

# 有限维空间映射下 SVM 分类机的半正定核函数<sup>\*</sup>)

原峰山<sup>1,2</sup> 朱思铭<sup>1</sup>

(中山大学数学与计算科学学院 广州 510275)<sup>1</sup> (广州航海高等专科学校现代化教学中心 广州 510725)<sup>2</sup>

**摘 要** 本文针对支持向量分类机的特点,分析了在有限维空间中满足特征空间映射的条件和结果。提出了有限维空间映射下满足半正定条件的核函数确定方法。指出,只要存在对称函数  $K$ ,如果这个函数形成的 Gram 矩阵是半正定的,就必然能够将原输入空间映射成有限维的特征空间,从而在映射空间的内积可以用这个核函数  $K$  表示。

**关键词** 有限维空间,映射空间,核函数

## Positive Semidefinite Kernel Function of SVM Classification Machines under Finite-dimension Space Mapping

YUAN Feng-Shan<sup>1,2</sup> ZHU Si-Ming<sup>1</sup>

(School of Mathematics and Computing Science, Zhongshan (Sun Yat-sen) University, Guangzhou 510275)<sup>1</sup>

(Guangzhou Maritime College, Guangzhou 510725)<sup>2</sup>

**Abstract** With regard to feature of Support Vector Classification Machines, conditions and results to satisfy mapping of feature space in finite-dimension are analyzed. The method to determine kernel function satisfied with positive semidefinite condition in mapping space of finite-dimension is put forward. The conclusion is that if Gram matrix which is formed by symmetrical function  $K$  is positive semidefinite, the original imported space is certainly mapped into a finite-dimension feature space in which the inner product can be expressed by the kernel function  $K$ .

**Keywords** Finite-dimension space, Mapping space, Kernel function

### 1 支持向量分类机中的相似性与内积

在机器学习问题中,常遇到分类问题,它也是支持向量机理论需要解决的最重要的问题之一。对此类问题求解的实质,就是从已有的数据中发现一些规律,找出数据中的相似性。在支持向量机中,相似程度是用内积来表达和评估的,可描述如下<sup>[1]</sup>。

给定训练集:

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l$$

其中,

$x_i \in X = R^n$ ,  $y_i \in Y = \{1, -1\}$ ,  $i = 1, \dots, l$ 。  $x_i$  为输入,  $y_i$  为输出。在构成的支持向量分类机系统中,对于新的输入,经过系统的处理,可以将输入进行适当的划分。这种划分由输出决定,分成正类  $y=1$  和负类  $y=-1$ 。当新的输入与正类相似,则  $y=1$ ;与负类相似,则  $y=-1$ 。凡是相似的输入都应该导致相同的输出。应该指出的是,上述的系统在数学上就对应于一个决策函数,用于判断输入数据的归属,即称为分类。

任何一个输入向量  $x = (x_1, x_2, \dots, x_n)^T$  归属于哪一类,关键是看它与那一类的向量如  $x' = (x'_1, x'_2, \dots, x'_n)^T$  的相似度逼近如何。这种度量可以用内积的关系来表示,如用向量  $x$  与  $x'$  的夹角  $\alpha$  来表示,那么:

$$\cos \alpha = \frac{(x \cdot x')}{\sqrt{(x \cdot x) \cdot (x' \cdot x')}} \quad (1)$$

$\alpha$  越小,两个向量的相似度越高。在式(1)中,内积是表示向量相似度的最基本要素。

对于支持向量分类机而言,根据其决策函数来最后决定输入向量的归属,有两种情况<sup>[1]</sup>:

(1)线性支持向量分类机。决策函数为:

$$f(x) = \text{sgn}(w^* \cdot x + b^*) \quad (2)$$

$$w^* \cdot x = \sum_{i=1}^l y_i a_i x_i \cdot x \quad (3)$$

$$b^* = y_i - \sum_{i=1}^l y_i a_i^* (x_i \cdot x_j) \quad (4)$$

显然,最终决策值依赖于内积  $(x_i \cdot x_j)$  和  $(x_i \cdot x)$ 。

(2)非线性支持向量机。决策函数为:

