

苏州大学学位论文独创性声明

本人郑重声明：所提交的学位论文是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含其他个人或集体已经发表或撰写过的研究成果，也不含为获得苏州大学或其它教育机构的学位证书而使用过的材料。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人承担本声明的法律责任。

论文作者签名： 高晓 日期： 2012.5.28



苏州大学学位论文使用授权声明



本人完全了解苏州大学关于收集、保存和使用学位论文的规定，即：学位论文著作权归属苏州大学。本学位论文电子文档的内容和纸质论文的内容相一致。苏州大学有权向国家图书馆、中国社科院文献信息情报中心、中国科学技术信息研究所（含万方数据电子出版社）、中国学术期刊（光盘版）电子杂志社送交本学位论文的复印件和电子文档，允许论文被查阅和借阅，可以采用影印、缩印或其他复制手段保存和汇编学位论文，可以将学位论文的全部或部分内容编入有关数据库进行检索。

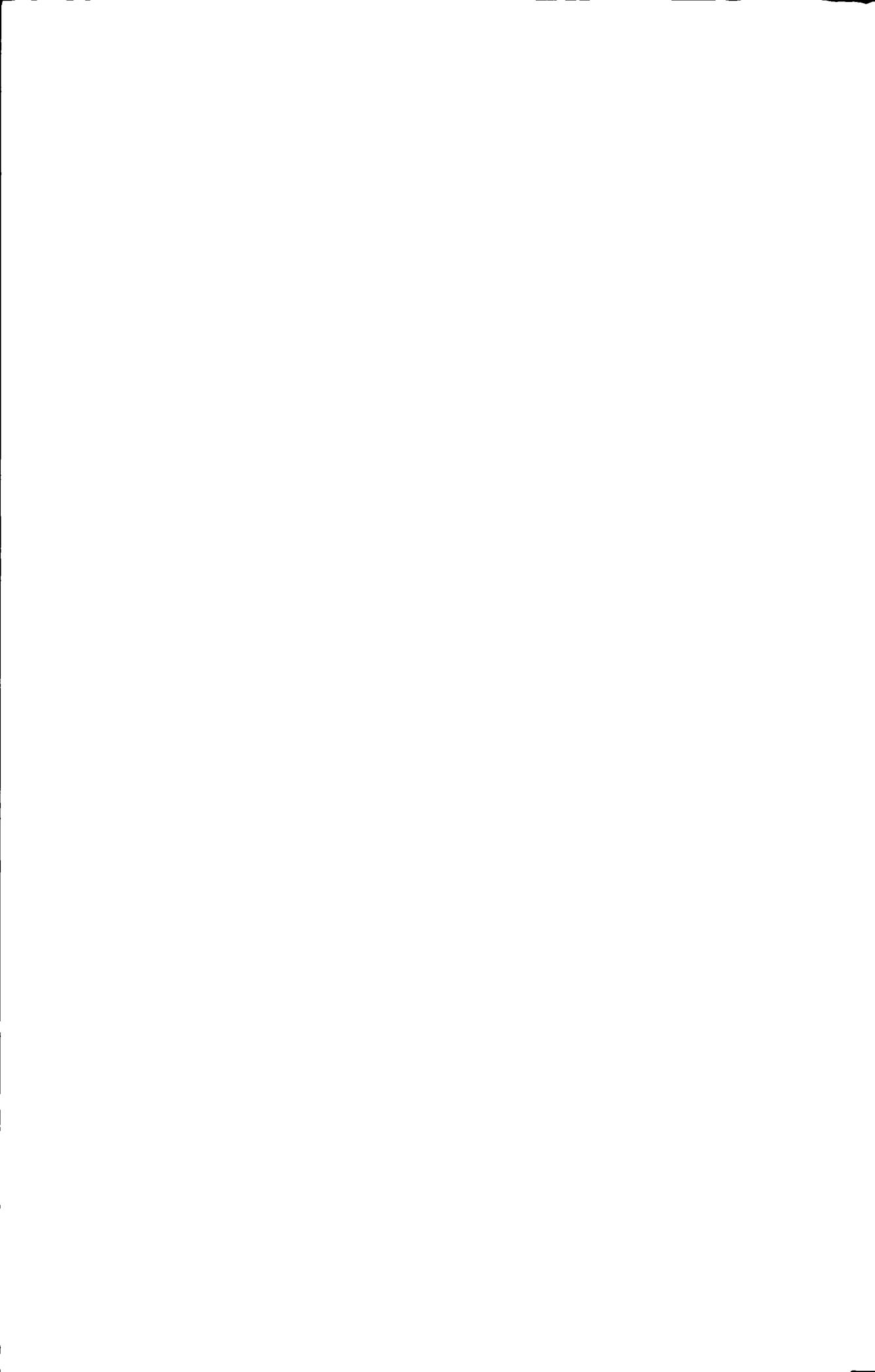
涉密论文

本学位论文属 _____ 在 _____ 年 _____ 月解密后适用本规定。

非涉密论文

论文作者签名： 高聪 日期： 2012.5.28

导师签名： [Signature] 日期： 2012.5.28



李群均值学习算法及应用研究

中文摘要

本文在李群均值计算的基础上,提出了 Lie-Fisher 学习算法,其主要思想是在李群流形上寻找一个由总体样本内均值的李代数元素决定的单参数子群,这个单参数子群是原李群上的一条测地线。然后定义了样本到测地线投影的概念,同时将李群样本向该测地线投影,并尽可能使投影后各类别间的散度与类内散度比值最大化,从而实现非线性李群空间的类别判别。文章也根据矩阵李群内积空间理论设计了基于矩阵李群的核函数,并借助 KFDA 算法思想和李群均值设计实现了 KLieDA 算法。通过大量实验表明,基于李群均值的 Lie-Fisher 学习算法与 KNN、FLDA、Lie-Mean 等算法相比,具有较好的分类效果,以及基于李群核的 KLieDA 及 SVM 算法在手写体分类中具有优良的识别率。

本文的主要贡献与创新点在于(1)提出的 Lie-Fisher 学习算法解决了李群流形空间中用测地线进行分类的问题。(2)推导了常用李群核函数,该核函数能方便结合现有的基于核的各种算法,如 KFDA、SVM、KPCA 等算法。(3)提出的 KlieDA 算法解决了非线性矩阵样本通过核理论进行分类的问题。(4)通过实验表明,手写体图像的区域协方差特征虽然位于流形空间内,但其拥有非常好的线性分布的特点。

关键词:李群机器学习;李群均值学习算法;李群核;分类器;视觉不变特征

作者:高聪

指导教师:李凡长

Research on Lie Group Mean Learning Algorithm and Its Application

Abstract

This paper based on Lie group mean computation theory, and proposed Lie-Fisher algorithm, which mainly finds a one-parameter sub group decided by a Lie algebra element of intrinsic mean of all samples, and this one-parameter sub group is a geodesic on the original Lie group. Moreover, we defined the projection of sample to geodesic, and projected all samples to a geodesic. In order to implement the discrimination in nonlinear Lie group space after projection, the ratio of between-class variance to within-class variance is maximized. In accordance with the matrix Lie group dot project theory, we also designed kernel functions based on matrix Lie group, and further proposed KLieDA algorithm with the help of KFDA and these kernel functions. A large number of experiments showed that Lie-Fisher based on Lie group mean is better than KNN, FLDA algorithms in classification performance, and both KLieDA and SVM based on Lie group kernels also have excellent recognition rate of handwritten classification.

The contributions and innovations of this paper are as follows. (1) The proposed Lie-Fisher learning algorithm solves the problem that geodesic is used to classify on Lie group manifold. (2) The several deduced Lie group kernel functions can be easily combined with the existing kernel-based algorithms, such as KFDA, SVM, and KPCA. (3) The proposed KlieDA algorithm solves the problem of nonlinear matrix samples' classification through the kernel theory. (4) The experiments demonstrate that the region covariance features of the handwritten images are located in Lie group manifold space, but they have a very good linear distribution characteristic.

Keywords: Lie group, Lie group mean algorithm, Lie group kernel, Classifier, Visual invariant feature

Written by Gao Cong

Supervised by Li Fanzhang

目 录

第一章 引言	1
1.1 课题研究意义	1
1.2 李群学习研究进展	1
1.3 李群相关视觉不变特征研究进展	5
1.4 问题的提出	7
1.5 本文内容安排	8
第二章 矩阵群学习算法	10
2.1 相关基本概念	10
2.1.1 矩阵群的定义	10
2.1.2 矩阵群度量空间	11
2.1.3 矩阵群实例	13
2.2 矩阵群学习算法	16
2.3 实例分析	18
2.3.1 KNN 矩阵样本分类	18
2.3.2 K-means 矩阵样本聚类	20
2.4 本章小结	22
第三章 李群学习相关理论	23
3.1 李群	23
3.2 矩阵李群实例	26
3.3 李群上的高斯分布	30
3.3.1 R^+ 的高斯分布	31
3.3.2 $SO(2)$ 的高斯分布	31
3.3.3 $SO(3)$ 的高斯分布	33
3.4 李群内均值计算	34
3.5 本章小节	38

第四章 李群均值相关学习算法	39
4.1 FLDA 算法	39
4.2 李群空间中的 Fisher 投影	40
4.3 Lie-Fisher 判别分析	42
4.3.1 Lie-Fisher 算法的第一形式	42
4.3.2 Lie-Fisher 算法的第二形式	44
4.3.3 Lie-Fisher 算法二种形式的区别	46
4.3.4 时间复杂度和空间复杂度分析	47
4.4 本章小结	47
第五章 李群核学习算法	48
5.1 SVM 算法原理	48
5.1.1 线性可分情况下的 SVM 分析	48
5.1.2 线性不可分情况下的 SVM 分析	49
5.2 KFDA 算法原理	50
5.3 核函数	51
5.4 李群核函数	53
5.4.1 李群 RBF 核	53
5.4.2 李群多项式核	54
5.4.3 李群线性核	55
5.4.4 李群感知器核	55
5.5 基于李群核函数的 KLieDA 算法	56
5.6 本章小节	57
第六章 实例应用	58
6.1 Lie-Fisher 算法实验分析	58
6.1.1 人工数据集	58
6.1.2 手写体识别	61

6.1.3 协方差李群特征的 Lie-Fisher 手写体识别	63
6.2 KLieDA 算法及李群核 SVM 实验分析	67
6.2.1 基于李群核 SVM 的手写体识别	67
6.2.2 基于李群核 KLieDA 算法的手写体识别	69
6.3 本章小结	72
第七章 结论与展望	73
参考文献	75
攻读学位期间参与的科研项目与发表（录用）论文情况	81
科研项目	81
发表（录用）的论文	81
致 谢	82

第一章 引言

1.1 课题研究意义

李群机器学习(Lie Group Machine Learning, LML)作为机器学习领域的一种新的学习方法,一方面继承流形学习的优点,另一方面借用李群的思想,形成了具有创新特色的学习范式。自2004年提出至今,已引起加拿大、爱尔兰、芬兰、意大利、美国等国内外同行的广泛关注。该方法和流形学习方法相比有明显优势,从李群的概念可以看出,它包含了微分流形和群的内容;微分流形包含了拓扑流形和微分结构。这套理论系统,既给我们提供了描述数据的几何表示方法,又给出了具体的代数求解方案。如:群保持了系统的完备性,微分提供了具体的代数计算方法,流形给出了几何表示方法,这正符合认知理论中的定性和定量表示相结合的认知模式。另外,从认知过程来看,人脑在认识客观世界中的任何对象时,首先关注的是表征问题的一个稳定点,然后依次进行分析图像的结构。对这种认知模式,在李群结构中,最小生成元就是这个稳定点,只要找到这个最小生成元,就可以利用李群方法对图像进行了分析。因此,李群机器学习既符合学习认知规律,又满足计算机解决现实问题的条件。

均值计算问题是机器学习中非常重要的一个子问题,现有大量的学习算法都离不开均值计算的过程。然而,大量的实际数据分布于几何结构为流形空间的李群中,常规学习算法中的均值计算在许多并非分布于欧氏空间的数据来说是不正确的,因此研究李群均值学习算法具有非常重要的意义。本课题是国家自然科学基金重点项目支持的研究内容之一,主要针对李群均值的性质、基于李群均值的分类器算法设计等问题进行研究。

1.2 李群学习研究进展

国内,由李凡长教授领导的李群机器学习研究小组在近些年做出了许多卓有成效的成果。例如,文献[61~63]给出了李群机器学习的4个公理,分别是(1)李群机器学习泛化能力假设公理;(2)李群机器学习的对偶假设公理;(3)李群机器学习的划分独

立性假设公理；(4) 李群机器学习一致性假设公理。并且按照文献[61~66]，可将李群机器学习模型分为代数模型和几何模型，这也就是李群机器学习方法和现有机器学习方法的主要区别之一，如流形学习、统计学习、支持向量机学习、关系学习、决策树学习、贝叶斯学习等。其中李群机器学习的代数模型将输入数据 R 看做变换，通过 $\theta: R \rightarrow G$ 变换成单参数子群 G ，利用 G 中的右平移变换映射 $R_g: R \times G \rightarrow G: (t, g) \rightarrow g \cdot \theta(t)$ ，它是流形 G 上的一个可微变换群。同样可得左平移变换。由此可以得到：{单参数子群} 和 {左不变流形} 是等价的。再通过左不变流形和左不变向量场之间的等价关系及单参数子群和李子代数之间的关系有 4 类等价系，即 $g \cong \{\text{左不变向量场}\} \cong \{\text{左不变流形}\} \cong \{\text{单参数子群}\}$ 。李群机器学习的几何模型主要涉及李群的一些几何性质，如平移不变性、旋转不变性、测地线性性质等。文献[67~68]介绍了李群机器学习中的 Dynkin 图的相关算法。另外，李群机器学习的分类器设计也是非常重要的一部分研究成果。例如，文献[69~72]分别给出了量子群、辛群分类器的设计。

2008 年，Risi Kondor 在他的博士论文《Group Theoretical Methods in Machine Learning》中就群在机器学习中的理论与应用作了详尽的探索和阐述，得出了许多有意义的结论^[1]。例如，得出了在群作用下核不变的对称性理论；从傅立叶谱的角度得出了平移不变正定核在群上的性质；从 spectral graph 理论的角度介绍了群上散度核的概念以及它们的含义；得出用以学习排列的 PerMceptron 算法；给出了一个基于傅立叶变换的身份管理和多目标跟踪的数学框架；介绍了非对称谱 (skew spectrum)，并证明其与双向谱 (bispectrum) 的等价性；从一种新的双向谱用以表示图像的旋转和平移不变性。可见，代数理论尤其是群论在机器学习中同样起到了很大的作用，而且文章最后指出了许多仍有待解决和深入探讨的问题，并大大看好代数学在机器学习领域中的发展前景。下面我们就李群在机器学习相关领域内的研究状况作一下小结。

李群用于维数约简领域：2003 年，由 Fletcher 证明了 m -rep 模型为李群^[2,3]，同时给出了李群样本的内均值算法，基于该李群内均值算法，给出了主测地线分析 PGA 降维算法。利用 PGA 算法，作者将用一组 m -rep 结点表示的 3D 医学器官模型进行降维。PGA

算法思想源于欧氏空间的 PCA 算法, 但它在李群所在的流形空间中进行降维, 是一种非线性降维算法。其主要思想是将 m -rep 李群样本映射到相应的李代数线性空间, 再计算其主成份方向。同年, Fletcher 等人又讨论了李群上的高斯分布在形状统计学中的应用^[9]。2008 年, M. Tournier 应用该 PGA 算法将运动数据进行压缩, 在压缩率达到 1: 50 以上时, 损失率仍能保持在 5% 左右, 而且每帧解压时间小于 1ms^[7]。可见李群在运动数据建模与压缩上的优势。

李群用于分类: 2006 年, 文献[20]提出将区域图像的协方差作为图像特征。协方差矩阵是正定对称矩阵, 构成张量空间, 文章推导了该张量空间中点与点之间的测地线距离计算公式, 并基于该度量给出了分类与检测算法。而文献[25]则从理论上证明这个对称正定阵构成的张量空间在附加一种运算后构成李群。并推导了该李群结构的许多性, 特别是推导出了在该结构下李群样本的均值计算公式, 这对后来对称正定阵作为图像特征的算法推导具有很大的意义。2008 年 Pradeep 通过对二维图像中的汽车点云构造李群样本, 并应用李群均值和 PGA 算法对构造的样本进行分类, 以达到对实际汽车车体分类的目的^[10]。通过实验表明, 基于李群的方法相比线性方法具有显著优势。由此可见, 李群在分类算法上尤有较好理论和应用价值。

李群用于解释原理与挖掘规律: 2002 年, Simone Fiori 在《Unsupervised Neural Learning on Lie Group》一文中深度探讨了李群和非监督神经网络的权值学习之间的联系^[6]。2007 年 Xu Miao 在《Learning the Lie Groups of Visual Invariance》中将二维图像的仿射变换构建在仿射李群框架之上, 并通过 EM 算法学习这个李群, 并在文中证明了基于图像的一阶泰勒展开式即为李群方法的特例^[26]。2008 年, Risi Kondor 探讨了快速傅立叶变换和李群之间的联系^[1]。2009 年, Dahua Lin 等三位来自 MIT 的学者用李群揭示了视觉动态现象, 探讨了李代数和线性动态过程的内在联系^[15]。2011 年, 文献[24]将李群应用于视频编码的领域, 通过李群生成元对频率图像帧间的局部变换进行学习, 以此对视频中帧间的临时相关性进行建模, 通过与标准编码比较的实验结果显示, 李群方法在高码率视频编码中具有更小的失真率, 在低码率视频中也仍具有很强的竞争性。

李群应用于视觉跟踪: 2002年, 由 Tom Drummond 和 Roberto Cipolla 利用李群对三维模型的变换进行建模, 研究了对复杂三维结构的实时可视化跟踪问题^[33]。随后 2004年, 由 Eduardo 等人研究了仿射变换结合投影变换构成的 8 维李群, 并应用于跟踪^[4,5]。2006 至 2010 年期间, Tuzel 等人连续在 CVPR 上发表多篇基于李群李代数研究视觉跟踪与检测问题中视觉不变特征估计与检测的文章^[16,19]。其间, 他们还就提出的图像协方差矩阵, 即正定对称矩阵所张成的张量空间构造为李群, 并提出相应的均值计算算法与分类和检测算法^[20], 取得了非常好的效果, 并被其它文章再次应用^[11,17,21], 得到了进一步发展。

均值计算是机器学习众多算法中必要的计算之一。然而在不同的空间和不同的数据集中均值的计算方法不尽相同, 特别是在李群这个非线性的流形拓扑空间中, 均值计算问题变得非常复杂。再由于李群, 特别是仿射群、特殊正交群以及正定对称群在众多问题中的重大应用价值, 使得李群的内均值计算也成为非常重要的研究问题。文献[32]详细分析了流形上的内均值和外均值的差异。文献[2,3]分析并给出了基于迭代方法的李群内均值计算方法。文献[9]分析了高斯分布在李群上的特性并应用到形状统计问题中。文献[12]详细分析了旋转群的均值问题, 并给出了优化后的计算方法。2004年, 文献[22]详细分析了由正定对称矩阵构成的张量群的均值问题, 并给出了正定对称阵的李群表示, 求解了正定对称阵的均值计算公式, 这对于正定对称阵应用广泛的理论具有非常重大的意义。例如上面述及的图像协方差特征表示即为对称阵形式, 2006年的文献[25]又在此基础上进行了概括与总结, 对正定对称阵的均值计算给出了完整的证明与描述。Fiori 在李群的均值计算方面也做了深入的探索^[13,14], 他从最原始的微分几何的角度, 即 Fréchet 均值理论出发, 推导了几种不同矩阵李群的均值计算方法: 1) . 对称正定矩阵群均值可直接在对应代数空间计算得到, 这和文献[21,25]的结果一致; 2) . 特殊正交群可通过迭代算法计算, 且计算结果与满足 Fréchet 定理的最小值是线性关系, 这和文献[2,3]给出的结果又不谋而合。

以上是对李群在相关领域的应用进行简要介绍, 下面对李群相关视觉不变特征研究进展进行分析。

1.3 李群相关视觉不变特征研究进展

视觉不变特征一直是计算机视觉研究的方向之一。而其视觉不变研究主要就是视觉不变特征的研究,是生物学和计算机视觉的基础性问题^[26]。计算机视觉中不变量的重要性早在 20 世纪 60 年代就已被提出^[42]。最初,强调的是光度量属性的不变性以及以此提供了发展出边缘检测算子的动力。视觉不变特征包括颜色不变特征、形状不变特征、几何不变特征、纹理不变特征、变换不变特征等等。

1999 年, Lowe 教授提出了著名的 SIFT 算法^[49],这是一种基于尺度空间的,对图像平移、旋转、缩放、甚至仿射变换保持不变性的图像局部特征。其主要思想是提取局特征,在尺度空间寻找极值点,提取位置,尺度,旋转不变量。该算法目前在军车、工业和民用方面都得到了不同程度的应用。典型应用如:图像匹配、视频跟踪、笔记鉴定、图像拼接、手势识别、三维建模等。虽然 SIFT 算法在视觉不变特征提取方面拥有非常好的优势,但并不完美,仍存在实时性不高、对边缘模糊的目标无法准确提取特征等缺点。自算法问世以来,不断对其进行发展与改进,先后提出了 PCA-SIFT^[74]、CSIFT^[75]、SURF^[76]、ASIFT^[77]等算法。其中 PCA-SIFT 算法通过 PCA 降维,可有效化简 SIFT 算子的 128 维描述子,由 Y. Ke 在 2004 年提出。CSIFT,即彩色尺度特征不变变换,可以针对彩色图像进行图像的不变特征提取,由 Farag 在 2006 年提出。SURF 即 Speeded Up Robust Features,被认为是 SIFT 的增强算法,其计算量小,运算速度快,提取的特征点几乎与标准 SIFT 相同,由 Bay 在 2006 年提出。ASIFT 抗反射 SIFT 变换,可以抵抗强仿射情况,提取的特征点远多于 SIFT 算法,由 J. M. Morel 于 2009 年提出。

2005 年,Navneet Datal 和 Bill Triggs 在文献[34]中提出了著名的 HOG 图像特征。HOG 特征类似于 SIFT 特征,作者将该算法用于识别图像中的行人,得到了非常好的效果。Datal 在 2006 年的博士论文^[36]中又对该算法进行了更为详细深入的研究,并扩展到视频流中的人物检测。之后 HOG 算法被用于许多识别与检测问题中,但同时标准的 R-HOG 和 C-HOG 特征的计算比较耗时,因此在 HOG 算法的基础之上,衍生出一些新的算法,如文献[35]提出的 LPP-HOG,用于快速行人检测;2008 年, Tuzel 等人在文献[19]中采取去掉标准 HOG 中相邻的 block 之间的重叠,以此减少大量的计算开销,再利

用李群李代数在仿射变换建模与学习中的优势, 将这类 HOG 特征用于图像跟踪和检测, 将李群作用在跟踪图像上, 构建一个运动变换空间, 在该空间中计算相应的类 HOG 特征, 实现了高精度、实时的跟踪算法和仿射人脸识别算法, 效果尤为卓越。2009 年, Mohamed 等人又将 HOG 应用于手势跟踪^[46]。

2006 年, Alessio Plebe 利用层次的 SOM 自组织映射神经网络来学习同一对象在不同视角下的不变量^[27], 从实验结果给出的图像中可以看出同一对象即使是在有 30 度至 60 度的视角差异, 一般的算法已很难识别时, 该方法仍能学到同一视觉对象的不变特征。同年, 文献[37]也应用了类似竞争型神经网络的方法, 提出“连续变换学习”的框架, 其本质上是一种将空间上相似的模式映射到竞争学习系统中相同突触后的神经元的工作模式, 随着输入模式沿着连续变换(例如平移、旋转等)的空间中运动, 活动突触就被调整到突触后的神经元的集合上, 因为相同刺激的其它变换会被先前学习的样例覆盖, 一个常规突触后神经元集合就会被新变换激活, 于是在相同的突触后神经元上学习新的活动样例就变得便利可行了, 最后就完成了同一对象不同变换下的不变量的学习。2008 年, Charles 等人的两篇文章^[28, 38]也是应用神经网络, 提出了一种层次、概率的产生式模型, 从自然图像与视频中学习变换不变量, 取得了非常好的效果。

