

稀疏学习、分类与识别

焦李成 尚荣华 刘芳 杨淑媛 著
侯彪 王爽 马文萍



科学出版社

稀疏学习、分类与识别

焦李成 尚荣华 刘芳 杨淑媛 著
侯彪 王爽 马文萍

科学出版社

北京

内 容 简 介

本书对近年来稀疏学习、分类与识别领域常见的理论及技术进行了较为全面的阐述和总结,并结合作者多年的研究成果,对相关理论及技术在应用领域的实践情况进行了展示和报告。全书从稀疏学习、分类与识别三个方面展开介绍,主要内容包含如下几个方面:以学习数据的有效表示为主题,通过挖掘数据本身固有的结构,如几何结构、稀疏与低秩结构等信息来更有效地学习数据的表示;从经典的压缩感知理论框架出发,讨论压缩感知的基本理论、方法和应用的发展概况,并侧重介绍基于过完备字典的结构化压缩感知;上述方法在图像解译中的应用。

本书可为计算机科学、信息科学、人工智能、自动化技术等领域及交叉领域中从事量子计算、进化算法、机器学习及相关应用研究的技术人员提供参考,也可作为相关专业研究生和高年级本科生教材。

图书在版编目(CIP)数据

稀疏学习、分类与识别/焦李成等著. —北京:科学出版社,2017.3
ISBN 978-7-03-052347-1

I. ①稀… II. ①焦… III. ①机器学习 IV. ①TP181

中国版本图书馆 CIP 数据核字 (2017) 第 053215 号

责任编辑:宋无汗 赵鹏利/责任校对:赵桂芬
责任印制:张 倩/封面设计:陈 敬

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

北京通州皇家印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2017年3月第 一 版 开本:720×1000 1/16

2017年3月第一次印刷 印张:20 1/4

字数:410 000

定价:138.00 元

(如有印装质量问题,我社负责调换)

前 言

随着现代传感器、多媒体技术、计算机通信及网络技术的飞速发展和广泛应用,人们经常需要存储、处理与分析规模更大、维数更高、结构更复杂的数据,如人脸图像数据、监控视频数据、生物信息数据等。海量数据为人们提供了更多的信息,但与此同时也对计算机计算和处理数据、存储数据以及传输数据等能力提出了更高的要求。高维数据不但会显著地增加计算和存储代价,而且使得推理、学习和识别等任务无法完成,并对传统的机器学习与统计分析理论提出了严峻的挑战,如导致所谓的“维数灾难”,也就是说为保证学习仍能获得良好的性能,样本集的大小需随着问题维数(变量或特征数目)的增加呈指数增长。

幸运的是,近年来发展起来的机器学习理论与方法为解决上述问题提供了一些帮助,并取得了一些成功的应用。机器学习的研究是根据生理学、认知科学等对人类学习机理的了解,建立人类学习过程的计算模型或认知模型,发展各种学习理论和学习方法,研究通用的学习算法并进行理论分析,建立面向任务的具有特定应用的学习系统。经过二十多年的发展,机器学习已应用于人工智能的各个分支领域,如专家系统、自动推理、自然语言理解、计算机视觉和智能机器人等领域。

压缩感知作为一种新的信号采样理论,一经提出就引起了学术界和工业界的广泛关注,现已被广泛应用于信息理论、图像压缩、模式识别、医疗成像、无线通信等领域。压缩感知是对可压缩或在某个变换域下稀疏的信号,以远低于信号 Nyquist 频率的采样率对信号进行采样,为了保持信号的原始结构,利用非自适应且与变换基不相关的观测矩阵将高维稀疏信号线性投影到一个低维子空间上,然后通过求解一个优化问题就可从少量的投影中以高概率重构原始信号。压缩感知理论极大地降低了信号的采样、存储以及传输的代价,已经带领信号处理进入一个新的时代,同时它的理论和方法也为高维及高阶复杂数据分析与处理指出了一条新的途径,极大地促进了数学理论和工程应用的结合,并将在大规模和复杂数据的处理中发挥重要作用。