$$f(x) = \text{sgn}(\sum_{i=1}^l y_i a_i^* K(x_i, x) + b^*) \quad (5)$$

$$b^* = y_i - \sum_{i=1}^l y_i a_i^* K(x_i \cdot x_j) \quad (6)$$

其中,  $K(x_i \cdot x_j) = (\Phi(x_i) \cdot \Phi(x_j))$

$$K(x_i \cdot x) = (\Phi(x_i) \cdot \Phi(x)), i, j = 1, \dots, l$$

在这里,最终决策函数值依赖于变换后的 Hilbert 空间的内积。上述实例说明,基于支持向量机的分类,无论是线性的还是非线性的,判断新的输入向量的归属取决于它与训练集的相似度,而这种相似度的最好描述也是用内积来表达的。

### 2 有限维空间映射下的分类问题

#### 2.1 非线性问题的分类及核函数

在我们面临的大量实际分类问题中,绝大多数都是非线性的,而这类问题的解决往往比较困难。自从 Vapnik 提出支持向量机理论以来,非常巧妙地解决了非线性分类问题。他指出,对于  $x$  所在的输入空间  $X$ ,虽然它是非线性空间,但是可以把它映射到另一个空间  $H$ ,称为 Hilbert 空间,也称为特

<sup>\*</sup>) 基金项目:国家自然科学基金资助项目(编号:10371135)。原峰山 高级工程师,博士生,研究方向:人工智能与计算机网络;朱思铭 教授,博士生导师,研究方向:应用数学,人工智能与计算机网络。

征空间,它通常是无穷维的,对应的是完备的内积空间。映射选定后,由其内积对应一个函数  $K(x, x')$ , 就可以使用支持向量机原理进行计算, 确定分类决策函数, 从而转化为线性空间的问题, 实现分类问题的求解。但由于 Vapnik 定义的 Hilbert 空间是无穷维的, 真正应用时比较困难。本文将提出有限维空间映射的支持向量分类机的概念。

基于支持向量机的非线性分类, 是指的其分类决策函数的非线性关系, 本质上就是确定函数  $K(x, x')$ , 即对应于式 (5) 中的函数。这个函数必须满足以下定义<sup>[1]</sup>:

**定义 1** 设  $X$  是  $R^n$  中的一个子集。称定义在  $X \times X$  上的函数  $K(x, x')$  是有限维空间下的核函数。如果存在着从  $X$  到某一个有限维空间的映射

$$\begin{aligned} X &\subset R^n \rightarrow R^m \\ \Phi: x &\rightarrow \Phi(x) \end{aligned}$$

使得  $K(x, x') = (\Phi(x) \cdot \Phi(x'))$ , 其中  $(\cdot)$  表示内积。

由上述定义可知, 用支持向量机解决非线性分类问题, 关键是找到或确定出这个核函数。

## 2.2 有限维空间下的 SVM 分类机的映射

Vapnik 在其代表性著作《统计学习理论的本质》一书<sup>[2]</sup>曾经提出了基于支持向量机的 Mercer 条件, 指出对前述的输入空间  $K(x, y)$  一般定义为核函数。根据 Hilbert-Schmidt 理论, 只要满足 Mercer 条件, 在非线性的特征空间中, 找到一定的核函数, 就能够把输入空间中线性不可分问题映射到一个特征空间而成为线性可分的问题, 使非线性问题得到了简化, 从而可以有效地解决构造支持向量机的问题。但是, 这个理论被映射的空间是无穷维的, 而在实际问题中面临的问题都是有限维空间而并非无穷维。在这里将主要描述有限维空间条件下的映射关系和核函数的特性。

对于 1 节中定义的训练集, 若输入是有限实数点集, 设  $K(x, x')$  为对称函数, 并且对  $x_i, x_j \in X$  定义了  $m^2$  个函数值, 形成的矩阵也是对称的, 即  $(K(x_i, x_j))_{i,j=1}^m$  为对称矩阵。  $K(x_i, x_j), i, j=1, 2, \dots, m$ , 在有限维条件下 ( $m$  维) 实现定义 1 的映射, 函数  $K(x, x')$  可用下述关系表达:

$$K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j))$$

$i, j=1, 2, \dots, m$ 。在满足映射实现之前, 即上述关系实现之前的函数暂称为  $K$  函数, 由  $K$  函数形成的对称矩阵称为  $K$  函数矩阵。