同是 2006 年, 由 Tuzel 等人首先提出了用图像的协方差矩阵作为图像不变特征^[20], 并用于图像分类与对象检测。图像的协方差矩阵计算不仅速度快, 而且是一种维数低, 又能通过积分图像方式大大降低图像任意子图协方差特征的计算开销, 最关键的是能很好表征图像性质的一种视觉不变特征。由于图像协方差矩阵的这些优势, 同年 Tuzel 等人又将其应用于实时跟踪^[16], 取得了非常好的效果。随后, 通过加入几何粒子过滤^[17], 在该论文的基础上加以改进和增强, 达到了更好的效果。2007 年, Tuzel 等人将协方差应用于行人检测, 并提出了一种新的分类方法^[23], 将该方法应用于 INRIA 行人数据库, 与著名的 HOG 特征方法^[31]进行对比, 误分率明显低于后者。2008 年, 文献[49]应用图像协方差特征结合 AdaBoost 算法提出了快速行人检测的算法框架。文献[50]则利用协方差特征的对称特性, 将有效元素拉成向量后应用 SVM 进行场景分类。2011 年, Ando 等人继续将协方差特征用于识别实况足球电视中运动员制服上的背号。

2007年, Xu Miao 等人在文献[26]中提出了基于李群理论的非监督方法用来学习视觉不变量。不同于传统的为了提取不变量而损失部分变换信息的方法, 对不同图像中经过仿射变换后的同一对象进行李群仿射运动建模, 并对模型中的李群变换生成元参数使用 EM 算法进行迭代估计。以前的基于图像一阶泰勒级数展开式被证明为是李群方法的特殊形式。利用图像的基于矩阵指数泛化模型, 能够处理任意大尺度的变换。论文中的实验表明, 通过包含 6 类仿射变换的人工数据集学到的李生成元非常符合分析预测的仿射生成元。同时作者还演示了该算法可以从自然图像序列中发现新的变换生成元。所学到的生成元可同时被用于图像的生成和估计, 也因此提供了获取视觉不变量的一个基础。

在视觉不变量研究中, 最早也一直保持着作为研究热点的就是图像不变矩的研究, 早在 1962 年, Hu 就将矩不变量用于模式识别^[39]。2010 年, Gianfranco Doretto 等人提出了一种新的视觉不变特征, 即区域矩^[18], 与上文指出的区域协方差特征进行了全面的比较, 除区域矩是欧氏空间的特征, 可以使用传统欧氏几何中的线性分类器, 而无需做相对较大的开销用于计算测地线距离外, 在航拍视频的小图像检测的实验中分别在质量、数量上也优于后者, 也是一种计算快速, 抗平移和旋转不变性好, 可以被广泛推广到其他应用中的视觉不变特征。

1.4 问题的提出

显然, 李群在处理变换问题具有先天优势。例如, 图形图像变换中最为常见的 6 种二维仿射变换矩阵可以构成仿射群 $A(2)$, 图形图像在二维空间中的旋转变换矩阵可以构成二维特殊正交群 $SO(2)$, 三维空间的旋转变换矩阵可以构成 $SO(3)$, 物理、图像和信号处理中应用广泛的协方差矩阵可以构成 n 维对称正定矩阵群 $Sym^+(n)$, 以及最为常见的 n 维实空间 R^n , n 维酉群 $U(n)$ 以及特殊酉群 $SU(n)$ 。其它常用的还有 $SL(n)$ 、 $GL(n)$ 、 $Sp(n)$ 等。由此可见, 李群学习方法在视觉不变特征的研究领域的应用越来越广, 但仍存在下述许多问题。

视觉不变特征方面具有代表性的 SIFT、HOG、Region Covariance、Region Moment 等算法虽然取得了长足进展，但和人类视觉特征相比还有相当大的差距。近年来李群尽管也在这个领域发挥过作用，但总体还有许多问题要做，例如：

文献[2, 3, 9, 10, 20]提取李群特征作为训练及测试样本，但其相应的算法本质上是将这些样本嵌入到线性欧氏空间中进行处理。而应用李群学习模型进行处理的研究中，其所处理的特征往往又不是李群样本^[1, 5, 8, 11, 19]，而是亮度、HOG 等各种特征，这并不是说这些特征不好，而是引发了一个问题：如果采用了李群作为学习算法的理论，如果能同时处理李群的样本，是否更加合理，并且也可以更好得发挥李群作为学习算法的优势呢？

不论是图像的区域协方差特征^[20]还是 HOG 特征^[34]都属于统计特征，而统计特征是对出现的训练样本集使用统计的方法进行分布规律的估计^[79]或是求 MAP^[79]，如果训练样本集比较稀疏或分布不均，不能表示样本的分布规律，那么统计方法也将失去用武之地。而文献[27, 28, 37, 38]通过神经网络来训练视觉不变量，学习到的不变量需要一个复杂的网络来表示，没有统一的数学形式，因此如果想对学习到的不变特征用于其他用途可能会受到一定的限制。

文献[10]提出了一种对二维图像的点云构造李群的算法，应用此方法可以对任何二维图像进行图像特征李群化，但普通图像使用该方法进行点采样来构造李群，效率非常低，而且如果采样点没有覆盖有效的图像区域或过于稀疏，该李群特征势必不能很好表达图像所具有的用与其它图像判别的性质。

仿射李群 $A(2)$ 包括了 6 种基本变换，但不能处理透视变换，因此基于仿射变换的研究^[4, 5, 10, 15, 19, 26]只能处理不具有透视效果或弱透视的问题。而视觉信息的采集，尤其是图像、照片等往往包含透视效果。

1.5 本文内容安排

本文重点分析李群学习算法，从李群均值学习的理论及其在视觉不变特征中的应用两个侧重点来详细分析和探讨。全文的主要内容安排如下：

第一章，引言。介绍并分析了最近李群学习算法以及视觉不变特征的研究现状，并对这些算法进行了分析比较，对它们的优缺点作了简要的概括，针对这些算法的特点，提出了一些疑问。

第二章，矩阵群相关理论及学习算法。介绍矩阵群的相关概念，给出矩阵群相关学习算法，并通过实例，分析了矩阵群学习算法的性能。

第三章，李群学习相关理论。在第二章的基础上，将矩阵群上升为矩阵李群，先介绍矩阵李群的相关理论基础，介绍常见常用的各矩阵李群，并给出这些矩阵李群的理论性质，分析这些李群高斯分布的形式和性质。再对几种常用的李群学习算法进行理论上的详细分析，最后通过实验来验证算法的性能与相关指标

第四章，李群均值相关学习算法及应用。本章首先对样本集的内均值与外均值作了详细的分析。根据欧氏空间中的 Fisher 向量投影思想，探索了李群流形空间中的测地线投影问题，并基于李群均值计算算法，设计实现了两个 Lie-Fisher 算法。

第五章，李群核算法研究及应用。本章将结合核方法，对 KFDA 和 SVM 算法进行深入解析，并结合矩阵李群理论，推导设计了基于李群矩阵样本的核函数，进一步设计实现李群核 KLieDA 算法。

第六章，实例应用。本章将对文中设计的算法应用于几何形状分类与手写体分类的实例中。本章中还设计了一个手写体李群特征提取方法。通过大量的实验图表，验证了两个 Lie-Fisher 算法以及基于李群核的 KLieDA 和 SVM 算法的优良性能。并进一步分析了手写体的协方差特征蕴含的偏向于线性可分的特性。

第七章，总结与展望。这一章将对全文进行总结，并对李群在视觉不变特征中的理论与应用的研究进行展望。

第二章 矩阵群学习算法

本章将讨论矩阵群的相关理论，并分析基于矩阵群学习算法的相关技术。矩阵群是矩阵李群的理论基础，更是矩阵李群均值学习算法的关键内容，内容安排如下：2.1 节介绍矩阵群相关基本概念，包括相关定义、度量空间以及实例。2.2 节介绍矩阵群学习算法，设计了基于矩阵样本的 KNN 和 K-means 算法。2.3 节应用 2.2 节的算法进行实验，为了数据可视化，样本采用 2 维矩阵。2.4 节为本章小结。

2.1 相关基本概念

如果 k 表示一个（交换）域。在绝大多数情况下，我们关心的这个域为 $k = \mathbf{R}$ （实数）或 $k = \mathbf{C}$ （复数）。本文只涉及实数域，即 $k = \mathbf{R}$ 。别外，在一般的机器学习问题的框架下，往往会使框架配上合适的范数，因此本节也会对矩阵范数作简单的讨论，这将在 2.1.2 节作介绍。

2.1.1 矩阵群的定义

命 $M_{m,n}(k)$ 为 $m \times n$ 的矩阵，其每一个元素都属于 k 。用 A_{ij} 或 a_{ij} 表示 $m \times n$ 阶矩阵中的第 (i, j) 个元素，于是可以写成：

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

我们也可以使用特殊的表示方法： $M_n(k) = M_{n,n}(k)$ ， $k^n = M_{n,1}(k)$ 。 $M_{m,n}(k)$ 是一个 k -向量空间，该空间附带矩阵的加法与标量乘法。零向量表示为 $O_{m,n}$ ，即为一个 $m \times n$ 阶零元素构成的矩阵，也可以简单用 O 来表示，其具体维数大小视所出现的上下文而定。矩阵：

$$(E^r)_{ij} = \delta_r \delta_{ij} = \begin{cases} 1 & \text{if } i=r \text{ and } j=s, \\ 0 & \text{otherwise,} \end{cases}$$

构成了 $M_{m,n}(k)$ 的一组基, 因此, k -向量空间的维数是:

$$\dim_k M_{m,n}(k) = mn$$

当 $n=1$ 时, 将 $k^n = M_{n,1}(k)$ 的标准基向量表示为:

$$e_r = E^{r1} \quad (r=1, \dots, m)$$

一旦把问题限定在维数为 n^2 的 k -向量空间中, 那么 $M_n(k)$ 就是一个代数环, 该环具有两个分别为矩阵加法与方阵乘法的运算, 并且有零元 $O_n = O_{n,n}$, 以 $n \times n$ 阶单位矩阵 I_n 为幺元, 除非 $n=1$, 否则 $M_n(k)$ 是不可交换的。 $M_n(k)$ 也是有限维 k -代数的一个重要的实例。

在矩阵论中, 矩阵行列式是一个映射 $\det: M_n(k) \rightarrow k$, 它拥有如下几个性质:

- (1) 对任意 $A, B \in M_n(k)$, 有 $\det(AB) = \det A \det B$ 。
- (2) $\det(I_n) = 1$ 。
- (3) 当且仅当 $A \in M_n(k)$, $\det A \neq 0$ 时, A 是可逆的。

下面给出两个非常重要的矩阵群, 即一般线性群 $GL_n(k)$:

$$GL_n(k) = \{A \in M_n(k) : \det A \neq 0\}$$

以及特殊线性群 $SL_n(k)$:

$$SL_n(k) = \{A \in M_n(k) : \det A = 1\} \subseteq GL_n(k)$$

集合 $GL_n(k)$ 和集合 $SL_n(k)$, 在矩阵乘法下构成群, 更进一步, 有 $SL_n(k) \leq GL_n(k)$, 即 $SL_n(k)$ 是 $GL_n(k)$ 的一个子群。

本文在后述应用中的矩阵群都是 $GL_n(k)$ 的子群, 而且是 $GL_n(R)$ 的子群。

2.1.2 矩阵群度量空间

在本节, 我们将在 $M_n(k)$ 上定义一个范数 $|||$ 。值得注意的是, 我们选择一个具体的范数主要是为了乘法性质的便利性。实际上, 其它任何一个向量空间的范数都将在 $M_n(k)$ 上给定一个等价的度量拓扑。 $M_n(k)$ 上具很多有价值的范数, 详细内容可以参考文献[45]。

度量中最基本的要素是空间中向量的长度, 定义一个向量 $x = [x_1, \dots, x_n]^T \in k^n$ 长度为:

$$|x| = \sqrt{|x_1|^2 + \dots + |x_n|^2} \quad (2.1)$$

这是向量空间 k^n 中众多范数中的一个可选范数。对于任意 $A \in M_n(k)$, 考虑如下集合:

$$S_A = \left\{ \frac{|Ax|}{|x|} : 0 \neq x \in k^n \right\}$$

有了子集:

$$S_A^1 = \{ |Ax| : x \in k^n, |x| = 1 \} \subseteq S_A$$

如果 $x \neq 0$, 有:

$$\frac{|Ax|}{|x|} = |Ax'|$$

其中 $x' = (1/|x|)x$, 长度 $|x'| = 1$ 。子集 $\{x \in k^n : |x| = 1\} \subseteq k^n$ 是封闭且有界的, 因此也是紧致的, 那么实值函数:

$$\{x \in k^n : |x| = 1\} \rightarrow \mathbb{R}; x \mapsto |Ax|$$

是有界的, 可得其上确界为:

$$\sup S_A = \sup S_A^1 = \max S_A^1 = \max S_A$$

这意味着定义了如下实数:

$$\|A\|_{\mathbb{R}} = \max S_A = \max S_A^1$$

此处的范数 $\|\cdot\|: M_n(k) \rightarrow \mathbb{R}$ 被称为 $M_n(k)$ 上的操作或上确界范数。更一般的 k -代数上范数的相关内容可参考文献[44]的定义 4.31。

矩阵空间中的范数有很多, 但不是任意一个 $M_n(k) \rightarrow \mathbb{R}$ 映射都能被用作范数, 因为映射 $\|\cdot\|: M_n(k) \rightarrow \mathbb{R}$ 必须满足下列性质:

- (i) 如果 $t \in k, A \in M_n(k)$, 那么 $\|tA\| = |t|\|A\|$ 。
- (ii) 如果 $A, B \in M_n(k)$, 那么 $\|AB\| \leq \|A\|\|B\|$ 。
- (iii) 如果 $A, B \in M_n(k)$, 那么 $\|A+B\| \leq \|A\| + \|B\|$ 。
- (iv) 如果 $A \in M_n(k)$, 那么 $\|A\| = 0$ 当且仅当 $A = 0$ 。
- (v) $\|I_n\| = 1$ 。

矩阵群中应用最为广泛的范数是 Hilbert-Schmidt 范数, 它在欧氏空间也叫做 Frobenius 范数:

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} \quad (2.2)$$

很明显, Frobenius 范数满足上述 5 条性质。以后不导致含意混淆, 就用 $\|A\|$ 代替 $\|A\|_F$ 。关于范数理论的其它内容可以参考文献[44, 45]。

2.1.3 矩阵群实例

一个群 $G \leq GL_n(k)$ 同时也是 k -矩阵群, 为了方便, 通常称 G 为 $GL_n(k)$ 的子群。在介绍矩阵群的一些常见的实例之前, 我们先来看看它们的一些性质。

(1) 若 $G \leq GL_n(k)$ 是一个矩阵子群, 那么封闭子群 $H \leq G$ 是 $GL_n(k)$ 的一个矩阵子群。

(2) 若 G 是一个矩阵子群, 如果有 $H \leq K, K \leq G$, 那么 H 是一个 G 的矩阵子群。

以上性质的证明从略, 具体参考文献[44]第 16 页。

给定一个矩阵子群 $G \leq GL_n(k)$, 限制 $GL_n(k)$ 上的行列式为如下映射:

$$\det_G : G \rightarrow k^*; \det_G A = \det A$$

不导致理解上的歧义, 以后均用 \det 代替 \det_G , \det_G 是一个连续的群同态。当 $k = \mathbb{R}$ 时, 约定:

$$R^+ = \{t \in R: t > 0\}, R^- = \{t \in R: t < 0\}, R^* = R^+ \cup R^-$$

注意到这里 R^+ 是 $GL_1(R) = R^*$ 的一个子群, 且是个开集, 而 R^- 也是个开集, 因此 R^+ 和 R^- 是闭开子集, 也就是封闭的开集。对于 $G \leq GL_n(k)$, 有

$$\det_G^{-1} R^+ = G \cap \det^{-1} R^+$$

以及

$$G = \det_G^{-1} R^+ \cup \det_G^{-1} R^-$$

因此 G 是以下两个闭开子集的非交并集:

$$G^+ = \det_G^{-1} R^+, G^- = \det_G^{-1} R^-$$

由于 $I_n \in G^+ = \det_G^{-1} R^+$, 因此 G^+ 将永不为空。确实, G^+ 是 G 的一个封闭子集, 也因此是 $GL_n(R)$ 的一个矩阵子群。当 $G^- \neq \emptyset$, 空间 G 是非连通的, 这是因为它是两个不相交了集的并集。当 $G^- = \emptyset$, $G = G^+$ 时, 可能连通也可能不连通。

本文研究的矩阵群及涉及到的数学域都是实数域, 即 $k = R$ 。这在本节开始已经有说明, 于是下面的符号中将不再出现域描述符 k , 默认域为实数域 R 。经此约定后, 可使符号简化以便理解: 去掉域符号 k , 再将维数符号 n 移到原 k 的位置。例如用 $GL(n)$ 代替原符号 $GL_n(R)$, 用 $M(m, n)$ 代替 $M_{m,n}(R)$, 用 $M(n)$ 代替 $M_n(R)$ 。

2.1.3.1 正交群 $O(n)$ 、特殊正交群 $SO(n)$

一个 $n \times n$ 矩阵 A 是正交矩阵, 当其每一列向量是标准正交的, 即:

$$\sum_{l=1}^n A_{lj} A_{lk} = \delta_{jk}, 1 \leq j, k \leq n$$

这里的 δ_{jk} 称作 Kronecker delta, 当 $j = k$ 时 $\delta_{jk} = 1$, $j \neq k$ 时 $\delta_{jk} = 0$ 。另一种描述是: 当 A 能保持内积时是正交阵, 即对于 R^n 中所有的向量 x 和 y , 有 $\langle x, y \rangle = \langle Ax, Ay \rangle$, 尖括号表示向量内积, 即 $\langle x, y \rangle = \sum_k x_k y_k$ 。另一个定义可能更加直接一点: $A^T A = I$, 即 $A^T = A^{-1}$ 。

由于 $\det A^T = \det A$, 当 A 是正交矩阵时, 可得 $\det(A^T A^{-1}) = \det(A)^2 = \det I = 1$ 。因此, 对所有的正交矩阵, 有 $\det A = \pm 1$ 。

这个公式说明正交矩阵一定是可逆矩阵。于是, 对正交矩阵 A 有:

$$\langle A^{-1}x, A^{-1}y \rangle = \langle A(A^{-1}x), A(A^{-1}y) \rangle = \langle x, y \rangle$$

更进一步, 两个正交矩阵的积也是正交的, 即两个正交阵 A 和 B 的内积 AB 是正交阵。那么, 所有 $n \times n$ 正交矩阵集合将构成一个群, 即正交群 $O(n)$, 它是 $GL(n)$ 的一个子群。

所有 $n \times n$ 阶并且行列式值为 1 的正交矩阵集合构成特殊正交群 $SO(n)$, 它是 $O(n)$ 的子群, 因此也是 $GL(n)$ 的子群。从几何的角度来看, $O(n)$ 的元素要么是旋转变换, 要么是旋转和反射两种变换的合成, 而 $SO(n)$ 的元素只有旋转。

2.1.3.2 酉群 $U(n)$ 、特殊酉群 $SU(n)$

一个 $n \times n$ 复矩阵 A 是酉矩阵, 当其每一列向量是标准正交的, 即:

$$\sum_{l=1}^n \overline{A_{lj}} A_{lk} = \delta_{jk}, \quad 1 \leq j, k \leq n$$

与 2.1.3.1 小节同理, 当 A 能保持内积时是酉矩阵。对于 C^n 中所有的向量 x 和 y , 有 $\langle x, y \rangle = \langle Ax, Ay \rangle$, 尖括号表示向量内积, 即 $\langle x, y \rangle = \sum_k \overline{x_k} y_k$ 。另一个定义可能更加直接一点: $A^* A = I$, 即 $A^* = A^{-1}$, 这里的 A^* 是 A 的伴随矩阵: $(A^*)_{jk} = \overline{A_{kj}}$ 。

由于 $\det A^* = \overline{\det A}$, 当 A 是酉矩阵时, 有 $\det(A^* A) = |\det A|^2 = \det I = 1$ 。因此, 对所有的正交矩阵, 有 $|\det A| = 1$ 。

这说明, 任何一个酉矩阵均是可逆的。与正交群类似, 所有 $n \times n$ 酉矩阵构成酉群 $U(n)$, 它是 $GL(n; C)$ 的子群。

所有 $n \times n$ 阶并且行列式值为 1 的酉矩阵集合构成特殊酉群 $SU(n)$ ，一个酉矩阵对任意的 θ 有行列式值 $e^{i\theta}$ ，因此相比 $SO(n)$ 对 $O(n)$ 的关系， $SU(n)$ 是比 $U(n)$ 更小的子集。

2.1.3.3 对称群 $Sym(n)$ 、对称正定群 $Sym^+(n)$

一个 $n \times n$ 矩阵 A 是对称矩阵，当且仅当它是下列集合中的元素：

$$Sym(n) = \{A \in M(n); A = A^T\}$$

而一个 $n \times n$ 矩阵 A 是正定对称矩阵，当且仅当它既是对称矩阵又是正定矩阵：

$$Sym^+(n) = \{A \in Sym(n); A > 0\}$$

按照普通的矩阵乘法规则， $Sym(n)$ 可以构成向量空间，而 $Sym^+(n)$ 只能张成一个流形空间，却不能构成向量空间，而且在常规矩阵乘法下不具有群结构。文献[25]指出，只要在上述集合上构建一种新的运算：

$$A_1 \odot A_2 = \exp(\log(A_1) + \log(A_2)) \quad (2.3)$$

便可构成一个群 (Sym^+, \odot) ，而且是李群，其中 \exp 和 \log 为矩阵的指数运算和对数运算，在李群理论中，分别被称为指数映射与对数映射，将在后面章节介绍。

上面主要介绍了几个常见的矩阵群，其中的 $O(n)$ 、 $SO(n)$ 、 $U(n)$ 、 $SU(n)$ 都是李群， $Sym^+(n)$ 在附加了新的矩阵运算后也可以构成李群，详细的证明可以参考文献[43]和[25]。

2.2 矩阵群学习算法

上节介绍了矩阵群的基本概念和矩阵群空间度量相关的知识，本节将在此基础上构造基于矩阵群向量空间的 KNN 算法及 K-means 算法。

$M_{m,n}(k)$ 是一个 k -向量空间，该空间附带矩阵加法与标量乘法，对该向量空间附加一个度量算子： $\rho(A, B)$ ，其中 $A, B \in M_{m,n}(k)$ ，该度量算子满足范数映射的 5 条性质。

具体的范数表达形式视具体情和具体的矩阵群而定，在 2.3 节我们会有实例分析进行演示。

矩阵群构成的空间中有了度量就有了距离，对某一具体问题，只要使用的度量合理，就可以应用这个度量来解决很多问题。作为机器学习问题中最基础的分类与聚类算法，下面将它们扩展到矩阵群空间中，对两个矩阵在空间中的距离用相关度量来表示。

算法 2-1: 基于矩阵群的 KNN 算法

输入：已知 c 个类别的 n 个矩阵样本 $\{x_i\}_1^n$ ，以及 m 个未知类别的样本 $\{y_i\}_1^m$ ，近邻个数 k 。

输出：未知样本集 $\{y_i\}_1^m$ 中每一个样本的具体类别。

过程：

1. 按某一取法从 m 个未知类别的样本集中取一样本 y_i 。
 2. 计算 y_i 与 $\{x_i\}_1^n$ 所有样本之间的度量 $\rho(y_i, x_j)$ 。
 3. 按第 2 步得到的度量升序排列，取前 k 个最小的度量对应的 $\{x_i\}_1^k$ 。
 4. 取 $\{x_i\}_1^k$ 中所属类别样本个数最多的类别，赋予样本 y_i 。
 5. 重复以上步骤，直至处理完每一个未知类别样本集 $\{y_i\}_1^m$ 中的样本。
-

上面的算法 2-1 给出了矩阵群空间的 KNN 算法，矩阵群空间的 K-means 算法由算法 2-2 给出。

算法 2-1 和算法 2-2 的思想利用了矩阵群的理论，方便地把传统向量空间的算法扩展到矩阵群空间中，利用这个思路，还可以构造更多矩阵样本集的算法。从时间复杂度和空间复杂来看，也并没有增加额外的开销。

算法 2-2: 基于矩阵群的 K-means 算法

输入: 矩阵样本集 $\{x_i\}_1^n$, 以及类别数 k

输出: 样本集 $\{x_i\}_1^n$ 中每个样本的类别归属

过程:

1. 在样本集对应的矩阵空间中随机生成 k 个样本中心 $\{z_j^i\}_1^k$, 其中 z 的下标表示第 j 轮迭代, 上标取值为 $1, \dots, n$, 表示第 i 个中心。
 2. 取样本集 $\{x_i\}_1^n$ 中一样本 x_i 到每个样本中心 $\{z_j^i\}_1^k$ 的距离 $\rho(x_i, z_j)$ 。
 3. 根据第 2 步得到的 k 个距离 $\{z_j^i\}_1^k$, 选最短距离的中心为 x_i 的类别。
 4. 重复 2~3 步, 直至处理完 $\{x_i\}_1^n$ 中每一个样本。
 5. 根据上述步骤得到样本集 $\{x_i\}_1^n$ 的类别归属, 重新计算每类中心 $\{z_{j+1}^i\}_1^k$ 。
 6. 如果 $\{z_{j+1}^i\}_1^k$ 和 $\{z_j^i\}_1^k$ 中对应的每个中心都相等, 则算法终止, 不然重复 1~5 步。
-

2.3 实例分析

这部分将通过几个实例来演示 2.2 节中两个算法的简单应用。第一个实验是将 KNN 算法应用于矩阵样本的分类问题, 第二个实验是将 K-means 算法应用于矩阵样本的聚类问题。

2.3.1 KNN 矩阵样本分类

首先人工生成 3 类矩阵样本集, 每类样本都是二维高斯分布下的随机样本, 只是每类高斯样本的均值不同, 于是样本是 $M(1,2)$ 矩阵群的子集。下面是二个维度上各均值分别取 $(0, 0)$ 、 $(4, 4)$ 、 $(2, 2.5)$, 方差均取 $\sigma^2 = 1$, 且每类样本各 200 个, 一

共 600 个矩阵样本，每类 200 个中分别取 100 个作为训练样本，另 100 个作为测试样本。那么 300 个训练样本的分布情况如下图 2-1 所示。

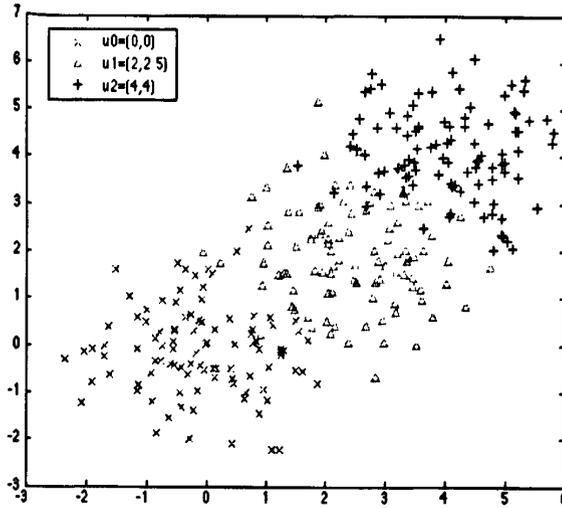


图 2-1. 3 个均值为 $(0, 0)$, $(2, 2.5)$, $(4, 4)$ 的二维高斯分布样本集，分别用 3 种不同符号和颜色表示，每个类别的样本个数均为 100。

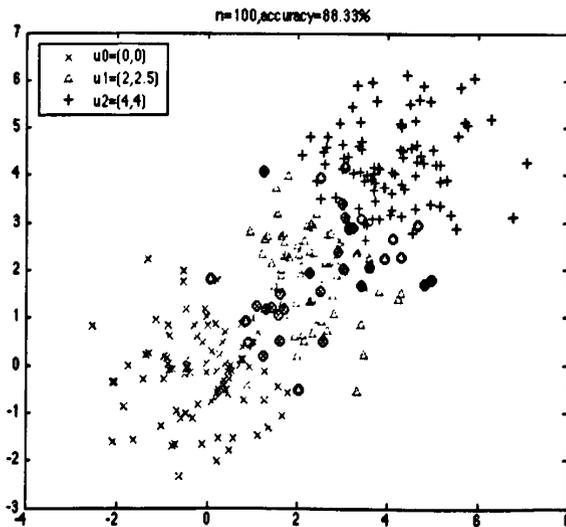


图 2-2. 对 3 类与图 2-1 同样参数配置下生成的二维高斯分布样本集采用 5-NN 分类结果，错误分类的样本用小圆圈圈出。

接下来,我们对 300 个测试样本执行算法 1-1 的过程。其中,参数取 $k=5$,而用于度量的范数取 $\rho(A,B)=\|A-B\|_F$, $A,B \in M(1,2)$,在对测试样本执行结果见图 1-2。图中为了对测试样本的类别归属情况看得更清楚,对训练样本不作显示,只显示测试样本,对于分类错误的样本用黑色小圆圈标记出来。

本实验里的样本属于 $M(1,2)$ 矩阵群的子集,其所在空间是 R^2 ,这是典型的向量空间。同样的方法我们还可以用于任意维数的矩阵群,这些矩阵群所构成的空间不一定是向量空间,可能是非线性的流形空间,而线性向量空间在理论上是流形空间的一种特殊情况。因此,本文将重点放在适应性更广的非线性流形空间上。

2.3.2 K-means 矩阵样本聚类

为了使矩阵样本数据的可视化,对 K-means 的矩阵聚类实验,仍采用二维 $M(1,2)$ 群的矩阵子群样本,实验使用算法 2-2 对 $M(1,2)$ 矩阵群进行聚类。

算法 2-2 中的样本集 $\{x_i\}_1^n$ 被实例化为 $M(1,2)$ 矩阵群,矩阵群空间的度量函数仍采用 $\rho(A,B)=\|A-B\|_F$, $A,B \in M(1,2)$,对于初始均值也随机取样于 $M(1,2)$ 空间。对于聚为一类的新样本集均值采用其所在线性空间的线性均值空间的计算公式:

$$\mu = \frac{1}{n_c} \sum_{i=1}^{n_c} x_i \quad (2.4)$$

其中 $x_i \in \{x_i\}_1^{n_c}$, n_c 表示第 c 类的样本数。因为 $M(1,2)$ 群所在空间为线性空间,因此采用式 (2.4) 是合适的。

实验首先生成不同均值的二维高斯分布 $M(1,2)$ 样本 500 个,分布形状为字母 ‘W’ 型,然后对这 500 个样本进行分别 $k=1, 2, 3$ 类的聚类。从实验结果图 2-3 中看出聚类的结果是符合样本在空间中的分布状态的,即保持同类聚集在一起。

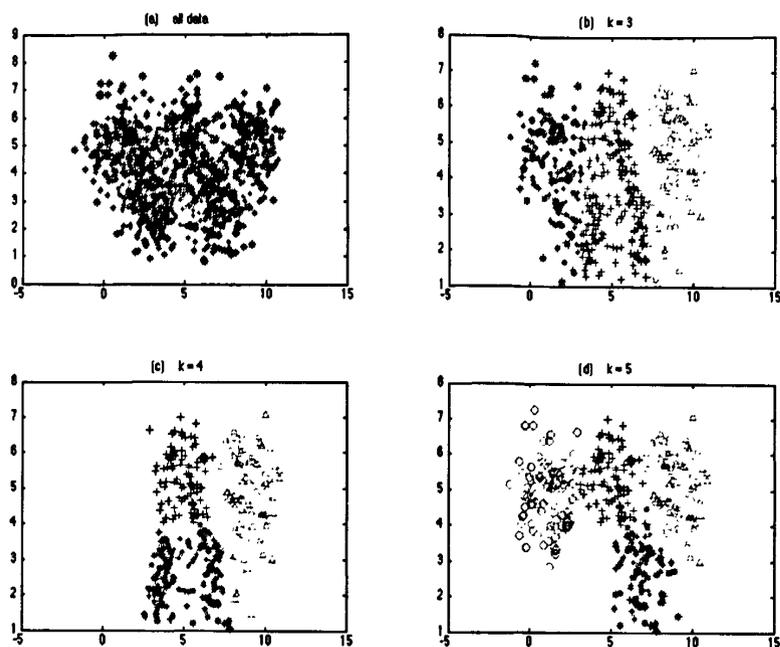


图 2-3. (a) 为 500 个 ‘W’ 型原始样本的分布情况。(b)~(d) 为对 (a) 的原始样本分别采用 $k=1, 2, 3$ 的聚类结果。

当实际样本分布空间不是向量空间，而是流形空间时，只需针对该流形空间的样本，设计新的度量函数 ρ ，就可以将算法应用在具体问题中。另外，以往大部分学习算法中的特征向量表示为 $M(1,n)$ 形式，是 $M(m,n)$ 的特征情况，这说明矩阵群学习算法在理论上包含了向量学习算法。而且矩阵在结构上拥有行和列双向结构，比向量的单向结构能表示更多的信息，Wang 等人已在矩阵样本集上作了一些研究工作^[29]，文献[30]将 PCA 与 FDA 算法扩展到了矩阵样本上。

2.4 本章小结

本章第一小节讨论矩阵群的相关理论并介绍了几个常用矩阵群。第二节以矩阵群为理论基础,给出了矩阵群的两个学习算法,分别是监督学习 KNN 和非监督学习 K-means 算法。第三小节对第二小节的两个算法分别给出了实例。

由于矩阵群的数学理论非常庞大,本章只对最基本的矩阵群定义和定理作了介绍,而且本文主要讨论的是它的一个子集,即李群。详细的李群理论及相关算法将在下一章进行介绍,在 2.1.3 节最后部分我们介绍了对称正定矩阵,该矩阵集在普通矩阵乘法下不具有群结构,但在附加了一个特殊运算后,构成李群 (Sym^+, \odot) 。

第三章 李群学习相关理论

上一章，我们介绍了矩阵群的基本理论，并给出了几个常用矩阵群的例子，目的是为了引出李群的相关理论。本章 3.1 节介绍矩阵李群，先给出矩阵李群的相关理论基础，包含代数群和微分几何两个层面的性质。3.2 节介绍常用矩阵李群的例子，给出这些矩阵李群的理论性质。3.3 节分析了几个常用李群在高斯分布下的密度函数形式，并讨论了它们在高斯分布下以最大似然估计来计算均值和方差的方法。3.4 给出李群均值计算算法。3.5 节为本章小结。

3.1 李群

李群是一种特殊的群结构，满足以下条件的 G 被称为李群：

(i) G 是一个拓扑群。

(ii) G 是一个微分流形。

(iii) G 的群运算及其逆运算是可微的，即： $G \times G \rightarrow G: (X, Y) \mapsto XY$ ，以及逆映射 $G \rightarrow G: X \mapsto X^{-1}$ 是可微的，其中 $(X, Y \in G)$ 。

在李群 G 的单位元 e 处的切空间 $T_e G$ 构成一个向量空间，称为李代数 \mathfrak{g} 。其上定义了一个二元运算符： $[\cdot, \cdot]: \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ ，称为李括号，符合以下条件：

(i) 双线性： $[\lambda X + \mu Y, Z] = \lambda[X, Z] + \mu[Y, Z]$ ($\forall \lambda, \mu \in G; \forall X, Y, Z \in \mathfrak{g}$)。

(ii) $[X, X] = 0, \forall X \in \mathfrak{g}$ 。

(iii) 雅可比 (Jacobi) 恒等式成立，即：

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0 \quad (\forall X, Y, Z \in \mathfrak{g})。$$

李群作为流形，其内任意两个元素之间的距离可通过连接这两个元素的测地曲线的长度来度量。从单位元出发，沿 $X \in \mathfrak{g}$ 方向存在唯一的一条测地线。

矩阵李群作为一种特殊的李群还具有其他一些特性。例如设 $G \subset GL(n, R)$ ， G 的李代数就是单位矩阵 I 处的切空间，并且李括号满足 $[X, Y] = XY - YX$ 。

指数映射是李群中另一个重要的理论, 定义 $\exp: \mathfrak{g} \rightarrow G$:

$$e^X = \sum_{n=0}^{\infty} \frac{1}{n!} X^n \quad (3.1)$$

它将李代数元素映射到李群中的元素。对于任意 $n \times n$ 阶矩阵 X 和 Y , 指数映射拥有以下性质:

1. $e^0 = I$ 。
2. $(e^X)^a = e^{aX}$ 。
3. 如果 X^{-1} 存在, 则有 $(e^X)^{-1} = e^{-X}$ 。
4. 对于任意复数 α, β , 有 $e^{(\alpha+\beta)X} = e^{\alpha X} e^{\beta X}$ 。
5. 如果 $XY = YX$, 那么有 $e^{X+Y} = e^X e^Y = e^Y e^X$ 。
6. 如果矩阵 A 可逆, 则 $e^{AXA^{-1}} = A e^X A^{-1}$ 。
7. $\|e^X\| \leq e^{\|X\|}$ 。

以及对数映射 \log :

$$\log(X) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} (X - I)^n \quad (3.2)$$

将李群中的元素映射到切空间。它有以下几个重要的性质:

1. 对于所有满足 $\|X - I\| < 1$ 的矩阵 X , 有 $e^{\log(X)} = X$ 。
2. 对于所有满足 $\|X\| < \log 2$, $\|e^X - I\| < 1$ 的矩阵 X , 有 $\log(e^X) = X$ 。

因为李群是一个微分流形, 那么李群中任意两个元素之间存在测地线。任意两个元素 X_1 和 X_2 之间最短的一条测地线的距离被定义为:

$$d(X_1, X_2) = \|\log(X_1^{-1} X_2)\| \quad (3.3)$$

其中 $\|\cdot\|$ 是 Frobenius 范数, 即上一章中的 (2.2) 式。矩阵李群的 Baker -

Campbell - Hausdorff (BCH) 公式:

$$\log(\exp(x_1)\exp(x_2)) = \sum_{n>0} \frac{(-1)^{n-1}}{n} \sum_{\substack{r_1+s_1>0 \\ |s_i| \leq n}} \frac{x_1^{r_1} x_2^{s_1} \cdots x_1^{r_n} x_2^{s_n}}{r_1! s_1! \cdots r_n! s_n!}$$

或它的另一个直观表达式:

$$\exp(x_1)\exp(x_2) = \exp\left(x_1 + x_2 + \frac{1}{2}[x_1, x_2] + \frac{1}{12}[x_1, [x_1, x_2]] - \frac{1}{12}[x_2, [x_1, x_2]] - \cdots\right) \quad (3.4)$$

BCH 公式类似于 Taylor 展开式, 可用前几项和作为计算结果的近似。

当给定 n 维李群 G 在单位元切空间 T_e 的一组基 T_1, T_2, \dots, T_n 时, 单位元邻域内的任何一个元素 X 可表示为以下形式:

$$X(\alpha_1, \alpha_2, \dots, \alpha_n) = \exp\left(\sum_{i=1}^n \alpha_i T_i\right) \quad (3.5)$$

在矩阵李群的框架下, (3.5) 式中的 T_1, T_2, \dots, T_n 就是李群的生成元矩阵, 并有 $[x_1, x_2] = x_1 x_2 - x_2 x_1$ 。

$GL(n)$ 的单参数子群是一个映射 $f: R \rightarrow GL(n)$, 它满足以下 3 个条件:

1. f 是连续的。
2. $f(0) = I$ 。
3. $f(t+s) = f(t)f(s)$, 对于任意 $t, s \in R$ 均成立。

下面从微分几何的角度分析李群和李代数之间的关系。如果 G 是一个矩阵李群, G 的李代数表示为 \mathfrak{g} , 它是所有满足 $e^{tX} \in G, t \in R$ 的矩阵 X 的集合, 而 e^{tX} 是李群 G 的一个单参数子群, 是在 G 的流形上的一条沿 X 方向的光滑曲线。现在, 在任一李群元素 e^{tX} 处对 t 求导可得: $\frac{d}{dt} e^{tX} = X e^{tX} = e^{tX} X$, 特别的, 有 $\left. \frac{d}{dt} e^{tX} \right|_{t=0} = X$, 根据指数映射的第一条性质 $e^0 = I$, 因此对于属于任意李群 G 的元素 e^{tX} , 当 t 时, X 可覆盖 G 对应的李代数 \mathfrak{g} 中的每一元素, 这说明李代数 \mathfrak{g} 是李群在单位元外的切空间。

对之与应的, 如果 $G \subset GL(n)$ 是一个矩阵李群, 对应的李代数为 \mathfrak{g} , 那么矩阵 X 在 \mathfrak{g} 内当且仅当在流形 $M(n)$ 内存在一条光滑曲线 γ , 使得: 1) $\gamma(t)$ 对于所有的 t 都在 G 内; 2) $\gamma(0) = I$; 3) $\left. \frac{d\gamma}{dt} \right|_{t=0} = X$ 。因此, \mathfrak{g} 是 G 在单位元处的切空间。

该命题的证明如下：如果 X 是 \mathfrak{g} 中的元素，命 $\gamma(t) = \exp(tX)$ ，并有 $\gamma(0) = I$ 以及 $dy/dt|_{t=0} = X$ ，那么，对于所有足够小的 t ， $\log(\gamma(t))$ 在 \mathfrak{g} 内。现在 \mathfrak{g} 是 $M(n)$ 的实子空间，因此也是 $M(n)$ 的拓扑闭子集，于是

$$\left. \frac{d \log(\gamma(t))}{dt} \right|_{t=0} = \lim_{h \rightarrow 0} \frac{\log(\gamma(h)) - 0}{h}$$

也在 \mathfrak{g} 内，并有

$$\log(\gamma(t)) = (\gamma(t) - I) - \frac{(\gamma(t) - I)^2}{2} + \frac{(\gamma(t) - I)^3}{3} + \dots$$

然后对上式两侧执行微分操作。右侧微分后的结果除了第一项，其他所有项均为 0。例如第二项求导后得 $-\frac{1}{2}[(dy/dt)(\gamma(t) - I) + (\gamma(t) - I)(dy/dt)]$ ，当 $t = 0$ 时该式为 0，因此我们得到

$$\left. \frac{d \log(\gamma(t))}{dt} \right|_{t=0} = \left. \frac{d\gamma}{dt} \right|_{t=0} \in \mathfrak{g}.$$

这样就从微分几何的角度证明了 \mathfrak{g} 是 G 在单位元处的切空间。

3.2 矩阵李群实例

在上一章已知 $GL(n)$ 、 $O(n)$ 、 $SO(n)$ 、 $U(n)$ 、 $SU(n)$ 、 $Sym^+(n)$ 等矩阵群同时也构成李群。 $GL(n)$ 在单位元处的切空间，即李代数 $\mathfrak{gl}(n) = T_1 GL(n) = M(n)$ ，另外 $GL(n)$ 的流形结构是 n^2 的。本节从矩阵李群和李代数的角度介绍几个常用矩阵李群。

(1) 仿射群 $Aff(n)$

n 维仿射群定义：

$$Aff(n) = \left\{ \begin{pmatrix} A & t \\ 0 & 1 \end{pmatrix} : A \in GL(n), t \in \mathbb{R}^n \right\} \leq GL(n+1)$$

很明显，它是 $GL(n+1)$ 的封闭子群，其维数为 $n^2 + n$ 。令 $x \in \mathbb{R}^n$ ， $(x, 1)^T \in \mathbb{R}^{n+1}$ ，那么通过下式：

$$\begin{pmatrix} A & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} Ax+t \\ 1 \end{pmatrix} \quad (3.6)$$

在 R^n 上获得了一个运算 $Aff(n)$, 该运算在 R^n 上定义了一个变换: $x \mapsto Ax+t$, 这个变换被称作仿射变换, 其相关的几何称为仿射几何。

$Aff(n)$ 对应的李代数 $\mathfrak{aff}(n)$ 是如下结构:

$$\mathfrak{aff}(n) = \left\{ \begin{pmatrix} U & v \\ 0 & 0 \end{pmatrix} : U \in \mathfrak{gl}(n), v \in R^n \right\}$$

在视觉不变研究中应用最多的仿射群是 $Aff(2)$, 它能表示二维图像的四种仿射变换: 平移、旋转、错切、缩放。利用仿射群对于大部分图像变换进行建模, 并借助对应的李代数工具, 能较好应用于跟踪或检测等问题。 $Aff(2)$ 的几何结构是一个 6 维流形, 作为群结构, 它的李代数向量空间 $\mathfrak{aff}(2)$ 有 6 个基向量, 即 6 个生成元:

$$T_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad T_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$T_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad T_5 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad T_6 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

这几个生成元可以通过对 6 个基本的仿射矩阵求导得到。根据这几个生成元, 利用 (3.5) 式可构造 $Aff(2)$ 中的任意元素, 这对于学习问题中训练样本的构造提供了一个很好的方法。

(2) $SO(2)$ 和 $SO(3)$

$SO(n)$ 在第二章已经有介绍, 这里针对 $SO(2)$ 和 $SO(3)$ 两个例子进行分析。同 $Aff(n)$ 一样, 在 R^2 中, $SO(2)$ 定义了一种运算, 和 $Aff(n)$ 不同的是, 前者结合了多种变换, 而 $SO(2)$ 只有一种单一变换, 即旋转。 $SO(2)$ 是 1 维流形, 对应的李代数 $\mathfrak{so}(2)$ 只有一个生成元:

$$\mathfrak{so}(2): \quad T_1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$SO(3)$ 是定义在 R^3 上的运算, 单从变换种类来说, 也是单一的旋转。几何结构是 3 维流形, 李代数 $\mathfrak{so}(3)$ 是 3 维实向量空间, 它有 3 个基向量:

$$\mathfrak{so}(3): \quad T_1 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad T_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

很明显, $\mathfrak{so}(2)$ 和 $\mathfrak{so}(3)$ 的基向量都是反对称矩阵, 因此由这些基向量线性组合构成的任何元素也都是反对称矩阵。通常对 $\mathfrak{so}(2)$ 和 $\mathfrak{so}(3)$ 元素有一种直观的表达方式, 即:

$$\mathfrak{so}(2): \quad A_\theta = \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}, \quad \mathfrak{so}(3): \quad A_v = \begin{pmatrix} 0 & -v_1 & v_2 \\ v_1 & 0 & -v_3 \\ -v_2 & v_3 & 0 \end{pmatrix}$$

对于任意 $\mathfrak{so}(2)$ 和 $\mathfrak{so}(3)$ 的元素, 相应的 $\theta \in [0, 2\pi)$, 以及 $v = (v_1, v_2, v_3) \in R^3$, 这里的 θ 即表示平面 R^2 中旋转的角度。对于 3D 旋转, 单位向量 $\bar{v} = v/\|v\|$ 是旋转轴, 绕此轴旋转的角度为 $\|v\|$ 。

$\mathfrak{so}(2)$ 的指数映射 $\exp(A_\theta) = R_\theta$, R_θ 是参数为 θ 的 R^2 平面内绕原点旋转的二维旋转矩阵, 表示为:

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

$\mathfrak{so}(3)$ 的指数映射可以表示为:

$$\exp(A_v) = \begin{cases} I_3, & \theta = 0 \\ I_3 + \frac{\sin \theta}{\theta} A_v + \frac{1 - \cos \theta}{\theta^2} A_v^2, & \theta \in (0, \pi) \end{cases}$$

其中, $\theta = \sqrt{\frac{1}{2} \text{tr}(A_v^T A_v)} = \|v\|$, 取值为 $[0, \pi)$ 区间。矩阵 $X \in SO(3)$ 的对数映射是

$\mathfrak{so}(3)$ 中的一个元素, 可表示为:

$$\log(X) = \begin{cases} 0, & \theta = 0 \\ \frac{\theta}{2\sin\theta}(X - X^T), & |\theta| \in (0, \pi) \end{cases}$$

其中, θ 满足 $\text{tr}(X) = 2\cos\theta + 1$ 。

另外还可以发现 $\mathfrak{so}(3)$ 的基向量 T_1 等于 $\mathfrak{aff}(2)$ 的基向量 T_3 , 这里基向量的顺序编号并不重要, 以及 $\mathfrak{so}(2)$ 的基是这该基向量 T_3 的左上 2×2 子矩阵, 这说明仿射变换中包含旋转变换的成份, 而且 $\mathfrak{so}(3)$ 的基向量 T_1 除了左上 $\mathfrak{so}(2)$ 的子阵外, 其余位置为 0, 也说明了 $SO(3)$ 包含平面旋转, 即 R^2 旋转的成份。

(3) $SU(n)$ 和 $SU(2)$

将 R^n 中的点积扩展到 C^n , 有:

$$x \cdot y = x \cdot y = \sum_{k=1}^n \overline{x_k} y_k$$

其中 $x = (x_1 \cdots x_n)^T$, $y = (y_1 \cdots y_n)^T$, 这里点运算 \cdot 不是复-线性的, 但对于 $u, v \in C$ 满足:

$$(ux) \cdot (vy) = \overline{uv}(x \cdot y)$$

该点积允许我们用下式来定义复向量的长度:

$$|x| = \sqrt{x \cdot x}$$

由于 $x \cdot x$ 是一个非负实数, 只当 $x = 0$ 为 0。那么矩阵 $A \in M_n(C)$ 是一个酉阵当且仅当:

$$Ax \cdot Ay = x \cdot y \quad (x, y \in C^n)$$

$SU(2)$ 的李代数 $\mathfrak{su}(2)$ 和 $\mathfrak{so}(3)$ 一样, 都是 3 维实向量空间, 它有下面 3 个基向量: 的李括号有以下性质

$$\mathfrak{su}(2): \quad T_1 = \frac{1}{2} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad T_2 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad T_3 = \frac{1}{2} \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

$\mathfrak{su}(2)$ 的李括号有以下性质:

$$[T_1, T_2] = T_3, [T_2, T_3] = T_1, [T_3, T_1] = T_2$$

(4) R^* 、 C^* 、 S^1 、 R 和 R^n

还有一些重要的群并不是纯粹的矩阵结构。例如基于乘法的非零实数集的 R^* 群同构于 $GL(n)$ ，因此 R^* 是一个矩阵李群。类似的，基于乘法的非零实数集的 C^* 群同构于 $GL_n(C)$ 。绝对值为1的复数的群 S^1 同构于 $U(1)$ 。

在加法运算下的 R 群通过映射 $x \rightarrow (e^x)$ 同构于 $GL(1)^+$ ，后者是 1×1 阶正定实矩阵。

R^n 群（基于向量加法）通过下面的映射同构于只含有正对角元素的实对角矩阵：

$$(x_1, \dots, x_n) \rightarrow \begin{pmatrix} e^{x_1} & & 0 \\ & \ddots & \\ 0 & & e^{x_n} \end{pmatrix}.$$

3.3 李群上的高斯分布

本节分析几个常用李群在流形结构上的高斯分布，复杂李群的高斯分布可参考文献[73]。

一个均值为单位元的李群，其高斯分布是定义在局部坐标上的热方程的解：

$$\begin{aligned} \frac{\partial f}{\partial t} &= \Delta f = \text{div}(\text{grad}(f)) \\ &= g^{ij} \left(\frac{\partial^2 f}{\partial x^i \partial x^j} - \Gamma_{ij}^k \frac{\partial f}{\partial x^k} \right) \end{aligned}$$

其中， g^{ij} 是黎曼度量的逆， Γ_{ij}^k 是Christoffel符号^[51]，那么 R^n 的高斯分布的密度函数为：

$$p(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}\right) \quad (3.7)$$

上式即为热方程在 R^n 中的解。

3.3.1 R^+ 的高斯分布

对于正实数在乘法下构成的李群 R^+ ，局部坐标通过对数给定。其上的热方程的解为对数密度函数：

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\log(x) - \log(\mu))^2}{2\sigma^2}\right) \quad (3.8)$$

给定独立且同为对数分布的样本 $x_1, \dots, x_n \in R^+$ ，它们的均值和方差的最大似然估计为：

$$\hat{\mu} = \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\log(x_i) - \log(\hat{\mu}))^2$$

这里的 $\hat{\mu}$ 为几何均值，也就是 R^+ 中测地线距离平方和最小的点，即 $\sum_{i=1}^n \log(\mu^{-1}x_i)^2$ 最小的点。

3.3.2 $SO(2)$ 的高斯分布

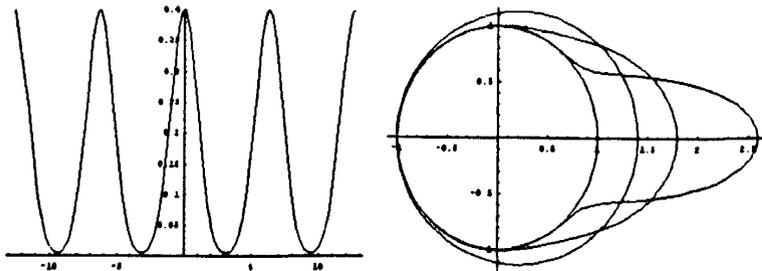


图 3-1. 左侧为 $[-\pi, \pi]$ 周期内的热方程的解，右侧为单位圆上 σ 分别为 1, 0.5 以及 0.25 时的 wrapped Gaussian。

$SO(2)$ 的参数是旋转角度 $\theta \in [0, 2\pi)$ 。因为 $SO(2)$ 通过映射 $\theta \mapsto e^{i\theta}$ 同构于 S^1 ，于是 $SO(2)$ 上均值和方差分别为 μ 和 σ 的高斯分布为：

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{+\infty} \exp\left(-\frac{(\theta - \mu - 2k\pi)^2}{2\sigma^2}\right) \quad (3.9)$$

(3.9) 式即为在 $[-\pi, \pi]$ 周期内的热方程的解, 也通常被称为 “wrapped Gaussian”, 如图 3-1 所示。

现在来推导均值和方差的最大似然估计。给独立且同为 (3.9) 式分布的样本集 $\theta_i \in [0, 2\pi), i=1, \dots, n$, 我们先假定 $\sigma=1$, 得到均值的最大似然估计为:

$$\hat{\mu} = \arg \max_{\mu \in [0, 2\pi)} \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{+\infty} \exp\left(-\frac{(\theta_i - \mu - 2k\pi)^2}{2}\right) \quad (3.10)$$

注意到二次指数函数是偶函数, 于是它的导函数为奇函数, 并且对于固定的 k 有:

$$\frac{\partial}{\partial \mu} \exp\left(-\frac{(\theta_i - \mu - 2k\pi)^2}{2}\right) = -\frac{\partial}{\partial \mu} \exp\left(-\frac{(\theta_i - \mu + 2k\pi)^2}{2}\right)$$

于是 (3.10) 式化简为只有 $k=0$ 的项, 因此最大化问题变为:

$$\hat{\mu} = \arg \max_{\mu \in [0, 2\pi)} \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\theta_i - \mu)^2}{2}\right) \quad (3.11)$$

(3.11) 式即为高斯分布下 $SO(2)$ 的均值的最大似然方程。于是有 $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \theta_i$ 。因为 θ_i 表示的多样性, 比如 θ_i 可以属于 $[0, 2\pi)$ 或者属于 $[-\pi, \pi)$, 这会使得 $\hat{\mu}$ 的算式出现歧义, 所以需要在具体问题中限定 θ_i 为其中一个区间。

对于方差的最大似然估计的推导, 考虑如下对数似然函数:

$$l(\sigma; \hat{\mu}, \theta_1, \dots, \theta_n) = \sum_{i=1}^n \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) + \log\left(\sum_{k=-\infty}^{+\infty} \exp\left(-\frac{(\theta_i - \hat{\mu} - 2k\pi)^2}{2\sigma^2}\right)\right)$$

对 σ 微分得:

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (\theta_i - \hat{\mu})^2 + \frac{1}{\sigma^3} \sum_{i=1}^n \left(\frac{\sum_{k=-\infty}^{+\infty} (2k\pi)^2 \exp\left(-\frac{(\theta_i - \hat{\mu} - 2k\pi)^2}{2\sigma^2}\right)}{\sum_{k=-\infty}^{+\infty} \exp\left(-\frac{(\theta_i - \hat{\mu} - 2k\pi)^2}{2\sigma^2}\right)} \right)$$

上式第三项虽收敛, 但不产生一个封闭形式的解。依据文献 [52] 可知, wrapped Gaussian 密度函数, 即 (3.9) 式当 $\sigma^2 \leq 2\pi$ 近似于和式中 $k=0$ 的项, 这符合图像处理问题中对角度的约束, 于是我们对上式只保留 $k=0$ 的项, 这样就有:

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (\theta_i - \hat{\mu})^2$$

令上式等于 0, 解出 σ , 得到近似方差的最大似然估计为:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\theta_i - \hat{\mu})^2 \quad (3.12)$$

3.3.3 $SO(3)$ 的高斯分布

与 $SO(2)$ 类似, 我们使用对数映射在 $SO(3)$ 上定义一个均值为 μ 的 wrapped Gaussian 分布。需要说明的一点是, 密度函数没有必要是 $SO(3)$ 上热方程的解。令 $u(x) = \Phi(\log(\mu^{-1}x)) \in \mathbb{R}^3$, 其中 $\Phi: \mathfrak{so}(3) \rightarrow \mathbb{R}^3$, $\Phi(A_i) = v$ 是一个标准同构。命 $\bar{u}(x) = u(x)/\|u(x)\|$, 依据 (3.9) 式, $SO(3)$ 的 wrapped Gaussian 密度函数为:

$$p(x) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}} \sum_{k=-\infty}^{+\infty} \exp\left(-\frac{1}{2} (u(x) - 2k\pi\bar{u}(x))^T \Sigma^{-1} (u(x) - 2k\pi\bar{u}(x))\right) \quad (3.13)$$

其中协方差结构定义为李代数 $\mathfrak{so}(3)$ 上的平形式, 表示为 3×3 的协方差矩阵 Σ 。

给定独立且同为 (3.13) 分布的样本集 $x_1, \dots, x_n \in SO(3)$, 下面推导最大似然估计下的均值和方差。针对均值, 不失一般性, 可以假设协方差是单位矩阵。联合密度由积密度给定:

$$p(\mu; x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{(2\pi)^3}} \sum_{k=-\infty}^{+\infty} \exp\left(-\frac{1}{2} \|u(x_i) - 2k\pi\bar{u}(x_i)\|^2\right) \quad (3.14)$$

注意到 $SO(3)$ 的测地线同构于 $SO(2)$, 并且密度函数 (3.13) 式约束一条测地线到 $SO(2)$ 上的 wrapped Gaussian。现在可以使用与上一小节相同的假设来表示消去和式中相应项后的导函数。最大化 $p(\mu, x_1, \dots, x_n)$ 等价于寻找 (3.15) 式的 $\hat{\mu}$:

$$\hat{\mu} = \arg \max_{\mu \in SO(3)} \prod_{i=1}^n \frac{1}{\sqrt{(2\pi)^3}} \exp\left(-\frac{1}{2}\|\mu(x_i)\|^2\right) \quad (3.15)$$

对上式求对数后，得最大似然估计的均值：

$$\hat{\mu} = \arg \max_{\mu \in SO(3)} \sum_{i=1}^n \|\Phi(\log(\mu^{-1}x_i))\|^2 \quad (3.16)$$

上式中的 $\|\Phi(\log(\mu^{-1}x_i))\|$ 是 μ 到 x 的黎曼距离。因此，均值的最大似然估计也是到所有样本点的测地线距离平方和最小的点，也就是 $SO(3)$ 的内均值。

类似于 $SO(2)$ 的情况，假定方差足够小，即 $\lambda \leq 2\pi$ ，这里的 λ 是 Σ 的特征值。使用与上一小节相同的假设，近似方差的最大似然估计为：

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \Phi(\log(\hat{\mu}^{-1}x_i))\Phi(\log(\hat{\mu}^{-1}x_i))^T \quad (3.17)$$

3.4 李群内均值计算

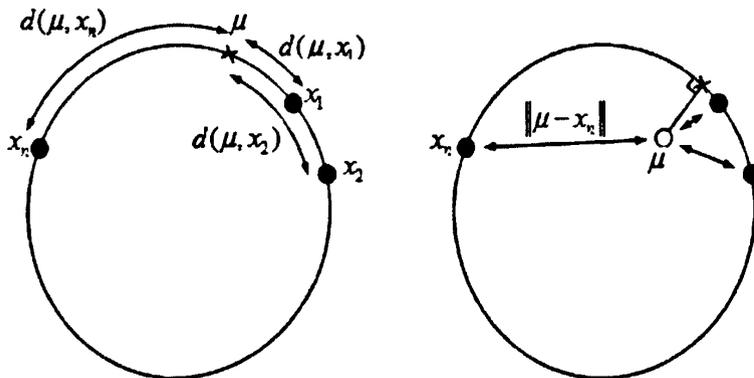


图 3-2. 图中左侧表示流形 S^1 上的内均值 μ ，各数据点到 μ 的距离是位于 S^1 上测地线距离（曲线长），所得的内均值也位于 S^1 上。右侧是直接计算欧氏距离所得的平均，该计算得到的 μ 叫做外均值（extrinsic mean），一般不在流形上，如右侧的 μ 位于 R^2 内，而不是 S^1 上。

在 R^d 中, n 个数据点集 $\{x_i\}_1^n \in R^d$ 的平均量可以通过 (2.4) 式计算得到。其中, 样本 x_i 可以是向量或矩阵, 所得的 μ 能保持与 $\{x_i\}_1^n$ 中每一个点的欧氏距离平方总和最小, 即 μ 可以表示为:

$$\mu = \arg \min_{x \in R^d} \sum_{i=1}^n \|x - x_i\|^2 \quad (3.18)$$

由于通常一个 d 维流形 M^d 并不是一个向量空间, 因此欧氏距离并不能代表流形上两点之间的实际距离, 所以 (3.18) 式不能直接用于流形上, 其结果如图 3-2 所示。然而, 由于 d 维流形 M^d 局部同胚于 R^d , 因此在 M^d 的局部领域内, 可以寻找一个同胚映射 $\Phi: M^d \rightarrow R^d$, 将 M^d 的局部嵌入到 R^d , 这样仍可应用 (3.18) 式计算得到在 R^d 中的 μ_Φ : $\mu_\Phi = \arg \min_{x \in M^d} \sum_{i=1}^n \|\Phi(x) - \Phi(x_i)\|^2$, 同时需要再定义一个映射 $\pi: R^d \rightarrow M^d$: $\mu = \pi(\mu_\Phi)$ 将 μ_Φ 映射回 M^d 上。省去 μ_Φ , 可得计算公式:

$$\mu = \pi \left(\arg \min_{x \in M^d} \sum_{i=1}^n \|\Phi(x) - \Phi(x_i)\|^2 \right) \quad (3.19)$$

在实际情况中, 为 M^d 寻找合适的 Φ 和 π 两个映射比较困难, 并且 M^d 并不一定由一个 Φ 就能全部嵌入到 R^d 中。一种更加合理的做法可以利用 M^d 上的黎曼距离代替 (3.19) 式中的 Φ 函数, 直接在 M^d 上计算任意 $x_i, x_j \in M^d$ 之间的距离, 而不是在欧氏空间中。

因此, 类似于欧氏空间 R^d , M^d 上某一点集的平均可以通过计算这样一个点 \bar{x} 来实现, 这个点要符合两个条件: (i) $\bar{x} \in M^d$, (ii) \bar{x} 与 $\{x_i\}_1^n \in M^d$ 中每一个点的测地线距离平方和最小。在一个紧致流形中, 一个由 n 个数据点集 $\{x_i\}_1^n$ 构成的内均值可定义为:

$$\mu = \bar{x} = \arg \min_{x \in M^d} \sum_{i=1}^n d(x, x_i)^2 \quad (3.20)$$

上式中 $d(\cdot, \cdot)$ 表示括号内两个点之间的测地线距离。

按上述内容, 李群中任意两元素之间的距离由 (3.3) 式给出, 于是将 (3.3) 式代入 (3.20) 式可得到:

$$\mu = \arg \min_{x \in G} \sum_{i=1}^n \|\log(x^{-1}x_i)\|^2 \quad (3.21)$$

利用一阶 BCH 公式来近似上式右边 $\|\cdot\|$ 内的部分得到:

$$\log(x^{-1}x_i) = -\log(x) + \log(x_i) \quad (3.22)$$

其中, m 和 m_i 分别是李群中的点 x 和 x_i 的李代数。要想获得比较精确的结果, 可以求前 n 阶近似, 但按照 BCH 公式, 后面的近似项计算量将变得非常大, 而且对总体近似值的贡献却更小。于是用一阶近似来代替 $\log(x^{-1}x_i)$, 内均值公式可简化为:

$$\begin{aligned} \hat{\mu} &= \arg \min_{x \in G} \sum_{i=1}^n \|\log(x) - \log(x_i)\|^2 \\ &= \arg \min_{m \in \mathfrak{g}} \sum_{i=1}^n \|-m + m_i\|^2 \end{aligned} \quad (3.23)$$

这就是最后得到的近似公式。现在的问题是需要寻找到满足上式条件的 $x \in G$ 。文献 [2] 给出了梯度下降法求解 $\hat{\mu}$ 的过程, 过程描述如下: 设计优化函数为

$f(x) = \frac{1}{2n} \sum_i d(x, x_i)^2$, 函数 f 的梯度为 $\nabla f(x) = -\frac{1}{n} \sum_i \log(x^{-1}x_i)$, 给定第 k 次迭代的内

均值的估计 μ_k , 第 $k+1$ 次迭代公式为: $\mu_{k+1} = \exp\left(\frac{\tau}{n} \sum_{i=1}^n \log(\mu_k^{-1}x_i)\right)$, 其中 τ 为步长。李

群内均值的具体的求解算法由算法 3-1 给出。

虽然梯度下降法只是局部收敛, 只能找到局部极小而不一定是全局最小的 $\hat{\mu}$, 但通过修改初始估计 μ_0 和步长 τ 仍可得到较理想的实际效果。这里 τ 的选择和李群 G 的流形结构有关, 文献 [47] 证明了对于球流形, $\tau=1$ 是合适的, 当 G 的流形结构是一个向量空间时, $\tau=1$ 的梯度下降就等价于求线性平均, 如果是 R^+ 、 R^d 、 $SO(3)$, $\tau=1$ 等价于求几何平均, 而且算法 3-1 能在单步迭代内收敛。对于一般情况下的李群, 如果当 $\tau=1$ 时的算法 3-1 不能收敛, 可适当使 τ 取更小的正数。按照梯度下降法的特点, 当 τ 取值过大时, 可能会越过极值点, 而取值过小则又会使收敛速度过慢。

算法 3-1: 求 n 个类别李群样本的内均值

输入: $\{x_{ij}\}_{i=1..c}^{j=1..n_i} \in G$, x_{ij} 表示分布在李群 G 上的第 i 类别的第 j 个样本, n_i 表示第 i 个分类中训练样本的个数, 一共有 c 个分类。

输出: $\mu_i, i=1, 2, \dots, c$, 即每个分类的内均值。

过程:

Do $i=1$

$k=0$

$\mu = x_{i1}$

Do

$$\Delta\mu = \frac{\tau}{n_i} \sum_{j=1}^{n_i} \log(\mu^{-1}x_{ij})$$

$\mu = \exp(\Delta\mu)$

$k = k + 1$

While $\|\Delta\mu\| > \varepsilon$ and $k < \text{Maximum_Iterations}$

$\mu_i = \mu$

$i = i + 1$

While $i \leq c$

由于内均值更本质的表现了一类事物的共性, 如果一个未知样本相比其它类别更接近于其中某一类别的内均值, 我们就认为这个未知样本最可能属于该类别, 因此, 我们还可以基于内均值设计一个简单的李群分类器: Lie-Mean 算法, 将训练样本按类别计算得到内均值后, 对于测试样本只需要计算该样本到各类内均值的测地线距离, 将类别判为最短距离的那个内均值所属的类:

$$i^* = \arg \min_{i=1, 2, \dots, c} \|\log(\mu_i^{-1}x)\| \quad (3.24)$$

算法 3-1 是基于迭代的李群内均值计算方法, 适合几何结构复杂的李群, 不方便直接求解内均值时使用。考虑到其他李群的流形结构特点, 可通过数学方法计算更为精确

的均值。例如，文献[22]给出了任意两个对称正定矩阵 $S_1, S_2 \in \text{Sym}^+(n)$ 保持仿射不变的度量：

$$d(S_1, S_2) = \left\| \log \left(S_1^{-\frac{1}{2}} S_2 S_1^{-\frac{1}{2}} \right) \right\| \quad (3.25)$$

在添加 (2.3) 式的运算构成李群结构之后，两个元素的距离又可以用下式表示：

$$d(S_1, S_2) = \left\| \log(S_2) - \log(S_1) \right\| \quad (3.26)$$

可以直接得到 $\text{Sym}^+(n)$ 群的新的均值计算公式：

$$\mu = \exp \left(\frac{1}{n} \sum_{i=1}^n \log(x_i) \right) \quad (3.27)$$

(3.27) 式称为 Log-Euclidean 均值。和算法 3-1 的不同之处在于：1) Log-Euclidean 均值计算的是全局均值。几何的观点看，就是将所有群元素样本映射到李代数线性空间后计算的几何平均，而算法 3-1 是局部均值，是直接在李群的流形空间内计算样本集的内均值；2) Log-Euclidean 均值的计算是一次求解，时间复杂度为 $O(n)$ ，而算法 3-1 是迭代算法，迭代次数依具体的流形结构和所给定的参数有关，当迭代次数为 k 时，时间复杂度为 $O(kn)$ 。

应用 (3.27) 式作为均值计算方法时，Lie-Mean 算法的类别判别可通过下面的 (3.28) 式完成：

$$i^* = \arg \min_{i=1,2,\dots,c} \left\| \log(\mu_i) - \log(x) \right\| \quad (3.28)$$

3.5 本章小节

本章介绍了矩阵李群的相关理论基础，包含代数群和微分几何两个层面的性质。分析了常用的矩阵李群的例子，给出这些矩阵李群的理论性质。分析了几个常用李群的高斯分布的密度函数形式，讨论了它们在高斯分布下以最大似然估计来计算均值和方差的方法。最后，结合文中的均值计算理论，给出李群均值计算算法，包括迭代计算法和直接求解两种方法。

第四章 李群均值相关学习算法

本章将在上一章中讨论的李群均值理论的基础上进一步分析李群李代数样本集的空间特性, 设计均值相关学习算法。内容安排为: 4.1 节分析经典线性判别算法 FLDA。4.2 节根据 FLDA 的思想, 分析了李群空间中 Fisher 投影的概念与意义。4.3 节推导了 Lie-Fisher 算法的两种形式, 并分析了它们的优缺点以及时空复杂度。4.4 节为本章小节。

4.1 FLDA 算法

FLDA (Fisher Linear Discriminant Analysis) 又称 LDA、FDA、Fisher 线性判别, 是线性空间中经典的分类判别分析算法^[41]。标准的 Fisher 投影是将样本点投影到一个比样本维数更低的超平面或直线上, 使得各类别之间的类间散度和类内散度比值最大。Fisher 投影的准则函数为:

$$J(\mathbf{v}) = \frac{\mathbf{v}^T \mathbf{S}_b \mathbf{v}}{\mathbf{v}^T \mathbf{S}_w \mathbf{v}} = \frac{|\tilde{\mathbf{S}}_b|}{|\tilde{\mathbf{S}}_w|} \quad (4.1)$$

其中 \mathbf{S}_b 和 \mathbf{S}_w 分别是投影前的各类样本间的散度程度总和与各类内散度总和。 $\tilde{\mathbf{S}}_b$ 和 $\tilde{\mathbf{S}}_w$ 是投影后各类样本间的散度程度总和与各类内散度总和, 定义如下:

$$\tilde{\mathbf{S}}_b = \sum_{i=1}^c n_i (\tilde{\boldsymbol{\mu}}_i - \tilde{\boldsymbol{\mu}})(\tilde{\boldsymbol{\mu}}_i - \tilde{\boldsymbol{\mu}})^T, \quad \tilde{\mathbf{S}}_w = \sum_{i=1}^c \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \tilde{\boldsymbol{\mu}}_i)(\mathbf{y}_{ij} - \tilde{\boldsymbol{\mu}}_i)^T$$

其中, c 是类别数, n_i 和 $\tilde{\boldsymbol{\mu}}_i$ 分别表示第 i 分类的样本个数和投影后的均值, $\tilde{\boldsymbol{\mu}}$ 为投影后总体样本均值, \mathbf{y}_{ij} 表示样本 \mathbf{x}_{ij} 投影后的结果。Fisher 投影的问题就是寻找使 (4.1) 式取最大值时的 \mathbf{v} 。在欧氏空间中, 向量 \mathbf{v} 的方向代表了需要投影的直线方向。

4.2 李群空间中的 Fisher 投影

李代数的生成元可用于生成李群流形上的一条测地线。例如，李群 G 对应的李代数上的元素 $v \in \mathfrak{g}$ ，通过一个单参数指数映射： $R \mapsto G: \exp(tv), t \in R$ ，可在李群流形上生成一条测地线 H_v ， H_v 是李群 G 上的一个单参数子群，它由李群 G 的李代数元素 v 唯一确定：

$$H_v \triangleq \{\exp(tv) \in G, t \in R\}$$

当给定一个确定的参数值 t ，也就确定了一个在测地线 H_v 上的点，这个点也同时属于 G 。于是，在李群的流形空间中找一条测地线的思路如图 4-1 所示：

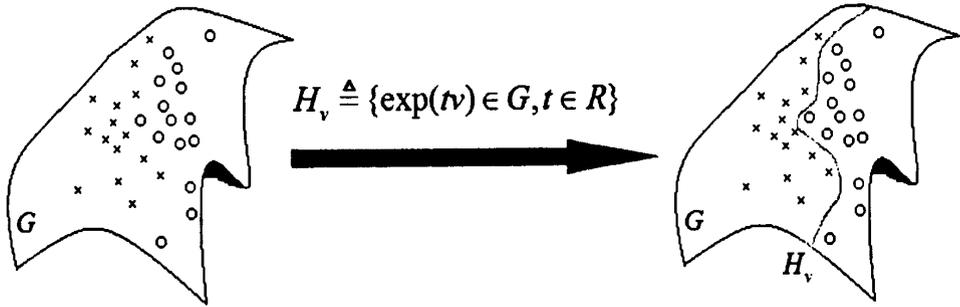


图 4-1. 左侧为李群 G 及其上分布的两类样本，分别为叉和圆圈表示，通过寻找一条测地线 H_v ，将两类样本分开。右侧的测地线即为通过算法寻找到的 H_v ，它是由 v 决定的 G 的一个单参数子群。

然后还需要定义李群上的点到一条测地线的投影的概念。设一个样本 $x \in G$ 和一条 G 上的测地线 H_v ，首先找到：

$$t^* = \arg \min d(x, \exp(tv)), t \in R$$

那么， $\exp(t^*v)$ 就是样本点 x 到 H_v 的投影点。从几何观点来看，这个投影点也就是 H_v 上所有点中，到 x 的测地线距离中最近的那个点，如图 4-2 所示。



图 4-2. 左侧, 将一个李群样本点连接至测地线 $\exp(tv)$ 上的点, 点与点连接形成一条测地线, 所有这样的测地线的长度中最短的那条 (图中为连接 $\exp(tv)$ 上圆点的测地线) 对应的位于 $\exp(tv)$ 上的点即为投影点 $\exp(t^*v)$ 。右侧将所有样本点用相同的方法投影到测地线上。

然后, 令 $y = \exp(t^*v)$ 是 x 到 H_v 的投影点, 那么一类样本 $\{x_{ij}\}_{j=1}^n$ 投影到 H_v 得到新的点集: $\{y_{ij}\}_{j=1}^n$, 投影后计算每类新样本的内均值:

$$\tilde{\mu}_i = \arg \min \sum_{j=1}^n d(y, y_{ij}) \quad (4.2)$$

以及总体样本的内均值为:

$$\tilde{\mu} = \arg \min \sum_{i=1}^c \sum_{j=1}^{n_i} d(y, y_{ij}) \quad (4.3)$$

上式中, c 表示类别个数, n_i 表示第 i 个类别的样本个数, y_{ij} 表示第 i 个类别中第 j 个样本投影后的新样本。

对于多类情况, 投影后各类样本的类间散度定义为:

$$\begin{aligned} \tilde{S}_b &= \sum_{i=1}^c n_i \cdot d^2(\tilde{\mu}_i, \tilde{\mu}) \\ &= \sum_{i=1}^c n_i \cdot \|\log(\tilde{\mu}^{-1} \tilde{\mu}_i)\|^2 \end{aligned} \quad (4.4)$$

投影后一类样本的类内散度定义为： $\tilde{s}_i^2 = \sum_{j=1}^{n_i} d^2(y_{ij}, \tilde{\mu}_i)$ ，那么将所有 \tilde{s}_i^2 之和定义为

总体样本的散度：

$$\begin{aligned} \tilde{S}_w &= \sum_{i=1}^c \tilde{s}_i^2 \\ &= \sum_{i=1}^c \sum_{j=1}^{n_i} d^2(y_{ij}, \tilde{\mu}_i) \end{aligned} \quad (4.5)$$

于是将 (4.2) ~ (4.5) 式代入 (4.1) 式并利用 BCH 公式第一阶近似，得到如下形式：

$$J(v) \approx \frac{\sum_{i=1}^c n_i \cdot \|\log(\tilde{\mu}_i) - \log(\tilde{\mu})\|^2}{\sum_{i=1}^c \sum_{j=1}^{n_i} \|\log(y_{ij}) - \log(\tilde{\mu}_i)\|^2} \quad (4.6)$$

相应的，Lie-Fisher 算法的问题归结为寻找满足下式的一条测地线 H_v ：

$$v^\circ = \arg \max J(v) \quad (4.7)$$

4.3 Lie-Fisher 判别分析

Lie-Fisher 判别分析的过程就是在李群流形上寻找一条分类测地线，将所有样本向这条测地线上投影后，使样本的投影点集具有最大化的类间散度和类内散度的比值。

4.3.1 Lie-Fisher 算法的第一形式

上一小节根据 Fisher 投影的理论推导了李群非线性空间中的样本的 Fisher 投影方向的计算公式。但实际上，由于函数 $J(v)$ 和参数 v 并没有显式的解析表达形式，对 (4.6) 式直接求解是比较困难的。因此我们将李群上的样本点映射到对应的李代数空间，在线性的李代数空间给出求解近似的 Lie-Fisher 投影测地线：

首先，将李群空间的样本映射到李代数空间构成新样本集，并计算以下各个成份。

新样本集的总体均值：

$$\mu = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} \log(M_{ij}) \quad (4.8)$$

以及各类均值:

$$\mu_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \log(M_{ij}) \quad (4.9)$$

根据上两式, 再利用标准 Fisher 的方法求解类间散度:

$$S_b = \sum_{i=1}^c n_i \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \log(M_{ij}) - \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} \log(M_{ij}) \right) \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \log(M_{ij}) - \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} \log(M_{ij}) \right)^T \quad (4.10)$$

和类内散度:

$$S_w = \sum_{i=1}^c \sum_{j=1}^{n_i} \left(\log(M_{ij}) - \frac{1}{n_i} \sum_{j=1}^{n_i} \log(M_{ij}) \right) \left(\log(M_{ij}) - \frac{1}{n_i} \sum_{j=1}^{n_i} \log(M_{ij}) \right)^T \quad (4.11)$$

综合上述计算公式, 最终求出投影方向 \mathbf{v} , 这样得到的近似 $\hat{J}(\mathbf{v})$ 计算公式为:

$$\hat{J}(\mathbf{v}) = \frac{\mathbf{v}^T \sum_{i=1}^c n_i \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \log(M_{ij}) - \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} \log(M_{ij}) \right) \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \log(M_{ij}) - \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} \log(M_{ij}) \right)^T \mathbf{v}}{\mathbf{v}^T \sum_{i=1}^c \sum_{j=1}^{n_i} \left(\log(M_{ij}) - \frac{1}{n_i} \sum_{j=1}^{n_i} \log(M_{ij}) \right) \left(\log(M_{ij}) - \frac{1}{n_i} \sum_{j=1}^{n_i} \log(M_{ij}) \right)^T \mathbf{v}} \quad (4.12)$$

对于两类分类问题, 可直接得到 \mathbf{v} (假设 S_w 可逆):

$$\begin{aligned} \mathbf{v} &= S_w^{-1}(\mu_1 - \mu_2) \\ &= \left(\sum_{i=1}^2 \sum_{j=1}^{n_i} \left(\log(M_{ij}) - \frac{1}{n_i} \sum_{j=1}^{n_i} \log(M_{ij}) \right) \left(\log(M_{ij}) - \frac{1}{n_i} \sum_{j=1}^{n_i} \log(M_{ij}) \right)^T \right)^{-1} \\ &\quad \cdot \left(\frac{1}{n_1} \sum_{j=1}^{n_1} \log(M_{1j}) - \frac{1}{n_2} \sum_{j=1}^{n_2} \log(M_{2j}) \right) \end{aligned} \quad (4.13)$$

综合上述内容, 得到算法 4.1。

算法 4-1: 求 Lie-Fisher 投影测地线

输入: $\{M_{ij}\}_{i=1..c}^{j=1..n_i} \in G$, M_{ij} 表示第 i 类别的第 j 个样本, n_i 表示第 i 个分类中训练样本的个数, 一共有 c 个分类。

输出: Lie-Fisher 投影测地线 H_ν 对应的李代数 ν

过程:

1. 对每一个李群样本 M_{ij} 做李代数映射: $x_{ij} = \log(M_{ij})$ 。

2. 计算每类新样本的均值: $\mu_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, i=1,2,\dots,c$, 以及总体均值:

$$\mu = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} x_{ij}, i=1,2,\dots,c。$$

3. 计算 S_b 和 S_w : $S_b = \sum_{i=1}^c n_i (\mu_i - \mu)(\mu_i - \mu)^T$, $S_w = \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T$ 。

4. 求解 $S_b \nu = \lambda S_w \nu$, 得到的 ν 即可按 $\exp(t\nu), t \in R$ 生成 Lie-Fisher 的测地线 H_ν 。

按上述算法求解得到 ν 后, 对测试样本 T 的类别判别可通过下式完成:

$$i^* = \arg \min_{i=1,2,\dots,c} \left\| \nu^T \log(T) - \tilde{\mu}_i \right\| \quad (4.14)$$

其中, $\tilde{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \nu^T \log(M_{ij})$ 是第 i 类李群样本在对应的李代数空间中往方向 ν 投影后所得的外均值。

4.3.2 Lie-Fisher 算法的第二形式

算法 4.1 是将所有李群样本映射到相应的李代数空间后, 在线性的李代数空间求解外均值, 从而计算 ν 。但是, 对离散的低维李群样本做对数映射后, 再求解均值可能会使实际样本均值的计算误差变大。

因此, 可考虑先在李群流形空间中求得总体样本内均值 μ 和各类别内均值 μ_i , 在李群空间中完成相应的运算后将结果映射到李代数空间再求解 ν 。例如, 对各类别去掉

总体均值可以用 $\mu^{-1}\mu_i$ 来完成, 所有第 i 类样本按 $\mu_i^{-1}M_{ij}$ 方式去掉类内均值, 然后通过
对数映射将上述计算结果映射到李代数空间, 再计算 v 。于是得到

$$S_b = \sum_{i=1}^c n_i \log(\mu^{-1}\mu_i) \log(\mu^{-1}\mu_i)^T, \quad S_w = \sum_{i=1}^c \sum_{j=1}^{n_i} \log(\mu_i^{-1}M_{ij}) \log(\mu_i^{-1}M_{ij})^T。$$

这样, 可以得到另一个近似 $\hat{J}(v)$:

$$\hat{J}(v) = \frac{v^T \sum_{i=1}^c n_i \log(\mu^{-1}\mu_i) \log(\mu^{-1}\mu_i)^T v}{v^T \sum_{i=1}^c \sum_{j=1}^{n_i} \log(\mu_i^{-1}M_{ij}) \log(\mu_i^{-1}M_{ij})^T v} \quad (4.15)$$

对于两类问题, 可直接得到 v (假设 S_w 可逆):

$$v = \left(\sum_{i=1}^2 \sum_{j=1}^{n_i} \log(\mu_i^{-1}M_{ij}) \log(\mu_i^{-1}M_{ij})^T \right)^{-1} \log(\mu_1^{-1}\mu_2) \quad (4.16)$$

综上所述, 我们得到新的 Lie-Fisher 算法, 即算法 4-2:

算法 4-2: 求 Lie-Fisher 投影测地线

输入: $\{M_{ij}\}_{i=1, \dots, c}^{j=1, \dots, n_i} \in G$, M_{ij} 表示第 i 个类别的第 j 个样本, n_i 表示第 i 个分类中训练样本的个数, 一共有 c 个分类。

输出: Lie-Fisher 投影测地线 H_v 对应的李代数 v 。

过程:

1. 用算法 1 分别计算全体李群样本 M_{ij} 的内均值 μ , 以及各类别样本内均值 $\mu_i, i=1, \dots, c$ 。

2. 计算 S_b 和 S_w : $S_b = \sum_{i=1}^c n_i \log(\mu^{-1}\mu_i) \log(\mu^{-1}\mu_i)^T$,

$$S_w = \sum_{i=1}^c \sum_{j=1}^{n_i} \log(\mu_i^{-1}M_{ij}) \log(\mu_i^{-1}M_{ij})^T。$$

3. 求解 $S_b v = \lambda S_w v$, 得到的 v 即可按 $\exp(tv), t \in \mathbb{R}$ 生成 Lie-Fisher 的测地线 H_v 。

对被测试李群样本 T 的类别判别和算法 2 不同, 需要先将测试样本映射到李代数空间, 往 v 方向投影后再映射回李群空间, 然后判断所得结果到每一类内均值的李代数在 v 方向投影后的新均值点的测地线距离, 即判别公式如下:

$$i^* = \arg \min_{i=1, \dots, c} \left\| \log \left(\left(\exp(v^T \log(T)) \right)^{-1} \exp(v^T \log(\mu_i)) \right) \right\| \quad (4.17)$$

考虑到计算问题, 显然 (4.17) 式也可以用 BCH 一阶公式来近似, 这样就可以得到判别公式:

$$i^* = \arg \min_{i=1, \dots, c} \left\| v^T (\log(\mu_i) - \log(T)) \right\| \quad (4.18)$$

4.3.3 Lie-Fisher 算法二种形式的区别

从上述两个形式 Lie-Fisher 的推导过程来看, 算法 4-1 对李群样本不论其分布特点, 先将它们全部做对数映射, 即映射到李代数线性空间, 然后再利用李代数空间的线性性质, 计算均值与投影方向。算法 4-2 与算法 4-1 最大的区别是均值的计算不在李代数空间内完成, 而是在李群的非线性空间内求得。

直接分析可知, 算法 4-1 有以下几个优点:

1. 计算速度快。只在第一步将样本映射到李代数空间, 需对所有样本做对数映射, 开销会稍大, 但后续步骤都在线性空间中完, 效率较高。
2. 对低维李群样本判别精度高。因为先对样本映射到李代数空间, 当李群样本的维数较高时, 李代数空间的欧氏距离计算会更难表示出样本的差异。
3. 对分布于平坦李群流形的样本判别精度高。因为当李群流形几何结构越平坦越接近于超平面时, 其映射后的李代数样本将和原样本更能保持等距与等角的特性。

相应的, 算法 4-1 也存在以下缺点:

1. 高维样本的均值计算误差可能增大。
2. 局部李群流形领域内分布不平坦的样本精度可能变差。

算法 4-1 的缺点和优点正好是算法 4-2 的优点与缺点。算法 4-2 中内均值的计算方法如果采用上一章的算法 3-1, 那么, 也有可能因为算法 3-1 中的梯度下降法导致的

局部极小值问题而使得本身在高维李群流形下占优的算法 4-2 而显现劣势。比如算法 4-1 在样本的局部李群流形领域分布不平坦时,可能分类精度会降低,而算法 4-2 能很好处理这个问题,但如果内均值的计算使之陷入不平坦流形的局部极小值点,也同样不能很好发挥出算法 4-2 的优势。

对于算法 4-2 的内均值计算问题,对具体的情况可以具体分析。比如除了算法 3-1,可采用 (3.27) 式的 Log-Euclidean 均值来代替,对于图像的协方差特征来说,这会是一个比较好的选择方案,不仅在计算时间上变成 $O(n)$,而且计算得到的均值使算法 4-2 具有更高的分类精度,详细的比较与分析在第七章中讲述。

4.3.4 时间复杂度和空间复杂度分析

下面对 Lie-Fisher 算法两种形式的时间和空间复杂度进行分析。针对 c 个类别的样本集,在算法 4-1 中,第一步至第四步的计算量分别为: $n, n+n, c+n, 1$,因此其总的时间复杂度为 $O(n)$ 。算法 4-2 的第二步和第三步的时间复杂度分别同算法 4-1 中的第三和第四步,而第一步中内均值计算则由算法 3-1 的复杂度决定,假设每个内均值计算的迭代次数均为 k 次,则算法 4-2 中第一步的时间代价为 $kn + \sum_{i=1}^c kn_i = 2kn$,因此算法 4-2 最终时间复杂度为 $O(2kn+n)$ 。很明显,两种形式的空间复杂度均为 $O(n)$ 。

4.4 本章小结

本章首先分析了线性空间 Fisher 判别的原理,然后介绍李群非线性流形空间通过李代数映射到线性空间的方法,并通过分析李群样本分布的流形结构。尝试在李群流形上寻找一个由总体样本内均值的李代数元素决定的单参数子群,这个单参数子群是原李群上的一条测地线。文中定义了样本到测地线投影的概念,同时将李群样本向该测地线投影,并尽可能使投影后各类别间的散度与类内散度比值最大化,从而实现非线性李群空间的类别判别。推导设计了两个 Lie-Fisher 算法,并从流形几何结构的角度的分析了两个 Lie-Fisher 算法各自的优缺点和时空复杂度。

第五章 李群核学习算法

本章基于李群数据样本, 结合矩阵点积空间理论, 分析并设计了相应的矩阵李群核函数。进一步实现了李群核算法 KLieDA。内容安排为: 5.1 节介绍 SVM 算法原理。5.2 节对 KFDA 算法的原理进行分析。5.3 节介绍核函数。5.4 节分析矩阵李群的度量函数, 并推导出 RBF、线性等 5 个矩阵李群核函数。5.5 节基于上节推导的李群核函数, 设计了李群核算法 KLieDA。5.6 节为本章小结。

5.1 SVM 算法原理

支持向量机 SVM^[55,56]是由 Vapnik 提出的基于统计理论的机器学习方法。其基本分类思想可以总结为两个要点: (i) SVM 针对线性可分情况进行分析, 对于线性不可分情况, 通过非线性映射函数将低维输入样本映射到高维特征空间中, 使之变为线性可分, 从而在这个高维空间中使用线性算法对样本进行分类。(ii) SVM 方法基于结构风险最小化理论, 在特征空间中按最大分类间隔构建最优分割超平面, 使学习器得到全局最优, 整个样本空间的期望风险以某个概率满足一定上界。

SVM 的目标就是要根据结构风险最小化原理, 构造一个目标函数将两类模式或样本尽可能地区分开来。下面从线性可分与线性不可分两种情况分析 SVM。

5.1.1 线性可分情况下的 SVM 分析

n 维样本集在线性可分情况下, 总意味着存在一个超平面使得训练样本完全分开, 该超平面可描述为: $\mathbf{v} \cdot \mathbf{x} + b = 0$, 其中, \mathbf{v} 为 n 维向量, b 为偏移量。最优分割超平面是满足一类数据与这个超平面距离最近的向量与超平面的间隔距离最大化。这个问题可以通过解下式的二次优化问题来获得:

$$\min \Phi(\mathbf{v}) = \frac{1}{2} \|\mathbf{v}\|^2 \quad (5.1)$$

满足约束条件: $y_i(v \cdot x_i + b) \geq 1, i=1, 2, \dots, n$ 。在特征数较大的情况下, 可以将此二次规划问题转化为它的对偶问题:

$$\max V(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (5.2)$$

满足约束条件:

$$\sum_{i=1}^n \alpha_i y_i = 0, \alpha_i \geq 0, i=1, 2, \dots, n \quad (5.3)$$

$$v^* = \sum_{i=1}^n \alpha_i y_i x_i \quad (5.4)$$

$$b^* = y_i - v \cdot x_i \quad (5.5)$$

这里 $\alpha = (\alpha_1, \dots, \alpha_n)$ 是 Lagrange 乘子, v^* 是最优超平面的法向量, b^* 是最优超平面的偏移量, 在这类优化问题的求解与分析中, KKT 条件将起到很重要的作用, 在 (5.5) 式中, 其解必须满足:

$$\alpha_i \{y_i(v \cdot x + b) - 1\} = 0, i=1, 2, \dots, n \quad (5.6)$$

从 (5.3) 式可知, 那些 $\alpha_i = 0$ 的样本对分类没有任何作用, 只有 $\alpha_i > 0$ 的样本才能影响分类, 是这些样本支撑起了最优分割超平面, 于是这些样本被称为支持向量。至此, 我们可以得到最终分类函数为:

$$f(x) = \sum_{i=1}^n y_i \alpha_i (x \cdot x_i) + b^* \quad (5.7)$$

根据 $f(x)$ 的符号来确定未知样本 x 的类别归属。

5.1.2 线性不可分情况下的 SVM 分析

对于样本集线性不可分的情况, 可以把样本集映射到高维特征空间 \mathcal{F} , 并在 \mathcal{F} 中运用原空间中满足一定条件的函数来实现内积运算, 将原始空间的非线性分类问题转换成目标空间的线性分类问题。根据泛函分析的有关理论, 只要一种核函数满足 Mercer

条件^[59]，它就对应某一空间中的内积。因此，只需选择适当的内积函数就可以解决这种线性不可分的问题。此时的目标函数为：

$$\max V(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) \quad (5.8)$$

其相应的分类函数为：

$$f(x) = \sum_{i=1}^n y_i \alpha_i K(x \cdot x_i) + b^* \quad (5.9)$$

5.2 KFDA 算法原理

核方法^[54]是一种非常有用的计算理论，它最初是在支持向量机 SVM^[55,56]中被用到。之后，被广泛应用到现有的线性算法中，例如基于核的 KPCA 算法^[57]，基于核的 KICA 算法^[58]等。下文将重点介绍的 KFDA 算法^[31]。

由于 FDA 算法只能对线性可分的数据进行有效处理，对于其它非线性分布或未知复杂分布的数据来说，其分类性能大打折扣。例如，对于图 5-1 中 6 种分布的样本集的分类问题，要将星号样本和加号样本分开，线性 FDA 算法就不能处理，因为永远都找不到一条直线将这些样本投影到该直线后，保持良好的区分性。

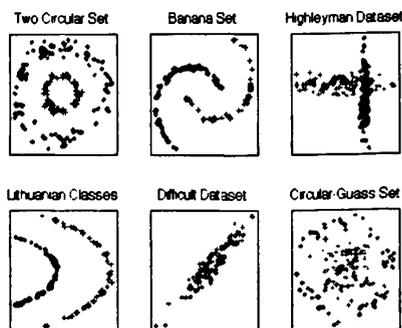


图 5-1. 6 类线性不可分样本集。上排从左至右分别为两个同心环分布、Banana 分布、Highleyman 分布的数据集，下排从左至右分别为 Lithuanian 分布、Difficult 分布、环形分布内嵌高斯分布的数据集。

因此,不能直接对这些看似非线性的样本用同等维度的空间中的线性算法来分类。假设这些样本在某个更高维度空间中是线性的,然后在这个高维度的空间中继续使用线性算法进行分类,这就是 KFDA 算法的动机。

假设 Φ 是一个非线性映射,它将样本映射到更高维的特征空间 \mathcal{F} 。为了在 \mathcal{F} 中寻找线性判别的投影方向,需要最大化下式:

$$J(v) = \frac{v^T S_B^\Phi v}{v^T S_W^\Phi v} \quad (5.10)$$

其中, $v \in \mathcal{F}$, S_B^Φ 和 S_W^Φ 是空间 \mathcal{F} 中相应的类间散度矩阵和类内散度矩阵。例如在两类分类问题中,它们分别定义如下:

$$S_B^\Phi = (m_1^\Phi - m_2^\Phi)(m_1^\Phi - m_2^\Phi)^T \quad (5.11)$$

$$S_W^\Phi = \sum_{i=1,2} \sum_{x \in \mathcal{X}_i} (\Phi(x) - m_i^\Phi)(\Phi(x) - m_i^\Phi)^T \quad (5.12)$$

其中, $m_i^\Phi = \frac{1}{l_i} \sum_{j=1}^{l_i} \Phi(x_j^i)$, l_i 为第 i 类样本个数, x_j^i 为第 i 类中第 j 个样本, \mathcal{X}_i 为第 i 类样本集合。

5.3 核函数

在 5.1 节中描述的目标特征空间 \mathcal{F} 的维数可以非常高,甚至是无穷维,因此直接求解是不可能的。为了克服这个困难,可以采用和 KPCA 以及 SVM 中类似的方法,即无需将数据显式映射到高维空间,代替的方案是寻找一个可以对训练样本直接得到其点积 $(\Phi(x) \cdot \Phi(y))$ 的函数。通过有效得计算这些点积从而解决原始问题,这可以通过使用 Mercer 核来实现^[56]。核 $k(x, y)$ 用于计算空间 \mathcal{F} 中特征向量的点积,即 $k(x, y) = (\Phi(x) \cdot \Phi(y))$ 。对于 k 的很多选择,已被证明是有效的。例如,在 SVM 或 KPCA 中的高斯 RBF 核 $k(x, y) = \exp(-\|x - y\|^2 / c)$, 又如多项式核 $k(x, y) = (x \cdot y)^d$, 其中的 c 和 d 为相应的正常数。

为了在特征空间 \mathcal{F} 中计算 Fisher 判别投影, 首先需要一个新的对应于 (5.10) 式的公式, 这个公式对输入样本的点积替换为核函数。从再生核的理论中知道任何方案 $v \in \mathcal{F}$ 必需处于由所有 \mathcal{F} 中的训练样本张成的空间中。因此 v 有如下形式:

$$v = \sum_{i=1}^l \alpha_i \Phi(x_i) \quad (5.13)$$

使用 (5.13) 式以及 m_i^Φ 的定义, 有:

$$\begin{aligned} v^T m_i^\Phi &= \frac{1}{l_i} \sum_{j=1}^l \sum_{k=1}^l \alpha_j k(x_j, x_k^i) \\ &= \alpha^T M_i \end{aligned} \quad (5.14)$$

上式中 $M_i = \frac{1}{l_i} \sum_{k=1}^l k(x_j, x_k^i)$, 这里用核函数替换了点积。根据 (5.10) 式和 (5.14) 式, 可以得到:

$$v^T S_B^\Phi v = \alpha^T M \alpha \quad (5.15)$$

上式中, $M = (M_1 - M_2)(M_1 - M_2)^T$ 。类似的, 使用 (5.13) 式以及 m_i^Φ 的定义, 有:

$$v^T S_W^\Phi v = \alpha^T N \alpha \quad (5.16)$$

上式中的 $N = \sum_{j=1,2} K_j (I - 1_{l_j}) K_j^T$, K_j 为 $l \times l_j$ 阶矩阵, 并且 $(K_j)_{nm} = k(x_n, x_m^j)$, 这个 $(K_j)_{nm}$ 也被称为第 j 类的核矩阵, I 为单位阵, 1_{l_j} 是所有元素均为 $1/l_j$ 的矩阵。

结合 (5.15) 式和 (5.16) 式, 即可在 \mathcal{F} 中得到 Fisher 线性判别:

$$J(\alpha) = \frac{\alpha^T M \alpha}{\alpha^T N \alpha} \quad (5.17)$$

此时, 可通过求解 $N^{-1}M$ 的特征值问题或计算 $\alpha = N^{-1}(M_1 - M_2)$ 得到解决。新样本 x 投影到 v 可以通过下式实现:

$$(v \cdot \Phi(x)) = \sum_{i=1}^l \alpha_i k(x_i, x) \quad (5.18)$$

数值问题及其正规化：从 l 个样本估计 l 维度的协方差矩阵，可能会使矩阵 N 变得不正定。为此，可简单加上单位矩阵的倍数，例如，将 N 替换为 N_μ ：

$$N_\mu = N + \mu I \quad (5.19)$$

上述问题也可视为以下不同的形式：(i) 它使问题的数值化更加稳定，对于足够大的 μ ， N_μ 将会变为正定。(ii) 这个方法类似于文献[60]，减少了基于样本特征值估计的基数量。(iii) 它强化了在 $\|\alpha\|^2$ 上的正规化，是具有小膨胀系数的合适方案。尽管这在正规化的问题上真实影响还未知，但在 SVM^[55] 中显示出确实有联系。另外，还有一个可以使用的正规化 N 的方法，即使用惩罚项 $\|v\|^2$ ，这类似于 SVM (加上全核矩阵 $K_{ij} = k(x_i, x_j)$)。

5.4 李群核函数

对于矩阵李群样本集，由于矩阵本身没有点积运算，而群运算中的矩阵乘法，其结果仍是矩阵，因此，基于向量的核函数不能用于矩阵李群样本集，需要设计新的核函数。

5.4.1 李群 RBF 核

高斯 RBF 核函数是应用最多的核函数之一。之所以被称为径向基函数，是因为它能对图 5-1 中同心环分布的非线性数据集进行很好的高维空间线性映射。基于向量的 RBF 核函数的形式为：

$$k(x, y) = \exp\left(\frac{-\|x - y\|^2}{2\sigma^2}\right) \quad (5.20)$$

上式分子部分对两个向量 x, y 的差向量求模长，对于普通矩阵李群不存在矩阵减法运算。但是，在李群所在的流形空间，从一个元素中减去另一个元素的成份可以通过群运算与群元素的求逆运算实现，例如， $x^{-1}y$ ， $x, y \in G$ 。但是除了交换李群，一般李群

的群运算是不可交换的，即 $\|x^{-1}y\| \neq \|y^{-1}x\|$ 。但是对矩阵李群，有 $\|\log(x^{-1}y)\| = \|\log(y^{-1}x)\|$ ，其中 $\|\cdot\|$ 为 Frobenius 范数。从另一个角度看， $\|x-y\|$ 是线性空间中两个点 x, y 之间的欧氏距离， $\|\log(x^{-1}y)\|$ 是李群流形空间中两个点之间的测地线距离。从 x 出发到 y 的测地线距离 $\|\log(x^{-1}y)\|$ 和从 y 出发到 x 的测地线距离 $\|\log(y^{-1}x)\|$ 是相等的。另外，任意两个李群元素之间的测地线距离是非负的，这就满足了核矩阵 K 为对称正定矩阵的要求。

根据上面的分析，可以得到李群矩阵 RBF 核函数：

$$k(x, y) = \exp\left(\frac{-\|\log(x^{-1}y)\|^2}{2\sigma^2}\right) \quad (5.21)$$

上式将非线性李群样本映射到无限维空间中，并完成点积计算。如果样本在原李群流形空间中的分布符合径向特点，那么使用该核函数将得到理想的样本线性化结果。对于 (5.21) 中分子部分利用 BCH 公式第一阶进行近似，可以得到近似的李群 RBF 核函数：

$$k(x, y) = \exp\left(\frac{-\|\log(y) - \log(x)\|^2}{2\sigma^2}\right) \quad (5.22)$$

5.4.2 李群多项式核

多项式核在向量特征空间中的形式为： $k(x, y) = (x \cdot y + 1)^d$ 。向量的点积运算可交换，但矩阵的乘法运算不可交换，因此，基于向量的多项式核函数也不能直接用于矩阵李群样本集。

由矩阵论可知，在 $n \times n$ 矩阵张成的空间上可以定义一个内积：

$$\langle x, y \rangle = \text{tr}(y^H \cdot x) \quad (5.23)$$

其中， tr 为矩阵求迹运算，而 $y^H = \bar{y}^T$ ，即 y^H 是 y 的共轭转置矩阵：

$$y^H = \begin{pmatrix} \overline{a_{11}} & \cdots & \overline{a_{1n}} \\ \vdots & \ddots & \vdots \\ \overline{a_{n1}} & \cdots & \overline{a_{nn}} \end{pmatrix} = y^T \quad (5.24)$$

由 (5.23) 式定义的矩阵内积, 可以推导出很多矩阵空间的几何量。例如矩阵空间中两个矩阵向量的夹角可以定义为 $\theta = \arccos(\text{tr}(y^H \cdot x) / (\|x\| \cdot \|y\|))$ 。同时, 该矩阵内积符合向量内积所具有的所有性质, 因此也满足 $\langle x, y \rangle = \text{tr}(y^H \cdot x) = \text{tr}(x^H \cdot y) = \langle y, x \rangle$ 。有了这个性质, 就容易推导出基于矩阵李群的多项式核函数:

$$k(x, y) = (\text{tr}(y^H \cdot x) + 1)^d \quad (5.25)$$

上式中, d 为任意正实数, 通常取正整数。

5.4.3 李群线性核

向量空间线性核函数为 $k(x, y) = x \cdot y$, 根据 (5.23) 式定义的矩阵内积, 可以直接得到李群线性核函数:

$$k(x, y) = \text{tr}(y^H \cdot x) \quad (5.26)$$

5.4.4 李群感知器核

应用最多的感知器函数是 Sigmoid 函数和双曲正切函数 \tanh 。双曲正切函数的形式为 $\tanh(b(x \cdot y) + c)$, $b > 0, c < 0$, 对应于 (5.23) 式, 得到李群感知器核函数:

$$k(x, y) = \tanh(b(\text{tr}(y^H \cdot x)) + c) \quad (5.27)$$

同理, Sigmoid 函数为 $\frac{1}{1 + e^{-x}}$, 得到 Sigmoid 形式的李群核函数:

$$k(x, y) = \frac{1}{1 + e^{c(\text{tr}(y^H \cdot x))}} \quad (5.28)$$

除 (5.22) 式采用李群测地线距离, 其余 (5.23) 式及后续核函数不仅可以用于矩阵李群, 也同样适用于一般的矩阵样本。

5.5 基于李群核函数的 KLieDA 算法

借助 5.4 小节推导的若干李群核函数可以设计出基于李群核函数的 KLieDA 算法, 多类问题的推导思路类似于 Lie-Fisher, 这里给出二分类的 KLieDA 算法, 即算法 5-1。

算法 5-1: 两类分类的李群核 KLieDA 分类算法

输入: $\{x_{ij}\}_{i=1,2}^{j=1..n_i} \in G$, x_{ij} 表示第 i 个类别的第 j 个样本, n_i 表示第 i 个分类中训练样本的个数。

输出: 每个未知样本的类别归属。

过程:

1. 对每一个李群样本 x_{ij} , 求出两类李群样本的均值 μ_1 和 μ_2 , 对一般李群使用算法 3-1, 对方差样本采用 (3.27) 式求 Log-Euclidean 均值。

2. 选择具体李群核函数, 求得两类李群样本的 $M_i = \frac{1}{n_i} \sum_{k=1}^{n_i} k(x_j, x_k^i)$, 并令

$$M = (M_1 - M_2)(M_1 - M_2)^T.$$

3. 计算 $N = \sum_{i=1,2} K_i(I - 1_{n_i})K_i^T$, 其中 K_i 为 $n \times n_i$ 阶矩阵, 且 $(K_i)_{nm} = k(x_n, x_m^i)$ 。

4. 计算 α : $\alpha = N^{-1}(M_1 - M_2)$ 。

5. 计算两类均值 μ_1 和 μ_2 到空间 \mathcal{F} 中的向量 v 上的投影:

$$\tilde{\mu}_1 = (v \cdot \Phi(\mu_1)) = \sum_{i=1}^c \sum_{j=1}^{n_i} \alpha_i k(x_{ij}, \mu_1), \quad \tilde{\mu}_2 = (v \cdot \Phi(\mu_2)) = \sum_{i=1}^c \sum_{j=1}^{n_i} \alpha_i k(x_{ij}, \mu_2).$$

6. 对未知样本 x , 先将其投影到空间 \mathcal{F} 中的向量 v 上:

$$\tilde{x} = (v \cdot \Phi(x)) = \sum_{i=1}^c \sum_{j=1}^{n_i} \alpha_i k(x_{ij}, x), \quad \text{再通过 } i^* = \arg \min_{i=1,2} |\tilde{x} - \tilde{\mu}_i| \text{ 得到其类别归属。}$$

算法 5-1 相比传统 KFDA 算法, 在保持了相同的时空复杂度下有两点创新: (i) 使用矩阵李群核函数, 能处理矩阵样本; (ii) 算法的第 1 和第 6 步中先计算原样本的均

值以及均值在 \mathcal{F} 空间中投影, 对于未知样本映射至 \mathcal{F} 空间中后, 只判断与每类均值点的距离, 计算量较小。

KLieDA 算法和 Lie-Fisher 算法都能对非线性矩阵李群样本进行分类。从分类过程来看, KLieDA 是通过将非线性可分样本映射到高维空间, 使之线性可分, 在分类期间采用李群均值加速完成未知样本的类别判别, 整个分类过程并不在样本所在的李群流形空间中完成。而 Lie-Fisher 是利用李群均值以及样本到测地线投影的方法在样本所在流形空间中寻找到一条分类测地线完成样本分类, 原始样本的分布特点直接决定学习到的测地线的分类性能。从计算开销来看, KLieDA 由于需要计算核矩阵, 而核矩阵的计算时间复杂度是 $O(n^2)$, 比 Lie-Fisher 计算两个散度矩阵的 $O(n)$ 明显要大, 空间复杂度也因为核矩阵的原因达到 $O(n^2)$, Lie-Fisher 只需 $O(n)$ 。因此, 对于大样本集的分类问题, KLieDA 算法将会明显比 Lie-Fisher 慢。从算法应用角度来看, Lie-Fisher 是学习到一条固定的测地线进行分类, 而 KLieDA 则可根据不同的样本集分布特点选择不同的李群核函数, 使算法达到更好分类结果, 因此 KLieDA 算法比 Lie-Fisher 算法在应用上更具灵活性。

5.6 本章小节

本章主要是分析了矩阵李群在高维点积空间中的核函数问题。介绍了李群李代数的相关理论与流形测地线距离, 并定义了矩阵内积, 进一步推导出相应的运用于矩阵李群的核函数。基于李群核函数设计了 KLieDA 算法, 该算法不仅能处理向量样本, 同时也能处理矩阵样本。利用文章定义的 5 个李群核函数, SVM、KPCA、KFDA 等算法将都能处理矩阵样本。

第六章 实例应用

本章将通过四个实验验证本文提出的算法的效用性。内容安排如下：6.1 节为 Lie-Fisher 两个算法应用在人工数据集和 MNIST 手写体上的分类实验，同时设计了一个从手写体中提取李群特征的方法。6.2 节为 KLieDA 算法和基于李群核的 SVM 算法结合区域图像协方差特征在 MNIST 数据集以及 USPS 数据集上的分类。6.3 为本章小结。

6.1 Lie-Fisher 算法实验分析

这部分我们将通过两个实验来验证 Lie-Fisher 算法的性能。第一个实验是对人工几何形状数据集的分类，第二个是对 MNIST 手写体的分类。由于本文的均值计算与分类算法是在样本的流形空间中完成的，为了进行算法比较，我们将和在欧式空间中分类的 FLDA 算法，在流形空间中分类的 Lie-Mean，以及使用测地线距离的 KNN 算法一起对人工和真实数据集进行分类比较，使得到的实验结果尽可能全面表达 Lie-Fisher 算法的特点。第二个实验先对 MNIST 手写体数字构造协方差李群特征，再使用算法 4-1 和算法 4-2 进行分类。

6.1.1 人工数据集

几何图形的分类问题在基本的图形图像处理与机器视觉问题中具有重要的意义。实际问题中的几何图形可能存在形体残缺或大量的噪声，针对这个问题，可通过随机采样和高斯噪声分别模拟残缺和噪声，这里设计两种基本几何图形，并使用 Lie-Fisher 算法对它们进行分类，并与其他方法进行比较。

首先用生成点云的方式构造两类数据集：第一类是单位正方形，第二类是单位圆，在每个单位正方形以及每个单位圆上分别随机取 n 个点构成点云，这 n 个点组合在一起构成单个样本，而不是一个点为一个样本。然后对这些点云添加相应方差的高斯噪声形成两个含噪声的新样本。如图 6-1 所示，以该方法构造 50 个圆和 50 个正方形，即 100 个训练样本。

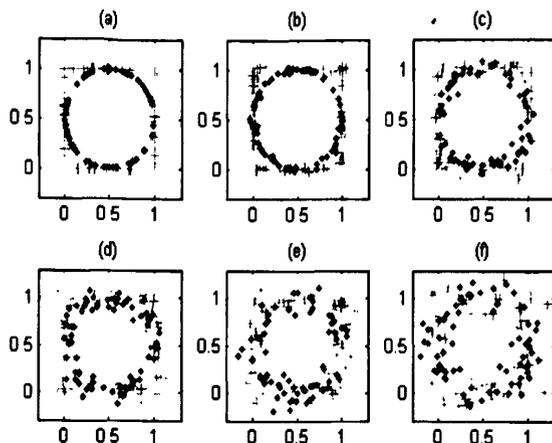


图 6-1. (a) 是由在单位正方形上和单位圆上各随机取 70 个点构成的两个样本, 注意这里 70 个点构成的正方形为一个单一样本, 70 个点的圆也是一个单一样本, (b)~(f) 是对原始样本(a)分别添加了均值为 0, 方差为 0.02, 0.04 直到 0.10 的二维高斯随机噪声后形成的两个新样本。

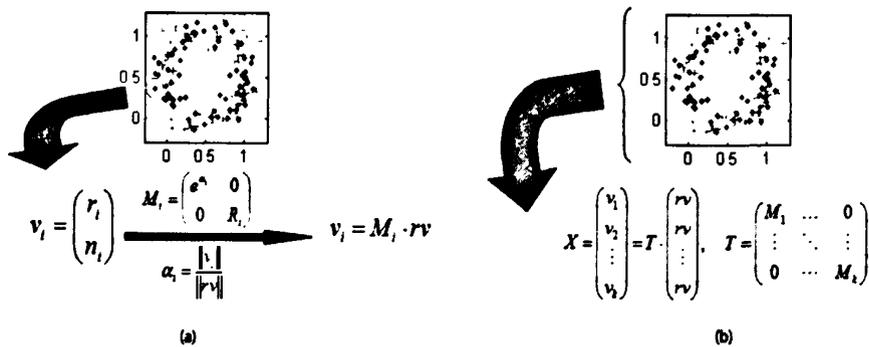


图 6-2. (a) 将带高斯噪声的圆样本中的一个点 v_i 通过参考向量 rv 用李群 M_i 来表示。(b) 对样本中的每个点都采用 (a) 的方法, 将一个完整的样本 X 提取李群特征 T 。

样本集的构造过程采用文献[10]的方法, 因为构成样本的点云位于 R^2 中, 连接原点和点云中的任意一点可构成一个二维向量 v_i 。将这个向量变换到一个指定的参考向量 rv , 不失一般性可令 $rv = (1,1)$, 对于每一个向量, 我们分别用两个参数, 即向量的模长和向量的方向来表示 $v_i = (r_i, n_i)^T$, 其中 r_i 表示向量 v_i 的模长 $\|v_i\|$, n_i 是 v_i 的方向且

$\|n_i\|=1$ 。那么将向量 v_i 变换到向量 rv ，可通过一个变换 M_i 分别对模长进行缩放，对方向进行旋转来实现： $v_i = M_i \cdot rv$ ，其中 $M_i = \begin{pmatrix} e^{\alpha_i} & 0 \\ 0 & R_i \end{pmatrix}$ ， M_i 内的 $\alpha_i = \|v_i\|/\|rv\|$ 是缩放系数， R_i 是一个 R^2 空间中的旋转矩阵，即 $SO(2)$ ，由于 rv 是固定的，因此 v_i 和 M_i 确定了一一对应的关系，即对每一个点 v_i 可用 M_i 来表示。样本的构建过程如图 6-2 所示，对一个由 k 个点构成的样本，可构建为：

$$X = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{pmatrix} = T \cdot \begin{pmatrix} rv \\ rv \\ \vdots \\ rv \end{pmatrix}, \quad T = \begin{pmatrix} M_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & M_k \end{pmatrix}$$

现在，用 T 来表示样本 X 。因为每个 M_i 是结构为 $R \times SO(2)$ 的李群， T 又是 k 个 M 的直积形式，因此， T 构成一个 $2k$ 维的李群。对李群 T 求李代数元素可得：

$$\log(T) = \begin{pmatrix} \log(M_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \log(M_k) \end{pmatrix}, \quad \text{其中, } \log(M_i) = \alpha_i \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \theta_i \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

α_i 上文已给出， θ_i 是向量 v_i 和参考向量 rv 的夹角： $\theta_i = \arccos(v_i \cdot rv / (\|v_i\| \cdot \|rv\|))$ 。

对样本集中每一个样本用上述方法提取特征 T ，应用 Lie-Fisher 算法以及 FLDA、KNN、Lie-Mean 等算法进行分类实验，结果见表 6-1。

表 6-1. 样本在不同噪声下各算法的分类准确率(%)

σ^2	Lie-Mean	5-NN	FLDA	Lie-Fisher (算法 4-1)	Lie-Fisher (算法 4-2)
0.00	100.00	76.67	96.67	100.00	100.00
0.02	94.00	76.00	94.00	100.00	100.00
0.04	91.67	85.33	95.00	100.00	100.00
0.06	85.00	78.67	91.67	95.00	95.00
0.08	84.00	76.67	88.33	93.33	93.33
0.10	82.67	84.67	93.33	91.34	91.33
0.20	80.00	80.00	83.33	90.00	90.00

表 6-1 是两类样本在不同方差高斯随机噪声干扰后的分类结果。上述数据在训练样本为 $n=100$ ， $k=70$ ，测试样本为 30 的配置下重复运行 5 次得到的平均识别率。此时由

于每个 M 是 2 维李群, 因此, T 构成维数为 140 的李群。Lie-Mean 算法是在李群流形空间中计算两类样本的均值, 按 (3.24) 式判别测试样本的类别, 5-NN 是直接在李群流形空间按测地线距离最近邻的 5 个样本判定类别, 这里的 FLDA 算法和向量样本的 FLDA 方法相似, 不同之处是将矩阵 T 作为线性样本, 均值计算采用 (3.3) 式, 即利用李群样本的外均值作为 Fisher 运算中的均值, 然后运用向量形式的 FLDA 算法求解矩阵 v , 计算测试样本的类别用投影后样本到投影后各类样本中心的距离来决定:

$$i^* = \arg \min_{i=1,2,\dots,c} \|v^T T - \tilde{\mu}_i\|.$$

从表中数据可以看出, Lie-Mean 算法随着噪声加大, 识别率也有随之下降。Lie-Fisher 算法 4-1 和算法 4-2 性能几乎一致, 且明显好于其他算法, 对于噪声没有其她几种算法敏感, 是一种相对稳定的、抗噪能力较好的算法。5-NN 与其他算法相比效果较差。