自 20 世纪 60 年代以来,随着航空航天领域的快速发展,人们获取和处理信息的能力获得极大的提高,进入构建天地一体化观测系统的新阶段。根据成像光谱仪的光谱分辨率,遥感成像技术可分多光谱遥感、高光谱遥感以及超光谱遥感;相应的,遥感图像分为多光谱图像、高光谱图像和超光谱图像。由于高光谱遥感

图像具有波谱覆盖范围广、光谱分辨率高、信噪比高等优势,在众多领域具有巨大的应用潜力和需求。但是,相对于遥感数据获取技术的快速发展,遥感信息的分析、处理和认识能力表现出明显的不足和滞后,不能真正地实现遥感信息的价值,也无法满足人们的需求。因此,提出遥感图像的分析模型和方法,挖掘遥感数据中的信息,提高遥感图像分析识别的精度,是当代遥感技术领域的重点和难点。针对不同的应用,需要与之对应的处理技术,常见的高光谱图像分析技术有以下几种:图像校正、图像去噪、特征表示、目标检测、变化检测、解混合以及图像分类等。高光谱图像分析技术不仅限于以上几种,各种处理技术之间相互联系,其中图像分类是各种应用中涉及的关键技术之一。

分类是从冗余复杂的数据中抽出类别归属信息的过程,是人们从遥感图像上提取有用信息的重要途径之一。目前的高光谱图像分析系统已经无法满足人们对快速、高精度、大规模的遥感数据分析处理的需求。探索新的分类模型,提出新的分类方法,实现高效高性能的遥感数据分析系统,对经济发展和社会进步都具有重大的意义。

从1996年开始,在国家“973”计划项目(2013CB329402、2006CB705707)、国家“863”计划项目(863-306-ZT06-1、863-317-03-99、2002AA135080、2006AA01Z107、2008AA01Z125和2009AA12Z210)、国家自然科学基金创新研究群体科学基金项目(61621005)、国家自然科学基金重点项目(60133010、60703107、60703108、60872548和60803098)及面上项目(61371201、61271302、61272279、61473215、61373111、61303032、61271301、61203303、61522311、61573267、61473215、61571342、61572383、61501353、61502369、61271302、61272282、61202176、61573267、61473215、61573015、60073053、60372045和60575037)、国家部委科技项目资助项目(XADZ2008159和51307040103)、高等学校学科创新引智计划(“111”计划)(B07048)、国家自然科学基金重大研究计划项目(91438201和91438103)以及教育部“长江学者和创新团队发展计划”项目(IRT_15R53和IRT0645)、陕西省自然科学基金项目(2007F32和2009JQ8015)、国家教育部高等学校博士点基金项目(20070701022和200807010003)、中国博士后科学基金特别资助项目(200801426)、中国博士后科学基金资助项目(20080431228和20090451369)及教育部重点科研项目(02073)的资助下,对稀疏学习、分类与识别进行了较为系统的研究和探讨。

鉴于稀疏学习、分类与识别展现的广阔前景,以及对社会各个方面的重要影响,本书作者在该领域进行了深入而有成效的研究工作,通过十多年的探索研究,取得了一些成果,并在广泛的应用领域进行了尝试。从稀疏学习、分类与识别的角度,对很多复杂问题提出了新颖的解决思路和方法。基于前面的工作,结合国内外的发展动态,本书集合了当前稀疏学习、分类与识别的很多相关内容。不仅

包括稀疏学习、分类与识别以及交叉领域的基础理论介绍，更加入了许多最新技术在不同领域的应用工作解析。

本书是西安电子科技大学智能感知与图像理解教育部重点实验室、智能感知与计算教育部国际联合实验室、国家“111”计划创新引智基地、国家“2011”信息感知协同创新中心、“大数据智能感知与计算”陕西省 2011 协同创新中心、智能信息处理研究所近十年来集体智慧的结晶。特别感谢保铮院士多年来的悉心培养和指导；感谢中国科学技术大学陈国良院士，IEEE 计算智能学会副主席、英国伯明翰大学姚新教授，英国埃塞克斯大学张青富教授，英国诺丁汉大学屈嵘教授的指导和帮助；感谢国家自然科学基金委员会信息科学部的大力支持；感谢西安电子科技大学田捷教授、高新波教授、石光明教授、梁继民教授的帮助；感谢尚凡华、林乐平、殷飞、张二磊、孟洋、袁一璟、张玮桐、王文兵、刘驰旻、都炳琪、文爱玲、刘欢、常姜维、刘永坤、兰雨阳等智能感知与图像理解教育部重点实验室全体成员所付出的辛勤劳动；感谢作者家人的大力支持和理解。