### 2.2.1 $K$ 函数的特征值和特征向量

(1)  $K$  函数对应的特征值  $\lambda_i$  为实数且大于零。

证明:

先证明  $\lambda_i$  为实数。设  $\lambda_i$  是  $K$  的特征值,  $\eta$  是  $K$  的属于  $\lambda_i$  的一个特征向量, 则

$$K\eta = \lambda_i \eta$$

若令  $\bar{\eta}$  表示  $\eta$  的共轭向量, 即  $\bar{\eta}$  的每个分量  $\bar{\eta}_i$  都是  $\eta$  的相应分量  $\eta_i$  的共轭复数 ( $i=1, 2, \dots, n$ ), 于是有:

$$\overline{K\eta} = \overline{\lambda_i \eta}$$

但由于  $K$  为实对称矩阵, 所以

$$\begin{aligned} (\bar{\eta})^T (K\eta) &= (\bar{\eta})^T K^T \eta \\ &= (K\bar{\eta})^T \eta = (\overline{K\eta})^T \eta \\ &= \overline{\lambda_i} (\bar{\eta})^T \eta \end{aligned}$$

又  $(\bar{\eta})^T (K\eta) = (\bar{\eta})^T \lambda_i \eta = \lambda_i (\bar{\eta})^T \eta$

因此  $\lambda_i (\bar{\eta})^T \eta = \overline{\lambda_i} (\bar{\eta})^T \eta$

又  $\because \eta \neq 0, \therefore (\bar{\eta})^T \eta \neq 0$ , 有  $\lambda_i = \overline{\lambda_i}$ , 即  $\lambda_i$  为实数。

再证明  $\lambda_i > 0$ , 即证明满足条件特征值  $\lambda_i$  是非负的。

用反证法。如果有一个负的特征值  $\lambda_i$  对应着特征向量  $u_i$ ,  $U$  为正交矩阵,  $\Lambda$  为对角矩阵, 那么特征空间中的点应为<sup>[3]</sup>:

$$z = \sum_{i=1}^n u_i \Phi(x_i) = \sqrt{\Lambda} U^T u_i$$

当  $z \neq 0$ , 有二阶范数:

$$\begin{aligned} \|z\|^2 &= (z \cdot z) = u_i^T U \sqrt{\Lambda} \sqrt{\Lambda} U^T u_i \\ &= u_i^T U \Lambda U^T u_i = u_i^T K u_i = \lambda_i < 0 \end{aligned}$$

结论矛盾, 并且与内积空间的几何特性矛盾。

(2)  $K$  函数转换为  $R^m$  的内积空间不同特征值的特征向量必正交。

证明:  $\because K$  为实对称矩阵,  $R^m$  是内积空间中的变换, 即:

$$R^m(\alpha) = k\alpha, k, \alpha \in R^m$$

$$\therefore (R^m(\alpha), \beta) = (k\alpha, \beta) = k(\alpha, \beta)$$

又设:  $\lambda, u$  是  $K$  的两个相异特征值,  $\alpha, \beta$  是分别属于  $\lambda, u$  的特征向量, 而由于  $K$  是对称矩阵, 所以

$$\beta^T (K\alpha) = \beta^T K^T \alpha$$

$$= (K\beta)^T \alpha = \alpha K \beta$$

$$\therefore \lambda(\beta, \alpha) = u(\alpha, \beta)$$

即  $\lambda(\alpha, \beta) = u(\alpha, \beta)$

$\because \lambda \neq u, \therefore (\alpha, \beta) = 0$ , 即  $\alpha$  与  $\beta$  正交。

### 2.2.2 $K$ 函数的半正定展开特性

(1) 满足 Gram 矩阵的  $K$  函数矩阵必为半正定的。

**定义<sup>[1]</sup>** 对于给定的  $K$  函数, 它是在  $X \times X$  上的连续对称函数, 且  $x_1, \dots, x_l \in X$ , 称第  $i$  行第  $j$  列元素  $K_{ij} = K(x_i, x_j)$  的  $l \times l$  矩阵  $K$  为关于  $x_1, \dots, x_l$  的 Gram 矩阵。

