Advances in Applied Mathematics 应用数学进展, 2021, 10(12), 4535-4541
Published Online December 2021 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2021.1012483

Hans汉斯

SVM分类理论与算法的若干新进展

岳朝霞[1]，刘　甲[2]

[1]内蒙古大学数学科学学院，内蒙古　呼和浩特
[2]辽宁师范大学数学学院，辽宁　大连

收稿日期：2021年11月27日；录用日期：2021年12月17日；发布日期：2021年12月31日

摘　要

支持向量机(Support Vector Machine, SVM)一直是处理二分类问题的重要工具，被广泛应用于机器学习、图像处理、生物信息等众多领域。自从1995年Vapnik和Cortes提出以来，经过多年的发展，国内外产生了丰富的研究成果。为了进一步拓宽有关SVM问题的研究，本文对近年来SVM的发展从模型和算法两方面进行梳理。

关键词

支持向量机，二分类，模型，算法

Some Advances in Theory and Algorithm for SVM

Zhaoxia Yue[1], Jia Liu[2]

[1]School of Mathematics, Inner Mongolia University, Huhhot Inner Mongolia
[2]School of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: Nov. 27th, 2021; accepted: Dec. 17th, 2021; published: Dec. 31st, 2021

Abstract

Support Vector Machine has always been an important tool to deal with binary classification problems, and is used in many fields widely, such as machine learning, image processing, and biological information. Since the proposal of Vapnik and Cortes in 1995, after years of development, rich research results have been produced at home and abroad. In order to further broaden the research on SVM issues, this paper sorts out the development of SVM in recent years from two aspects: model and algorithm.

文章引用: 岳朝霞, 刘甲. SVM 分类理论与算法的若干新进展[J]. 应用数学进展, 2021, 10(12): 4535-4541.
DOI: 10.12677/aam.2021.1012483

Keywords

SVM, Binary Classification, Model, Algorithm

1. 引言

支持向量机(Support Vector Machines, SVM)是一种以统计学理论以及结构风险最小化的原理为基础的机器学习方法。该方法在解决样本分类问题、线性、非线性以及高维模式识别问题中有许多优势。此外，支持向量机也能有效地处理回归问题，如时间序列分析问题等。

支持向量机(Support Vector Machine)的原理是寻找一个满足分类要求的最优分类超平面，使得该超平面不仅能够保证分类的准确性，还可以使超平面两侧的空白区域最大化。该原理用数学语言表述即通过训练大量的数据来找到一个最优超平面 $\langle w,x\rangle+b=w_1x_1+\cdots+w_nx_n+b=0$，其中 $w\in R^n$ 和 $b\in R$ 通过训练集来估计。使得对于任意新输入的向量 x，如果 $\langle w,x\rangle+b>0$，则可以预测 x 的标签 $y=1$；若 $\langle w,x\rangle+b<0$，则 $y=-1$。从而达到分类的目的。

自从 1995 年 Vapnik 和 Cortes 提出以来[1]，支持向量机一直是国内外学者研究的热点，在机器学习、图像处理、生物信息等众多领域只要是二分类问题大多都会想到支持向量机。由于其适用范围广的特点，多数涉及到分类的问题，研究者都会用支持向量机的分类方法来做模型评估或算法效果对比。如 Wang 等为了说明所提出的方法更有利于疾病的筛选，把每一种方法都与 SVM 分类模型[2]进行了对比。诸如此类的例子数不胜数，涉及到多个研究领域，在此就不进行一一列举。足以可见，它的应用之广以及它的重要程度。

在 SVM 蓬勃发展的近二十年的时间里，大量优秀的成果不断涌现，但是作者发现近年来很少有从数学角度出发来分析模型与算法两方面的综述类文章，为此激发了写这篇文章的想法。考虑到此类文章星罗棋布、文章篇幅限制以及作者的个人能力有限，本文仅回顾一些经典研究成果，并着重从数学角度出发，分析函数的性质、模型的特点等。

