稀疏贝叶斯模型与相关向量机学习研究

杨国鹏[1] 周 欣[2] 余旭初[1]
(信息工程大学测绘学院 郑州 450052)[1] (信息工程大学信息工程学院 郑州 450002)[2]

摘 要 虽然支持向量机在模式识别的相关领域得到了广泛应用,但它自身固有许多不足之处。相关向量机是在稀疏贝叶斯框架下提出的稀疏模型,模型没有规则化系数,核函数不要求满足Mercer条件。相关向量机不仅具备良好的泛化能力,而且还能够得到具有统计意义的预测结果。首先介绍了稀疏贝叶斯回归和分类模型,通过参数推断过程,将相关向量机学习转化为最大化边缘似然函数估计,并分析了3种估计方法,给出了快速序列稀疏贝叶斯学习算法流程。

关键词 稀疏贝叶斯模型,相关向量机,支持向量机
中图法分类号 TP751 **文献标识码** A

Research on Sparse Bayesian Model and the Relevance Vector Machine

YANG Guo-peng[1] ZHOU Xin[2] YU Xu-chu[1]
(Institute of Surveying and Mapping, Information Engineering University, Zhengzhou 450052, China)[1]
(Institute of Information Engineering, Information Engineering University, Zhengzhou 450002, China)[2]

Abstract The support vector machine is successfully applied in many fields of pattern recognition, but there ere several limitations thereof. The relevance vector machine is a Bayesian treatment, its mathematics model doesn't have regularization coefficient, and its kernel functions don't need to satisfy Mercer's condition. The relevance vector machine can present the good generalization performance, and its predictions are probabilistic. We introduced the sparse Bayesian modes for regression and classification, regarded the relevance vector machine learning as the maximization of marginal likelihood through the model parameters inference, then we described three kinds of training methods and presented the flow of the fast sequential sparse Bayesian learning algorithm.

Keywords Sparse bayesian model, Relevance vector machine, Support vector machine

1 引言

自20世纪90年代中期开始,核方法在支持向量机(Support Vector Machine,SVM)中得到成功应用,人们开始了利用核函数将经典的线性特征提取与分类方法推广到更一般情况的研究,取得了许多重要的研究成果,因此核方法被成为继经典统计线性分析、神经网络与决策树非线性分析之后的第三次模式分析方法的变革[1,2]。同时,SVM、稀疏核主成份分析[3]等的广泛应用,引起了人们研究"稀疏"学习模型的兴趣。

稀疏学习模型具有的一般形式为

$$y(x)=\sum_{m=1}^{M} w_m \phi_m(x) \tag{1}$$

它是相对于权值向量 $w=(w_1,\cdots,w_M)^T$ 的线性模型,$y(x)$能够逼近实变量函数或判别函数。假定训练样本集$\{x_n, t_n\}_{n=1}^{N}$,稀疏模型通过将权值向量 w 的多数元素设置为零来控制模型复杂度,避免过学习现象,减小模型预测的计算量。

SVM采用基于统计学习理论的结构风险最小化原则。对于两类分类问题,SVM先将样本投影到高维线性可分空间,构造具有低VC维的最优分类超平面来作为判决面,使得线性可分的两类样本之间的间隔最大[4],其数学模型为

$$y(x;w)=\sum_{i=1}^{N} w_i K(x,x_i)+w_0 \tag{2}$$

式中,$K(x,x_i)$为核函数,是定义在训练样本点的基函数。式(1)中 w 包含偏移量 w_0,因此 $M=N+1$。

SVM自身存在着许多不足之处,主要表现在:① 基函数个数基本上随训练样本集的规模成线性增长,模型稀疏性有限;② 预测结果不具有统计意义,无法直接获取预测结果的不确定性;③ 核函数参数和规则化系数通常需要通过交叉验证等方法来获得,增加了模型训练的计算量;④ 核函数 $K(x,x_i)$必须满足Mercer条件。

2000年,Tipping提出一种稀疏概率模型即相关向量机(Relevance Vector Machines,RVM)[5]来弥补SVM的不足,它是在贝叶斯框架下,进行回归估计获得预测值的分布,以得到一个基于核函数的稀疏解。2003年,Tipping设计了快速序列稀疏贝叶斯学习算法,显著提高了模型训练速度[6];2005年,Thayananthan将该模型推广,解决了多元输出回归和多类分类的训练问题[7]。最近,开始了RVM在文本识别[8]、影像分类[9]、时序分析[10]等领域的应用研究。