**定理 1** 若  $K$  是关于  $x_1, \dots, x_l$  的 Gram 矩阵, 则对称  $K$  矩阵是半正定的。

证明: 设有任意非零向量  $\alpha$ , 已知  $K$  为对称矩阵, 所以对任意矩阵  $M$ , 有  $K = M^T M$

$$\alpha^T K \alpha = \alpha^T M^T M \alpha$$

$$\therefore = (M\alpha)^T M\alpha = (M\alpha \cdot M\alpha)$$

$$= \|M\alpha\|^2 \geq 0$$

$\therefore K$  是一个半正定矩阵。

(2) 当  $K$  函数构成的矩阵关于  $x_1, \dots, x_l \in X$  的 Gram 矩阵半正定, 则  $K$  函数可以实现到  $R^m$  空间的映射。

**定理 2** 若  $K$  满足定理 1, 则存在着  $m$  个非负的特征值和  $K$  转换为  $R^m$  内积空间的相互正交的单位特征向量:

$$\lambda_i; u_i = (u_{i1}, u_{i2}, \dots, u_{im})^T$$

$$\|u_i\| = 1, i=1, 2, \dots, m, \text{ 对特征映射:}$$

$$\Phi: x_i \rightarrow (\sqrt{\lambda_i} u_{ij})_{j=1}^m \in R^m, i=1, \dots, m,$$

有:

$$K(x_i, x_j) = \sum_{i=1}^m \lambda_i u_{ij} u_{ij}, i, j=1, 2, \dots, m$$

证明: 矩阵  $K$  是对称的,  $\therefore$  存在着正交矩阵  $U = (u_1, u_2, \dots, u_m)$  和对角矩阵  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ , 使得:

$$K = U \Lambda U^T$$

$$\text{即 } K_{ij} = K(x_i, x_j) = (U \Lambda U^T)_{ij}$$

$$\text{而 } (U \Lambda U^T)_{ij} = \sum_{t=1}^m u_{it} \lambda_t u_{jt} = \sum_{t=1}^m \lambda_t u_{it} u_{jt}$$

$$\therefore K_{ij} = K(x_i, x_j) = (U \Lambda U^T)_{ij}$$

$$= \sum_{t=1}^m u_{it} \lambda_t u_{jt} = \sum_{t=1}^m \lambda_t u_{it} u_{jt}$$

其中  $u_t = (u_{t1}, u_{t2}, \dots, u_{tm})^T$  是矩阵  $K$  的第  $t$  个特征向量, 它对应的特征向量是  $\lambda_t$ 。这就说明对映射:

$\Phi: x_i \rightarrow (\sqrt{\lambda_i} u_{in})_{i=1}^n \in R^n, i=1, \dots, n$ , 满足

$$(\Phi(x_i) \cdot \Phi(x_j)) = \sum_{i=1}^m \lambda_i u_{in} u_{jn} = K(x_i, x_j)$$

当上式成立时,就存在着从  $X=\{x_1, \dots, x_m\}$  到  $R^m$  的映射,满足

$$K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j))$$

由上述分析可知,对有限输入实数点集的情况,若被映射的空间有限维( $m$  维),只要满足  $K$  函数矩阵关于  $x_1, \dots, x_l \in X$  的 Gram 矩阵半正定即可;只要找到对应的核函数,就能实现这种映射,并且可以将高位空间的点积运算转化为核函数的运算,从而使问题的解决得到简化。

### 3 有限维空间映射下分类问题的误差

在 2.2 节中曾提到著名的 Mercer 定理通过核映射可以实现到一个无穷维 Hilbert 空间的映射。但在应用中,没有必要也不可能做这种映射,实际上采用的是有限维空间的映射,但这种映射是存在误差的,只要它在允许的范围内就可以了。