由于作者水平有限，书中不妥之处在所难免，恳请读者批评指正。

作 者

2016 年 10 月 28 日

目 录

前言

第 1 章 引言	1
1.1 机器学习理论	1
1.1.1 维数约简	2
1.1.2 稀疏与低秩	2
1.1.3 半监督学习	4
1.2 压缩感知理论	5
1.2.1 压缩感知的研究意义	5
1.2.2 压缩感知的理论框架	6
1.2.3 压缩感知的重构算法介绍	8
1.3 高光谱遥感技术	9
1.3.1 遥感技术	9
1.3.2 高光谱遥感技术发展现状	10
1.3.3 高光谱遥感技术的应用	13
参考文献	15
第 2 章 机器学习理论基础	19
2.1 维数约简的研究进展	19
2.1.1 子空间分割	19
2.1.2 稀疏表示	21
2.1.3 矩阵恢复与填充	21
2.1.4 非线性降维	22
2.2 半监督学习与核学习的研究进展	23
2.2.1 半监督学习	23
2.2.2 非参数核学习	24
参考文献	25
第 3 章 快速密度加权低秩近似谱聚类	29
3.1 引言	29
3.2 背景与相关工作	30
3.2.1 谱聚类算法	30
3.2.2 近邻传播算法	30
3.2.3 Nyström 方法	31
3.3 全局距离测度与采样算法	33

3.3.1	全局距离	33
3.3.2	快速采样算法	34
3.4	快速两阶段谱聚类框架	35
3.4.1	采样阶段	36
3.4.2	正交化的密度加权近似谱聚类阶段	36
3.5	算法分析	39
3.5.1	采样算法比较	39
3.5.2	有效性分析	40
3.5.3	快速近邻搜索	41
3.5.4	复杂度分析	42
3.6	实验结果	42
3.6.1	双螺旋线数据	42
3.6.2	实际数据	44
3.6.3	评价指标	44
3.6.4	比较算法	45
3.6.5	聚类结果	46
3.6.6	参数稳定性分析	48
3.6.7	谱嵌入	50
	参考文献	50
	附录	52
第 4 章	双图正则非负矩阵分解	54
4.1	引言	54
4.2	相关工作	55
4.2.1	非负矩阵分解	55
4.2.2	图正则非负矩阵分解	56
4.2.3	双正则联合聚类	57
4.3	双图正则非负矩阵分解方法	57
4.3.1	数据图与特征图	57
4.3.2	DNMF 模型	58
4.3.3	迭代更新规则	59
4.3.4	收敛性分析	60
4.4	双图正则非负矩阵三分解	60
4.4.1	DNMTF 模型	61
4.4.2	迭代规则	61
4.4.3	收敛性分析	62
4.4.4	复杂度分析	63
4.5	实验	63
4.5.1	比较算法	63
4.5.2	UCI 数据	64

4.5.3	图像数据	66
4.5.4	稳定性分析	69
4.5.5	雷达高分辨距离像数据	70
	参考文献	72
	附录 A (定理 4.1 的证明)	74
	附录 B (定理 4.2 的证明)	75
第 5 章	学习鲁棒低秩矩阵分解	77
5.1	引言	77
5.2	相关工作及研究进展	78
5.3	鲁棒低秩矩阵分解框架	80
5.3.1	单子空间模型	80
5.3.2	多子空间模型	80
5.4	基于交替方向法的迭代算法	81
5.4.1	引入辅助变量	81
5.4.2	迭代求解算法	82
5.4.3	求解单子空间模型	85
5.4.4	拓展应用于矩阵填充	85
5.4.5	复杂度分析	87
5.5	实验	87
5.5.1	人工数据聚类	87
5.5.2	人脸聚类	90
5.5.3	背景建模	93
5.5.4	图像修复	94
	参考文献	95
第 6 章	学习谱表示应用于半监督聚类	98
6.1	引言	98
6.2	图的创建与谱表示	99
6.2.1	对称偏好图	99
6.2.2	图拉普拉斯谱嵌入	100
6.3	问题模型与求解	101
6.3.1	目标函数	101
6.3.2	问题求解	102
6.4	算法	103
6.4.1	半监督聚类	103
6.4.2	直推式分类	104
6.4.3	复杂度分析	106
6.5	实验	106
6.5.1	比较算法与参数设置	106
6.5.2	人工数据集	107