到稿日期:2009-08-25 返修日期:2009-11-10
杨国鹏(1982—),男,博士生,主要研究方向为图像处理与模式识别、高光谱遥感应用,E-mail:yangguopeng@hotmail.com;**周 欣**(1983—),女,博士生,主要研究方向为信号处理及模式识别;**余旭初**(1963—),男,教授,博士生导师,主要研究方向为遥感影像军事应用、图像处理与模式识别。

RVM最初用以处理回归问题，通过Laplace逼近可以将分类问题转化为回归问题。本文首先介绍稀疏贝叶斯回归和分类模型；然后通过参数推断将RVM学习转化为最大化边缘似然函数估计；最后分析了3种参数估计方法，并描述了快速序列稀疏贝叶斯学习的算法流程。

2 稀疏贝叶斯回归模型

2.1 模型描述

对于一维目标函数的回归问题[11]，假定训练样本集为$\{x_n,t_n\}_{n=1}^N$，$x_n\in\mathbb{R}^d$，$t_n\in\mathbb{R}$，依据概率论观点，假设目标值t_n获取时常附带有误差ε_n，则目标值可以表示为$t_n=y(x_n;w)+\varepsilon_n$。

在稀疏贝叶斯框架里，假定误差ε_n服从独立的零均值Gauss分布，即$p(t_n|x)=N(t_n|y(x_n;w),\sigma^2)$，方差$\sigma^2$通常需要估计得到。基函数采用定义在训练样本向量的核函数，即$\phi_i(x)=K(x,x_i)$。由于这里不要求基函数为正定的，因此核函数不必要满足Mercer条件。不失一般性，稀疏贝叶斯模型采用式(1)所示的表达形式。

假设训练样本独立同分布，则训练样本集的似然函数可以表示为

$$p(t|w,\sigma^2)=(2\pi\sigma^2)^{-N/2}\exp\{-\frac{1}{2\sigma^2}\|t-\Phi w\|^2\}$$

式中，$t=(t_1,\cdots,t_N)^T$为目标向量，$w=(w_0,\cdots,w_N)^T$为参数向量，$\Phi=[\phi(x_1),\phi(x_2),\cdots,\phi(x_M)]$为基函数，而且$\phi(x_n)=[1,K(x_n,x_1),K(x_n,x_2),\cdots,K(x_n,x_N)]^T$。

如果直接最大化似然函数来估计参数向量w，会导致模型过学习。假设参数w_i服从均值为0方差为α_i^{-1}的Gauss条件概率分布，因此

$$p(w|\alpha)=\prod_{i=0}^{M}N(w_i|0,\alpha_i^{-1})$$

式中，α是决定权值w先验分布的超参数。

由于Gauss正态分布方差倒数的共轭概率分布为Gamma分布[11]，因此假设α_i与σ^2的超先验概率分别为

$$p(\alpha)=\prod_{i=0}^{N}\mathrm{Gamma}(\alpha_i|a,b)$$

$$p(\sigma^2)=\mathrm{Gamma}(\sigma^2|c,d)$$

式中，$\mathrm{Gamma}(\alpha|a,b)=\Gamma(a)^{-1}b^a\sigma^{a-1}e^{-b\alpha}$

$$\Gamma(a)=\int_0^\infty t^{a-1}e^{-t}\mathrm{d}t$$

由于不存在先验知识，假定$a=b=c=d=0$获取一致超先验，因此，存在

$$p(w_i)=\int N(w_i|0,\alpha_i^{-1})\mathrm{Gamma}(\alpha_i|a,b)\mathrm{d}\alpha_i$$

这样为每一个权值(或基函数)配置独立的超参数是稀疏贝叶斯模型最显著的特点，也是导致模型具有稀疏性的根本原因[12]。由于这种先验概率分布是一种自动相关判定(Automatic Relevance Determination，ARD)先验分布，模型训练结束后，非零权值的基函数所对应的样本向量被称为相关向量，这种学习机被称为相关向量机。

