

Xian-Da Zhang

A Matrix Algebra Approach to Artificial Intelligence

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>

© Springer Nature Singapore Pte Ltd. 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Contents

Part I Introduction to Matrix Algebra

1	Basic Matrix Computation	3
1.1	Basic Concepts of Vectors and Matrices	3
1.1.1	Vectors and Matrices	3
1.1.2	Basic Vector Calculus	6
1.1.3	Basic Matrix Calculus	7
1.2	Sets and Linear Mapping	10
1.2.1	Sets	10
1.2.2	Linear Mapping	12
1.3	Norms	14
1.3.1	Vector Norms	14
1.3.2	Matrix Norms	18
1.4	Random Vectors	20
1.4.1	Statistical Interpretation of Random Vectors	20
1.4.2	Gaussian Random Vectors	24
1.5	Basic Performance of Matrices	27
1.5.1	Quadratic Forms	27
1.5.2	Determinants	27
1.5.3	Matrix Eigenvalues	30
1.5.4	Matrix Trace	32
1.5.5	Matrix Rank	33
1.6	Inverse Matrices and Moore–Penrose Inverse Matrices	35
1.6.1	Inverse Matrices	35
1.6.2	Left and Right Pseudo-Inverse Matrices	39
1.6.3	Moore–Penrose Inverse Matrices	40
1.7	Direct Sum and Hadamard Product	42
1.7.1	Direct Sum of Matrices	43
1.7.2	Hadamard Product	43

1.8	Kronecker Products	46
1.8.1	Definitions of Kronecker Products	46
1.8.2	Performance of Kronecker Products	47
1.9	Vectorization and Matricization	48
1.9.1	Vectorization and Commutation Matrix	48
1.9.2	Matricization of Vectors	50
	References	53
2	Matrix Differential	55
2.1	Jacobian Matrix and Gradient Matrix	55
2.1.1	Jacobian Matrix	55
2.1.2	Gradient Matrix	57
2.1.3	Calculation of Partial Derivative and Gradient	59
2.2	Real Matrix Differential	61
2.2.1	Calculation of Real Matrix Differential	62
2.2.2	Jacobian Matrix Identification	64
2.2.3	Jacobian Matrix of Real Matrix Functions	68
2.3	Complex Gradient Matrices	71
2.3.1	Holomorphic Function and Complex Partial Derivative ...	72
2.3.2	Complex Matrix Differential	75
2.3.3	Complex Gradient Matrix Identification	82
	References	87
3	Gradient and Optimization	89
3.1	Real Gradient	89
3.1.1	Stationary Points and Extreme Points	90
3.1.2	Real Gradient of $f(\mathbf{X})$	93
3.2	Gradient of Complex Variable Function	95
3.2.1	Extreme Point of Complex Variable Function	95
3.2.2	Complex Gradient	98
3.3	Convex Sets and Convex Function Identification	100
3.3.1	Standard Constrained Optimization Problems	100
3.3.2	Convex Sets and Convex Functions	102
3.3.3	Convex Function Identification	104
3.4	Gradient Methods for Smooth Convex Optimization	106
3.4.1	Gradient Method	106
3.4.2	Projected Gradient Method	108
3.4.3	Convergence Rates	110
3.5	Nesterov Optimal Gradient Method	112
3.5.1	Lipschitz Continuous Function	113
3.5.2	Nesterov Optimal Gradient Algorithms	115
3.6	Nonsmooth Convex Optimization	118
3.6.1	Subgradient and Subdifferential	119
3.6.2	Proximal Operator	123
3.6.3	Proximal Gradient Method	128

3.7	Constrained Convex Optimization	132
3.7.1	Penalty Function Method	133
3.7.2	Augmented Lagrange Multiplier Method	135
3.7.3	Lagrange Dual Method	137
3.7.4	Karush–Kuhn–Tucker Conditions	139
3.7.5	Alternating Direction Method of Multipliers	144
3.8	Newton Methods	147
3.8.1	Newton Method for Unconstrained Optimization	147
3.8.2	Newton Method for Constrained Optimization	150
	References	153
4	Solution of Linear Systems	157
4.1	Gauss Elimination	157
4.1.1	Elementary Row Operations	158
4.1.2	Gauss Elimination for Solving Matrix Equations	159
4.1.3	Gauss Elimination for Matrix Inversion	160
4.2	Conjugate Gradient Methods	162
4.2.1	Conjugate Gradient Algorithm	163
4.2.2	Biconjugate Gradient Algorithm	164
4.2.3	Preconditioned Conjugate Gradient Algorithm	165
4.3	Condition Number of Matrices	167
4.4	Singular Value Decomposition (SVD)	171
4.4.1	Singular Value Decomposition	171
4.4.2	Properties of Singular Values	173
4.4.3	Singular Value Thresholding	174
4.5	Least Squares Method	175
4.5.1	Least Squares Solution	175
4.5.2	Rank-Deficient Least Squares Solutions	178
4.6	Tikhonov Regularization and Gauss–Seidel Method	179
4.6.1	Tikhonov Regularization	179
4.6.2	Gauss–Seidel Method	181
4.7	Total Least Squares Method	184
4.7.1	Total Least Squares Solution	184
4.7.2	Performances of TLS Solution	189
4.7.3	Generalized Total Least Square	191
4.8	Solution of Under-Determined Systems	193
4.8.1	ℓ_1 -Norm Minimization	193
4.8.2	Lasso	196
4.8.3	LARS	197
	References	199
5	Eigenvalue Decomposition	203
5.1	Eigenvalue Problem and Characteristic Equation	203
5.1.1	Eigenvalue Problem	203
5.1.2	Characteristic Polynomial	205

5.2	Eigenvalues and Eigenvectors	206
5.2.1	Eigenvalues	206
5.2.2	Eigenvectors	208
5.3	Generalized Eigenvalue Decomposition (GEVD).....	210
5.3.1	Generalized Eigenvalue Decomposition	210
5.3.2	Total Least Squares Method for GEVD.....	214
5.4	Rayleigh Quotient and Generalized Rayleigh Quotient.....	214
5.4.1	Rayleigh Quotient.....	215
5.4.2	Generalized Rayleigh Quotient.....	216
5.4.3	Effectiveness of Class Discrimination	217
	References	220

Part II Artificial Intelligence

6	Machine Learning	223
6.1	Machine Learning Tree	223
6.2	Optimization in Machine Learning	226
6.2.1	Single-Objective Composite Optimization	226
6.2.2	Gradient Aggregation Methods	230
6.2.3	Coordinate Descent Methods	232
6.2.4	Benchmark Functions for Single-Objective Optimization	236
6.3	Majorization-Minimization Algorithms	241
6.3.1	MM Algorithm Framework.....	242
6.3.2	Examples of Majorization-Minimization Algorithms	244
6.4	Boosting and Probably Approximately Correct Learning	247
6.4.1	Boosting for Weak Learners	248
6.4.2	Probably Approximately Correct Learning	250
6.5	Basic Theory of Machine Learning	253
6.5.1	Learning Machine.....	253
6.5.2	Machine Learning Methods	254
6.5.3	Expected Performance of Machine Learning Algorithms	256
6.6	Classification and Regression	256
6.6.1	Pattern Recognition and Classification.....	257
6.6.2	Regression	259
6.7	Feature Selection	260
6.7.1	Supervised Feature Selection	261
6.7.2	Unsupervised Feature Selection	264
6.7.3	Nonlinear Joint Unsupervised Feature Selection	266
6.8	Principal Component Analysis	269
6.8.1	Principal Component Analysis Basis	269
6.8.2	Minor Component Analysis	270
6.8.3	Principal Subspace Analysis.....	271

6.8.4	Robust Principal Component Analysis	276
6.8.5	Sparse Principal Component Analysis	278
6.9	Supervised Learning Regression	281
6.9.1	Principle Component Regression	282
6.9.2	Partial Least Squares Regression	285
6.9.3	Penalized Regression	289
6.9.4	Gradient Projection for Sparse Reconstruction	293
6.10	Supervised Learning Classification	296
6.10.1	Binary Linear Classifiers	296
6.10.2	Multiclass Linear Classifiers	298
6.11	Supervised Tensor Learning (STL)	301
6.11.1	Tensor Algebra Basics	301
6.11.2	Supervised Tensor Learning Problems	307
6.11.3	Tensor Fisher Discriminant analysis	309
6.11.4	Tensor Learning for Regression	311
6.11.5	Tensor K-Means Clustering	314
6.12	Unsupervised Clustering	314
6.12.1	Similarity Measures	316
6.12.2	Hierarchical Clustering	320
6.12.3	Fisher Discriminant Analysis (FDA)	324
6.12.4	K-Means Clustering	327
6.13	Spectral Clustering	330
6.13.1	Spectral Clustering Algorithms	330
6.13.2	Constrained Spectral Clustering	333
6.13.3	Fast Spectral Clustering	335
6.14	Semi-Supervised Learning Algorithms	337
6.14.1	Semi-Supervised Inductive/Transductive Learning	338
6.14.2	Self-Training	340
6.14.3	Co-training	341
6.15	Canonical Correlation Analysis	343
6.15.1	Canonical Correlation Analysis Algorithm	344
6.15.2	Kernel Canonical Correlation Analysis	347
6.15.3	Penalized Canonical Correlation Analysis	352
6.16	Graph Machine Learning	354
6.16.1	Graphs	354
6.16.2	Graph Laplacian Matrices	358
6.16.3	Graph Spectrum	360
6.16.4	Graph Signal Processing	363
6.16.5	Semi-Supervised Graph Learning: Harmonic Function Method	368
6.16.6	Semi-Supervised Graph Learning: Min-Cut Method	371
6.16.7	Unsupervised Graph Learning: Sparse Coding Method	377
6.17	Active Learning	378
6.17.1	Active Learning: Background	379
6.17.2	Statistical Active Learning	380

6.17.3	Active Learning Algorithms	382
6.17.4	Active Learning Based Binary Linear Classifiers	383
6.17.5	Active Learning Using Extreme Learning Machine.....	384
6.18	Reinforcement Learning	387
6.18.1	Basic Concepts and Theory.....	387
6.18.2	Markov Decision Process (MDP)	391
6.19	Q-Learning	394
6.19.1	Basic Q-Learning	394
6.19.2	Double Q-Learning and Weighted Double Q-Learning	396
6.19.3	Online Connectionist Q-Learning Algorithm.....	399
6.19.4	Q-Learning with Experience Replay	400
6.20	Transfer Learning	402
6.20.1	Notations and Definitions.....	403
6.20.2	Categorization of Transfer Learning	406
6.20.3	Boosting for Transfer Learning	410
6.20.4	Multitask Learning	411
6.20.5	EigenTransfer	414
6.21	Domain Adaptation	417
6.21.1	Feature Augmentation Method	418
6.21.2	Cross-Domain Transform Method	421
6.21.3	Transfer Component Analysis Method	423
	References	427
7	Neural Networks	441
7.1	Neural Network Tree	441
7.2	From Modern Neural Networks to Deep Learning	444
7.3	Optimization of Neural Networks	445
7.3.1	Online Optimization Problems	446
7.3.2	Adaptive Gradient Algorithm	447
7.3.3	Adaptive Moment Estimation	449
7.4	Activation Functions.....	452
7.4.1	Logistic Regression and Sigmoid Function.....	452
7.4.2	Softmax Regression and Softmax Function	454
7.4.3	Other Activation Functions	456
7.5	Recurrent Neural Networks	460
7.5.1	Conventional Recurrent Neural Networks	460
7.5.2	Backpropagation Through Time (BPTT)	463
7.5.3	Jordan Network and Elman Network	467
7.5.4	Bidirectional Recurrent Neural Networks	468
7.5.5	Long Short-Term Memory (LSTM).....	471
7.5.6	Improvement of Long Short-Term Memory	474
7.6	Boltzmann Machines	477
7.6.1	Hopfield Network and Boltzmann Machines	477
7.6.2	Restricted Boltzmann Machine.....	481

7.6.3	Contrastive Divergence Learning	484
7.6.4	Multiple Restricted Boltzmann Machines	487
7.7	Bayesian Neural Networks	490
7.7.1	Naive Bayesian Classification	490
7.7.2	Bayesian Classification Theory	491
7.7.3	Sparse Bayesian Learning	493
7.8	Convolutional Neural Networks	496
7.8.1	Hankel Matrix and Convolution	498
7.8.2	Pooling Layer	504
7.8.3	Activation Functions in CNNs	507
7.8.4	Loss Function	510
7.9	Dropout Learning	514
7.9.1	Dropout for Shallow and Deep Learning	515
7.9.2	Dropout Spherical K-Means	518
7.9.3	DropConnect	520
7.10	Autoencoders	524
7.10.1	Basic Autoencoder	525
7.10.2	Stacked Sparse Autoencoder	532
7.10.3	Stacked Denoising Autoencoders	535
7.10.4	Convolutional Autoencoders (CAE)	538
7.10.5	Stacked Convolutional Denoising Autoencoder	539
7.10.6	Nonnegative Sparse Autoencoder	540
7.11	Extreme Learning Machine	542
7.11.1	Single-Hidden Layer Feedforward Networks with Random Hidden Nodes	543
7.11.2	Extreme Learning Machine Algorithm for Regression and Binary Classification	545
7.11.3	Extreme Learning Machine Algorithm for Multiclass Classification	549
7.12	Graph Embedding	551
7.12.1	Proximity Measures and Graph Embedding	552
7.12.2	Multidimensional Scaling	557
7.12.3	Manifold Learning: Isometric Map	559
7.12.4	Manifold Learning: Locally Linear Embedding	560
7.12.5	Manifold Learning: Laplacian Eigenmap	563
7.13	Network Embedding	567
7.13.1	Structure and Property Preserving Network Embedding	568
7.13.2	Community Preserving Network Embedding	569
7.13.3	Higher-Order Proximity Preserved Network Embedding	573
7.14	Neural Networks on Graphs	576
7.14.1	Graph Neural Networks (GNNs)	577
7.14.2	DeepWalk and GraphSAGE	579
7.14.3	Graph Convolutional Networks (GCNs)	583

7.15	Batch Normalization Networks	588
7.15.1	Batch Normalization	588
7.15.2	Variants and Extensions of Batch Normalization.....	593
7.16	Generative Adversarial Networks (GANs)	598
7.16.1	Generative Adversarial Network Framework.....	598
7.16.2	Bidirectional Generative Adversarial Networks	602
7.16.3	Variational Autoencoders	604
	References	607
8	Support Vector Machines	617
8.1	Support Vector Machines: Basic Theory	617
8.1.1	Statistical Learning Theory	618
8.1.2	Linear Support Vector Machines	621
8.2	Kernel Regression Methods	624
8.2.1	Reproducing Kernel and Mercer Kernel	624
8.2.2	Representer Theorem and Kernel Regression	628
8.2.3	Semi-Supervised and Graph Regression	630
8.2.4	Kernel Partial Least Squares Regression.....	632
8.2.5	Laplacian Support Vector Machines	633
8.3	Support Vector Machine Regression.....	635
8.3.1	Support Vector Machine Regressor	635
8.3.2	ϵ -Support Vector Regression	636
8.3.3	ν -Support Vector Machine Regression.....	639
8.4	Support Vector Machine Binary Classification.....	641
8.4.1	Support Vector Machine Binary Classifier.....	642
8.4.2	ν -Support Vector Machine Binary Classifier	645
8.4.3	Least Squares SVM Binary Classifier.....	647
8.4.4	Proximal Support Vector Machine Binary Classifier	649
8.4.5	SVM-Recursive Feature Elimination	651
8.5	Support Vector Machine Multiclass Classification.....	653
8.5.1	Decomposition Methods for Multiclass Classification.....	653
8.5.2	Least Squares SVM Multiclass Classifier.....	656
8.5.3	Proximal Support Vector Machine Multiclass Classifier...	659
8.6	Gaussian Process for Regression and Classification	660
8.6.1	Joint, Marginal, and Conditional Probabilities	661
8.6.2	Gaussian Process	661
8.6.3	Gaussian Process Regression	663
8.6.4	Gaussian Process Classification	666
8.7	Relevance Vector Machine	667
8.7.1	Sparse Bayesian Regression	668
8.7.2	Sparse Bayesian Classification	672
8.7.3	Fast Marginal Likelihood Maximization.....	673
	References	677

