



信息与计算科学丛书 — 59

支持向量机的算法设计与分析

杨晓伟 郝志峰 著



科学出版社

“十二五”国家重点图书出版规划项目



信息与计算科学丛书 59

支持向量机的算法 设计与分析



杨晓伟 郝志峰 著

科学出版社

北京

内 容 简 介

支持向量机的研究是近十余年机器学习、模式识别和数据挖掘领域中的研究热点,受到了计算数学、统计、计算机、自动化和电信等有关学科研究者的广泛关注,取得了丰硕的理论成果,并被广泛地应用于文本分类、图像处理、语音识别、时间序列预测和函数估计等领域.本书首先介绍了核函数的概念;然后从几何直观的角度介绍了建立二分类模型和回归模型过程中所取得的理论成果;最后对于分解算法、最小二乘支持向量机、多分类、模糊支持向量机、在线学习和大规模分类相关的优秀成果进行了归纳和整理,从数学上对相关算法的原理进行了详细分析.本书的内容既包括支持向量机的最新进展,也包括作者的多年研究成果.作者希望本书能够有助于对机器学习、模式识别和数据挖掘感兴趣的读者更加快速地了解支持向量机的最新研究动态,能够有助于读者理清算法的本质,从而使读者能够在已有研究成果的基础之上更加有效地开展工作.

本书可作为数学、统计、计算机、电信、自动化等有关专业的高年级本科生和研究生教材,也可作为相关领域的教师和科研工作者的参考书.

图书在版编目(CIP)数据

支持向量机的算法设计与分析/杨晓伟,郝志峰著. —北京:科学出版社, 2013

(信息与计算科学丛书; 59)

ISBN 978-7-03-037869-9

I. ①支… II. ①杨… ②郝… III. ①向量计算机-算法设计 ②向量计算机-算法分析 IV. ①TP301.6

中国版本图书馆 CIP 数据核字(2013) 第 128845 号

责任编辑:李 欣/责任校对:胡小洁

责任印制:钱玉芬/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>

北京佳信达欣艺术印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2013 年 6 月第 一 版 开本: B5 (720 × 1000)

2013 年 6 月第一次印刷 印张: 13 3/4

字数: 260 000

定价: **68.00 元**

(如有印装质量问题,我社负责调换)

《信息与计算科学丛书》序

20 世纪 70 年代末, 由已故著名数学家冯康先生任主编, 科学出版社出版了一套《计算方法丛书》, 至今已逾 30 册. 这套丛书以介绍计算数学的前沿方向和科研成果为主旨, 学术水平高、社会影响大, 对计算数学的发展、学术交流及人才培养起到了重要的作用.

1998 年教育部进行学科调整, 将计算数学及其应用软件、信息科学、运筹控制等专业合并, 定名为“信息与计算科学专业”. 为适应新形势下学科发展的需要, 科学出版社将《计算方法丛书》更名为《信息与计算科学丛书》, 组建了新的编委会, 并于 2004 年 9 月在北京召开了第一次会议, 讨论并确定了丛书的宗旨、定位及方向等问题.

新的《信息与计算科学丛书》的宗旨是面向高等学校信息与计算科学专业的高年级学生、研究生以及从事这一行业的科技工作者, 针对当前的学科前沿, 介绍国内外优秀的科研成果. 强调科学性、系统性及学科交叉性, 体现新的研究方向. 内容力求深入浅出, 简明扼要.

原《计算方法丛书》的编委和编辑人员以及多位数学家曾为丛书的出版做了大量工作, 在学术界赢得了很好的声誉, 在此表示衷心的感谢. 我们诚挚地希望大家一如既往地关心和支持新丛书的出版, 以期为信息与计算科学在新世纪的发展起到积极的推动作用.

石钟慈

2005 年 7 月

前 言

20 世纪 90 年代中期, 基于统计学习理论, Vapnik 提出了支持向量机 (support vector machine, SVM) 模型. 其在手写体邮政编码识别中的成功应用引起了模式识别、数据挖掘、机器学习、数学、统计等相关领域国内外研究人员的广泛关注. 近十余年来, 研究者在支持向量机的理论研究和算法实现方面都取得了突破性的进展, 涌现出了一批优秀的科研成果. 目前, 支持向量机被广泛地应用于基于内容的视频检索、网页分类、文本分类、光学字符识别、信号处理和生物信息处理等领域.