2.2 参数推断

根据贝叶斯公式，若已知模型参数的先验概率分布$p(w,\alpha,\sigma^2)$，那么训练样本集的后验概率为

$$p(w,\alpha,\sigma^2|t)=\frac{p(t|w,\alpha,\sigma^2)p(w,\alpha,\sigma^2)}{p(t)}$$

假定待测样本为x_*，则相应的预测值t_*的分布为

$$p(t_*|t)=\int p(t_*|w,\alpha,\sigma^2)p(w,\alpha,\sigma^2|t)\mathrm{d}w\mathrm{d}\alpha\mathrm{d}\sigma^2$$

由于模型参数的后验分布$p(w,\alpha,\sigma^2|t)$不能通过积分直接获取，故将其分解为

$$p(w,\alpha,\sigma^2|t)=p(w|t,\sigma,\sigma^2)p(\alpha,\sigma^2|t)$$

由于$p(t|\alpha,\sigma^2)=\int p(|t|w,\alpha)p(|w|\alpha)\mathrm{d}w$可以积分得到，权值向量$w$后验分布可以表示为

$$p(w|t,\alpha,\sigma^2)=\frac{p(t|w,\sigma^2)p(w|\alpha)}{p(t|\alpha,\sigma^2)}$$

$$=(2\pi)^{-(N+1)/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(w-\mu)^T\Sigma^{-1}(w-\mu)\}$$

因此，后验概率分布的均值$\mu=\sigma^{-2}\Sigma\Phi^Tt$，方差$\Sigma=(\sigma^{-2}\Phi^T\Phi+A)^{-1}$，$A=\mathrm{diag}(\alpha_0,\alpha_1,\cdots,\alpha_N)$。

RVM学习问题就可以转化为超参数后验分布$p(\alpha,\sigma^2|t)\propto p(t|\alpha,\sigma^2)p(\alpha)p(\sigma^2)$关于$\alpha$和$\sigma^2$的最大值问题。在一致超先验分布的情况下，最大化$p(\alpha,\sigma^2|t)$问题可以表示为

$$\log(p(t|\alpha,\sigma^2))=\log(\int p(t|w,\sigma^2)p(w|\alpha)\mathrm{d}w)$$

$$=-\frac{1}{2}[N\log(2\pi)+\log|C|+t^TC^{-1}t]$$

式中，$C=\sigma^2I+\Phi A^{-1}\Phi^T$。

在贝叶斯模型理论中，$p(\alpha,\sigma^2|t)$被称为边缘似然函数，通过最大化边缘似然函数$p(\alpha,\sigma^2|t)$来估计α和σ^2的方法被称为第II类型最大似然参数估计方法[12]。

2.3 回归预测

假设第II类型最大似然参数估计值为α_{MP}和σ_{MP}^2，那么待测样本x_*的预测值t_*的分布为

$$p(t_*|t,\alpha_{MP},\sigma_{MP}^2)=\int p(t_*|w,\sigma_{MP}^2)p(w|t,\alpha_{MP},\sigma_{MP}^2)\mathrm{d}w$$

由于被积函数是两个Gauss正态分布的乘积，因此预测值t_*同样服从Gauss正态分布，即

$$p(t_*|t,\alpha_{MP},\sigma_{MP}^2)=N(t_*|y_*,\sigma_*^2)$$

式中，$y_*=\mu^T\phi(x_*)$，$\sigma_*^2=\sigma_{MP}^2+\phi(x_*)^T\Sigma\phi(x_*)$。

那么样本x_*的预测值t_*的均值为$y(x_*;\mu)$。基函数权值向量为后验概率均值μ，具有许多零元素。因此RVM具有很强的稀疏性，我们把非零权值的基函数所对应的样本向量称为“相关向量”。

3 稀疏贝叶斯分类模型

3.1 模型描述

对于分类问题，基函数的线性组合需要经过S形函数映射，即$y(x;w)=\sigma(w^T\phi(x))$，S形函数表示$\sigma(z)=1/(1+e^{-z})$。与稀疏贝叶斯归回模型类似，假设参数$w_i$服从均值为0方差为$\alpha_i^{-1}$的Gauss条件概率分布。