9	Evolutionary Computation	681
9.1	Evolutionary Computation Tree	681
9.2	Multiobjective Optimization	683
9.2.1	Multiobjective Combinatorial Optimization	684
9.2.2	Multiobjective Optimization Problems	686
9.3	Pareto Optimization Theory	691
9.3.1	Pareto Concepts	691
9.3.2	Fitness Selection Approach	698
9.3.3	Nondominated Sorting Approach	700
9.3.4	Crowding Distance Assignment Approach	702
9.3.5	Hierarchical Clustering Approach	703
9.3.6	Benchmark Functions for Multiobjective Optimization	704
9.4	Noisy Multiobjective Optimization	707
9.4.1	Pareto Concepts for Noisy Multiobjective Optimization	708
9.4.2	Performance Metrics for Approximation Sets	712
9.5	Multiobjective Simulated Annealing	714
9.5.1	Principle of Simulated Annealing	714
9.5.2	Multiobjective Simulated Annealing Algorithm	718
9.5.3	Archived Multiobjective Simulated Annealing	721
9.6	Genetic Algorithm	723
9.6.1	Basic Genetic Algorithm Operations	724
9.6.2	Genetic Algorithm with Gene Rearrangement Clustering	729
9.7	Nondominated Multiobjective Genetic Algorithms	733
9.7.1	Fitness Functions	733
9.7.2	Fitness Selection	734
9.7.3	Nondominated Sorting Genetic Algorithms	736
9.7.4	Elitist Nondominated Sorting Genetic Algorithm	739
9.8	Evolutionary Algorithms (EAs)	740
9.8.1	(1 + 1) Evolutionary Algorithm	741
9.8.2	Theoretical Analysis on Evolutionary Algorithms	742
9.9	Multiobjective Evolutionary Algorithms	743
9.9.1	Classical Methods for Solving Multiobjective Optimization Problems	743
9.9.2	MOEA Based on Decomposition (MOEA/D)	745
9.9.3	Strength Pareto Evolutionary Algorithm	749
9.9.4	Achievement Scalarizing Functions	754
9.10	Evolutionary Programming	758
9.10.1	Classical Evolutionary Programming	758
9.10.2	Fast Evolutionary Programming	759
9.10.3	Hybrid Evolutionary Programming	761
9.11	Differential Evolution	763
9.11.1	Classical Differential Evolution	763
9.11.2	Differential Evolution Variants	765

9.12	Ant Colony Optimization	767
9.12.1	Real Ants and Artificial Ants	768
9.12.2	Typical Ant Colony Optimization Problems.....	771
9.12.3	Ant System and Ant Colony System	773
9.13	Multiobjective Artificial Bee Colony Algorithms	776
9.13.1	Artificial Bee Colony Algorithms	776
9.13.2	Variants of ABC Algorithms	778
9.14	Particle Swarm Optimization	780
9.14.1	Basic Concepts	780
9.14.2	The Canonical Particle Swarm	781
9.14.3	Genetic Learning Particle Swarm Optimization	783
9.14.4	Particle Swarm Optimization for Feature Selection	785
9.15	Opposition-Based Evolutionary Computation	788
9.15.1	Opposition-Based Learning	788
9.15.2	Opposition-Based Differential Evolution	790
9.15.3	Two Variants of Opposition-Based Learning	791
	References	795
	Index	805