_i} \cdot \frac{\partial z_i}{\partial x_j}$$

where $f_i(x) = \sigma(z_i)$.

First term: Since σ is applied element-wise:

$$\frac{\partial \sigma(z_i)}{\partial z_i} = \sigma'(z_i)$$

Second term: Since $z_i = \sum_{k=1}^{d} W_{ik} x_k + b_i$:

$$\frac{\partial z_i}{\partial x_j} = W_{ij}$$

Therefore:

$$\frac{\partial f_i(x)}{\partial x_j} = \sigma'(z_i) W_{ij}$$

In matrix form, define $D = \mathrm{Diag}\big(\sigma'(z)\big) \in \mathbb{R}^{m \times m}$, a diagonal matrix:

$$D = \begin{pmatrix} \sigma'(z_1) & 0 & \cdots & 0 \\ 0 & \sigma'(z_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma'(z_m) \end{pmatrix}$$

Then: