

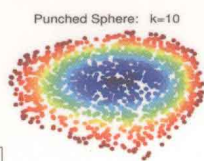
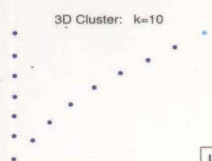
Geometric Structure of High-Dimensional Data and Dimensionality Reduction

高维数据几何结构及降维

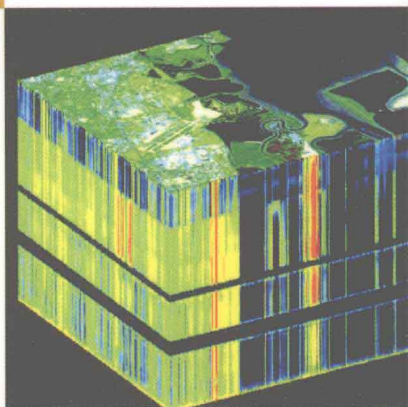
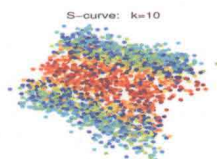
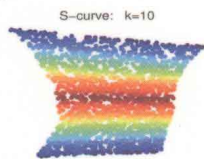
(英文版)

Jianzhong Wang

王建忠



LLE



高等教育出版社
HIGHER EDUCATION PRESS

Geometric Structure of High-Dimensional Data and Dimensionality Reduction

Journal of Statistical Software

[2022]



新一代信息科学与技术丛书

Geometric Structure of High-Dimensional Data and Dimensionality Reduction

高维数据几何结构及降维（英文版）

王建忠 Jianzhong Wang

 高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

Author
Prof. Jianzhong Wang
Department of Mathematics and Statistics
Sam Houston State University
1901 Avenue I, Huntsville
TX 77382-2206, USA
E-mail: jzwang@shsu.edu

图书在版编目(CIP)数据

高维数据几何结构及降维=Geometric Structure of High-Dimensional Data and Dimensionality Reduction: 英文/王建忠编著. —北京: 高等教育出版社, 2012. 1

ISBN 978-7-04-031704-6

I. ①高… II. ①王… III. ①统计数据—统计分析(数学)—英文 IV. ①O212.1

中国版本图书馆 CIP 数据核字(2011)第184611号

策划编辑 刘 英 责任编辑 刘 英 封面设计 张 楠 版式设计 王 莹
责任校对 胡美萍 责任印制 朱学忠

出版发行	高等教育出版社	咨询电话	400-810-0598
社 址	北京市西城区德外大街 4 号	网 址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印 刷	涿州市星河印刷有限公司	网上订购	http://www.landaco.com
开 本	787mm×1092mm 1/16		http://www.landaco.com.cn
印 张	23.5	版 次	2012 年 1 月第 1 版
字 数	540 千字	印 次	2012 年 1 月第 1 次印刷
购书热线	010-58581118	定 价	69.00 元

本书如有缺页、倒页、脱页等质量问题, 请到所购图书销售部门联系调换

版权所有 侵权必究

物 料 号 31704-00

Not for sale outside the mainland of China
仅限中国大陆地区销售

To my wife Huilan and my sons Chunhui and Chunze

Preface

Many objects in our world can be electronically represented with high-dimensional data — speech signals, images, videos, electrical text documents. We often need to analyze a large amount of data and process them. However, due to the high dimension of these data, directly processing them using regular systems may be too complicated and unstable to be feasible. In order to process high-dimensional data, dimensionality reduction technique becomes crucial. Dimensionality reduction is a method to represent high-dimensional data by their low-dimensional embeddings so that the low-dimensional data can be effectively used either in processing systems, or for better understanding. This technique has proved an important tool and has been widely used in many fields of data analysis, data mining, data visualization, and machine learning.

Principal component analysis (PCA) and classical multidimensional scaling (CMDS) are two traditional methods based on linear data models. However, they are not effective if data do not well reside on superplanes. Recently, many methods of nonlinear dimensionality reduction, also called manifold learning, have been developed based on nonlinear data models, which assume each observed high-dimensional data resides on a low-dimensional manifold. A nonlinear method *learns* the underlying manifold from the data and then represents the data by its low-dimensional manifold coordinates.

This monograph is an introductory treatise on dimensionality deduction with an emphasis on nonlinear methods. It includes a comprehensive introduction to linear methods, and then focuses on the geometric approach to nonlinear dimensionality reduction. The approach adopts manifold embedding technique that embeds a high-dimensional dataset into a low-dimensional manifold. The main idea of this approach can be outlined as follows. The geometry of the underlying manifold is learned from the neighborhood structure of the data, which defines the data similarity. Then a kernel is created to represent the learned manifold geometry and the spectral decomposition of the kernel leads to the manifold embedding. The different nonlinear methods adopt different manifold learning models. I hope that this book will illustrate this idea effectively.

Giving consideration to various readers, I try to keep a balance between

intuitive description and strictly mathematical deduction in this book. Each method is first described with the intuitive idea. Then the formulation is derived with necessary details. Finally, the algorithm is presented and interpreted. Graphs are used to illustrate the capability and validity, as well as the advantages and shortcomings, of each method. Several applications, such as face recognition, document ordering and classification, objective visualization, are introduced or outlined. I have no ambition to completely cover these applications. I hope the introduction to them will help readers understand the concepts and contents of dimensionality reduction. The technical mathematical justification is put in the last section of each chapter. The readers who are only interested in the applications of dimensionality reduction can skip, for example, Chapter 2, Section 3.3, and all justification sections.

This monograph first presents a general introduction. It then presents the material in three parts. In Part I, the geometry of data is discussed. In this part, there is a brief review of manifold calculus, followed by a preliminary presentation of spectral graph theory. These are the main tools in the study of the geometry of data. The last chapter of Part I discusses the data models and strategies of dimensionality reduction. In Part II, linear methods are discussed. Besides PCA and CMDS, the random projection methods are also introduced. The main part of this monograph is Part III, in which several popular nonlinear methods are introduced, such as Isomaps, maximum variance unfolding (MVU), locally linear embedding (LLE), local tangent space alignment (LTSA), Laplacian eigenmaps (Leigs), Hessian locally linear embedding (HLLE), and diffusion maps (Dmaps). This part ends with a chapter on fast dimensionality reduction algorithms, which have attracted more interest recently, with the increasing availability of very large data sets.

The book is primarily intended for computer scientists, applied mathematician, statisticians, and data analysts. It is also accessible to other audiences who want to apply dimensionality reduction technique to understanding and treating the complex data in their own fields, such as economists and geophysics.

I am very grateful to my colleague and friend Dr. Barry Friedman who has read the manuscript. To the editorial office, particularly to Dr. Ying Liu, I wish to express my appreciation of their efficient assistance and friendly cooperation.

Jianzhong Wang
Huntsville, Texas
February, 2011

Acronyms

CMDS	classical multidimensional scaling
CSDP	C semidefinite programming
DR	dimensionality reduction
DWT	discrete wavelet transform
EM-PCA	expectation-maximization principle component analysis
FAT	fast anisotropic transformation
FFT	fast Fourier transform
GAT	greedy anisotropic transformation
HEEL	Hessian locally linear embedding
HSI	hyperspectral image
IRAT	interpolative randomizing anisotropic transformation
LLE	locally linear embedding
LMVU	landmark maximum variance unfolding
LTSA	local tangent space alignment
MDS	multidimensional scaling
MVU	maximum variance unfolding
PCA	principle component analysis
PRAT	projective randomizing anisotropic transformation
RAT	randomizing anisotropic transformation
RP	random projection
SDE	semidefinite embedding
SDP	semidefinite programming
SLLE	supervised locally linear embedding
SVD	singular value decomposition
Dmaps	diffusion maps
Isomaps	isometric maps
JL-embedding	Johnson and Lindenstrauss embedding
LB operator	Laplace Beltrami operator
Leigs	Laplacian eigenmaps
i.i.d.	independent and identically distributed
o.n.	orthonormal

xviii **Acronyms**

o.g.	orthogonal
psd	positive semidefinite
pd	positive definite

Symbols

$\mathfrak{D}_{m,n}$	real $m \times n$ diagonal matrices
$\mathfrak{D}_n$	real $n \times n$ diagonal matrices
$\mathfrak{M}_{m,n}$	real $m \times n$ matrices
$\mathfrak{M}_{m,n}(r)$	real $m \times n$ matrices of rank r
$\mathfrak{M}_n$	real $n \times n$ matrices
$\mathfrak{M}_n(r)$	real $n \times n$ matrices of rank r
$\mathfrak{O}_{m,n}$	real $m \times n$ matrices with o.n. columns if $m \geq n$, and with o.n. rows if $m < n$
$\mathfrak{O}_n$	real $n \times n$ o.g. matrices
$\mathfrak{S}_n$	$n \times n$ psd matrices
$\mathcal{SP}_n(r)$	$n \times n$ psd matrices of rank r
$\mathfrak{P}_n$	$n \times n$ pd matrices
$\mathfrak{R}_{m,n}$	$m \times n$ random matrices
$\mathfrak{R}_{m,n}^1$	$m \times n$ Gaussian random matrices, random matrices of Type 1
$\mathfrak{R}_{m,n}^2$	$m \times n$ random matrices of Type 2
$\mathfrak{R}_{m,n}^3$	$m \times n$ random matrices of Type 3
$\mathbb{R}^n$	space of real n -vectors, n -dimensional Euclidean space
$\mathbb{R}_+$	nonnegative real numbers
$\mathbb{Z}$	integer numbers
$\mathbb{Z}_+$	nonnegative integer numbers
$\mathbb{N}$	natural numbers

Contents

Chapter 1 Introduction	1
1.1 Overview of Dimensionality Reduction	1
1.2 High Dimension Data Acquisition	3
1.2.1 Collection of Images in Face Recognition	3
1.2.2 Handwriting Letters and Digits	5
1.2.3 Text Documents	6
1.2.4 Hyperspectral Images	6
1.3 Curse of the Dimensionality	8
1.3.1 Volume of Cubes and Spheres	9
1.3.2 Volume of a Thin Spherical Shell	9
1.3.3 Tail Probability of the Multivariate Gaussian Distributions	10
1.3.4 Diagonals of Cube	11
1.3.5 Concentration of Norms and Distances	11
1.4 Intrinsic and Extrinsic Dimensions	12
1.4.1 Intrinsic Dimension Estimation	12
1.4.2 Correlation Dimension	15
1.4.3 Capacity Dimension	16
1.4.4 Multiscale Estimation	16
1.5 Outline of the Book	17
1.5.1 Categories of DR Problems	17
1.5.2 Scope of This Book	18
1.5.3 Other Topics Related to This Book	21
1.5.4 Artificial Surfaces for Testing DR Algorithms	22
References	24

Part I Data Geometry

Chapter 2 Preliminary Calculus on Manifolds	29
2.1 Linear Manifold	29
2.1.1 Subspace and Projection	30
2.1.2 Functions on Euclidean Spaces	32
2.1.3 Laplace Operator and Heat Diffusion Kernel	34
2.2 Differentiable Manifolds	35
2.2.1 Coordinate Systems and Parameterization	35
2.2.2 Tangent Spaces and Tangent Vectors	38
2.2.3 Riemannian Metrics	43
2.2.4 Geodesic Distance	45
2.3 Functions and Operators on Manifolds	46
2.3.1 Functions on Manifolds	46
2.3.2 Operators on Manifolds	47
References	49
 Chapter 3 Geometric Structure of High-Dimensional Data	51
3.1 Similarity and Dissimilarity of Data	51
3.1.1 Neighborhood Definition	52
3.1.2 Algorithms for Construction of Neighborhood	56
3.2 Graphs on Data Sets	60
3.2.1 Undirected Graphs	60
3.2.2 Directed Graphs	64
3.2.3 Neighborhood and Data Graphs	66
3.3 Spectral Analysis of Graphs	67
3.3.1 Laplacian of Graphs	67
3.3.2 Laplacian on Weighted Graphs	74
3.3.3 Contracting Operator on Weighted Graph	75
References	76
 Chapter 4 Data Models and Structures of Kernels of DR	79
4.1 Data Models in Dimensionality Reduction	79
4.1.1 Input Data of First Type	79
4.1.2 Input Data of Second Type	81
4.1.3 Constraints on Output Data	82
4.1.4 Consistence of Data Graph	83
4.1.5 Robust Graph Connection Algorithm	84
4.2 Constructions of DR Kernels	87
4.2.1 DR Kernels of Linear Methods	87

4.2.2 DR Kernels of Nonlinear Methods	88
4.2.3 Conclusion	90
References	91

Part II Linear Dimensionality Reduction

Chapter 5 Principal Component Analysis	95
5.1 Description of Principal Component Analysis	95
5.1.1 Geometric Description of PCA	95
5.1.2 Statistical Description of PCA	99
5.2 PCA Algorithms	100
5.2.1 Matlab Code for PCA Algorithm	101
5.2.2 EM-PCA Algorithm	102
5.2.3 Testing PCA Algorithm on Artificial Surfaces	104
5.3 Applications of PCA	106
5.3.1 PCA in Machine Learning	107
5.3.2 PCA in Eigenfaces	108
5.3.3 PCA in Hyperspectral Image Analysis	112
References	113
Chapter 6 Classical Multidimensional Scaling	115
6.1 Introduction of Multidimensional Scaling	115
6.1.1 Data Similarities and Configuration	115
6.1.2 Classification of MDS	117
6.2 Euclidean Matric and Gram Matrices	118
6.2.1 Euclidean Distance Matrices	119
6.2.2 Gram Matrix on Data Set	120
6.2.3 Relation between Euclidean Distance and Gram Matrix	120
6.3 Description of Classical Multidimensional Scaling	123
6.3.1 CMDS Method Description	123
6.3.2 Relation between PCA and CMDS	126
6.3.3 Weighted Graphic Description of CMDS	126
6.4 CMDS Algorithm	126
References	128
Chapter 7 Random Projection	131
7.1 Introduction	131
7.1.1 Lipschitz Embeddings	131
7.1.2 JL-Embeddings	132

7.2	Random Projection Algorithms	133
7.2.1	Random Matrices and Random Projection	133
7.2.2	Random Projection Algorithms	135
7.3	Justification	136
7.3.1	Johnson and Lindenstrauss Lemma	136
7.3.2	Random Projection based on Gaussian Distribution	138
7.3.3	Random Projection based on Types 2 and 3	141
7.4	Applications of Random Projections	144
7.4.1	Face Recognition Experiments with Random Projection	145
7.4.2	RP Applications to Image and Text Data	146
	References	147

Part III Nonlinear Dimensionality Reduction

Chapter 8	Isomaps	151
8.1	Isomap Embeddings	151
8.1.1	Description of Isomaps	151
8.1.2	Geodesic Metric on Discrete Set	153
8.1.3	Isomap Kernel and its Constant Shift	153
8.2	Isomap Algorithm	156
8.2.1	Algorithm Description	156
8.2.2	Matlab Code of Isomap	158
8.3	Dijkstra's Algorithm	160
8.3.1	Description of Dijkstra's Algorithm	160
8.3.2	Matlab Code of Dijkstra's Algorithm	163
8.4	Experiments and Applications of Isomaps	169
8.4.1	Testing Isomap Algorithm on Artificial Surfaces	169
8.4.2	Isomap Algorithm in Visual Perception	172
8.4.3	Conclusion	174
8.5	Justification of Isomap Methods	175
8.5.1	Graph Distance, S-distance, and Geodesic Distance	175
8.5.2	Relation between S-distance and Geodesic Distance	176
8.5.3	Relation between S-distance and Graph Distance	177
8.5.4	Main Result	178
	References	179
Chapter 9	Maximum Variance Unfolding	181
9.1	MVU Method Rescription	181
9.1.1	Description of the MVU Method	183

9.1.2	MVU Algorithm	185
9.2	Semidefinity Programming	186
9.2.1	CSDP	187
9.2.2	SDPT3	188
9.3	Experiments and Applications of MVU	188
9.3.1	Testing MVU Algorithm on Artificial Surfaces	188
9.3.2	MVU Algorithm in Sensor Localization	190
9.4	Landmark MVU	194
9.4.1	Description of Landmark MVU	194
9.4.2	Linear Transformation from Landmarks to Data Set	195
9.4.3	Algorithm for Landmark Linear Transformation	198
9.4.4	Construction of Kernel of Landmark MVU	199
9.4.5	Experiments of LMVU	200
9.4.6	Conclusion	200
	References	201
Chapter 10	Locally Linear Embedding	203
10.1	Description of Locally Linear Embedding	203
10.1.1	Barycentric Coordinates	204
10.1.2	LLE Method	205
10.1.3	LLE Algorithm	207
10.2	Experiments and Applications of LLE	210
10.2.1	Experiments on Artificial Surfaces	210
10.2.2	Conclusion	213
10.3	Applications of LLE	214
10.3.1	LLE in Image Ordering	214
10.3.2	Supervised LLE	215
10.4	Justification of LLE	216
10.4.1	Invariance Constraint	216
10.4.2	Condition for Weight Uniqueness	218
10.4.3	Reduction of the DR Data to LLE Model	219
	References	220
Chapter 11	Local Tangent Space Alignment	221
11.1	Description of Local Tangent Space Alignment	221
11.1.1	Tangent Coordinates and Manifold Coordinates	221
11.1.2	Local Coordinate Representation	223
11.1.3	Global Alignment	225
11.2	LTSA Algorithm	226

11.2.1	LTSA Algorithm Description	226
11.2.2	Matlab Code of LTSA	227
11.3	Experiments of LTSA Algorithm	229
11.3.1	Test LTSA on Artificial Surfaces	229
11.3.2	Conclusion	233
	References	233
Chapter 12	Laplacian Eigenmaps	235
12.1	Description of the Laplacian Eigenmap Method	235
12.1.1	Approximation of Laplace-Beltrami Operator	236
12.1.2	Discrete form of Laplace-Beltrami Operator	237
12.1.3	Minimization Model for DR Data Set	238
12.1.4	Construction of General Leigs Kernels	239
12.2	Laplacian Eigenmaps Algorithm	240
12.2.1	Steps in Leigs Algorithm	240
12.2.2	Matlab Code of Leigs Algorithm	241
12.3	Implementation of Leigs Algorithm	243
12.3.1	Experiments on Artificial Surfaces	243
12.3.2	Conclusion	247
	References	247
Chapter 13	Hessian Locally Linear Embedding	249
13.1	Description of Hessian Locally Linear Embedding	249
13.1.1	Hessian on Manifold	250
13.1.2	Hessian on Tangent Space	251
13.1.3	Construction of Hessian Functional	252
13.1.4	Construct of HLLE DR Kernel	255
13.2	HLLE Algorithm	256
13.2.1	HLLE Algorithm Description	256
13.2.2	Matlab Code of HLLE	257
13.3	Implementation of HLLE	260
13.3.1	Experiments on Artificial Surfaces	260
13.3.2	Conclusion	265
	References	265
Chapter 14	Diffusion Maps	267
14.1	Description of DR Method of Diffusion Maps	267
14.1.1	Diffusion Operator on Manifold	268
14.1.2	Normalization of Diffusion Kernels	269
14.2	Diffusion Maps Algorithms	271

14.2.1	Dmaps DR Algorithm Description	271
14.2.2	Dmaps Algorithm of Graph-Laplacian Type	272
14.2.3	Dmaps Algorithm of Laplace-Beltrami Type	274
14.2.4	Dmaps Algorithm of Self-tuning Type	276
14.3	Implementation of Dmaps for DR	279
14.3.1	Implementation of Dmaps of Graph-Laplacian Type	280
14.3.2	Implementation of Dmaps of Laplace-Beltrami Type	284
14.3.3	Implementation of Dmaps of Self-turning Type	286
14.3.4	Conclusion	287
14.4	Diffusion Maps and Multiscale Diffusion Geometry	287
14.4.1	Construction of General Diffusion Kernels	287
14.4.2	Diffusion Distances	289
14.4.3	Diffusion Maps as Feature Extractors	292
14.5	Implementation of Dmaps for Feature Extraction	293
14.5.1	Feature Extracted from 3-dimensional Toroidal Helix	293
14.5.2	Reordering Face Images	293
14.5.3	Image Parameters Revealing	295
14.5.4	Feature Images of Hyperspectral Image Cube	296
	References	298
Chapter 15	Fast Algorithms for DR Approximation	299
15.1	Low-rank Approximation and Rank-revealing Factorization	299
15.1.1	Rank-revealing Factorization	300
15.1.2	Fast Rank-revealing Algorithms	302
15.1.3	Nyström Approximation	307
15.1.4	Greedy Low-rank Approximation	309
15.2	Randomized Algorithm for Matrix Approximation	311
15.2.1	Randomized Low-rank Approximation	311
15.2.2	Randomized Interpolative Algorithm	312
15.2.3	Randomized SVD Algorithm	314
15.2.4	Randomized Greedy Algorithm	315
15.3	Fast Anisotropic Transformation DR Algorithms	316
15.3.1	Fast Anisotropic Transformation	316
15.3.2	Greedy Anisotropic Transformation	318
15.3.3	Randomized Anisotropic Transformation	319