Xian-Da Zhang

A Matrix Algebra Approach to Artificial Intelligence

Springer

A Matrix Algebra Approach to Artificial Intelligence

Xian-Da Zhang

A Matrix Algebra Approach to Artificial Intelligence

Springer

Xian-Da Zhang (Deceased)
Department of Automation
Tsinghua University
Beijing, Beijing, China

ISBN 978-981-15-2769-2 ISBN 978-981-15-2770-8 (eBook)
https://doi.org/10.1007/978-981-15-2770-8

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Human intelligence is the intellectual prowess of humans, which is marked by four basic and important abilities: learning ability, cognition (acquiring and storing knowledge) ability, generalization ability, and computation ability. Correspondingly, artificial intelligence (AI) also consists of four basic and important methods: machine learning (learning intelligence), neural network (cognitive intelligence), support vector machines (generalization intelligence), and evolutionary computation (computational intelligence).

The development of AI is built on mathematics. For example, multivariant calculus deals with the aspect of numerical optimization, which is the driving force behind most machine learning algorithms. The main math applications in AI are matrix algebra, optimization, and mathematical statistics, but the latter two are usually described and applied in the form of matrix. Therefore, matrix algebra is a vast mathematical tool of fundamental importance in most AI subjects.

The aim of this book is to provide the solid matrix algebra theory and methods for four basic and important AI fields, including machine learning, neural networks, support vector machines, and evolutionary computation.

Structure and Contents

The book consists of two parts.

Part I (Introduction to Matrix Algebra) provides fundamentals of matrix algebra and contains Chaps. 1 through 5. Chapter 1 presents the basic operations and performances of matrices, followed by a description of vectorization of matrix and matricization of vector. Chapter 2 is devoted to matrix differential as an important and effective tool in gradient computation and optimization. Chapter 3 is concerned with convex optimization theory and methods by focusing on gradient/subgradient methods in smooth and nonsmooth convex optimizations, and constrained convex optimization. Chapter 4 describes singular value decomposition (SVD) together with Tikhonov regularization and total least squares for solving over-determined matrix equations, followed by the Lasso and LARS methods for solving underdetermined matrix equations. Chapter 5 is devoted to the eigenvalue decomposition (EVD), the generalized eigenvalue decomposition, the Rayleigh quotient, and the generalized Rayleigh quotient.

Part II (Artificial Intelligence) focuses on machine learning, neural networks, support vector machines (SVMs), and evolutionary computation from the perspective of matrix algebra. This part is the main body of the book and consists of the following four chapters.

Chapter 6 (Machine Learning) presents first the basic theory and methods in machine learning including single-objective optimization, feature selection, principal component analysis and canonical correlation analysis together with supervised, unsupervised, and semi-supervised learning and active learning. Then, this chapter highlights topics and advances in machine learning: graph machine learning, reinforcement learning, Q-learning, and transfer learning.

Chapter 7 (Neural Networks) describes optimization problem, activation functions, and basic neural networks. The core part of this chapter are topics and advances in neural networks: convolutional neural networks (CNNs), dropout learning, autoencoders, extreme learning machine (ELM), graph embedding, network embedding, graph neural networks (GNNs), batch normalization networks, and generative adversarial networks (GANs).

Chapter 8 (Support Vector Machines) discusses the support vector machine regression and classification, and the relevance vector machine.

Chapter 9 (Evolutionary Computation) is concerned primarily with multiobjective optimization, multiobjective simulated annealing, multiobjective genetic algorithms, multiobjective evolutionary algorithms, evolutionary programming, differential evolution together with ant colony optimization, artificial bee colony algorithms, and particle swarm optimization. In particular, this chapter highlights also topics and advances in evolutionary computation: Pareto optimization theory, noisy multiobjective optimization, and opposition-based evolutionary computation.

Part I uses some revised content and materials from my book *Matrix Analysis and Applications* (Cambridge University Press, 2017), but there are considerable differences in content and book objective. This book is concentrated on the applications of the matrix algebra approaches in AI in Part II (561 pages of text), compared to Part I which is only 205 pages of text. This book is also related to my previous book *Linear Algebra in Signal Processing* (in Chinese, Science Press, Beijing, 1997; Japanese translation, Morikita Press, Tokyo, 2008) in some ideas.

Features and Contributions

- The first book on matrix algebra methods and applications in artificial intelligence.
- Introduces the machine learning tree, the neural network tree, and the evolutionary tree.

- Presents the solid matrix algebra theory and methods for four core AI areas: Machine Learning, Neural Networks, Support Vector Machines, and Evolutionary Computation.
- Highlights selected topics and advances of machine learning, neural networks, and evolutionary computation.
- Summarizes about 80 AI algorithms so that readers can further understand and implement AI methods.

Audience

This book is widely suitable for scientists, engineers, and graduate students in many disciplines, including but not limited to artificial intelligence, computer science, mathematics, engineering, etc.

Acknowledgments

I want to thank more than 40 of my former doctoral students and more than 20 graduate students for their cooperative research in intelligent signal and information processing and pattern recognition.

I am grateful to the several anonymous internal and external experts invited by Springer for their comments and suggestions. I am very grateful to my editor, Dr. Celine Lanlan Chang, Executive Editor, Computer Science, for her patience, understanding, and help in the course of my writing this book.

Finally, I would also like to thank my wife Xiao-Ying Tang for her consistent understanding and support for my work, teaching, and research in the past near 50 years.

Beijing, China Xian-Da Zhang
November, 2019

A Note from the Family of Dr. Zhang

It is with a heavy heart we share with you that the author of this book, our beloved father, passed away before this book was published. We want to share with you some inspirational thoughts about his life's journey and how passionate he was about learning and teaching.

Our father endured many challenges in his life. During his high school years, the Great Chinese Famine occurred, but hunger could not deter him from studying. During his college years, our father's family was facing poverty, so he sought to help alleviate this by undergoing his studies at a military institution, where free meals were provided. Not long after, his tenacity to learn would be tested again, as the Cultural Revolution started, closing all the universities in China. He was assigned to work in a remote factory, but his perseverance to learn endured as he continued his education by studying hard secretly during off-hours.

After all the universities were reopened, our father left us to continue his study in a different city. He obtained his PhD degree at the age of 41 and became a professor at Tsinghua University in 1992.

Our father has taught and mentored many students both professionally and personally throughout his life. He is even more passionate in sharing his ideas and knowledge through writing as he believes that books, with its greater reach, will benefit many more people. He had been working for more than 12 h a day before he was admitted into the hospital. He planned to take a break after he finishes three books this year. Unfortunately, he could not handle such a heavy workload that he asked our help in editing this book in our last conversation.

Our father has lived a life with purpose. He discovered his great passion when he was young: learning and teaching. For him, it was even something to die for. He told our mom once that he would rather live fewer years to produce a high-quality book. Despite the numerous challenges and hardships he faced throughout his life, he never stopped learning and teaching. He self-studied Artificial Intelligence in the past few years and completed this book before his final days.

We sincerely hope you will enjoy reading his final work, as we believe that he would undoubtedly have been very happy to inform and teach and share with you his latest learning.

Fremont, CA, USA Yewei Zhang and Wei Wei (Daughter and Son In Law)
Philadelphia, PA, USA Zhang and John Zhang (Son and Grand Son)
April, 2020

Contents

List of Notations

$\forall$	For all
$\mid$	Such that
$\ni$	Such that
$\exists$	There exists
$\nexists$	There does not exist
$\wedge$	Logical AND
$\vee$	Logical OR
$\lvert A \rvert$	Cardinality of a set A
$A \Rightarrow B$	"condition A results in B" or "A implies B"
$A \subseteq B$	A is a subset of B
$A \subset B$	A is a proper subset of B
$A = B$	Sets $A = B$
$A \cup B$	Union of sets A and B
$A \cap B$	Intersection of sets A and B
$A \cap B = \emptyset$	Sets A and B are disjoint
$A + B$	Sum set of sets A and B
$A - B$	The set of elements of A that are not in B
$X \setminus A$	Complement of the set A in the set X
$A \succ B$	Set A dominates set B: every $\mathbf{f}(\mathbf{x}_2) \in B$ is dominated by at least one $\mathbf{f}(\mathbf{x}_1) \in A$ such that $\mathbf{f}(\mathbf{x}_1) <_{IN} \mathbf{f}(\mathbf{x}_2)$ (for minimization) or $\mathbf{f}(\mathbf{x}_1) >_{IN} \mathbf{f}(\mathbf{x}_2)$ (for maximization)
$A \succeq B$	A weakly dominates B: for every $\mathbf{f}(\mathbf{x}_2) \in B$ and at least one $\mathbf{f}(\mathbf{x}_1) \in A$ $\mathbf{f}(\mathbf{x}_1) \leq_{IN} \mathbf{f}(\mathbf{x}_2)$ (for minimization) or $\mathbf{f}(\mathbf{x}_1) \geq_{IN} \mathbf{f}(\mathbf{x}_2)$ (for maximization)
$A \succ\succ B$	A strictly dominates B: every $\mathbf{f}(\mathbf{x}_2) \in B$ is strictly dominated by at least one $\mathbf{f}(\mathbf{x}_1) \in A$ such that $f_i(\mathbf{x}_1) <_{IN} f_i(\mathbf{x}_2), \forall i = \{1, \ldots, m\}$ (for minimization) or $f_i(\mathbf{x}_1) >_{IN} f_i(\mathbf{x}_2), \forall i = \{1, \ldots, m\}$ (for maximization)
$A \parallel B$	Sets A and B are incomparable: neither $A \succeq B$ nor $B \succeq A$
$A \rhd B$	A is better than B: every $\mathbf{f}(\mathbf{x}_2) \in B$ is weakly dominated by at least one $\mathbf{f}(\mathbf{x}_1) \in A$and $A \neq B$

$\mathrm{AGG}_{\mathrm{mean}}(z)$	Mean aggregate function of z
$\mathrm{AGG}_{\mathrm{LSTM}}(z)$	LSTM aggregate function of z
$\mathrm{AGG}_{\mathrm{pool}}(z)$	Pooling aggregate function of z
$\mathbb{C}$	Complex numbers
$\mathbb{C}^n$	Complex n-vector
$\mathbb{C}^{m\times n}$	Complex $m \times n$ matrix
$\mathbb{C}[x]$	Complex polynomial
$\mathbb{C}[x]^{m\times n}$	Complex $m \times n$ polynomial matrix
$\mathbb{C}^{I\times J\times K}$	Complex third-order tensors
$\mathbb{C}^{I_1\times\cdots\times I_N}$	Complex N-order tensor
$\mathbb{K}$	Real or complex number
$\mathbb{K}^n$	Real or complex n-vector
$\mathbb{K}^{m\times n}$	Real or complex $m \times n$ matrix
$\mathbb{K}^{I\times J\times K}$	Real or complex third-order tensor
$\mathbb{K}^{I_1\times\cdots\times I_N}$	Real or complex N-order tensor
$G(V, E, \mathbf{W})$	Graph with vertex set V, edge set E and adjacency matrix $\mathbf{W}$
$\mathcal{N}(v)$	Neighbors of a vertex (node) v
$\mathrm{PReLU}(z)$	Parametric rectified linear unit activation function of z
$\mathrm{ReLU}(z)$	Rectified linear unit activation function of z
$\mathbb{R}$	Real number
$\mathbb{R}^n$	Real n-vectors ($n \times 1$ real matrix)
$\mathbb{R}^{m\times n}$	Real $m \times n$ matrix
$\mathbb{R}[x]$	Real polynomial
$\mathbb{R}[x]^{m\times n}$	Real $m \times n$ polynomial matrix
$\mathbb{R}^{I\times J\times K}$	Real third-order tensors
$\mathbb{R}^{I_1\times\cdots\times I_N}$	Real N-order tensor
$\mathbb{R}_+$	Nonnegative real numbers, nonnegative orthant
$\mathbb{R}_{++}$	Positive real number
$\sigma(z)$	Sigmoid activation function of z
$\mathrm{softmax}(z)$	Softmax activation function of z
$\mathrm{softplus}(z)$	Softplus activation function of z
$\mathrm{softsign}(z)$	Softsign activation function of z
$\tanh(z)$	Tangent (tanh) hyperbolic activation function of z
$T : V \to W$	Mapping the vectors in V to corresponding vectors in W
$T^{-1} : W \to V$	Inverse mapping of the one-to-one mapping $T : V \to W$
$X_1 \times \cdots \times X_n$	Cartesian product of n sets $X_1, \ldots, X_n$
$\{(\mathbf{x}_i, y_i = +1)\}$	Set of training data vectors $\mathbf{x}_i$ belonging to the classes (+)
$\{(\mathbf{x}_i, y_i = -1)\}$	Set of training data vectors $\mathbf{x}_i$ belonging to the classes (−)
$\mathbf{1}_n$	n-dimensional summing vector with all entries 1
$\mathbf{0}_n$	n-dimensional zero vector with all zero entries
$\mathbf{e}_i$	Base vector whose ith entry equal to 1 and others being zero
$\mathbf{x} \sim N(\bar{\mathbf{x}}, \mathbf{\Gamma}_x)$	Gaussian vector with mean vector $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{\Gamma}_x$
$\|\mathbf{x}\|_0$	ℓ_0-norm: number of nonzero entries of vector $\mathbf{x}$
$\|\mathbf{x}\|_1$	ℓ_1-norm of vector $\mathbf{x}$
$\|\mathbf{x}\|_2$	Euclidean form of vector $\mathbf{x}$

$\|\mathbf{x}\|_p$	ℓ_p-norm or Hölder norm of vector $\mathbf{x}$
$\|\mathbf{x}\|_*$	Nuclear norm of vector $\mathbf{x}$
$\|\mathbf{x}\|_\infty$	ℓ_∞-norm of vector $\mathbf{x}$
$\langle \mathbf{x}, \mathbf{y} \rangle$	Inner product of vectors $\mathbf{x}$ and $\mathbf{y}$
$d(\mathbf{x}, \mathbf{y})$	Distance or dissimilarity between vectors $\mathbf{x}$ and $\mathbf{y}$
$N_\epsilon(\mathbf{x})$	ϵ-neighborhood of vector $\mathbf{x}$
$\rho(\mathbf{x}, \mathbf{y})$	Correlation coefficient between two random vectors $\mathbf{x}$ and $\mathbf{y}$
$\mathbf{x} \in A$	$\mathbf{x}$ belongs to the set A, i.e., $\mathbf{x}$ is an element or member of A
$\mathbf{x} \notin A$	$\mathbf{x}$ is not an element of the set A
$\mathbf{x} \circ \mathbf{y} = \mathbf{x}\mathbf{y}^H$	Outer product of vectors $\mathbf{x}$ and $\mathbf{y}$
$\mathbf{x} \perp \mathbf{y}$	Vector orthogonal
$\mathbf{x} > 0$	Positive vector with components $x_i > 0, \forall i$
$\mathbf{x} \geq 0$	Nonnegative vector with components $x_i \geq 0, \forall i$
$\mathbf{x} \geq \mathbf{y}$	Vector elementwise inequality $x_i \geq y_i, \forall i$
$\mathbf{x} \succ \mathbf{x}'$	$\mathbf{x}$ domains (or outperforms) $\mathbf{x}'$: $\mathbf{f}(\mathbf{x}) < \mathbf{f}(\mathbf{x}')$ for minimization
$\mathbf{x} \succ \mathbf{x}'$	$\mathbf{x}$ domains (or outperforms) $\mathbf{x}'$: $\mathbf{f}(\mathbf{x}) > \mathbf{f}(\mathbf{x}')$ for maximization
$\mathbf{x} \succeq \mathbf{x}'$	$\mathbf{x}$ weakly dominates $\mathbf{x}'$: $\mathbf{f}(\mathbf{x}) \leq \mathbf{f}(\mathbf{x}')$ for minimization
$\mathbf{x} \succeq \mathbf{x}'$	$\mathbf{x}$ weakly dominates $\mathbf{x}'$: $\mathbf{f}(\mathbf{x}) \geq \mathbf{f}(\mathbf{x}')$ for maximization
$\mathbf{x} \succ\succ \mathbf{x}'$	$\mathbf{x}$ strictly dominates $\mathbf{x}'$: $f_i(\mathbf{x}) < f_i(\mathbf{x}'), \forall\, i$ for minimization
$\mathbf{x} \succ\succ \mathbf{x}'$	$\mathbf{x}$ strictly dominates $\mathbf{x}'$: $f_i(\mathbf{x}) > f_i(\mathbf{x}'), \forall\, i$ for maximization
$\mathbf{x} \parallel \mathbf{x}'$	$\mathbf{x}$ and $\mathbf{x}'$ are incomparable, i.e., $\mathbf{x} \not\succeq \mathbf{x}' \wedge \mathbf{x}' \not\succeq \mathbf{x}$
$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}')$	$f_i(\mathbf{x}) = f_i(\mathbf{x}'),\ \forall i = 1, \ldots, m$
$\mathbf{f}(\mathbf{x}) \neq \mathbf{f}(\mathbf{x}')$	$f_i(\mathbf{x}) \neq f_i(\mathbf{x}')$, for at least one $i \in \{1, \ldots, m\}$
$\mathbf{f}(\mathbf{x}) \leq \mathbf{f}(\mathbf{x}')$	$f_i(\mathbf{x}) \leq f_i(\mathbf{x}'),\ \forall i = 1, \ldots, m$
$\mathbf{f}(\mathbf{x}) < \mathbf{f}(\mathbf{x}')$	$\forall\, i = 1, \ldots, m : f_i(\mathbf{x}) \leq f_i(\mathbf{x}') \ \wedge \exists j \in \{1, \ldots, m\} : f_j(\mathbf{x}) < f_j(\mathbf{x}')$
$\mathbf{f}(\mathbf{x}) \geq \mathbf{f}(\mathbf{x}')$	$f_i(\mathbf{x}) \geq f_i(\mathbf{x}'),\ \forall i = 1, \ldots, m$
$\mathbf{f}(\mathbf{x}) > \mathbf{f}(\mathbf{x}')$	$\forall\, i = 1, \ldots, m : f_i(\mathbf{x}) \geq f_i(\mathbf{x}') \ \wedge \exists j \in \{1, \ldots, m\} : f_j(\mathbf{x}) > f_j(\mathbf{x}')$
$\mathbf{f}(\mathbf{x}_1) <_{IN} \mathbf{f}(\mathbf{x}_2)$	Interval order relation: $\forall i = 1, \ldots, m : \underline{f}_i(\mathbf{x}_1) \leq \underline{f}_i(\mathbf{x}_2) \wedge \overline{f}_i(\mathbf{x}_1) \leq \overline{f}_i(\mathbf{x}_2) \wedge \exists j \in \{1, \ldots, m\} : \underline{f}_j(\mathbf{x}_1) \neq \underline{f}_j(\mathbf{x}_2) \vee \overline{f}_j(\mathbf{x}_1) \neq \overline{f}_j(\mathbf{x}_2)$
$\mathbf{f}(\mathbf{x}_1) >_{IN} \mathbf{f}(\mathbf{x}_2)$	Interval order relation: $\forall i = 1, \ldots, m : \underline{f}_i(\mathbf{x}_1) \geq \underline{f}_i(\mathbf{x}_2) \wedge \overline{f}_i(\mathbf{x}_1) \geq \overline{f}_i(\mathbf{x}_2) \wedge \exists j \in \{1, \ldots, m\} : \underline{f}_j(\mathbf{x}_1) \neq \underline{f}_j(\mathbf{x}_2) \vee \overline{f}_j(\mathbf{x}_1) \neq \overline{f}_j(\mathbf{x}_2)$
$\mathbf{f}(\mathbf{x}_1) \leq_{IN} \mathbf{f}(\mathbf{x}_2)$	Weak interval order relation: $\forall i \in \{1, \ldots, m\} : \underline{f}_i(\mathbf{x}_1) \leq \underline{f}_i(\mathbf{x}_2) \wedge \overline{f}_i(\mathbf{x}_1) \leq \overline{f}_i(\mathbf{x}_2)$
$\mathbf{f}(\mathbf{x}_1) \geq_{IN} \mathbf{f}(\mathbf{x}_2)$	Weak interval order relation: $\forall i \in \{1, \ldots, m\} : \underline{f}_i(\mathbf{x}_1) \geq \underline{f}_i(\mathbf{x}_2) \wedge \overline{f}_i(\mathbf{x}_1) \geq \overline{f}_i(\mathbf{x}_2)$
$\mathbf{A}^T$	Transpose of matrix $\mathbf{A}$
$\mathbf{A}^H$	Complex conjugate transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	Inverse of nonsingular matrix $\mathbf{A}$
$\mathbf{A}^\dagger$	Moore–Penrose inverse of matrix $\mathbf{A}$

$\mathbf{A}^*$	Conjugate of matrix $\mathbf{A}$
$\mathbf{A} \succ 0$	Positive definite matrix
$\mathbf{A} \succeq 0$	Positive semi-definite matrix
$\mathbf{A} \prec 0$	Negative definite matrix
$\mathbf{A} \preceq 0$	Negative semi-definite matrix
$\mathbf{A} > 0$	Positive (or elementwise positive) matrix
$\mathbf{A} \geq 0$	Nonnegative (or elementwise nonnegative) matrix
$\mathbf{A} \geq \mathbf{B}$	Matrix elementwise inequality $a_{ij} \geq b_{ij}, \forall i, j$
$\mathbf{I}_n$	$n \times n$ Identity matrix
$\mathbf{O}_n$	$n \times n$ Null matrix
$\|\mathbf{A}\|$	Determinant of matrix $\mathbf{A}$
$\\|\mathbf{A}\\|_1$	Maximum absolute column-sum norm of matrix $\mathbf{A}$
$\\|\mathbf{A}\\|_2 = \\|\mathbf{A}\\|_{\text{spec}}$	Spectrum norm of matrix $\mathbf{A}$
$\\|\mathbf{A}\\|_F$	Frobenius norm of matrix $\mathbf{A}$
$\\|\mathbf{A}\\|_\infty$	Max norm of $\mathbf{A}$: absolute maximum of all entries of $\mathbf{A}$
$\\|\mathbf{A}\\|_{\mathbf{G}}$	Mahalanobis norm of matrix $\mathbf{A}$
$\\|\mathbf{A}\\|_*$	Nuclear norm, called also the trace norm, of matrix $\mathbf{A}$
$\mathbf{A} \oplus \mathbf{B}$	Direct sum of an $m \times m$ matrix $\mathbf{A}$ and an $n \times n$ matrix $\mathbf{B}$
$\mathbf{A} \odot \mathbf{B}$	Hadamard product (or elementwise product) of $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{A} \oslash \mathbf{B}$	Elementwise division of matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$
$\langle \mathbf{A}, \mathbf{B} \rangle$	Inner (or dot) product of $\mathbf{A}$ and $\mathbf{B}$: $\langle \mathbf{A}, \mathbf{B} \rangle = \langle \text{vec}(\mathbf{A}), \text{vec}(\mathbf{B}) \rangle$
$\rho(\mathbf{A})$	Spectral radius of matrix $\mathbf{A}$
$\text{cond}(\mathbf{A})$	Condition number of matrix $\mathbf{A}$
$\text{diag}(\mathbf{A})$	Diagonal function of $\mathbf{A} = [a_{ij}]$: $\sum_{i=1}^n \|a_{ii}\|^2$
$\mathbf{Diag}(\mathbf{A})$	Diagonal matrix consisting of diagonal entries of $\mathbf{A}$
$\text{eig}(\mathbf{A})$	Eigenvalues of the Hermitian matrix $\mathbf{A}$
$\text{Gr}(n, r)$	Grassmann manifold
$\text{rvec}(\mathbf{A})$	Row vectorization of matrix $\mathbf{A}$
$\text{off}(\mathbf{A})$	Off function of $\mathbf{A} = [a_{ij}]$: $\sum_{i=1, i \neq j}^m \sum_{j=1}^n \|a_{ij}\|^2$
$\text{tr}(\mathbf{A})$	Trace of matrix $\mathbf{A}$
$\text{vec}(\mathbf{A})$	Vectorization of matrix $\mathbf{A}$

List of Figures

List of Tables

List of Algorithms