6.5.3	向量型数据	108
6.5.4	图结构数据	113
6.5.5	半监督聚类应用	115
6.5.6	直推式分类应用	116
	参考文献	117
第 7 章	应用低秩矩阵填充学习数据表示	120
7.1	引言	120
7.2	学习谱表示框架	122
7.2.1	核矩阵填充	122
7.2.2	提升矩阵学习模型	123
7.3	特征值迭代阈值算法	123
7.3.1	改进的不动点算法	124
7.3.2	加速策略	125
7.3.3	半监督聚类	127
7.3.4	推广到分类问题	127
7.3.5	复杂度分析	128
7.4	收敛性分析	128
7.5	实验	129
7.5.1	学习谱表示	129
7.5.2	比较算法与参数设置	130
7.5.3	向量型数据	131
7.5.4	图结构数据	134
7.5.5	分类应用	136
	参考文献	137
	附录 A (定理 7.2 的证明)	140
	附录 B (定理 7.3 的证明)	140
	附录 C (定理 7.4 的证明)	140
	附录 D (定理 7.6 的证明)	141
第 8 章	结合约束与低秩核学习的半监督学习	143
8.1	引言	143
8.2	符号与相关工作	145
8.3	复合信息半监督学习框架	147
8.3.1	基本框架	147
8.3.2	核范数正则模型	147
8.4	半监督学习算法	148
8.4.1	改进的不动点迭代算法	149
8.4.2	连续性策略和 BB 步长技术	151
8.4.3	标签传播	152
8.5	算法分析	154

8.5.1	收敛性分析	154
8.5.2	合法核	154
8.5.3	复杂度分析	155
8.5.4	归纳分类	155
8.6	实验	155
8.6.1	比较算法与参数设置	156
8.6.2	交叉螺旋线数据	156
8.6.3	实际数据	157
8.6.4	直推式分类	158
8.6.5	归纳分类	162
	参考文献	163
第 9 章	基于子空间类标传播和正则判别分析的单标记图像人脸识别	166
9.1	引言	166
9.2	正则判别分析和稀疏保持判别分析	167
9.3	子空间类标传播	168
9.4	基于子空间类标传播和正则判别分析的半监督维数约简	169
9.5	相关方法比较	171
9.6	实验	172
	参考文献	177
第 10 章	基于双线性回归的单标记图像人脸识别	178
10.1	引言	178
10.2	LDA 和 RDA 简述	180
10.2.1	LDA	180
10.2.2	RDA	180
10.3	双线性回归	181
10.3.1	基于子空间假设的类标传播	181
10.3.2	学习稀疏表示结构	182
10.3.3	稀疏保持正则项	183
10.3.4	基于双线性回归的半监督维数约简	184
10.3.5	核 DLR	185
10.3.6	计算复杂性分析	186
10.4	相关方法比较	187
10.5	实验	188
10.5.1	数据库介绍	188
10.5.2	实验设置	189
10.5.3	实验结果与讨论	190
10.5.4	DLR 方法的进一步探索	194
	参考文献	196

第 11 章 基于旋转扩展和稀疏表示的鲁棒遥感图像目标识别	200
11.1 引言	200
11.2 基于稀疏表示的识别	201
11.2.1 稀疏表示	201
11.2.2 稀疏表示用于识别	202
11.3 基于旋转扩展和稀疏表示的遥感目标识别	203
11.4 实验结果与分析	205
11.4.1 基于 RETSRC 的遥感图像目标识别	206
11.4.2 旋转扩展倍数对识别性能的影响	207
参考文献	208
第 12 章 压缩感知理论基础	210
12.1 压缩感知概述	210
12.1.1 基于字典的稀疏表示	210
12.1.2 压缩观测	213
12.1.3 结构化稀疏重构模型	215
12.2 稀疏重构方法	218
12.2.1 凸松弛方法	218
12.2.2 贪婪方法	220
12.2.3 其他重构方法	221
12.2.4 基于自然计算优化方法的稀疏重构	222
参考文献	223
第 13 章 基于分块策略和过完备字典的非凸压缩感知框架	232
13.1 引言	232
13.2 基于过完备字典的分块压缩感知框架	233
13.2.1 分块压缩感知	233
13.2.2 过完备字典	233
13.2.3 结构化压缩感知模型	234
13.3 基于 Ridgelet 过完备字典的图像稀疏表示	234
13.4 结构化重构模型	237
13.4.1 基于图像自相似性的结构稀疏先验	237
13.4.2 基于图像块方向结构估计的重构模型	238
13.5 非凸重构策略	238
参考文献	239
第 14 章 基于协同优化的稀疏重构	240
14.1 引言	240
14.2 基于过完备字典的协同压缩感知	241
14.2.1 基于过完备字典的结构稀疏先验	241
14.2.2 基于协同优化的稀疏重构策略	241

14.2.3 相关工作	242
14.3 基于过完备字典的协同重构模型	244
14.3.1 基于字典的分块稀疏重构	244
14.3.2 基于结构稀疏模型的协同重构	244
14.3.3 基于自回归模型的协同重构	245
14.4 CR_CS 协同重构算法	247
参考文献	254
第 15 章 基于过完备字典的方向结构估计模型及重构方法	256
15.1 引言	256
15.2 基于方向结构估计的重构模型	257
15.2.1 基于过完备字典的方向结构估计	257
15.2.2 稀疏字典的优化学习	257
15.2.3 基于方向结构估计的进化重构策略	258
15.3 相关工作	260
15.4 方向指导的稀疏字典优化及结构稀疏重构模型	260
15.4.1 方向指导的稀疏字典优化学习	260
15.4.2 基于稀疏子字典的结构稀疏重构模型	262
15.5 基于方向结构估计的非凸重构方法	263
15.5.1 基于字典的结构类型判定及方向结构估计	264
15.5.2 基于遗传优化的光滑图像块重构	266
15.5.3 基于遗传和克隆选择优化的非光滑图像块重构	267
15.6 仿真实验及结果分析	271
参考文献	275
第 16 章 基于光谱信息散度与稀疏表示的高光谱图像分类	276
16.1 高光谱图像分类的研究现状与挑战	276
16.1.1 高光谱图像分类研究现状	276
16.1.2 高光谱图像分类存在的挑战	277
16.2 研究动机	278
16.3 光谱信息散度	279
16.4 基于 SID 的稀疏表示分类方法	280
16.5 基于 SID 的联合稀疏表示分类方法	281
16.6 实验结果和分析	282
16.6.1 三种测度的比较	283
16.6.2 稀疏表示分类方法的性能比较	284
16.6.3 参数影响分析	285
16.6.4 收敛性证明	287
参考文献	287

第 17 章 基于多特征核稀疏表示学习的高光谱图像分类.....	289
17.1 引言	289
17.2 基于多特征加权联合的稀疏表示分类方法.....	291
17.3 基于多特征加权联合的核稀疏表示非线性分类方法.....	293
17.4 实验结果与分析	296
17.4.1 实验基本设置	296
17.4.2 AVIRIS 数据的实验结果	297
17.4.3 ROSIS 数据的实验结果	304
参考文献	307

第 1 章 引 言

1.1 机器学习理论

机器学习 (machine learning) 是当前人工智能主要的研究发展方向之一。机器学习与认知科学、心理学、计算机科学等许多学科都有着密切的联系, 涉及领域比较广, 已经成功地运用于许多实际问题, 并取得了不错的学习效果, 如自动驾驶汽车、疾病预测、下棋和语音识别等^[1]。在解决这些实际问题的过程中, 机器学习技术被深入地进行分析和研究, 得到了迅速发展, 并产生了很多优秀的学习算法, 如常用的八大机器学习算法: 决策树算法^[2]、随机森林算法^[3]、人工神经网络算法^[4]、支持向量机算法^[5]、Boosting 与 Bagging 算法^[6, 7]、关联规则算法^[8, 9]、贝叶斯学习算法^[10, 11]以及 EM 算法^[12]。