对于有限维映射导致的误差,应有如下表述。若  $K$  是一个核函数,则存在一个映射  $\Phi$ ,使得在相应的 Hilbert 空间  $H$  有:

$$K(x, x') = (\Phi(x) \cdot \Phi(x'))$$

并且对任意的  $\epsilon > 0$ , 存在一个从  $R^n$  到  $R^m$  的映射  $\Phi_p$  和正整数  $p$  ( $p \in N$  依赖于  $\epsilon$ ), 使得

$$|K(x, x') - (\Phi_p(x) \cdot \Phi_p(x'))| < \epsilon$$

也就是说,当  $K$  可以  $R^m$  中的  $\Phi_p$  内积表示,在有限维空间中可以以误差不超过  $\epsilon$  的精度逼近。

现在以一个实际的分类实例来说明有限维分类的误差问题。

在二维平面上取 100 个向量作为实验数据,决定向量特征的参数为坐标值和两类不同的颜色,利用 Chih-Jen Lin<sup>[4]</sup> 开发的 LIBSVM 软件,可以对以上数据进行分类。本例取核函数为径向基函数(RBF),表达式可写成以下形式:

$$K(x, y) = \exp(-\gamma \|x - y\|^2), \gamma > 0$$

对核函数起主要影响的参数是  $\gamma$ 。对原数据进行 SVM 处理后,生成新的数据模型,构成其对偶形式。决定支持向量机训练效果的是参数  $C$  和  $\gamma$ , 取不同的参数得到不同的分类精度。为了较好地比较和函数引起的误差,取相同  $C$  值,  $\gamma$  的参数取值为不同的数值。当取不同的组合时,做支持向量机分类处理,取得的结果如表 1。

表 1

| 组别  | C     | $\gamma$ | 迭代次数 | 分类误差 |
|-----|-------|----------|------|------|
| 第一组 | $2^3$ | $2^5$    | 43   | 14%  |
| 第二组 | $2^3$ | $2^{-3}$ | 45   | 9%   |
| 第三组 | $2^3$ | $2^{-1}$ | 83   | 2%   |
| 第四组 | $2^3$ | 2        | 84   | 2%   |
| 第五组 | $2^3$ | $2^5$    | 600  | 0    |

从表 1 中可以看出,当  $C$  值一定( $2^3$ )、 $\gamma$  逐渐变大时,分类精度逐渐提高,误差减小,而且变化比较明显,但是运算时迭代的次数也大为提高。尤其应该注意的是,当迭代次数过大而分类误差为 0 时,分类间隔线形状明显扭曲,说明分类时

确定判别决策函数的难度加大,对构造分类型 SVM 并不利。以表 1 的第三组和第五组数据为例,虽然分类误差由 2% 减少到 0,但运算时迭代的次数却由 83 次上升到 600 次,在解决实际问题时并不一定是最好的选择。图 1 和图 2 比较形象地表明了核函数为径向基函数,  $\gamma$  分别为  $2^{-1}$  和  $2^5$ , 其中图 1 是由第三组数据作出的分类图形,而图 2 是由第五组数据作出的图形,从图中可以明显地看到分隔线形状的扭曲。



图 1



图 2

**结论** 对于非线性分类问题,实际上是构造一个非线性的分类器。一般情况下是使用一个非线性映射将数据变换到一个特征空间,公认的是 Hilbert 空间,然后在这个特征空间形成一个实际的线性分类器。在构造的分类器中,其决策函数可以表达为对偶形式,分类的决策规则取决于核函数。

Hilbert 空间条件下的核函数决定方法通常是被映射到无穷维空间,依据是 Mercer 定理。理论上讲,它能够无误差地对样本分类。但是,在实际的分类问题中做不到这一点,而且也没有必要。本文提出了有限维空间映射下满足半正定条件的核函数确定方法。指出,只要存在对称函数  $K$ ,如果这个函数形成的 Gram 矩阵是半正定的,就必然能够将原输入空间映射成有限维的特征空间,从而在映射空间的内积可以用这个核函数表示。此外,还对上述结论提出的相应算法引起的误差进行了分析。

### 参考文献

- 1 邓乃扬,田英杰. 数据挖掘中的新方法—支持向量机. 北京:科学出版社,2004
- 2 Vapnik V N. 统计学习理论的本质. 张学工译. 北京:清华大学出版社,2000
- 3 Cristianini N, Shawe-Taylor J. 支持向量机导论. 李国正,王猛,曾华军,译. 北京:电子工业出版社,2004
- 4 <http://www.csie.ntu.edu.tw/~cjlin/libsvm>