6.1.2 手写体识别

和以往的特征提取和表示方法不同, 本小节将利用上一小节中李群特征表示方法, 对二维手写体图像的形状信息提取李群特征, 并进行分类。

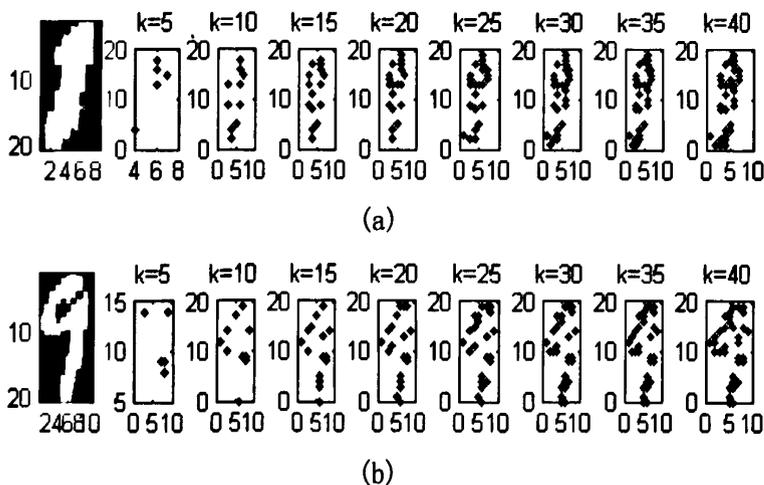


图 6-3. 手写体样本构造方法, (a) 为手写体 1, (b) 为手写体 9

本实验采用 MNIST 数据集^[90]，该数据集中的每个手写数字是 28×28 的 256 阶灰度图片。实验对数字 1 和 9 进行分类，首先对原始灰度图像用阈值灰度 100 进行二值化，然后对图像随机采样 k 个点，使这 k 个点都位于代表手写数字的像素上。如图 6-3 所示，最左侧图像为原始手写数字取其最小有效外接矩形后的结果，右侧 8 个图是在二值化处理后，随机采样 k 点形成的点云， k 从 5 取到 40，间隔为 5。

首先从手写体数据集中的 train1 和 train9 中各随机挑选 200 个，一共 400 个数字作为训练集，从 test1 和 test9 中各随机挑选 100 个，共 200 个数字做为测试集，用上述方法随机取 k 个点形成点云，将点云中的点按坐标从上至下，从左至右排序，以使得同一类数字图像的点云相对稳定。使用 6.1.1 小节描述的方法对每一个排序后的点云构造 $k \cdot (R \times SO(2))$ 维的李群元素 T 。应用 Lie-Mean、李群 FLDA、Lie-Fisher 算法 4-1 和算法 4-2 进行分类。每次样本都是随机选择，重复运行 5 次，得到结果如图 6-4 所示。

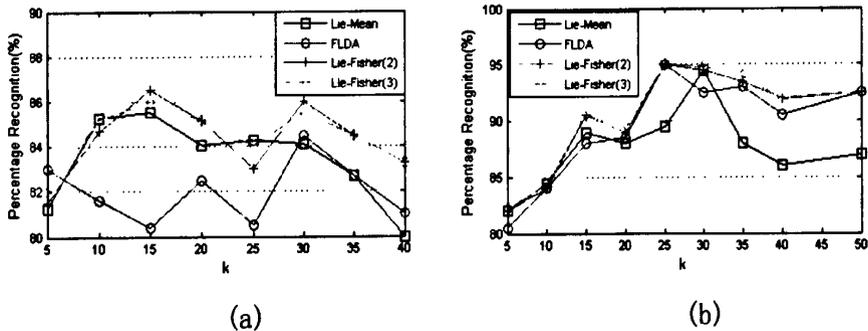


图 6-4. 四种算法在不同的 k 值下对手写体的分类结果。

上图中，Lie-Fisher(2)代表算法 4-1，Lie-Fisher(3)代表算法 4-2。(a)是未对点云中的点进行排序的结果，很明显 Lie-Fisher 算法比其他两个算法分类效果要好，当 k 取值小于 20 时，算法 4-2 的性能略好于算法 4-1，基于李群内均值的 Lie-Mean 算法也明显要优于 FLDA。(b)是对点云进行排序后的结果，随着 k 值从 5 增长到 25 时，Lie-Fisher 算法 4-1 和算法 4-2 达到最高识别率 95%，两者性能非常接近，在 k 取值小于 40 时，算法 4-2 的性能要比算法 4-1 好出 0.1%至 1%，当 k 取值大于 40 时，算法 4-

1 和算法 4-2 分类效果则基本一致。从总体来看, 两个 Lie-Fisher 算法也明显优于其他两种算法。

6.1.3 协方差李群特征的 Lie-Fisher 手写体识别

本节仍采用 MNIST 手写体数据集和 Lie-Fisher 分类算法, 不同的是, 对手写体提取的特征不是点云李群特征, 而是图像协方差李群特征。

6.1.3.1 图像协方差特征

对于常规图像, 它是灰度值或三维彩色分量的集合, 可用 I 表示, 每个像素的属性可以表示为 $I(x, y)$ 。该方法也同样可以用于表示其他图像格式, 比如红外图像。用 F 表示从图像 I 中提取的图像特征:

$$F(x, y) = \phi(I, x, y) \quad (6.1)$$

上式中的函数 ϕ 可以表示任何有关图像属性的映射, 比如灰度值、各彩色通道分量值、梯度值等。 $\phi(I, x, y)$ 的值域可以张成 d 维实空间, 对于所有像素对应的 $\phi(I, x, y)$ 值均为 d 维实空间中的一个点。取所关心的图像 I 中的局部矩形区域 $R \subseteq F$, 区域内的所有像素对应的 ϕ 值可以表示为 $\{z_i\}_{i=1}^n$, z_i 为 d 维实向量, n 为区域 R 内像素个数。对图像区域的向量集 $\{z_i\}_{i=1}^n$ 求其协方差矩阵:

$$C_R = \frac{1}{n+1} \sum_{i=1}^n (z_i - \mu)(z_i - \mu)^T \quad (6.2)$$

上式中的 μ 为向量集 $\{z_i\}_{i=1}^n$ 的均值。

图像的区域协方差特征拥有很多优良的特性: 特别是可以使用积分图像方法, 计算速度快, 方便计算图像中任一子区域的协方差特征; 分类性能好; 特征维数低, 且维数不受图像尺寸的影响。图像协方差特征的详细介绍可参考文献[20]。

本实验对 MNIST 数据库中的图像, 取 $R = F$, 以及 $\phi(I, x, y)$ 为:

$$\phi(I, x, y) = \left(x, y, I(x, y), \left| \frac{\partial}{\partial x} I(x, y) \right|, \left| \frac{\partial}{\partial y} I(x, y) \right| \right)^T \quad (6.3)$$

$$\phi(I, x, y) = \left(x, y, I(x, y), \left| \frac{\partial}{\partial x} I(x, y) \right|, \left| \frac{\partial}{\partial y} I(x, y) \right|, \sqrt{\left| \frac{\partial}{\partial x} I(x, y) \right|^2 + \left| \frac{\partial}{\partial y} I(x, y) \right|^2} \right)^T \quad (6.4)$$

$$\phi(I, x, y) = \left(x, y, I(x, y), \left| \frac{\partial}{\partial x} I(x, y) \right|, \left| \frac{\partial}{\partial y} I(x, y) \right|, \left| \frac{\partial^2}{\partial x^2} I(x, y) \right|, \left| \frac{\partial^2}{\partial y^2} I(x, y) \right| \right)^T \quad (6.5)$$

$$\phi(I, x, y) = \left(x, y, I(x, y), \left| \frac{\partial}{\partial x} I(x, y) \right|, \left| \frac{\partial}{\partial y} I(x, y) \right|, \left| \frac{\partial^2}{\partial x^2} I(x, y) \right|, \left| \frac{\partial^2}{\partial y^2} I(x, y) \right|, \tan^{-1} \left(\left| \frac{\partial}{\partial x} I(x, y) \right| / \left| \frac{\partial}{\partial y} I(x, y) \right| \right) \right)^T \quad (6.6)$$

(6.3) ~ (6.6) 式中 $I(x, y)$ 为坐标 (x, y) 处的亮度值, $\frac{\partial}{\partial x} I(x, y)$ 为 (x, y) 处的横向

一阶梯度, $\frac{\partial^2}{\partial x^2} I(x, y)$ 为二阶梯度, $\sqrt{\left| \frac{\partial}{\partial x} I(x, y) \right|^2 + \left| \frac{\partial}{\partial y} I(x, y) \right|^2}$ 为一阶梯度向量的模长,

$\tan^{-1} \left(\left| \frac{\partial}{\partial x} I(x, y) \right| / \left| \frac{\partial}{\partial y} I(x, y) \right| \right)$ 为一阶梯度向量比值的反正切。这样建模后, 对 (6.3) ~

(6.6) 四个函数, 可以分别得到的 z_i 为 5 维、6 维、7 维和 8 维实向量。对于 (6.6)

式, 计算得到的 C_R 为 8×8 的实对称矩阵, 因为 C_R 是对称矩阵, 因此只有 $\sum_{i=1}^n i \Big|_{n=8} = 36$

维, 是比许多图像特征更低维的特征。协方差矩阵是实对称正定矩阵, 关于这类矩阵的空间结构、度量、均值等性质的研究已取得不少成果^[14, 22, 25]。

图像的协方差矩阵张成了一个张量空间, 每个矩阵为该空间内的一个点, 对于任意两个点 C_1, C_2 之间的距离由文献[53]给出:

$$\rho(C_1, C_2) = \sqrt{\sum_{i=1}^n \ln^2 \lambda_i(C_1, C_2)} \quad (6.7)$$

其中, $\{\lambda_i(C_1, C_2)\}_{i=1}^n$ 为 C_1, C_2 的广义特征值, 可通过下式计算得到:

$$\lambda_1 C_1 x_i - C_2 x_i = 0, \quad i=1, \dots, d \quad (6.8)$$

满足 $x_i \neq 0$ 则为广义特征向量，可以证明 (6.7) 式的 ρ 满足第二章的关于度量的几条公理。此度量公式同样也满足正定矩阵的李群结构，可以从正定矩阵的李代数推导得到。

6.1.3.2 协方差特征手写体识别

同实验 6.1.2，本节对手写体 1 和 9 进行分类。对 (6.3) ~ (6.6) 式构造的协方差样本 C_R 集采用算法 4-1 和算法 4-2 进行分类，对算法 4-2 中的协方差样本间度量采用 (6.7) 式。首先，分别从训练集中随机选择两类训练样本各 100 个，以及从测试集中随机选择两类测试样本也各 100 个，使用 (6.3) ~ (6.6) 式构造协方差矩阵，然后对每一种情况分别运行 10 次。实验中，对算法 4-2 中的均值计算采用 Log-Euclidean 平均，即 (3.27) 式，算法 4-1 不变。得到结果如图 6-5 所示。

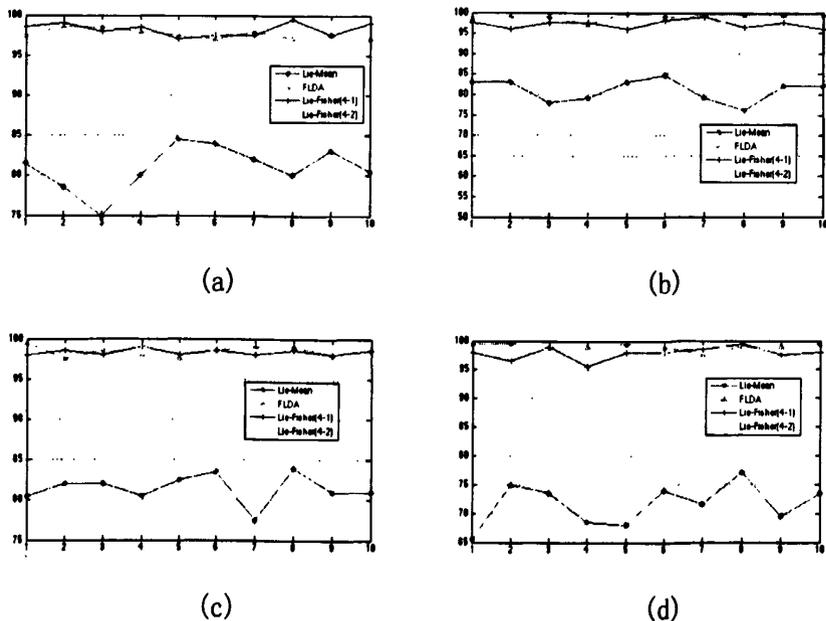


图 6-5. (a) ~ (d) 分别为采用 (4.21) ~ (4.24) 式的实验结果。横坐标为运行迭代次数，纵坐标为分类准确度。Lie-Fisher(4-1) 代表算法 4-1，Lie-Fisher(4-2) 代表算法 4-2。

从图 6-5 的实验结果可以清楚地看出, 采用 (6.3) 式时算法 4-1 在大多数情况下的分类精度是最高的, 采用 (6.5) 式时算法 4-2 和基于矩阵的 FLDA 算法基本持平。但在采用 (6.4) 以及 (6.6) 式时, 两个 Lie-Fisher 算法分类准确度略低于 FLDA 算法, 说明此时的样本分布基本处于线性可分, 两个 Lie-Fisher 算法在求样本均值和李群李代数映射时损失了精度, 导致分类精度下降。图 6-5 (b) 相比 (a), 因为采用了一阶梯度向量长度的算了, 使得算法 4-2 的分类精度变得十分不稳定, 而 FLDA 和算法 4-1 则比较稳定且精度较高。这很可能是因为加入该算了后, 使得生成的协方差流形结构总体接近线性但局部不平坦, 使得算法 4-2 尽管采用 Log-Euclidean 平均, 但在计算 (4.17) 式时使用的是标准的李群距离 (3.3) 式而不是仿射不变距离 (3.25) 式, 计算过程中产生的误差在多次李群李代数之间映射时被放大。

另外, 图 6-5 (a) 中对 z_i 只采用了 5 维向量, 计算速度更快。算法 4-1 在主频为 1.6 的 intel i5 双核 CPU, 内存 4G 以及 Matlab 2011b 的配置下, 200 个训练样本, 200 个测试样本的总训练加分类时间小于 1 秒, 而分类精度基本都能维持在 98% 左右, 如图 6-6。因此, Lie-Fisher 算法 4-1 适合应用于速度要求比较高的场合, 如在目标跟踪、快速检测等领域。

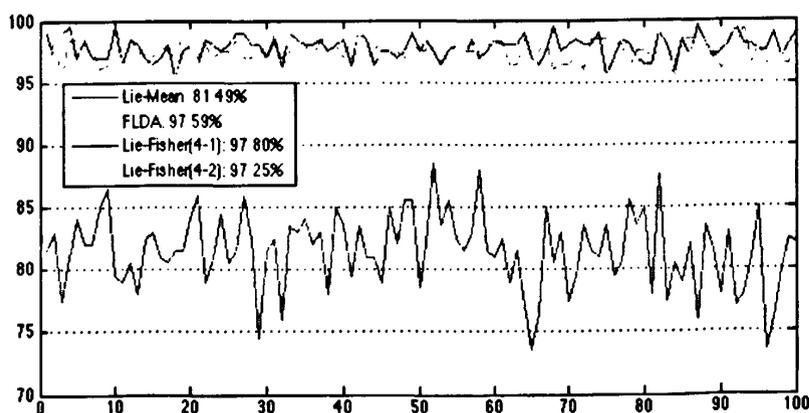


图 6-6. 采用 (6.3) 式计算 100 次后绘制的分类精度图。横坐标为计算次数, 纵坐标为分类精度。

下图为算法 4-1 分类错误的测试样例。



图 6-7. 算法 4-1 错分样例。图中错分样例均为手写数字 1。

6.2 KLieDA 算法及李群核 SVM 实验分析

这部分将使用上一章中推导的李群核函数，配合 SVM 算法以及 KLieDA 算法展开实验。6.2.1 小节利用李群核 SVM 算法在 MNIST 以及 USPS 手写体数据集上进行分类实验。6.2.2 小节利用 KLieDA 算法在上述两个数据集上进行分类实验。

6.2.1 基于李群核 SVM 的手写体识别

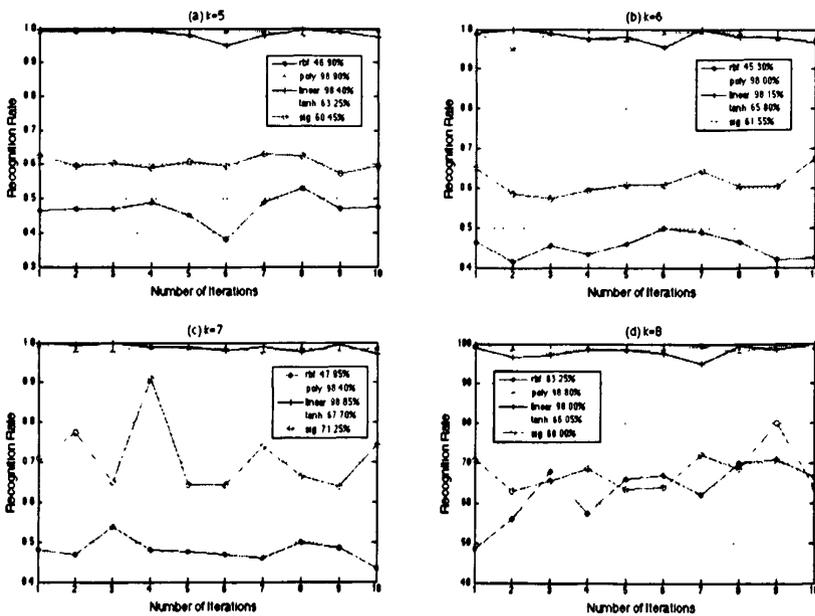


图 6-8. (a) ~ (d) 是对 MNIST 数据集中的数字 1 和 9 进行分类的结果。纵坐标是分类识别率，横坐标是运行次数。

采用 (6.3) ~ (6.6) 四个公式分别构造图像的协方差特征, 然后采用 (5.21) 式, (5.25) 式, (5.26) ~ (5.28) 式作为核函数配合使用 SVM 算法, 对 MNIST 数据集中的数字 1 和 9 进行分类。本实验中的 SVM 算法采用 libSVM2.8 配合 Matlab7.0 完成。对于李群 RBF 核, 取 $2\sigma^2 = 10^9$, (5.25) 式中取 $d = 2$, (5.27) 式中取 $b = 10^{-9}$ 以及 $c = 0$, (5.28) 式中取 $c = -10^{-9}$ 。对每个类别分别随机抽取 100 个作为训练, 100 个做为测试, 于是实验的配置是 200 个训练样本, 200 个测试样本。分别构造 5~8 维 z_i 下的协方差样本, 实验得到的结果如图 6-8 所示。标题中的 k 值代表图像像素的 z_i 值, 分别取从 5 至 8, 并构造的协方差样本, 然后进行 SVM 分类得到的结果。

下面用同样的方法在 USPS 数据集上进行实验。USPS 手写体数据集是美国邮政部门收集整理的手写体数据集, 包含 0~9 十个数字, 每个数字对应 16*16 分辨率的图片, 每个图片像素均为 256 阶灰度, 且每个数字都有 1100 个样本。图 6-9 为其部分样本的显示效果。李群核 SVM 在 USPS 上的实验结果见图 6-10, 参数含意同图 6-8。

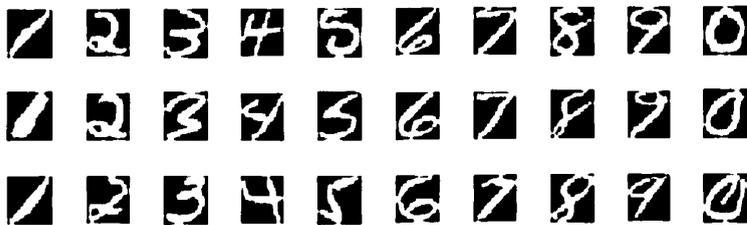


图 6-9. USPS 数据集样本图像, 从左右至右, 分别是数字 1 至 0 的图像。

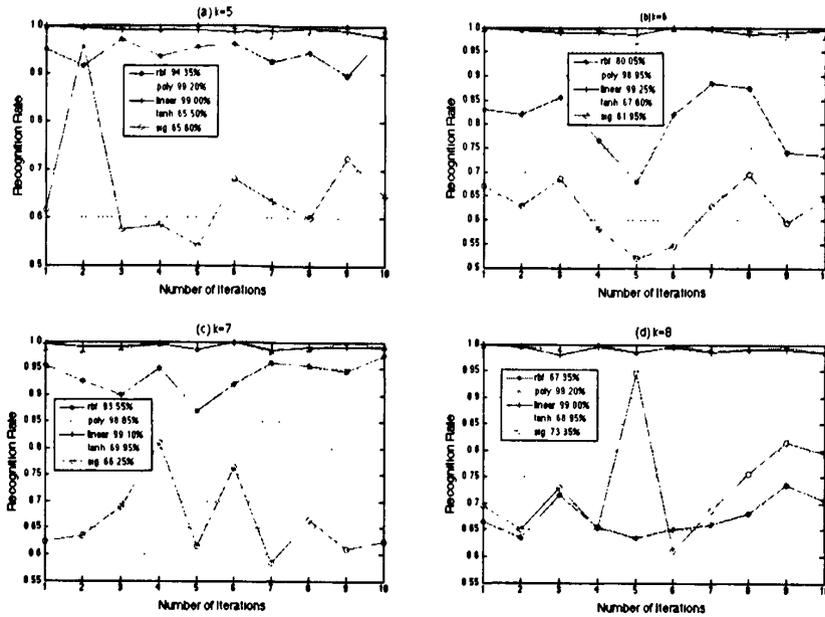


图 6-10. (a) ~ (d) 是对 USPS 数据集中的数字 1 和 9 进行分类的结果。纵坐标是分类识别率，横坐标是运行次数。

从图 6-10 中，我们得到明显的结果是：线性核和多项式核的分类准确率明显比其他核要高很多。而线性核和取 $d=2$ 的多项式核两者之间差别不大。

6.2.2 基于李群核 KLieDA 算法的手写体识别

使用 KLieDA 算法在 USPS 数据集上进行分类，算法的运行环境与核函数参数配置同上一小节。KLieDA 算法结合文中推导的五个核函数在 USPS 的数字 1 和 9 上的具体分类结果如图 6-11 所示。

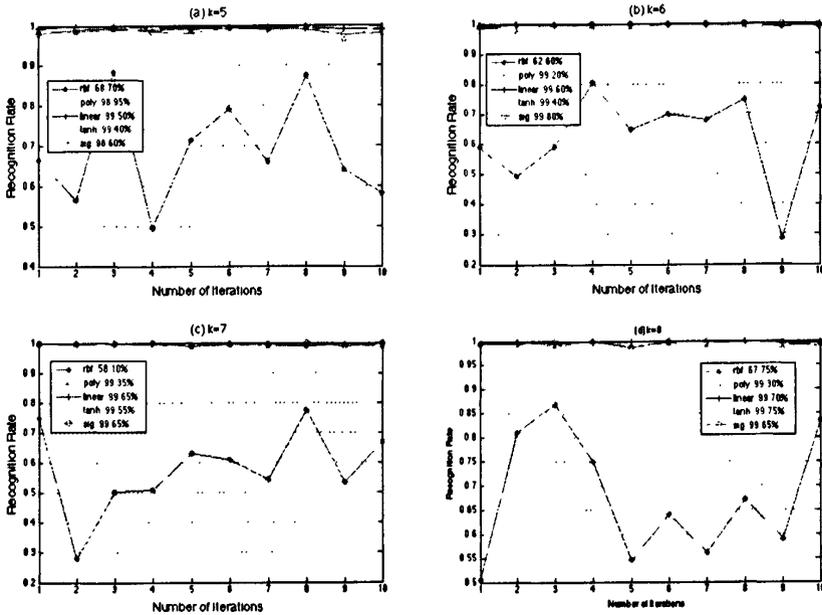


图 6-11. (a) ~ (d) 是使用 KLieDA 算法对 USPS 数据集中的数字 1 和 9 进行分类的结果。纵坐标是分类识别率，横坐标是运行次数。

从图 6-11 中明显看出，在采用相同的核函数及参数的情况下，两个感知器核在 KLieDA 算法上比 SVM 算法更有效，而且表现出了非常高的精确度。对于指定 k 值的情况下，KLieDA 算法要比 SVM 的分类精度更高。