针对两类分类问题，假设样本独立同分布，那么样本集的似然函数可以表示为

$$P(t|w)=\prod_{n=1}^{N}\sigma\{y(x_n;w)\}^{t_n}[1-\sigma\{y(x_n;w)\}]^{1-t_n}$$

式中，$t_n\in\{0,1\}$为目标值。

3.2 参数推断

针对分类问题，后验概率密度$p(w|t,\alpha)$和边缘似然函数$p(t|\alpha)$都无法通过积分求解，需要采用MacKay[13]提出的La-

place 逼近方法近似。

当 α 一定时，就是要找到最大后验概率估计值 w_{MP}，即 Gauss 近似分布的众数位置 μ_{MP}。由于

$$p(w|t,\alpha)=\frac{p(t|w)p(w|\alpha)}{p(t|\alpha)}$$

因此关于 w 的最大后验概率估计等价于最大化

$$\begin{aligned}&\log\{p(w|t,\alpha)\}\\&=\log\{p(t|w)\}+\log\{p(w|\alpha)\}-\log\{p(t|\alpha)\}\\&=\sum_{n=1}^{N}[t_n\log y_n+(1-t_n)\log(1-y_n)]-\frac{1}{2}w^TAw+\text{const}\end{aligned}\tag{3}$$

式中，$y_n=\sigma\{y(x_n;w)\}$。

式(3)为带惩罚的对数似然函数，可以通过迭代再加权最小二乘法求解。式(3)关于 w 的梯度向量和海赛矩阵分别为

$$\nabla_w\log p(w|t,\alpha)|_{w_{MP}}=\Phi^T(t-y)-Aw$$

$$\nabla_w\nabla_w\log p(w|t,\alpha)|_{w_{MP}}=-(\Phi^TB\Phi+A)$$

式中，$B=\text{diag}(\beta_1,\beta_2,\cdots,\beta_N)$，$\beta_n=\sigma\{y(x_n)\}[1-\sigma\{y(x_n)\}]$。

Gauss 正态分布来逼近后验概率分布的 Laplace 方法，是对后验概率分布的众数位置 μ_{MP} 处的函数的二次逼近。在 IRLS 迭代收敛后，得到以权值向量 μ_{MP} 为中心的近似 Gauss 分布，均值 $\mu_{MP}=A^{-1}\Phi^T(t-y)$，方差 $\Sigma=(\Phi^TB\Phi+A)^{-1}$。

同样地，使用 Laplace 逼近方法可以将边缘似然函数近似表示为

$$p(t|\alpha)=\int p(t|w)\ p(w|\alpha)\mathrm{d}w\simeq p(t|w_{MP})p(w_{MP}|\alpha)(2\pi)^{M/2}|\Sigma|^{1/2}$$

如果令 $\hat{t}=\Phi w_{MP}+B^{-1}(t-y)$，则近似 Gauss 后验分布均值为 $\mu_{MP}=\Sigma\Phi^TB\hat{t}$，方差为 $\Sigma=(\Phi^TB\Phi+A)^{-1}$。近似的边缘似然函数对数的形式为

$$\log p(t|\alpha)=-\frac{1}{2}\{N\log(2\pi)+\log|C|+(\hat{t})^TC^{-1}\hat{t}\}$$

式中，$C=B+\Phi A^{-1}\Phi^T$。

比较分类和回归模型的后验概率分布和边缘似然函数对数形式可知，利用 Laplace 逼近方法已经将分类问题转化为回归问题，相应回归问题的目标向量为

$$\hat{t}=\Phi w_{MP}+B^{-1}(t-y)$$

因此，稀疏贝叶斯的回归和分类模型学习，最终都归结为第 II 类型最大似然参数估计。

3.3 多类分类器

上面研究的是二值分类问题的稀疏贝叶斯分类模型，在多类别分类情况下，假设共存在 K 个类别($K>2$)，随机样本服从独立同分布的多项式分布，此时最大似然函数可以表示为[14]

$$P(t|w)=\prod_{n=1}^{N}\prod_{k=1}^{K}\sigma\{y_k(x_n;w_k)\}^{t_{nk}}$$

这里采用 K 目标编码方法，分类器共有 K 个输出 $y_k(x_n;w)$，每个输出都有独自的参数向量 w_k 和超参数 α_k。也可以与 SVM 多类分类一样，通过不同方法将其分解成多个二类问题进行求解。

4 相关向量机学习

RVM 学习最终归结为第 II 类型最大似然参数估计问题，通过最大化边缘似然函数 $p(t|\alpha,\sigma^2)$ 来估计 α 和 σ^2。针对回归问题，通常采用以下 3 种方法[7]。

4.1 MacKay 迭代估计

使边缘似然函数 $p(t|\alpha,\sigma^2)$ 关于 α 的导数等于零，按照 MacKay 方法整理得[11]

$$\alpha_i^{new}=\gamma_i/\mu_i^2\tag{4}$$

式中，μ_i 是后验概率 $p(w|t,\alpha,\sigma^2)$ 的均值向量 μ 的第 i 个后验概率，$\gamma_i=1-\alpha_i\Sigma_{ii}$，$\Sigma_{ii}$ 是后验概率 $p(w|t,\alpha,\sigma^2)$ 的方差 Σ 的第 i 个对角元素。

使边缘似然函数 $p(t|\alpha,\sigma^2)$ 关于 σ^2 的导数等于零，得到迭代估计公式

$$(\sigma^2)^{new}=\|t-\Phi\mu\|/(N-\Sigma_i\gamma_i)\tag{5}$$

式中，N 为训练样本的个数。

RVM 学习过程，就是迭代使用式(4)与式(5)计算 α_i^{new} 和 $(\sigma^2)^{new}$，并随后更新后验概率 $p(w|t,\alpha,\sigma^2)$ 的统计量 Σ 和 μ，直到满足合适的收敛准则。

在实际估计过程中，许多估计值 α_i 趋于无限大，所以相应的权值 w_i 为零，与它对应的基函数将被删除，实现模型的稀疏化。

4.2 期望最大化迭代估计

期望最大化迭代估计[11,14]通过最大化边缘似然函数 $p(t|\alpha,\sigma^2)$ 估计 α 和 σ^2 时，将权值当作“隐含”变量，最大化期望为

$$E_{w|t,\alpha,\beta}[\log p(t|w,\beta)p(w|\alpha)p(\alpha)p(\beta)]$$

式中，$\beta=\sigma^{-2}$。$E_{w|t,\alpha,\beta}[\cdot]$ 表示在给定样本和隐藏变量时权值分布 $p(w|t,\alpha,\sigma^2)$ 的期望值。EM 迭代估计是通过以下重复迭代计算实现的。

对于 α，忽略本身对数独立性，等价于最大化

$$E_{w|t,\alpha,\beta}[\log p(w|\alpha)p(\alpha)]$$

求导得到更新公式 $\alpha_i=\frac{(1+2a)}{([w_i^2]+2b)}$，因此，根据后验概率存在 $[w_i^2]=E_{w|t,\alpha,\beta}[w_i^2]=\Sigma_{ii}+\mu_i^2$。

对于 σ^2，等价于最大化

$$E_{w|t,\alpha,\beta}[\log p(t|w,\beta)p(\beta)]$$

得到更新公式 $(\sigma^2)^{new}=\|t-\Phi\mu\|+(\sigma^2)^{old}\Sigma_i\gamma_i/N$。

4.3 自下而上的基函数选择

最大边缘似然估计超参数过程中，超参数更新需要计算后验权值的协方差矩阵，矩阵求逆需要计算复杂度 $O(M^3)$ 和存储空间 $O(M^2)$[11]，M 为基函数的个数。

*自下而上的基函数选择方法是 Tipping 于 2003 年提出*的快速序列稀疏贝叶斯学习算法[6]，基函数个数从 1 开始不断增加直至获取相关向量，而且 Φ 与 Σ 只包含当前模型中存在的基函数，因此该方法计算速度较其它两种要快得多。

由于边缘似然函数的对数 $L(\alpha)$ 与单个超参数 α_i 的相关性，$i\in\{1,\cdots,M\}$，将 C 分解为

$$C=C_{-i}+\alpha_i^{-1}\phi_i\phi_i^T$$

式中，C_{-i} 是 C 第 i 个基函数影响去除后的矩阵，满足

$$|C|=|C_{-i}||1+\alpha_i^{-1}\phi_i^TC_{-i}^{-1}\phi_i|$$

$$C_{-i}^{-1}=C^{-1}-\frac{C^{-1}\phi_i\phi_i^TC^{-1}}{\alpha_i+\phi_i^TC^{-1}\phi_i}$$

将边缘似然函数 $p(t|\alpha,\sigma^2)$ 的对数 $L(\alpha)$ 可以表示为

$$L(\alpha)=L(\alpha_{-i})+l(\alpha_i)$$

式中，$l(\alpha_i)=\frac{1}{2}\left[\log\alpha_i-\log(\alpha_i+s_i)+\frac{q_i^2}{\alpha_i+s_i}\right]$。

目标函数 $L(\alpha)$ 可分解为去除基函数 ϕ_i 后的边缘似然函

数 $L(\alpha_{-i})$ 与关于 α_i 的独立表达式 $l(\alpha_i)$。这里的 $s_i=\phi_i^T C_{-i}^{-1}\phi_i$，$q_i=\phi_i^T C_{-i}^{-1}t$。稀疏因子 s_i 用于度量基函数 ϕ_i 与模型中剩余所有基函数的重叠程度；质量因子用于度量去除基函数 ϕ_i 后对模型误差的校正。

通过分析 $l(\alpha_i)$ 表明，$L(\alpha)$ 关于 α_i 存在唯一最大值。当 $q_i^2>s_i$ 时，$\alpha_i=s_i^2/(q_i^2-s_i)$；当 $q_i^2<s_i$ 时，$\alpha_i=\infty$。通过采用这种方法，可以直接计算出所有的基函数 ϕ_i 对应的 s_i 和 q_i。

如果假设 $S_i=\phi_i^T C^{-1}\phi_i$ 和 $Q_i=\phi_i^T C^{-1}t$，则有 $s_i=\alpha_i S_i/(\alpha_i-S_i)$，$q_i=\alpha_i Q_i/(\alpha_i-S_i)$。当 $\alpha_i=\infty$ 时，$s_i=S_i$ 且 $q_i=Q_i$。

实际学习过程中，利用 Woodbury 恒等式获取 S_i 和 Q_i 非常方便[6]，有

$$S_i=\phi_i^T B\phi_i-\phi_i^T B\Phi\Sigma\Phi^T B\phi_i$$

$$Q_i=\phi_i^T B\hat{t}-\phi_i^T B\Phi\Sigma\Phi^T B\hat{t}$$

对于稀疏贝叶斯回归模型来说，$B=\sigma^2 I$，$\hat{t}=t$；对于稀疏贝叶斯分类模型来说，$B=\mathrm{diag}(\beta_1,\beta_2,\cdots,\beta_N)$，$\hat{t}=\Phi w_{MP}+B^{-1}(t-y)$。

针对回归问题，序列稀疏贝叶斯学习算法流程可以描述为[12]：

①对于回归问题，赋予 σ^{-2} 合适的初值。

②初始化 1 个基函数 ϕ_1，指定其超参数为 $\alpha_1=\|\phi_1\|^2/(\|\phi_1^T t\|^2/\|\phi_1\|^2-\sigma^2)$，其它所有超参数 α_j 为无穷大，即模型中只有基函数 ϕ_1。

③计算均值 μ 和方差 Σ，同时计算出所有基函数对应的 q_i 和 s_i。

④选择候选的基函数 ϕ_i。

⑤如果 $q_i^2>s_i$，并且 $\alpha_i<\infty$，则在模型中基函数 ϕ_i 已经存在，并更新超参数 $\alpha_i=s_i^2/(q_i^2-s_i)$。

⑥如果 $q_i^2>s_i$，并且 $\alpha_i=\infty$，则在模型中增加基函数 ϕ_i，并更新超参数 $\alpha_i=s_i^2/(q_i^2-s_i)$。

⑦如果 $q_i^2\leqslant s_i$，并且 $\alpha_i<\infty$，则在模型中删除基函数 ϕ_i，并更新 $\alpha_i=\infty$。

⑧对于回归问题，更新测量误差的方差 $\sigma^2=\|t-y\|^2/(N-M+\Sigma_i\alpha_i\Sigma_{ii})$。

⑨如果收敛，算法结束；否则，执行步骤③—步骤⑨。

结束语 RVM 是按照 SVM 模型的形式，在贝叶斯框架下提出的具有稀疏概率模型的学习机。RVM 不仅具有良好的泛化能力，可以达到与 SVM 相当的精度，还能够突破 SVM 固有的局限。

在 RVM 模型中，没有规则化系数，不需要通过交叉验证获取该参数。在 RVM 求解过程中，核函数不必满足 Mercer 条件。在 RVM 训练完成后，只有少数基函数的权值非零，比 SVM 更加稀疏。模型预测结果是用概率表示的，更够明确地分析预测结果的不确定性。

RVM 学习训练过程可以最终归结于最大边缘似然函数估计问题，自下而上的基函数选择是快速序列学习算法，较 MacKay 迭代估计和期望最大化迭代估计的计算速度更快。

参 考 文 献

[1] Shawe-Tsylor J, Cristianini N. Kernel Methods for Pattern Analysis [M]. London: Cambridge University Press, 2004: 47-82

[2] Vapnik V N. The nature of statistical learning theory [M]. Springer, 1995

[3] Tipping M E. Sparse kernel principal component analysis [M]. Advances in Neural Information Processing Systems. MIT Press, 2001

[4] Boser B E, Guyon I M, Vapnik V N. A training algorithm for optimal margin classifiers[C]// Proceedings Fifth Annual Workshop on Computational Learning Theory. 1992: 144-152

[5] Bishop C M, Tipping M E. Variational relevance vector machines[C]// Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann, 2000: 46-53

[6] Tipping M E, Faul A. Fast marginal likelihood maximization for sparse Bayesian models [C]// Proceedings Ninth International Workshop on Artificial Intelligence and Statistics. Key West, Florida, 2003

[7] Thayananthan A. Template-based Pose Estimation and Tracking of 3D Hand Motion [D]. Department of Engineering, University of Cambridge, September 2005

[8] Silva C, Ribeiro B. Scaling Text Classification with Relevance Vector Machines [C]// IEEE International Conference on Systems, Man and Cybernetics. 2006: 4186-4191

[9] Demir B, Erturk S. Hyperspectral Image Classification Using Relevance Vector Machines [J]. Geoscience and Remote Sensing Letters, IEEE, 2007: 586-590

[10] Nikolaev N, Tino P. Sequential relevance vector machine learning from time series [C]// IEEE International Joint Conference on Neural Networks. 2005: 1308-1313

[11] Tipping M E. Sparse Bayesian learning and the relevance vector machine [J]. Journal of Machine Learning Research, 2001: 211-244

[12] Bishop C M. Pattern Recognition and Machine Learning [M]. Springer, 2007

[13] MacKay D J C. The evidence framework applied to classification networks[J]. Neural Computation, 1992: 720-736

[14] Thayananthan A. Relevance Vector Machine based Mixture of Experts [R]. Department of Engineering, University of Cambridge, 2005

(上接第 164 页)

[4] Bertino E, Samarati P, Jajodia S. An Extended Authorization Model for Relational Databases [J]. IEEE Transactions on Knowledge and Data Engineering, 1997, 9(1): 85-101

[5] Sandhu R S, Coyne E J, Feinsein H L, et al. Role-based Access Control Models[J]. IEEE Computer, 1996, 29(2): 38-47

[6] Ferraiolo D F, Sandhu R, Gavrila S. Proposed NIST Standard for Role-based Access Control[J]. ACM Transactions on Information and System Security, 2001, 4(3): 224-274

[7] Ferraiolo D, Kuhn D R, Chandramouli R. Role-based Access Control[M]. Artech House, Computer Security Series, 2003

[8] Bertino E, Bettini C. An Access Control Model Supporting Periodicity Constraints and Temporal Reasoning[J]. ACM Transactions on Database Systems, 1998, 23(3): 231-285

[9] Bertino E, Bettini C, Ferrari E, et al. A Temporal Access Control Mechanism for Database Systems [J]. IEEE Transactions on Knowledge and Data Engineering, 1996, 8(1): 67-80

[10] Oracle. The Virtual Private Database in Oracle9iR2[EB/OL]. http://otn.oracle.com/deploy security /oracle9iR2/Pdf/ VPD9ir2twp.pdf, 2000

[11] Biba K J. Integrity Considerations for Secure Computer Systems [R]. ESD-TR-76-372. Bedford, Massachusetts: USAF Electronic Systems Division, Hanscom Air Force Base, 1977