近年来随着计算机及采样技术的发展, 人们可以越来越容易地获取海量的高维数据, 如何从这些数据中找出合理有效的信息并进行探索, 已成为机器学习、数据挖掘等领域研究的热点问题。高维数据对传统的机器学习与统计分析提出了严峻的挑战, 如导致所谓的“维数灾难”(curse of dimensionality)^[13], 也就是说为保证学习仍能获得良好的性能, 样本集的大小需随着问题维数(变量或特征数目)的增加呈指数增长。与之相关的另一个挑战问题为空空间现象^[14](empty space phenomenon), 即高维空间本质上是稀疏空间, 如标准正态分布 $\mathcal{N}(0, 1)$ 在只有一维变量时, $[-1, 1]$ 区间内包含接近 70% 的数据点。然而当变量维数增加到十维时, 以原点为球心的单位超球内只包含 0.02% 的数据。另外, 当样本数目远小于维数时, 将导致典型的小样本 (small sample size) 问题, 从而最终影响学习算法的推广能力^[15]。

大量认知科学的实验验证了很多高维数据确实存在较低的本征维数, 且分布于高维空间中的一个低维子流形上。例如, 在不同角度、不同光照情况下, 同一个人的图像集就是一个以姿态、尺度、光照等为参数的低维子流形。这也更加表明对高维数据进行维数约简具有必要性。人眼能在瞬间认出多年未曾谋面的老同学, 然而计算机识别却很难做到。神经生物学研究发现视感知系统具有某种特性的不变性, 且整个神经细胞群的触发率可由少量维度的变量来描述, 这也进一步表明视神经元的群体活动由内在的低维结构所控制^[16]。

1.1.1 维数约简

给定的数据 $X \in \mathbb{R}^{m \times n}$ 是由 n 个 m 维的数据向量 x_i 组成, 且该数据集的本征维数为 $\tilde{d}$ (一般情况下 $\tilde{d} = m$), 其中本征维数为嵌入在 D 维高维空间的数据集 X 分布或接近于低维子空间或流形的维数 $\tilde{d}$ 。维数约简的基本思想是通过线性或非线性变换把高维的数据集 X 映射到一个低维空间, 从而获得 d (一般 $d \geq \tilde{d}$) 维的数据表示 $Y \in \mathbb{R}^{n \times d}$, 同时尽可能地保持原高维数据的信息。

如此一来, 维数约简技术不仅囊括了经典的主成分分析(principal component analysis, PCA)^[17]和线性判别分析(linear discriminant analysis, LDA)^[18]等方法, 而且诸如压缩感知中的随机投影^[19]、图像下采样等策略也自然地归属于上述维数约简定义的范畴。维数约简通常分为特征提取(如 PCA 和 LDA)与特征选择(如图像下采样)两类方法。

维数约简可在很大程度上避免维数灾难, 使得学习任务(如分类或聚类)更加稳定、高效, 并产生更优的推广性能。实际中, 对于成千或上万甚至更高维的数据而言, 如何通过维数约简技术获得数据的有效表示已变得越来越重要, 也更具挑战性, 且要满足两个基本特性^[20]: 数据的维数得到一定程度的约简, 可有效地识别出数据的重要成分、内在结构特征及隐变量等; 另外, 通过将数据降维至二维或三维进行可视化, 人们可准确直观地感知与发现隐藏在数据中的内在结构与规律。

1.1.2 稀疏与低秩

压缩感知(compressed sensing, CS)与稀疏表示(sparse representation, SR)是由 Candès 等提出的一种新的理论框架^[21], 最早被用于从低维观测信号 $y \in \mathbb{R}^{d \times l}$ 中恢复出高维原始信号 $x \in \mathbb{R}^{m \times l}$ ($m \gg d$), 其优化问题如下所述:

$$\min \|x\|_0, \quad \text{s.t. } Ax = y \quad (1.1)$$

式中, $\|\cdot\|_0$ 表示 l_0 范数, 即向量中非零元素的个数; $A \in \mathbb{R}^{d \times m}$ 为观测矩阵。该框架现已被广泛应用于信号与图像处理领域, 如图像去噪、恢复(recovery)及超分辨率(super-resolution)重建等, 并取得了巨大成功。该理论框架表明: 当感兴趣的信号是可稀疏表示的或具有可压缩性时, 可以通过极少的采样或观测精确地重构该信号, 也就是说, 很多现实信号都拥有较多的冗余, 类似的说法还有奥卡姆剃刀(Ockham's razor)原理或最小描述长度(minimal description length)。稀疏表示已成为最近几年信号处理、机器学习、模式识别及计算机视觉等领域的一个研究热点。其实稀疏表示的概念早在1996年*Nature*中就有涉及, 将稀疏性正则引入到最小二乘问题中, 计算得到具有方向特性的图像块, 这样能很好地解释初级视皮层(V1)的工

作原理^[22]。另外，在同一年，著名的Lasso算法^[23]也被提出用于求解带有稀疏约束的最小二乘问题。

最近几年，衍生于压缩感知技术的低秩矩阵重建已成为机器学习、计算机视觉、信号处理、优化等领域最热的研究方向之一，并在图像与视频处理、计算机视觉、文本分析、多任务学习、推荐系统等方面得到了成功的应用^[24]。矩阵恢复或填充可看成压缩感知理论由一维信号到二维矩阵的推广^[25]。矩阵的稀疏性主要表现在两个方面：第一是矩阵元素的稀疏性，即矩阵非0元素的个数相对较少，也就是矩阵的 l_0 范数；第二是矩阵奇异值(若为对称矩阵，则为特征值)的稀疏性，即矩阵奇异值中的非0元素的个数相对较少，也就是秩函数值较小。先看矩阵奇异值的稀疏性，即通常假定待恢复或填充的矩阵为低秩的，可通过矩阵的某些线性运算的结果由如下的优化问题精确地重构该矩阵：

$$\min \text{rank}(X), \quad \text{s.t. } \mathcal{A}(X) = b \quad (1.2)$$

式中， $\text{rank}(\cdot)$ 为矩阵的秩函数； $\mathcal{A}(\cdot)$ 为一个线性算子。具体的低秩矩阵填充问题可表述为如下的形式：

$$\min_X \text{rank}(X), \quad \text{s.t. } P_\Omega(X) = P_\Omega(Z) \quad (1.3)$$

式中， Ω 为已知元素下标的集合。 $P_\Omega(Z)$ 定义为如下的形式：

$$P_\Omega(Z_{ij}) = \begin{cases} Z_{ij}, & (i, j) \in \Omega \\ 0, & \text{其他} \end{cases}$$

若同时考虑矩阵元素与矩阵奇异值的稀疏性，可得到两类最近几年非常流行的问题模型：鲁棒主成分分析(robust principal component analysis, RPCA)或稀疏加低秩矩阵分解(sparse and low-rank matrix decomposition)模型和低秩表示(low-rank representation, LRR)模型。鲁棒主成分分析模型可由如下的优化问题描述：

$$\min_{Z, E} \text{rank}(Z) + \lambda \|E\|_l, \quad \text{s.t. } X = Z + E \quad (1.4)$$

式中， $\lambda > 0$ 为正则参数； $\|\cdot\|_l$ 为一种特定的正则策略，如用于对高斯噪声建模的Frobenius范数^[26, 27]，即 $\|\cdot\|_F$ ，处理少量较大幅值噪声的 l_0 范数^[26, 28]，及可有效处理列噪声或奇异点的 $l_{2,0}$ 范数^[29, 30]等。上述三类不同的噪声分别如图1.1所示。

然而上述的式(1.4)隐式地假设观测数据的潜在结构为单独一个低秩线性子空间^[29, 31, 32]。很多实际数据都分布于多个线性子空间的并集中，且任何数据点属于某个子空间的关系也是未知的。最近，有一种低秩加稀疏矩阵分解的拓展模型被提出，并被称为低秩表示模型^[29, 30]，即结合子空间分割与噪声识别于一个框架中用于处理多子空间问题。该低秩表示模型有如下所述的形式：

$$\min_{Z, E} \text{rank}(Z) + \lambda \|E\|_l, \quad \text{s.t. } X = DZ + E \quad (1.5)$$