另外一个重要的发现是：线性核对协方差样本的分类精度最高，多项式核其次。由于在本文实验中，两个协方差矩阵的点积 $\langle x, y \rangle = \text{tr}(y^H \cdot x)$ 的值往往会很大。例如当 k 取 5 时， $\text{tr}(y^H \cdot x)$ 的大小基本在 10^8 左右，而 k 取 5 时的结果介于 10^8 和 10^9 之间，因此，考虑到 tanh 和 sig 核函数的特点，针对前者取 $b = 10^{-9}$ 以及后者 $c = -10^{-9}$ ，使这两个感知器函数的输入值基本维持在 [0,1] 区间内。此定义域内 tanh 和 sig 函数的曲线变化如图 6-12 所示，将感知器核的输入控制在 [0,1] 区间，可使相应的核函数值域处于较陡的梯度范围之内，有利于进行分类判别，避免进入函数对输入值不敏感的区域。

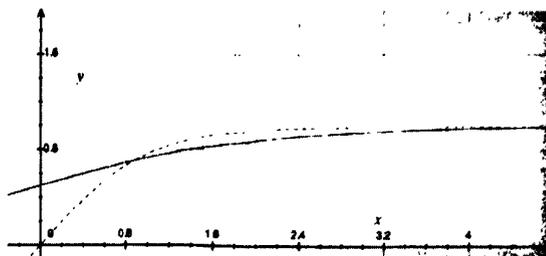


图 6-12. 图中虚线为 $\tanh(x)$ 函数的图形, 实线为 $1/(1+e^{-x})$ 函数的图形。

利用 KLieDA 算法对 USPS 数据集进行全 10 类进行分类。分类器构造采用 One-to-All 方案, 每个子分类器为 KLieDA 二分类器, 因此一共构造 9 个这样的分类器。每次随机抽取 0~9 中每类 60 个手写体构造 8 维协方差特征, 每类 30 个用于训练, 30 个用于测试, 一共 300 个训练样本, 300 个测试样本。重复执行算法 10 次, 得实验结果如图 6-13。得到的结果显示各李群核函数对 USPS 手写体的分类精度比较接近, 李群多项式核精度最高, 达到 90%, 李群 RBF 核虽然在 1 和 9 分类中应用不是太好, 但在 10 分类中表现出仅次于多项式核的精度, 达到 89.3%。

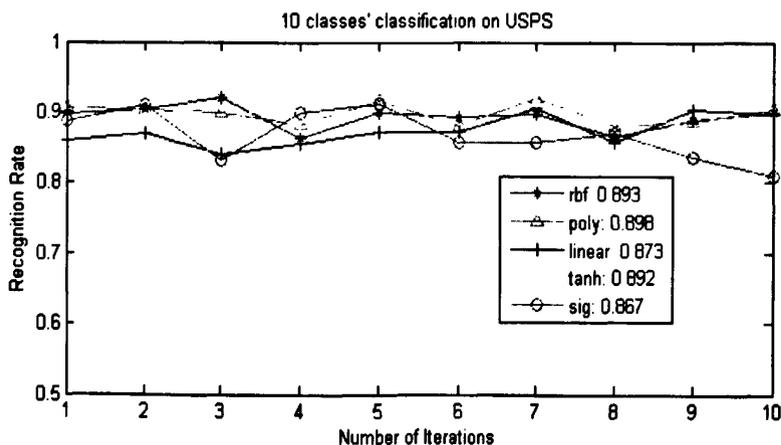


图 6-13. USPS 手写体数据集上应用 KLieDA 进行 10 分类的结果。

另外, 对 $k=5$, 多项式核参数 d 取 2~17 不等的值进行实验, 得到如图 6-13 的结果。

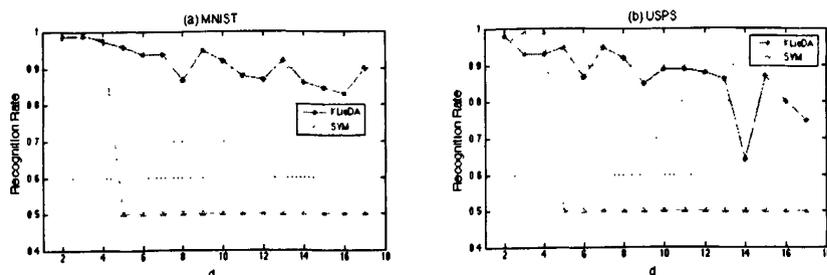


图 6-14. 对矩阵李群多项式核 (5.25) 式取 $d=2$ 至 17 时, KLieDA 和 SVM 算法的结果, 样本配置同上文。(a) 为在 MNIST 上的结果, (b) 为 USPS 数据集上得到的结果。

从图 6-14 中看出, 随着 d 取值的不断变大, KLieDA 的分类识别率总体呈下降趋势, 而 SVM 算法当 $d=5$ 时, 识别率急剧下降到 50%, 并在 $d \geq 5$ 时仍然处于该最差状态。

因此有理由相信, 李群线性核函数适用于分类手写体数字图像的核函数, 手写体数字图像协方差特征拥有非常好的线性可分特性。

6.3 本章小结

本章首先设计了一种手写体图像的李群特征提取方法, 该方法提取的李群特征保留了较好的几何形状信息。通过在人工数据集上添加不同程度高斯噪声的实验表明: Lie-Fisher 分类算法性能优于经典的 KNN 和 FLDA 分类器, 且比较稳定、抗噪能力强。对于手写体字符的识别实验表明, 在对随机采样的点云不进行排序的情况下, Lie-Mean 算法和两种 Lie-Fisher 算法明显优于 FLDA 算法。在排序后, Lie-Fisher 算法表现出较高的识别率, 且对样本维数收敛较快, 性能优于 FLDA。

第二部分实验是基于李群核函数的 KLieDA 算法和 SVM 算法对 MNIST 和 USPS 数据集进行分类。实验结果表明 KLieDA 算法分类精度较高, 算法性能较 SVM 强, 同时验证了李群核函数的有效性。另外, 通过实验分析了图像的协方差特征样本的分布特性, 虽然它们是分布于流形空间中的样本, 但仍然具有非常好的线性可分的特点。

第七章 结论与展望

7.1 结论

李群结是合了代数学和微分几何的数学工具，在物理学和数学两大理论科学中被广泛应用。但在机器学习中的应用却没有统计学来得多，虽然上世纪 90 年代就有李群在坐标计算和智能控制相关领域的应用，却没有成为一个体系，其应用也没有在学习问题上展开。由李凡长教授带领的研究小组首先提出了李群机器学习 (LML) 的概念，并出版了文献[40]，就李群在机器学习研究中的一些基本问题给出了答案与尝试。李群学习模型对机器学习中的众多问题都有很好的研究与应用价值，近些年来，李群机器学习已逐渐成为国内外诸多领域的研究和应用热点，并日益成为机器学习中一个重要的研究分支。

通过对视觉不变性的研究，可以对目标的识别、分类、标注、分割、重建、跟踪等众多问题进行处理。几乎机器学习中所有以图形图像以及视频为样本载体的研究问题都可以归纳到视觉不变问题中。利用李群特有的数学性质在视觉不变问题研究中具有非常大的意义，这不仅是一个有别于以往的一种新的机器学习方法，而且该方法已经得到了许多有效的证明，同时，目前对李群这个高深的数学工具的理解与应用还有很大的不足，仍具有较大的研究空间。

本文首先从李群李代数的几何结构出发，尝试在李群流形上寻找一条测地线，同时将李群样本向该测地线投影，并尽可能使投影后各类别间的散度与类内散度比值最大化，从而实现非线性李群空间的类别判别。然后通过研究矩阵论和核方法，结合李群测地线理论，设计实现了基于矩阵李群的核函数，将李群机器学习扩展到基于核方法的分类与降维算法中。本文提出的基于李群均值的学习算法 Lie-Fisher 以及 KLieDA 算法的分类效果相对传统方法具有较高的分类精度，其中 Lie-Fisher 算法解决了李群流形空间中测地线分类的问题，KLieDA 解决了非线性矩阵李群样本的高维空间映射与分类问题。

7.2 展望

李群机器学习是很大的一个理论框架，在这个框架之下有很多具体的学习子问题有待进一步研究与探索，例如分类、特征提取、降维、跟踪等等。虽然本文提出的算法解决了两个问题子问题，但仍有许多问题有待继续研究与挖掘，比如在李群大数据集下如何使算法运算量减少，降低运算复杂度问题，如何利用李群生成元思想对实际问题进行建模等。在未来，本文将进一步结合这些问题，在此基础上来设计相应的李群均值计算方法将会更加准确与高效，并考虑从几何拓扑的角度提取图像序列间的李群特征，从图像的几何不变性、曲率尺度空间不变性来构造李群，使算法有进一步的应用价值。

参考文献

- [1] Kondor R. Group Theoretical Methods in Machine Learning [D]. 2008.
- [2] Fletcher T, Lu C, Joshi S. Statistics of Shape via Principal Geodesic Analysis on Lie Groups [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2003, 1: 95-101.
- [3] Fletcher P T, Lu C, Pizer S M, Joshi S. Principal geodesic analysis for the study of nonlinear statistics of shape [C]. IEEE Transactions on Medical Imaging, 2004, 23(8): 995-1005.
- [4] Bayro C E, Ortegon A J. Template Tracking with Lie Algebras [C]. IEEE International Conference on Robotics and Automation, 2004, 5: 5183-5188.
- [5] Bayro C E, Ortegon A J. Lie Algebra Template Tracking [C]. International Conference on Pattern Recognition, 2004, 2: 56-59.
- [6] Fiori S. Unsupervised Neural Learning on Lie Group [J]. International Journal of Neural Systems, 2002, 12(3-4): 219-246.
- [7] Tournier M, Wu X, Courty N, Arnaud E, Reveret L. Motion Compression using Principal Geodesics Analysis [C]. ACM Conference on Computer and Communications Security, 2008, 28(2): 355-364.
- [8] Li G, Liu Y, Yin J, Shi Zelin. Projective Tracking Based on Second-order Optimization on Lie Manifolds [C]. Chinese Conference on Pattern Recognition, 2008: 1-6.
- [9] Fletcher P T, Joshi S, Lu C, Pizer S. Gaussian Distributions on Lie Groups and Their Application to Statistical Shape Analysis [C]. International Conference on Information Processing and Medical Imaging, LNCS 2732, 2003: 450-462.
- [10] Yarlagadda P, Ozcanli O, Mundy J. Lie Group distance based generic 3-d vehicle Classification [C]. International Conference on Pattern Recognition, 2008: 1-4.
- [11] Li G, Liu Y, Shi Z L, Yin J. Optimization on Lie Manifolds and Projective Tracking [C]. CSSE, 2008, 1: 768-771.
- [12] Maher M. Means and Averaging in the Group of Rotations [J]. SIAM Journal on Matrix Analysis and Applications, 2002, 24(1): 1-16.

- [13] Fiori S, Tanaka T. An Algorithm to Compute Averages on Matrix Lie Groups [J]. *IEEE Transactions on Signal Processing*, 2009, 57(12): 4734 - 4743.
- [14] Fiori S, Tanaka T. Learning Averages over the Lie Group of Symmetric Positive-Definite Matrices [C]. *International Joint Conference on Neural Networks*, 2009: 828-832.
- [15] Lin D H, Grimson E, Fisher J. Learning visual flows: A Lie algebraic approach [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2009: 747-754.
- [16] Porikli F, Tuzel O, Meer P. Covariance Tracking using Model Update Based on Lie Algebra [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2006: 728-735.
- [17] Chen D Q, Han J Q, Yu Z J. Visual Tracking Using Region Covariance And Geometric Particle Filtering [C]. *International Congress on Image and Signal Processing (CISP)*, 2010, 1: 381-386.
- [18] Doretto G, Yao Y. Region Moments: Fast invariant descriptors for detecting small image structures [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2010: 3019-3026.
- [19] Tuzel O, Porikli F, Meer P. Learning on Lie Groups for Invariant Detection and Tracking [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008: 1-8.
- [20] Tuzel O, Porikli F, Meer P. Region Covariance: A Fast Descriptor for Detection and Classification [C]. *European Conference on Computer Vision*, 2006: 589-600.
- [21] Ando R, Ohki H, Fujita Y. A Comparison with Covariance with Covariance Features on Player Uniform Number Recognition [C]. *Korea-Japan Joint Workshop on Frontiers of Computer Vision*, 2011: 1-4.
- [22] Pennec X, Fillard P, Ayache N. A Riemannian Framework for Tensor Computing [J]. *International Journal of Computer Vision*, 2006, 66(1): 41-66.
- [23] Tuzel O, Porikli F, Meer P. Human Detection via Classification on Riemannian Manifolds [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007: 1-8.
- [24] Wang C M, Jascha S D, Ivana T, Bruno A O. Lie Group Transformation Models for Predictive Video Coding [C]. *Data Compression Conference*, 2011: 83-92.

- [25] Arsigny V, Fillard P, Pennec X, Ayache N. Geometric means in a novel vector space structure on symmetric positive-definite matrices [J]. *SIAM Journal on Matrix Analysis and Applications*, 2007, 29(1): 328-347.
- [26] Xu M, Rajesh P N R. Learning the Lie Groups of Visual Invariance [J]. *Neural Computation*, 2007, 19(10): 2665-2693.
- [27] Plebe A. Learning Visual Invariance [C]. *The European Symposium on Artificial Neural Networks*, 2006: 73-76.
- [28] Cadieu C F, Olshausen B A. Learning Transformational invariants from Natural Movies [C]. *Neural Information Processing Systems*, 2008: 209-216.
- [29] Wang Z, Chen S C, Liu J, Zhang D Q. Pattern Representation in Feature Extraction and Classifier Design: Matrix Versus Vector [C]. *IEEE Transactions on Neural Networks*, 2008, 19(5): 758-769.
- [30] Chen S C, Zhu Y L, Zhang D Q, Yang J Y. Feature extraction approaches based on matrix pattern: MatPCA and MatFLDA [J]. *Pattern Recognition Letters*, 2005, 26(8): 1157-1167.
- [31] Mika S, Rätsch G, Weston J, Schölkopf B, Müller K R. Fisher discriminant analysis with kernels [C]. *Neural Networks for Signal Processing IX*, 1999: 41-48.
- [32] Brun A, Westin C F, Herberthson M, Knutsson H. Intrinsic and Extrinsic Means on the Circle - A Maximum Likelihood Interpretation [C]. *International Conference on Acoustics, Speech, and Signal Processing*, 2007, 3:1053-1056.
- [33] Drummond T, Cipolla R. Real-Time Visual Tracking of Complex Structures [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(7): 932-946.
- [34] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, 1: 886-893.
- [35] Wang Q J, Zhang R B. LPP-HOG: A New Local Image Descriptor for Fast Human Detection [C]. *KAMW*, 2008: 640-643.
- [36] Dalal N. Finding People in Images and Videos [D]. 2006.
- [37] Stringer S M, Perry G, Rolls E T, Proske J H. Learning invariant object recognition in the visual system with continuous transformations [J]. *Biological Cybernetics*, 2006, 94(2): 128-142.

- [38] Charles F C, Bruno A O. Learning Transformational Invariants from Time-Varying Natural Images [C]. Neural Information Processing Systems, 2008.
- [39] Hu M K. Visual pattern recognition by moment invariants [C]. IRE Transactions on Information Theory, 1962, 8:179-187.
- [40] 李凡长, 钱旭培等. 机器学习理论及应用 [M]. 中国科学技术大学出版社, 2009.
- [41] Richard O D, Peter E H, David G S. Pattern Classification, Second edition [M]. John Wiley & Sons, 2000.
- [42] Mundy J L, Zisserman A. Geometric Invariance in Computer Vision [M]. The MIT Press, 1992.
- [43] Hall B C. Lie Groups, Lie Algebras, and Representations: An Elementary Introduction [M]. Springer, 2000.
- [44] Baker A. Matrix Groups An Introduction to Lie Group Theory [M]. Springer, 2002.
- [45] Strang G. Linear Algebra and its Applications, Second edition [M]. Academic Press, 1980.
- [46] Kaaniche M B, Bremond F. Tracking HoG Descriptors for Gesture Recognition [C]. IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009: 140-145.
- [47] Buss S R, Fillmore J P. Spherical averages and applications to spherical splines and interpolation [J]. ACM Transactions on Graphics, 2001, 20: 95-126.
- [48] Lowe D G, Object Recognition from Local Scale-Invariant Features [C]. International Conference on Computer Vision, 1999, 2: 1150-1157.
- [49] Paisitkriangkrai S, Shen C H, Zhang J. Fast Pedestrian Detection Using a Cascade of Boosted Covariance Features [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(8):1140-1151.
- [50] Ge Y H, Yu J J. A Scene Recognition Algorithm Based on Covariance Descriptor [C]. International Conference on Computer and Information Sciences, 2008:838-842.
- [51] Lee J M. Riemannian Manifolds: An Introduction to Curvature [M]. Springer-Verlag, 1997.
- [52] Mardia K V. Directional Statistics [M]. John Wiley & Sons, 1999.
- [53] Förstner W, Moonen B. A metric for covariance matrices [J]. Ratio, 1999, 66: 113-128.

- [54] Aizerman M, Braverman E, Rozonoer L. Theoretical foundations of the potential function method in pattern recognition learning [J]. Automation and Remote Control, 1964, 25: 821-837.
- [55] Vapnik V. The nature of statistical learning theory [M]. New York Springer Verlag, 1995.
- [56] Schölkopf B, Burges C J C, Smola A J. Advances in Kernel Methods: Support Vector Learning [J]. Neuropsychologia, 1999, 49(7): 185-208.
- [57] Scholkopf B, Smola A, Muller K R. Nonlinear Component Analysis as a Kernel Eigenvalue Problem [J]. Neural Computation, 1998, 10(5): 1299-1319.
- [58] Cheng J, Liu Q, Lu H Q. Texture classification using kernel independent component analysis [C]. International Conference on Pattern Recognition, 2004, 1: 620-623.
- [59] Saitoh S. Theory of Reproducing Kernels and its Applications [M]. Longman Scientific & Technical, Harlow, England, 1988.
- [60] Friedman J. Regularized discriminant analysis [J]. Journal of the American Statistical Association, 1989, 84(405):165-175.
- [61] Xu H, Li F Z. Lie group machine Learning's axioms hypothesizes [C]. Proceedings of the 2006 IEEE International Conference on Granular Computing, 2006: 401-404
- [62] Li F Z. Research on the Lie group machine learning [C]. School of Computer Science and Technology, Suzhou University, Technical Report, 2004.
- [63] 许欢. 李群机器学习模型及其应用研究 [D]. 苏州大学, 2007.
- [64] Xu H, Li F Z. Study on Lie group machine learning [J]. Journal of Computational Information Systems, 2005, 1(4): 843-849.
- [65] 李凡长,康宇. 基于 Lie 群的机器学习理论框架 [J]. 云南民族大学学报, 2004, 13(4) : 251-255.
- [66] Li F Z, Xu H. The theory framework of Lie group machine learning (LML) [J]. Computer Technology and Application, 2007, 1(3): 62-80.
- [67] Xu H, Li F Z. Geometry algorithms of dynkin diagrams in Lie group machine learning [J]. Journal of Nanchang Institute of Technology, 2006, 25(2): 75-78.
- [68] Xu H, Li F Z. Algorithms of dynkin diagrams in Lie group machine learning [J]. Journal of Communication and Computer, 2006, 4(3): 13-17.

- [69] Xu H, Li F Z. The design of SU(n) classifier of Lie group machine learning (LML) [J]. Journal of Computational Information Systems, 2005, 1(4): 835-841.
- [70] Li F Z, Xu H. SO(3) Classifier of Lie group machine learning [J]. Journal of Communication and Computer, 2006, 4(3): 76-80.
- [71] 何书萍. 李群机器学习中的量子群分类器研究 [D]. 苏州大学, 2008.
- [72] 付会欣. 李群机器学习中的辛群分类器研究 [D]. 苏州大学, 2008.
- [73] Grenander U. Probabilities on Algebraic Structures [M]. John Wiley & Sons, 1963.
- [74] Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2004, 2: 506-513.
- [75] Abdel-Hakim A E, Farag A A. CSIFT: A SIFT Descriptor with Color Invariant Characteristics [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2006, 2: 1978-1983.
- [76] Bay H, Ess A, Tuytelaars T, Gool L V. SURF: Speeded Up Robust Features [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [77] Morel J M, Yu G. ASIFT: A New Framework for Fully Affine Invariant Image Comparison [J]. SIAM Journal on Imaging Sciences, 2009, 2(2): 438-469.
- [78] Mitchell T. Machine Learning [M], McGraw Hill, 1997.
- [79] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning [M]. Springer Series in Statistics, 2001.
- [80] Savich A W, Moussa M. Resource Efficient Arithmetic Effects on RBM Neural Network Solution Quality Using MNIST [C]. International Conference on Reconfigurable Computing and FPGAs, 2011: 35-40.

攻读学位期间参与的科研项目与发表(录用)论文情况

科研项目

1. 国家自然科学基金项目(60775045): 面向数据约简的机器学习新方法研究;
2. 东吴学者计划(14317360): 李群机器学习研究;
3. 国家自然科学基金项目(61033013): 基于认知模型的图像不变性特征理论和关键技术。

发表(录用)的论文

1. 高聪, 李凡长. 李群均值学习算法研究. 模式识别与人工智能(已录用), 2011. 同时, 本文还被 CRSSC-CWI-CGrC2011 联合会议评为优秀学生论文。
2. 沈程, 李凡长, 高聪, 何书萍. 李群彩色图像人脸跟踪. 计算机科学与探索(已录用), 2012.
3. 高聪, 李凡长. 李群核学习算法研究. 已投往 CRSSC-CWI-CGrC2012 联合会议。

致 谢

三年的研究生学习即将结束。回首这三年，我不仅在科研能力和专业技能上有了较大的提高，使我从一个对机器学习不甚了解的程度跨入了这个广阔而又有前景的领域。而且，在整个过程中学到了很多为人处世的道理。在这三年里，我得到了我的老师、同学、朋友和亲人的众多支持和帮助。在此，我衷心地向他们表达诚挚的谢意。

尊敬的李凡长教授、博导是我的导师，李教授知识渊博、治学严谨，使我深受感染；在毕业课题与论文写作期间，李教授更以其深厚的理论功底和敏锐的学术眼光给我以不断的启发和指导，使我能够顺利完成论文的写作！他平易近人、举重若轻，能够在谈笑间将高深的专业理论与深刻的处世哲学教给我们，教会了我如何紧扣国际学术前沿、如何进行科研工作、如何提高研究素质等，这些将是我一生的宝贵财富！在此，我要向李教授致以最衷心的感谢。

我要感谢实验室的王颖、陈明同学，特别是我的师兄范自立，他们诸多的帮助和大量宝贵的意见，使我们不仅在学习上共同进步，更分享了生活中的乐趣。

我要感谢，我的室友小军军同学和 188 高个的帅哥管延伟！前者还和我一起度过了三年时光，你们给我留下了许多美好的回忆。

我要感谢含辛茹苦养育我的父母以及一直关心我和理解我的亲朋好友，没有你们支持和理解，是不可能完成这三年的科研任务的。是你们给了我无限的温暖，给了我前进的力量，让我一直快乐地成长着，我为生活在这样一个和睦幸福的大家庭中感到骄傲。

我要感谢这三年来认识的每一位朋友，我们在一起学习生活了一千多个日日夜夜，分享快乐、相互鼓励、共同进步，建立了深厚的友谊，它将是我一生最宝贵的财富！

祝愿我的导师和我的亲朋好友们在今后的工作、生活中一切顺利。

李群均值学习算法及应用研究

作者：[高聪](#)
学位授予单位：[苏州大学](#)

本文链接：http://d.wanfangdata.com.cn/Thesis_Y2120815.aspx