自 2002 年 10 月以来, 作者一直从事支持向量机方面的研究工作, 在分解算法、最小二乘支持向量机稀疏化、多分类、带噪声的模式识别和大规模分类等方面开展了一些研究工作, 取得了一些有价值的研究成果. 自 2007 年 9 月以来, 作者开始在数学系计算数学专业和概率论与数理统计专业的研究生课程《机器学习》中讲述支持向量机的内容, 并形成了电子讲义. 经过六年的努力, 完成和完善了本书的内容. 本书的内容一定程度上是作者关于支持向量机研究的一种体会和总结, 希望本书的出版对机器学习、模式识别、数据挖掘、数学、统计等相关领域的研究者有所帮助.

本书各章的主要内容如下: 第 1 章介绍了核函数、二分类模型和回归模型. 第 2 章讨论了解支持向量机模型的流行方法——分解算法. 第 3 章讨论了最小二乘支持向量机的模型、求解算法和稀疏化问题. 第 4 章讨论了多分类算法. 第 5 章讨论了模糊支持向量机的模型和隶属度设置问题. 第 6 章讨论了支持向量机的在线学习算法. 第 7 章讨论了大规模线性分类算法和大规模非线性分类算法, 特别讨论了最新的基于局部学习的大规模非线性分类算法. 本书包括了作者多年的研究成果. 例如, 对于分解算法, 提出了扩展的 Lagrange 支持向量机; 对于最小二乘支持向量机, 基于增量学习和减量学习, 提出了自下而上的稀疏化算法; 对于多分类, 提出了基于一对多分割的二叉树支持向量机算法和基于一对一策略的嵌套算法; 对于模糊支持向量机, 一方面提出了基于核模糊 c -均值聚类和最远对策略的模糊支持向量机解决带噪声的分类问题, 另一方面也提出了基于孤立点删除的加权最小二乘支持向量机. 对于大规模分类问题, 提出了基于带标注聚类特征树和局部学习的支持向量机算法.

本书的出版得到了国家社科基金重大项目 (11&ZD156)、国家自然科学基金 (61273295, 61070033)、国家出版基金、广东省自然科学基金重点项目 (9251009001000005)、华南理工大学研究生院研究生重点课程建设项目 (yjzk2010007)

等基金的资助. 本书的出版得到了西安交通大学数学与统计学院徐宗本院士、大连理工大学数学科学学院吴微教授和吉林大学计算机科学与技术学院梁艳春教授的大力支持, 在此, 对他们表示感谢. 关于模糊支持向量机的内容, 作者在澳大利亚悉尼科技大学访问时与张广全教授和路节教授进行了大量的讨论, 得到了很大的启发. 本书的大部分内容在华南理工大学理学院数学系研究生《机器学习》课上使用过五年. 在讲课的过程中, 作者的研究生温雯、刘波、蔡瑞初、何丽芳、付春岚、李艳、吕浩然、肖法镇、谭亮军、余巧珍、郭腾蛟等提出了许多宝贵的意见, 在此对他们表示感谢. 另外, 吴广潮副教授帮助校对了本书第 7.7 节的内容. 借此机会, 也特别感谢廖芹教授, 她一直在鼓励出版这本研究生教材.

由于作者水平所限, 书中难免有不妥之处, 欢迎读者批评指正.

杨晓伟 郝志峰

华南理工大学

2013 年 3 月 8 日

目 录

《信息与计算科学丛书》序

前言

第 1 章 支持向量机的分类和回归模型	1
1.1 多项式核函数	1
1.2 Mercer 核	3
1.3 再生核 Hilbert 空间	7
1.4 正定核函数的构造	11
1.5 二分类问题的数学提法	13
1.6 平分最近点模型	14
1.7 最大间隔模型	17
1.8 平分最近点模型和最大间隔模型之间的关系	19
1.9 回归问题的数学提法	23
1.10 硬 ε -带超平面	24
1.11 基于分类的回归模型	25
参考文献	34
第 2 章 分解算法	36
2.1 无约束问题的提法	36
2.2 分解算法的提出	38
2.3 选块算法	39
2.4 SVM ^{Light} 算法	41
2.5 Platt 的 SMO 算法	43
2.6 Keerthi 的 SMO 改进算法	48
2.7 改进的 SMO 算法的收敛性	49
2.8 解回归问题的 SMO 算法	51
2.9 扩展的 Lagrange 支持向量机	56
参考文献	60
第 3 章 最小二乘支持向量机	62
3.1 最小二乘支持向量机	62
3.2 最小二乘隐空间支持向量机	65

3.3 基于矩阵模式的最小二乘支持向量机	66
3.4 最小二乘支持向量机的求解算法	68
3.5 最小二乘支持向量机的稀疏化算法	74
参考文献	87
第 4 章 多分类问题	89
4.1 一对多算法	89
4.2 一对一算法	91
4.3 基于决策树的支持向量机	92
4.4 嵌套算法	98
4.5 纠错输出编码支持向量机	102
4.6 一次求解算法	104
4.7 支持向量机分类-回归算法	107
参考文献	109
第 5 章 模糊支持向量机	112
5.1 单边加权模糊支持向量机	112
5.2 双边加权模糊支持向量机	116
5.3 基于加权间隔的模糊支持向量机	118
5.4 模糊支持向量机中的隶属度设置	120
5.5 加权稳健支持向量回归方法	123
5.6 基于 ε -不敏感学习的模糊系统	124
5.7 基于模糊 if-then 规则的 ε -间隔非线性分类器	127
5.8 基于核模糊 c -均值聚类和最远对策略的模糊支持向量机分类器	129
参考文献	134
第 6 章 支持向量机的在线学习算法	137
6.1 基于增量和减量学习的支持向量机算法	137
6.2 增量支持向量机分类算法	142
6.3 增量支持向量机回归算法	143
6.4 核递归最小二乘算法	147
6.5 基于结构风险最小化的在线核方法	150
6.6 快速的在线核分类器	153
参考文献	157
第 7 章 大规模分类	158
7.1 大规模线性支持向量机算法	158
7.2 基于低秩核矩阵表示的支持向量机算法	164
7.3 缩减支持向量机	172

7.4 核向量机	173
7.5 多核学习机	178
7.6 局部化支持向量机	186
7.7 基于带类标聚类特征树和局部学习的支持向量机分类算法	188
参考文献	200
索引	204
《信息与计算科学丛书》已出版书目	205

第1章 支持向量机的分类和回归模型

分类和回归问题是模式识别的两个基本问题,在实际中有着非常广泛的应用.在支持向量机的研究中,研究者对分类和回归问题开展了广泛的研究,取得了大量的研究成果.核函数在数学上表示为高维特征空间中的向量内积,巧妙地避免了机器学习和模式识别中的维数灾难问题,在处理非线性问题时起到了非常重要的作用.为了清楚地理解支持向量机的原理,在这一章中,首先介绍核函数的基本知识,然后从几何直观的角度出发,给出二分类问题的数学模型,最后从二分类的角度给出回归问题的数学模型.至于模型的求解问题,放在以后的章节中讨论.

1.1 多项式核函数

定义 1.1(核和正定核) 设 X 是 R^n 中的一个子集,称定义在 $X \times X$ 上的函数(对称函数) $K(x, z)$ 是核函数(正定核函数),如果存在一个从 X 到内积空间(Hilbert 空间) H 的映射 Φ

$$\Phi: x \mapsto \Phi(x) \in H \quad (1-1)$$

使得对任意的 $x, z \in X$,

$$K(x, z) = (\Phi(x) \cdot \Phi(z)) \quad (1-2)$$

都成立.其中 $(\cdot)$ 表示内积空间(Hilbert 空间) H 中的内积.

定义 1.2(d 阶多项式) 设 $x = ([x]_1, [x]_2, \dots, [x]_n)^T \in R^n$, 则称乘积 $[x]_{j_1} [x]_{j_2} \dots [x]_{j_d}$ 为 x 的一个 d 阶多项式,其中 $j_1, j_2, \dots, j_d \in \{1, 2, \dots, n\}$.

考虑二维空间中($x \in R^2$)的模式 $x = ([x]_1, [x]_2)^T$,如果把 $[x]_1[x]_2$ 和 $[x]_2[x]_1$ 看成两个不同的单项式,则其所有的 2 阶单项式为

$$[x]_1^2, [x]_2^2, [x]_1[x]_2, [x]_2[x]_1 \quad (1-3)$$

这 4 个单项式张成的是一个四维特征空间,称为 2 阶有序齐次多项式空间,记为 H .相应地可建立从原空间 R^2 到多项式空间 H 的非线性映射

$$C_2: x = ([x]_1, [x]_2)^T \mapsto C_2(x) = ([x]_1^2, [x]_2^2, [x]_1[x]_2, [x]_2[x]_1)^T \in H \quad (1-4)$$

同理,从 R^n 到 d 阶有序齐次多项式空间 H 的映射可表示为

$$C_d: x = ([x]_1, [x]_2, \dots, [x]_n)^T \mapsto C_d(x)$$

$$= ([x]_{j_1} [x]_{j_2} \cdots [x]_{j_d} | j_1, j_2, \cdots, j_d \in \{1, 2, \cdots, n\})^T \in H \quad (1-5)$$

这样的有序单项式 $[x]_{j_1} [x]_{j_2} \cdots [x]_{j_d}$ 的个数为 n^d , 即多项式空间 H 的维数 $n_H = n^d$. 如果在 H 中进行内积运算 $C_d(\mathbf{x}) \cdot C_d(\mathbf{z})$, 当 n 和 d 都不太小时, 多项式空间 H 的维数 n_H 会相当大. 譬如, 当 $n = 200$, $d = 5$ 时, 维数可达到上亿维. 显然, 在多项式空间 H 中直接进行内积运算将会引起“维数灾难”问题, 那么, 如何处理这个问题呢?

先考查 $n = d = 2$ 的情况, 计算多项式空间 H 中两个向量的内积

$$(C_2(\mathbf{x}) \cdot C_2(\mathbf{z})) = [x]_1^2 [z]_1^2 + [x]_2^2 [z]_2^2 + [x]_1 [x]_2 [z]_1 [z]_2 + [x]_2 [x]_1 [z]_2 [z]_1 = (\mathbf{x} \cdot \mathbf{z})^2 \quad (1-6)$$

若定义函数

$$K(\mathbf{x}, \mathbf{z}) = (\mathbf{x} \cdot \mathbf{z})^2 \quad (1-7)$$

则有

$$(C_2(\mathbf{x}) \cdot C_2(\mathbf{z})) = K(\mathbf{x}, \mathbf{z}) \quad (1-8)$$

即四维多项式空间 H 上的向量内积可以转化为原始二维空间上的向量内积的平方. 对于一般的从 R^n 到 d 阶有序多项式空间 H 的映射 (1-5) 也有类似的结论.

定理 1.1^[4] 考虑由式 (1-5) 定义的从 R^n 到多项式空间 H 的映射 $C_d(\mathbf{x})$, 则在空间 H 上的内积 $(C_d(\mathbf{x}) \cdot C_d(\mathbf{z}))$ 可表示为

$$(C_d(\mathbf{x}) \cdot C_d(\mathbf{z})) = K(\mathbf{x}, \mathbf{z}) \quad (1-9)$$

其中

$$K(\mathbf{x}, \mathbf{z}) = (\mathbf{x} \cdot \mathbf{z})^d \quad (1-10)$$

上述定理表明, 我们并不需要在高维的多项式空间 H 中直接做内积运算 $(C_d(\mathbf{x}) \cdot C_d(\mathbf{z}))$, 而利用式 (1-10) 给出的输入空间 R^n 上的二元函数 $K(\mathbf{x}, \mathbf{z})$ 来计算高维多项式空间中的内积.

在式 (1-5) 定义的映射中, 多项式空间 H 的分量由所有的 d 阶有序单项式组成. 如果把该多项式空间的分量扩充为所有不超过 d 阶的有序单项式, 便得到从 R^n 到有序多项式空间的映射 $\tilde{C}_d$

$$\begin{aligned} \tilde{C}_d : \mathbf{x} &= ([x]_1, [x]_2, \cdots, [x]_n)^T \mapsto \tilde{C}_d(\mathbf{x}) \\ &= ([x]_{j_1} \cdots [x]_{j_d}, \sqrt{d} [x]_{j_1} \cdots [x]_{j_{d-1}}, \cdots, \\ &\quad \sqrt{d} [x]_1, \cdots, \sqrt{d} [x]_n, 1 | j_1, j_2, \cdots, j_d \in \{1, 2, \cdots, n\})^T \end{aligned} \quad (1-11)$$

对于这个映射, 有如下定理:

定理 1.2^[4] 考虑由式 (1-11) 定义的从 R^n 到多项式空间 H 的映射 $\tilde{C}_d$, 则空间 H 上的内积 $(\tilde{C}_d(\mathbf{x}) \cdot \tilde{C}_d(\mathbf{z}))$ 可表为空间 R^n 上的内积 $(\mathbf{x} \cdot \mathbf{z})$ 的函数 $((\mathbf{x} \cdot \mathbf{z}) + 1)^d$, 即若定义两个变量 $\mathbf{x}$ 和 $\mathbf{z}$ 的函数

$$K(\mathbf{x}, \mathbf{z}) = ((\mathbf{x} \cdot \mathbf{z}) + 1)^d \quad (1-12)$$

则有

$$(\tilde{C}_d(\mathbf{x}) \cdot \tilde{C}_d(\mathbf{z})) = K(\mathbf{x}, \mathbf{z}) \quad (1-13)$$

根据定义 1.1, 称式 (1-12) 为 d 阶多项式核函数.

如果把式 (1-4) 中的 $[x]_1[x]_2$ 和 $[x]_2[x]_1$ 看做相同的单项式, 那么就可以把从 R^2 到四维多项式空间 H 的映射 (1-4) 简化为从 R^2 到三维多项式空间的映射

$$([x]_1[x]_2)^T \mapsto ([x]_1^2, [x]_2^2, [x]_1[x]_2)^T \quad (1-14)$$

将映射 (1-14) 调整为

$$\Phi_2(\mathbf{x}) = \Phi_2([x]_1, [x]_2) = ([x]_1^2, [x]_2^2, \sqrt{2}[x]_1[x]_2) \quad (1-15)$$

则相应的多项式空间称为 2 阶无序多项式空间, 并且有

$$(\Phi_2(\mathbf{x}) \cdot \Phi_2(\mathbf{z})) = (\mathbf{x} \cdot \mathbf{z})^2 \quad (1-16)$$

比较式 (1-4) 定义的变换 $C_2(\mathbf{x})$ 和式 (1-15) 定义的 $\Phi_2(\mathbf{x})$ 可以发现, 它们所映射到的多项式空间是不同的. 前者是一个四维多项式空间, 后者为一个三维多项式空间. 但是内积是相同的, 它们都可以表示为内积的函数 $K(\mathbf{x}, \mathbf{z}) = (\mathbf{x} \cdot \mathbf{z})^2$. 这说明: 多项式空间不是由核函数唯一确定的.

1.2 Mercer 核

Mercer 定理是支持向量机解决非线性模式识别问题的重要理论基础, 它是 Mercer 在 1909 年提出的^[1]. 1964 年, Aizerman、Braverman 和 Rozonoer 把核函数引入机器学习领域, 并运用 Mercer 定理把核函数解释为 Hilbert 空间中的内积^[2]. 1992 年, Boser、Guyon 和 Vapnik 首次把这种思想引入支持向量机领域^[3], 其工作为支持向量机的蓬勃发展奠定了坚实的理论基础. 在这一节中, 首先介绍半正定矩阵的特征展开, 然后介绍半正定积分算子的特征展开, 最后给出 Mercer 定理和 Mercer 核.

1.2.1 半正定矩阵的特征展开

给定向量集合 $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l\}$, 其中 $\mathbf{x}_i \in R^n, i = 1, 2, \dots, l$. 设 $K(\mathbf{x}, \mathbf{z})$ 是 $X \times X$ 上的对称函数, 定义

$$G_{ij} = K(\mathbf{x}_i, \mathbf{x}_j), \quad i, j = 1, 2, \dots, l \quad (1-17)$$

则称 $\mathbf{G} = (G_{ij})$ 是 $K(\mathbf{x}, \mathbf{z})$ 关于 X 的 Gram 矩阵. 那么, 当 Gram 矩阵 $\mathbf{G}$ 满足什么条件时, 函数 $K(\cdot, \cdot)$ 是一个核函数呢?

定义 1.3(矩阵算子) 定义在 R^l 上的矩阵算子 $\mathbf{G}$: 对 $\mathbf{u} = (u_1, u_2, \dots, u_l)^T \in R^l$, $\mathbf{G}\mathbf{u}$ 的分量由下式确定

$$[\mathbf{G}\mathbf{u}]_i = \sum_{j=1}^l K(\mathbf{x}_i, \mathbf{x}_j)u_j, \quad i = 1, 2, \dots, l \quad (1-18)$$

定义 1.4(特征值和特征向量) 考虑定义 1.2 给出的矩阵算子 $\mathbf{G}$. 称 $\lambda \in R$ 为它的特征值, 并称 $\mathbf{v}$ 为相应的特征向量, 如果

$$\mathbf{G}\mathbf{v} = \lambda\mathbf{v} \quad \text{且} \mathbf{v} \neq \mathbf{0} \quad (1-19)$$

定义 1.5(半正定性) 考虑定义 1.2 给出的矩阵算子 $\mathbf{G}$. 称它是半正定的, 如果 $\forall \mathbf{u} = (u_1, u_2, \dots, u_l)^T \in R^l$, 有

$$\mathbf{u}^T \mathbf{G} \mathbf{u} = \sum_{i,j=1}^l K(\mathbf{x}_i, \mathbf{x}_j)u_i u_j \geq 0 \quad (1-20)$$

引理 1.1^[4] 若定义 1.2 给出的矩阵算子 $\mathbf{G}$ 是半正定的, 则存在 l 个非负特征值 λ_t 和互相正交的单位特征向量 $\mathbf{v}_t$, 使得

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_{t=1}^l \lambda_t v_{ti} v_{tj}, \quad i, j = 1, 2, \dots, l \quad (1-21)$$

引理 1.2 若引理 1.1 的结论成立, 则存在着从 X 到 R^l 的映射 Φ , 使得

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)), \quad i, j = 1, 2, \dots, l \quad (1-22)$$

其中 $(\cdot)$ 是特征空间 R^l 的内积. 因而 $K(\cdot, \cdot)$ 是一个正定核函数.

证明 定义映射

$$\Phi: \mathbf{x}_i \mapsto \Phi(\mathbf{x}_i) = (\sqrt{\lambda_1} v_{1i}, \sqrt{\lambda_2} v_{2i}, \dots, \sqrt{\lambda_l} v_{li})^T \in R^l \quad (1-23)$$

直接验证可知引理 1.2 成立.

引理 1.3 若引理 1.2 的结论成立, 则矩阵 G 是半正定的.

证明 设 G 不是半正定的, 则一定存在与一个负特征值 λ_s 相对应的单位特征向量 v_s . 定义 R^l 中的向量 z

$$z = [\Phi(x_1), \Phi(x_2), \dots, \Phi(x_l)]v_s \quad (1-24)$$

则有

$$0 \leq \|z\|^2 = v_s^T [\Phi(x_1), \dots, \Phi(x_l)]^T [\Phi(x_1), \dots, \Phi(x_l)] v_s = v_s^T G v_s = \lambda_s \|v_s\|^2 = \lambda_s \quad (1-25)$$

显然, 这与 λ_s 是负特征值相矛盾. 因此 G 必须是半正定的.

定理 1.3^[4] 设 X 是有限集合 $X = \{x_1, x_2, \dots, x_l\}$, $K(x, z)$ 是定义在 $X \times X$ 上的对称函数. 则由定义 1.3 给出的矩阵算子 G 半正定, 等价于 $K(\cdot, \cdot)$ 可表示为

$$K(x_i, x_j) = \sum_{t=1}^l \lambda_t v_{ti} v_{tj}, \quad i, j = 1, 2, \dots, l \quad (1-26)$$

其中 $\lambda_t \geq 0$ 是矩阵

$$G = (K(x_i, x_j))_{i,j=1}^l \quad (1-27)$$

的特征值, $v_t = (v_{t1}, v_{t2}, \dots, v_{tl})^T$ 为对应于 λ_t 的特征向量, 也等价于 $K(x, z)$ 是一个正定核函数, 即 $K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j))$, 其中映射 Φ 由式 (1-23) 定义.

1.2.2 半正定积分算子的特征展开

设输入集合为 R^n 中的紧集 X , 并设 $K(x, z)$ 是 $X \times X$ 的连续对称函数. 当 $K(x, z)$ 满足什么条件时, 它是一个核函数呢?

定义 1.6(积分算子 T_K) 定义积分算子 T_K 为按下式确定的在 $L_2(x)$ 上的积分算子

$$T_K f = T_K f(\cdot) = \int_X K(\cdot, z) f(z) dz, \quad \forall f \in L_2(x) \quad (1-28)$$

定义 1.7(特征值和特征函数) 考虑定义 1.6 给出的积分算子 T_K , 称 λ 为它的特征值, φ 为相应的特征函数, 如果

$$T_K \varphi = \lambda \varphi \quad (1-29)$$

定义 1.8(半正定性) 考虑定义 1.6 给出的积分算子 T_K , 称它是半正定的, 如果对 $\forall f \in L_2(x)$, 有

$$\int_{X \times X} K(x, z) f(x) f(z) dx dz \geq 0 \quad (1-30)$$

引理 1.4^[4] 若定义 1.6 给出的积分算子 T_K 是半正定的, 则存在可数个非负特征值 λ_t 和相应的互相正交的单位特征函数 $\varphi_t(x)$, 使得 $K(\cdot, \cdot)$ 可表示为 $X \times X$ 上的一致收敛级数

$$K(x, z) = \sum_{t=1}^{\infty} \lambda_t \varphi_t(x) \varphi_t(z) \quad (1-31)$$

引理 1.5 若引理 1.4 的结论成立, 则存在着 $X \in R^n$ 到 Hilbert 空间 l_2 的映射 Φ , 使得

$$K(x, z) = (\Phi(x) \cdot \Phi(z)), \quad x, z \in X \quad (1-32)$$

其中 $(\cdot)$ 是 l_2 上的内积. 因而 $K(\cdot, \cdot)$ 是一个正定核函数.

证明 定义映射

$$\Phi: x \mapsto \Phi(x) = (\sqrt{\lambda_1} \varphi_1(x), \sqrt{\lambda_2} \varphi_2(x), \dots)^T \quad (1-33)$$

则可验证引理 1.5 成立.

由引理 1.5 和定义 1.8, 很容易得到下面的引理成立.

引理 1.6^[4] 若引理 1.5 的结论成立, 则积分算子 T_K 是半正定的.

定理 1.4(Mercer 定理) 令 X 是 R^n 上的一个紧集, $K(x, z)$ 是 $X \times X$ 上的连续实值对称函数. 则由定义 1.6 给出的积分算子 T_K 半正定

$$\int_{X \times X} K(x, z) f(x) f(z) dx dz \geq 0, \quad \forall f \in L_2(x) \quad (1-34)$$

等价于 $K(\cdot, \cdot)$ 可表示为 $X \times X$ 上的一致收敛级数

$$K(x, z) = \sum_{t=1}^{\infty} \lambda_t \varphi_t(x) \cdot \varphi_t(z) \quad (1-35)$$

其中 $\lambda_t > 0$ 是 T_K 的特征值, $\varphi_t \in L_2(x)$ 是对应 λ_t 的特征函数. 它也等价于 $K(x, z)$ 是一个正定核函数

$$K(x, z) = (\Phi(x) \cdot \Phi(z)) \quad (1-36)$$

其中映射 Φ 由式 (1-33) 定义.

定义 1.9(Mercer 核) 称函数 $K(x, z)$ 为 Mercer 核, 如果 $K(x, z)$ 是定义在 $X \times X$ 上的连续对称函数, 其中 X 是 R^n 的紧集, 且由定义 1.6 给出的积分算子是半正定的.

定理 1.5^[4] 设 X 为 R^n 上的紧集, $K(x, z)$ 是 $X \times X$ 上的连续对称函数, 则积分算子 T_K 半正定的充要条件是 $K(x, z)$ 关于任意的 $x_1, x_2, \dots, x_l \in X$ 的 Gram 矩阵半正定.

1.3 再生核 Hilbert 空间

再生核 Hilbert 空间是 Aronszajn 在 1950 年提出的^[5]. Schölkopf、Burges 和 Smola 在 1999 年给出了 Hilbert 空间的微分几何描述, 并且还给出了核函数满足 Mercer 定理的必要条件^[6]. 2004 年, 邓乃扬和田英杰从构造特殊映射的角度出发给出了核函数是正定核函数的充分条件和再生核 Hilbert 空间的定义^[4]. 在这一节中, 将详细地介绍文献 [4] 的结果.

定理 1.6 设 X 是 R^n 的子集. 若 $K(x, z)$ 是定义在 $X \times X$ 上的正定核, 则 $\forall x_1, x_2, \dots, x_l \in X$, 函数 $K(x, z)$ 关于 $x_1, x_2, \dots, x_l$ 的 Gram 矩阵都是半正定的.

证明 $K(x, z)$ 是定义在 $X \times X$ 上的正定核, 因此存在着从 X 到 Hilbert 空间 H 的映射 Φ , 使得

$$K(x, z) = (\Phi(x) \cdot \Phi(z)) \quad (1-37)$$

任取 $x_1, x_2, \dots, x_l \in X$, 构造 $K(\cdot, \cdot)$ 关于 $x_1, x_2, \dots, x_l$ 的 Gram 矩阵 $(K_{ij})_{i,j=1}^l = (K(x_i, x_j))_{i,j=1}^l$. 显然, 由式 (1-37) 可以断言, $\forall C_1, C_2, \dots, C_l \in R$, 有

$$\begin{aligned} \sum_{i,j} C_i C_j K(x_i, x_j) &= \sum_{i,j} C_i C_j \Phi(x_i) \cdot \Phi(x_j) \\ &= \left(\sum_i C_i \Phi(x_i) \cdot \sum_j C_j \Phi(x_j) \right) = \left\| \sum_i C_i \Phi(x_i) \right\|^2 \geq 0 \end{aligned} \quad (1-38)$$

这表明 $K(x, z)$ 关于 $x_1, x_2, \dots, x_l$ 的 Gram 矩阵是半正定的.

