Support Vector Machines
Algorithms and Financial Application

支持向量机算法及其金融应用

梁循◎著

知识产权出版社
全国百佳图书出版单位

支持向量机算法及其金融应用

梁 循 著

知识产权出版社
全国百佳图书出版单位

内容提要

本书综合了大量国内外的最新资料和作者的研究成果，有选择地介绍了支持向量机的一些算法和应用。全书从结构上分为三篇：第一篇介绍了支持向量机的一些概念；第二篇具体介绍了一些支持向量机的算法，包括支持向量机的结构修剪方法、粗略删除支持向量的方法、特征空间椭圆模式挖掘、在特征空间和经验图中的训练算法、奇偶校验问题的一些解法，并研究了支持向量机超曲面对两类样本的分隔问题；第三篇主要讨论了支持向量机的应用问题，包括互联网金融信息时间序列、股价预测等问题，以及互联网金融信息分析系统的介绍。

本书的读者可以是对支持向量机、数据挖掘感兴趣的计算机专业人士，也可以是对互联网金融信息挖掘感兴趣的领域专家，它可供支持向量机、数据挖掘、金融数据建模等领域的科技人员和高校师生研究参考。

责任编辑：江宜玲　　　　**责任出版**：卢运霞

图书在版编目（CIP）数据

支持向量机算法及其金融应用/梁循著. —北京：知识产权出版社，2012. 1

ISBN 978-7-5130-0908-9

Ⅰ. ①支… Ⅱ. ①梁… Ⅲ. ①支持向量机—算法理论—应用—金融 Ⅳ. ①F830. 49

中国版本图书馆 CIP 数据核字（2011）第 226057 号

支持向量机算法及其金融应用

ZHICHI XIANGLIANGJI SUANFA JIQI JINRONG YINGYONG

梁　循　著

出版发行：知识产权出版社

社　址：北京市海淀区马甸南村 1 号	**邮　编**：100088
网　址：http：//www. ipph. cn	**邮　箱**：bjb@ cnipr. com
发行电话：010-82000860 转 8101/8102	**传　真**：010-82000507/82000893
责编电话：010-82000860 转 8339	**责编邮箱**：jiangyiling@ cnipr. com
印　刷：知识产权出版社电子制印中心	**经　销**：新华书店及相关销售网点
开　本：787mm×1092mm　1/16	**印　张**：12. 5
版　次：2012 年 1 月第 1 版	**印　次**：2012 年 1 月第 1 次印刷
字　数：196 千字	**定　价**：38. 00 元

ISBN 978-7-5130-0908-9/F·484（3827）

前　言

本书介绍了一些支持向量机的算法与应用。本书作者先前出版的另外7本书籍《网络金融》、《数据挖掘算法与应用》、《互联网金融信息系统的设计与实现》、《电子商务理论与实践》、《网络金融信息挖掘导论》、《网络金融系统设计与实现案例集》和《互联网金融信息智能挖掘基础》之间的关系如图0－1所示。本书是它们的延续、补充和进一步系统化及理论归纳。

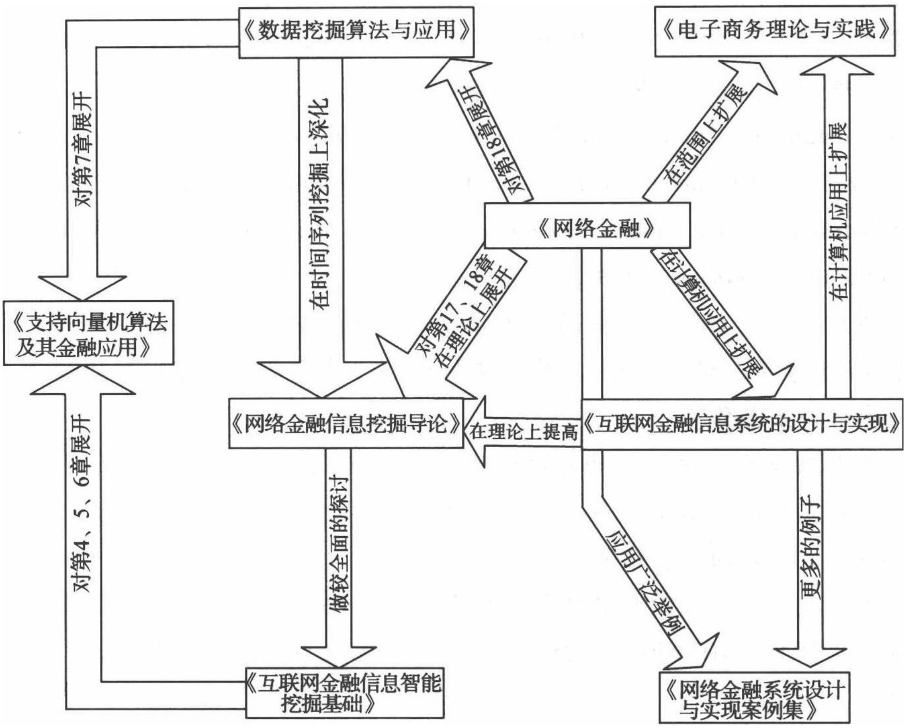

图0－1　《网络金融》、《数据挖掘算法与应用》、《互联网金融信息系统的设计与实现》、《电子商务理论与实践》、《网络金融信息挖掘导论》、《网络金融系统设计与实现案例集》、《互联网金融信息智能挖掘基础》和本书之间的关系

本书分为3篇。（1）绪论篇。介绍了和支持向量机相关的一些问题。（2）算法篇。介绍了支持向量机的结构修剪方法、粗略删除支持向量的方法、特征空间椭圆模式挖掘、在特征空间和经验图中的算法以及奇偶校验问题的三种解法，即SMO（Sequential Minimal Optimization）方法、直接使用梯度下降法以及大系统协调分解方法，列举了一些支持向量机超曲面分隔两类样本的情况。（3）应用篇。第9章结合随机过程的模型，讨论了互联网金融信息时间序列 W 分析问题及股价预测问题。第10章介绍了一个基于支持向量机的互联网金融信息分析系统。

本书作为学习支持向量机的参考资料，适用于对数据挖掘和支持向量机感兴趣的研究生和高年级本科生。由于国内已经有不少系统、全面介绍支持向量机理论算法的书籍，本书不再重复全面介绍有关支持向量机的所有算法与应用。国内目前比较著名的有关书籍有邓乃扬、田英杰教授的《支持向量机——理论、算法与应用》、《数据挖掘中的新方法——支持向量机》，以及Nello Cristianini和John Shawe-Taylor教授著 *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*、Vladimir N. Vapnik教授著 *Statistical Learning Theory*、John Shawe-Taylor和Nello Cristianini教授著 *Kernel Methods for Pattern Analysis*。本书中的符号方法，也受《支持向量机——理论、算法与应用》和《数据挖掘中的新方法——支持向量机》影响较重，如对向量 x 的第 i 个分量，记为 $[x]_i$。对支持向量机的算法系统，有兴趣的读者可以直接参考以上著作。

本书综合了大量国内外的最新资料及作者最近几年的研究成果，改写、简化和补充而成。本书的第9章、第10章的实验结果改编自作者的学生阮进、陈嬴、梁霞的实验结果。

本书也是中国人民大学科学研究基金（中央高校基本科研业务费专项资金资助）项目（10XNI029）和国家自然科学基金资助项目（70871001）的成果之一。

由于作者水平和时间的限制，书中一定存在不少缺点和错误，恳请读者批评指正。

梁　循

2011年8月

目 录

绪 论 篇

算 法 篇

应用篇

绪论篇

第 1 章　绪论

1.1　概　　述

近年来,统计学习理论发展迅速,已经形成了一套完整理论。从 20 世纪 90 年代中后期开始,基于核(kernel)函数的统计学习中的支持向量机(Support Vector Machine,SVM)取得了巨大成功,引起了同行的广泛兴趣。*IEEE Transactions on Neural Networks*(IEEE TNN)和 *Journal of Machine Learning Research*(JMLR)是影响因子相对较高、刊登论文集中在数据挖掘算法问题上的两个主要期刊。我们对刊登在这两个期刊上的 2000 ~ 2010 年的 SVM/kernel 的论文做了一个统计,发现其数量呈稳步上升趋势(见图 1 - 1)。此外,从内容上看,近几年,不论是在基础理论研究还是各种领域的应用上,SVM 的研究都在进一步深化和做实。

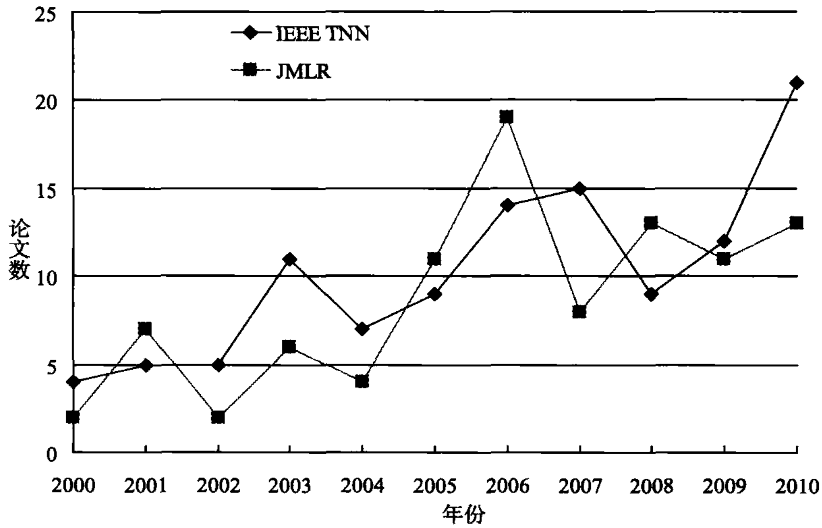

图 1 - 1　*IEEE Transactions on Neural Networks* 和 *Journal of Machine Learning Research* 中 SVM/kernel 的论文数

支持向量机是在统计学习理论的基础上发展起来的一种新的数据挖掘方法，在解决小样本、非线性和高维模式识别问题中表现出许多特有的优势。支持向量机是建立在最优超平面基础上的一种数据挖掘方法。通过引入核函数，支持向量机获得了很多的优点，如通过解二次规划问题，支持向量机算法巧妙地把数据映射到高维特征空间，并在特征空间中将分隔超平面正好放在距离两类都相等的位置，也就是说，可以使信息的各样本点在特征空间中得到最优拟合。此外，SVM 的训练过程没有局部极小点。

支持向量机适合于解决二分类模式识别问题。在理论上，支持向量机方法也已经得出了错误上界。支持向量机适用于小样本、非线性及高维模式识别领域，在模式识别、回归估计和天气预报等方面有着广泛的应用，与传统机器学习算法的性能相当。

与神经网络相比，SVM 有坚实的统计学基础，它具有以下 3 个优点。

(1) 以结构风险最小原理为基础，减小推广错误的上界，具有很好的推广性能，解决了神经网络的过适应现象。

(2) 问题的求解等价于线性约束的凸二次型规划问题，具有全局最优解，解决了神经网络的局部极小问题，且在一定条件下，可以有限步就收敛到最优点。

(3) 把原问题映射到高维空间，通过在高维空间构造线性决策函数来实现原问题的划分，引入核函数，解决了维数灾难问题。

考虑图 1－2 所示情况。图 1－2 中实心点和空心点分别表示不同的训练样本。图 1－2 中的实线显示了两个可能的分隔超平面，每个面都可以正确分隔两组数据。与实线平行的虚线表示分隔超平面平移后得到的平面，这种平移不会造成数据的分隔错误。平行线间的距离称为分类间隔或分类空隙(margin)。SVM 就是要在训练集中找到具有最大分类间隔的分隔超平面，即最优超平面。最优超平面使被它分隔的每一类数据最近的点与它之间的距离最大，即空隙最大，如图 1－2(b)的结果优于图 1－2(a)的结果，这个最优超平面能“最好”地分隔两个分类中的数据点，并可以获得最好的泛化能力。

可以证明，使分类空隙最大实际上就是使推广性的界中的置信范围最小，即使真实风险最小。在向高维空间作变换后，最优分类超平面就转换成更高维的最优分类超平面。

支持向量机的基础是统计学习理论，它主要是针对小样本统计问题，不仅考虑了对渐进性能的要求，而且追求在有限信息的条件下得到最优结果。基于

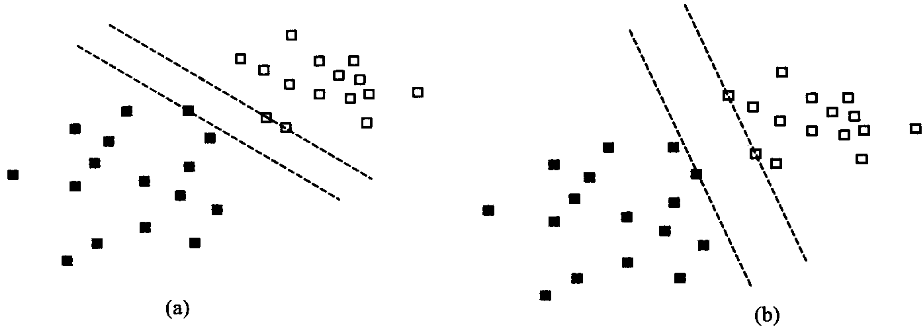

图 1－2 分隔超平面

统计学习理论的支持向量机方法建立在 Vapnik-Chervonenkis Dimension（VC 维）和结构风险最小化原理基础上，根据有限的样本建模。VC 维的直观意义如下：对一个指示函数集，如果存在 h 个样本能够决策函数集里的函数按照所有可能的 2^h 种形式分成两类，则称函数集能够把 h 个样本打散（shatter）。函数集的 VC 维就是它能够打散的最大样本数目 h。也就是说，如果存在一个有 h 个样本的样本集能被某函数集打散，而找不到一个有 $h+1$ 个样本的样本集能被该函数集打散，则函数集的 VC 维就是 h。若对任意数目的样本都有函数能将它们打散，则函数集的 VC 维就是 $+\infty$。下面我们给出函数集的 VC 维的定义。

定义 1－1 一个指示函数集的 VC 维等于，能够用该函数集以所有可能的 2^h 种形式分成不同两类的样本的最大数目（Vapnik，1998）。

例如，按照函数集的 VC 维定义，R^d 空间中的线性分类机

$$\sum_{i=1}^{d} w_i x_i + a$$

和

$$\text{sgn}\left(\sum_{i=1}^{d} w_i x_i + a\right)$$

的 VC 维均为 $d+1$，其中，$x \in R^d$，w_i，$a \in R$。

函数集能够打散的点越多，说明函数集的表达分辨能力越强，所以，函数集的 VC 维反映了函数集的学习分辨能力。一般地，函数集的 VC 维大表示该函数集学习机的学习容量大。目前，没有通用的关于任意函数集的 VC 维计算方法，只是对一些特殊的函数集我们可以计算其 VC 维。

以下，本章首先介绍线性支持向量机、非线性支持向量机，然后讨论核函数和模型选择等基本理论，最后介绍 Libsvm 仿真平台。

1.2 线性可分问题的SVM方法

假设有训练样本 $(x_1,y_1),\cdots,(x_l,y_l)$, $x_i\in\mathrm{R}^d$, $y_i\in\{-1,+1\}$, $i=1,\cdots,l$, 如果样本线性可分,则存在一个超平面,使得这两类样本完全分开,

$$w^Tx+b=0$$

其中,$w\in\mathrm{R}^d$,即

$$w^Tx_i+b<0,y_i=-1$$

$$w^Tx_i+b>0,y_i=+1$$

也就是说,调整 b,可以要求对应 $y=-1$ 的点 x^- 和对应 $y=+1$ 的点 x^+ 满足,

$$w^Tx^-+b=-1$$

$$w^Tx^++b=+1$$

在计算几何间隔 γ 时,先将 w 归一化,

$$\begin{aligned}\gamma &= \frac{1}{2}\left(<\frac{w}{\|w\|},x^+>-<\frac{w}{\|w\|},x^->\right)\\ &= \frac{1}{2\|w\|}(<w,x^+>-<w,x^->)\\ &= \frac{1}{\|w\|}\end{aligned}$$

其中,$<u,v>$表示 u,v 的内积,分类时要求几何间隔 γ 最大,也就是使 $\|w\|$ 最小,并且满足

$$0\leqslant y_i(w^Tx_i+b)-1$$

的条件。在两类样本中,过离最优分隔超平面最近的点,作和最优分隔超平面平行的两个超平面,位于这两个超平面上的样本,使上式等号成立,这些样本称为支持向量。

具体地说,在线性可分的情况下,求解最优超平面,即为求解二次型规划问题。

$$\min_w \frac{1}{2}\|w\|^2$$

$$\text{s.t. } 0\leqslant y_i(w^Tx_i+b)-1\ ,\ i=1,\cdots,l$$

对上面的问题,可由 Lagrangian 乘子法来解。设 $0\leqslant\alpha_i, i=1,\cdots,l$,

$$L = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l}\alpha_i[y_i(w^T x_i + b) - 1]$$

得

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{l}\alpha_i y_i = 0$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{l}\alpha_i y_i x_i = 0$$

求解之并代入原 Lagrangian 函数，可得

$$\begin{aligned}
L &= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l}\alpha_i[y_i(w^T x_i + b) - 1] \\
&= \frac{1}{2}\left(\sum_{i=1}^{l}\alpha_i y_i x_i\right)^T\left(\sum_{i=1}^{l}\alpha_i y_i x_i\right) - \sum_{i=1}^{l}\alpha_i\left\{y_i\left[\sum_{i=1}^{l}\alpha_i y_i (x_i)^T x_i + b\right] - 1\right\} \\
&= \frac{1}{2}\left(\sum_{i=1}^{l}\alpha_i y_i x_i\right)^T\left(\sum_{i=1}^{l}\alpha_i y_i x_i\right) - \left(\sum_{i=1}^{l}\alpha_i y_i x_i\right)^T\left(\sum_{i=1}^{l}\alpha_i y_i x_i\right) - b\sum_{i=1}^{l}\alpha_i y_i + \sum_{i=1}^{l}\alpha_i \\
&= \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\left(\sum_{i=1}^{l}\alpha_i y_i x_i\right)^T\left(\sum_{i=1}^{l}\alpha_i y_i x_i\right) \\
&= \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j x_i{}^T x_j
\end{aligned}$$

于是上述问题的对偶问题变为

$$\max_{\alpha_1,\cdots,\alpha_l}\sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j x_i{}^T x_j$$

$$\text{s. t. } \sum_{i=1}^{l}\alpha_i y_i = 0$$

$$0 \leqslant \alpha,\ i = 1,\cdots,l$$

注意，它只涉及样本的内积 $x_i^T x_j$。

在上面的规划中，通过训练算法可以得到最优的 α_i^*，代入相应的 Karush-Kuhn-Tucker(KKT)条件，可得 w^* 和 b^*。由 KKT 条件可以得出结论，对应 $\alpha_i^* = 0$ 的那些样本对 w^* 不起作用，只有那些 $0 < \alpha_i^*$ 的样本对 w^* 起作用，这些样本称为支持向量。

可以看出，SVM 的优点是网络隐节点数可以由算法确定，而且算法也没有局部极小问题。因而，

$$\begin{aligned}\| w^* \|^2 &= \left(\sum_{i=1}^{l}\alpha_i^* y_i x_i\right)^T\left(\sum_{i=1}^{l}\alpha_i^* y_i x_i\right)\\ &= \sum_{i=1}^{s}\alpha_i^* y_i \sum_{j=1}^{s}\alpha_j^* y_j x_i^T x_j\\ &= \sum_{i=1}^{s}\alpha_i^*(1-y_i b^*)\\ &= \sum_{i=1}^{s}\alpha_i^* - b^*\sum_{i=1}^{s}\alpha_i^* y_i\\ &= \sum_{i=1}^{s}\alpha_i^*\end{aligned}$$

即几何间隔为

$$\gamma = \frac{1}{\| w^* \|} = \frac{1}{\sqrt{\sum_{i=1}^{s}\alpha_i^*}}$$

对输入 x 进行测试时,可以使用

$$w^T x + b = \sum_{i=1}^{s}\alpha_i y_i < x_i, x > + b$$

来确定 x 的归类。

1.3 线性不可分问题的 SVM 方法

对于非线性可分问题,可以把样本 x 映射到某个高维空间 $\mathcal{H}$ 中去(见图 1-3),然后,在高维空间中,使用上面的方法。设该映射为 Φ, $x \to \Phi(x)$, 上节中的对偶问题变为

$$\begin{aligned}&\max_{\alpha_1,\cdots,\alpha_l}\sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \Phi^T(x_i)\Phi(x_j)\\ &\text{s.t.}\ \sum_{i=1}^{l}\alpha_i y_i = 0\\ &\qquad 0 \leqslant \alpha_i, i=1,\cdots,l\end{aligned}$$

注意,它只涉及样本变换到高维空间的内积 $\Phi^T(x_i)\Phi(x_j)$ 。相应地,决策函数变为

$$w^T\Phi(x) + b = \sum_{i=1}^{l}\alpha_i y_i \Phi^T(x_i)\Phi(x) + b$$

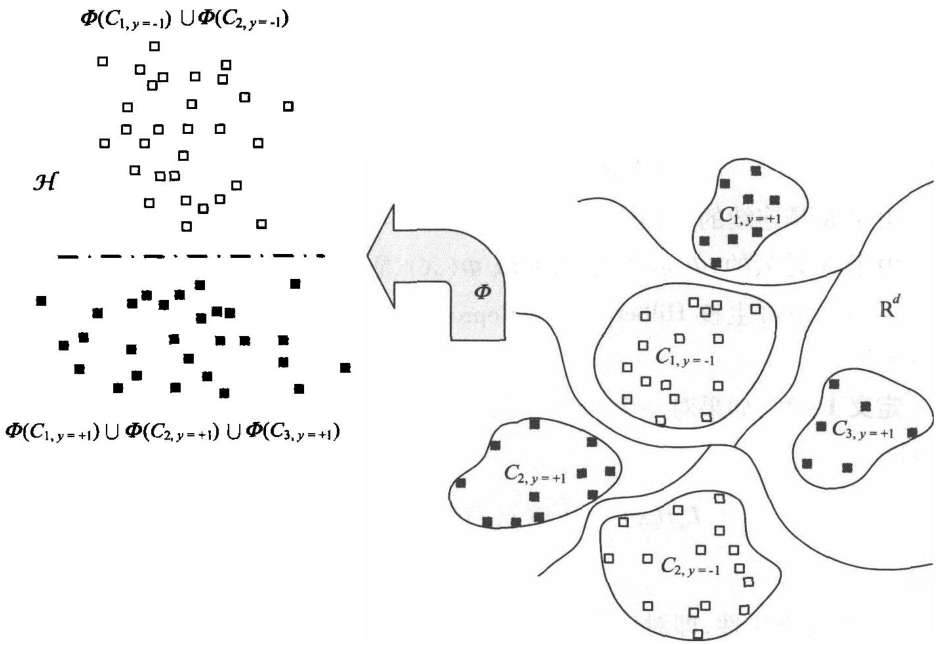

图 1－3 样本被映射到高维空间中去

可以看出，$\Phi^T(x_i)\Phi(x_j)$ 可以用原空间的函数来实现。事实上，我们注意定义变换后的内积即可，而不必实际进行这种变换。由 Hilbert-Schmidt 原理，只要核函数满足 Mercer 条件，它就对应某一空间中的内积（Vapnik，1998）。

引理 1－1（Mercer 条件） 设 $A\subseteq \mathcal{X}$ 为紧子集，$K(x,x')$ 为 $A\times A$ 上的连续实值对称函数，它在某个特征空间中内积运算的充分必要条件是，对于任意的不恒为 0 的函数 $g\in L^2(A)$，有

$$0\leqslant\int_{A\times A}K(x,x')g(x)g(x')\mathrm{d}x\mathrm{d}x'$$

事实上，

$$K(x,x')=\sum_{i=0}^{\infty}\lambda_i g_i(x)g_i(x')$$

其中，$g_i(x)$ 为算子 $T_K(\bullet)=\int_A K(\bullet,x')g(x')dx'$ 对应特征值 λ_i 的特征函数。

在定理条件下，$0\leqslant\lambda_i$。令 $\varphi_i(x)=\sqrt{\lambda_i}g_i(x)$，得

$$K(x,x')=\sum_i\varphi_i(x)\varphi_i(x')=<\Phi(x),\Phi(x')>$$

在实践中，定理的条件不难满足。满足定理条件的核成为 Mercer 核。