引理 1.7 若集合 S 由所有的下列元素组成

$$f(\cdot) = \sum_{i=1}^l \alpha_i K(\cdot, x_i) \quad (1-39)$$

其中 l 为任意的正整数, $\alpha_1, \alpha_2, \dots, \alpha_l \in R$, $x_1, x_2, \dots, x_l \in X$, 则 S 为一个向量空间.

证明 显然集合 S 中的元素对于加法和数乘运算封闭.

设

$$f(\cdot) = \sum_{i=1}^l \alpha_i K(\cdot, x_i), \quad g(\cdot) = \sum_{i=1}^l \beta_i K(\cdot, x_i), \quad h(\cdot) = \sum_{i=1}^l \gamma_i K(\cdot, x_i) \quad (1-40)$$

为集合 S 中的任意三个元素, c 和 d 是任意的实数. 则有

$$f(\cdot) + g(\cdot) = \sum_{i=1}^l \alpha_i K(\cdot, x_i) + \sum_{i=1}^l \beta_i K(\cdot, x_i) = \sum_{i=1}^l (\alpha_i + \beta_i) K(\cdot, x_i) = \sum_{i=1}^l (\beta_i + \alpha_i) K(\cdot, x_i)$$

$$= \sum_{i=1}^l \beta_i K(\cdot, \mathbf{x}_i) + \sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) = g(\cdot) + f(\cdot) \quad (1-41)$$

$$\begin{aligned} (f(\cdot) + g(\cdot)) + h(\cdot) &= \left(\sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) + \sum_{i=1}^l \beta_i K(\cdot, \mathbf{x}_i) \right) + \sum_{i=1}^l \gamma_i K(\cdot, \mathbf{x}_i) \\ &= \sum_{i=1}^l (\alpha_i + \beta_i + \gamma_i) K(\cdot, \mathbf{x}_i) \\ &= \sum_{i=1}^l (\alpha_i + (\beta_i + \gamma_i)) K(\cdot, \mathbf{x}_i) = \sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) + \left(\sum_{i=1}^l \beta_i K(\cdot, \mathbf{x}_i) + \sum_{i=1}^l \gamma_i K(\cdot, \mathbf{x}_i) \right) \\ &= f(\cdot) + (g(\cdot) + h(\cdot)) \end{aligned} \quad (1-42)$$

$$f(\cdot) + 0 = \sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) + 0 = \sum_{i=1}^l (\alpha_i + 0) K(\cdot, \mathbf{x}_i) = \sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) = f(\cdot) \quad (1-43)$$

$$f(\cdot) - f(\cdot) = \sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) - \sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) = \sum_{i=1}^l (\alpha_i - \alpha_i) K(\cdot, \mathbf{x}_i) = 0 \quad (1-44)$$

$$1 \times f(\cdot) = 1 \times \left(\sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) \right) = \sum_{i=1}^l (1 \times \alpha_i) K(\cdot, \mathbf{x}_i) = \sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) = f(\cdot) \quad (1-45)$$

$$c(df(\cdot)) = c \left(d \sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) \right) = (cd) \left(\sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) \right) = (cd)f(\cdot) \quad (1-46)$$

$$\begin{aligned} (c+d)f(\cdot) &= (c+d) \sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) = \sum_{i=1}^l (c+d) \alpha_i K(\cdot, \mathbf{x}_i) = \sum_{i=1}^l (c\alpha_i + d\alpha_i) K(\cdot, \mathbf{x}_i) \\ &= \sum_{i=1}^l c\alpha_i K(\cdot, \mathbf{x}_i) + \sum_{i=1}^l d\alpha_i K(\cdot, \mathbf{x}_i) \\ &= c \sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) + d \sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) = cf(\cdot) + df(\cdot) \end{aligned} \quad (1-47)$$

$$\begin{aligned} c(f(\cdot) + g(\cdot)) &= c \left(\sum_{i=1}^l \alpha_i K(\cdot, \mathbf{x}_i) + \sum_{i=1}^l \beta_i K(\cdot, \mathbf{x}_i) \right) \\ &= c \sum_{i=1}^l (\alpha_i + \beta_i) K(\cdot, \mathbf{x}_i) = \sum_{i=1}^l c(\alpha_i + \beta_i) K(\cdot, \mathbf{x}_i) \end{aligned}$$