2. 支持向量机基本思想概述

给定一个训练集 $\{(x_i,y_i):i=1,2,\cdots,m\}$，$x_i\in R^n$ 是输入向量，$y_i\in\{-1,1\}$ 是输出的标签。SVM 的目标则是训练一个超平面 $\langle w,x\rangle+b=w_1x_1+\cdots+w_nx_n+b=0$，通过训练集来估计 $w\in R^n$ 和 $b\in R$。对于任意新输入的向量 x，如果 $\langle w,x\rangle+b>0$，则可以预测 x 的标签 $y=1$；若 $\langle w,x\rangle+b<0$，则 $y=-1$；可以计算出分类间隔为 $2/\|w\|$。那么，为了找到最优超平面，有两种可能的情形。第一种，如果训练数据是线性可分的，如图 1。

那么直接求解以下凸二次规划就可以找到唯一的最优超平面

$$\min_{w\in R^n,b\in R}\frac{1}{2}\|w\|^2$$

$$\text{s.t}\quad y_i\left(\langle w,x_i\rangle+b\right)\geq 1,\ i\in N_m$$

$$N_m:=\{1,2,\cdots,m\}$$

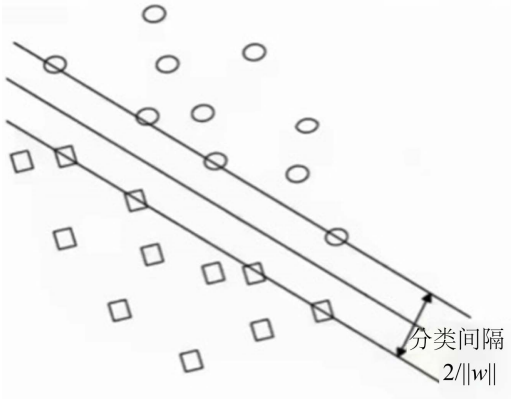

Figure 1. Data is linearly separable
图 1. 数据线性可分

关于该二次规划的求解，这里以二维空间为例，不失一般性，也可推广到其它维度。上述模型本身是一个凸二次规划问题，可以使用现有的优化计算包来计算，但我们这里选择更方便的方法，即拉格朗日乘子法，首先可得它的拉格朗日函数为

$$L(w,b,\alpha)=\frac{1}{2}\|w\|^2+\sum_{i=1}^{m}\alpha_i\left(1-y_i\left(w^{\mathrm{T}}x_i+b\right)\right)$$

其中，α 为拉格朗日乘子。

接下来，分别对 w 和 b 求偏导数，令所求得的偏导数等于零，可得

$$\begin{cases} w=\sum_{i=1}^{m}\partial_i y_i x_i \\ \sum_{i=1}^{m}\partial_i y_i=0 \end{cases}$$

把该结果代回上述模型就可以得到一个关于 α 的问题，且只含 α 一个未知变量，那么可以很容易解出 α，再代入 $w=\sum_{i=1}^{m}\partial_i y_i x_i$ 中可以求出 w，最后求出 b。最后可以得到

$$f(x)=w^{\mathrm{T}}x+b=\sum_{i=1}^{m}\partial_i y_i x_i^{\mathrm{T}}x+b$$

这里我们写出以上整个过程的 KKT 条件

$$\begin{cases} \alpha_i\geq 0 \\ y_i f(x_i)-1\geq 0 \\ \alpha_i\left(y_i f(x_i)-1\right)=0 \end{cases}$$

可以看到对于任意训练样本(x_i,y_i)，若 α_i 等于零，它不在求和项中出现，也就不会影响模型训练结果；若 α_i 大于零，该样本在边界上，是一个支持向量。所以，当支持向量机训练完成后，最终模型只与支持向量有关。

第二种，如果训练数据是非线性可分的，相应的，我们也需要用非线性模型来进行分类。

若数据类似图 2 中所示，明显可以看到不能通过直线来进行分割，但可以用一条曲线，也即非线性模型进行分割。而非线性模型往往不好求解，所以还是转化为线性模型求解。转化的方法就是将训练样本从原始空间映射到一个更高维的空间，使得样本在这个高维空间线性可分。即把向量 x 映射成特征向

量$\varphi(x)$，然后再根据第一种线性可分的求解方法—拉格朗日法进行求解即可。如图 3。

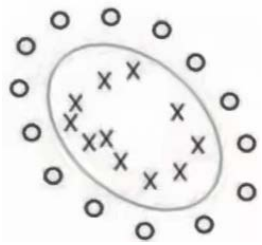

Figure 2. Data is non-linearly separable
图 2. 数据非线性可分

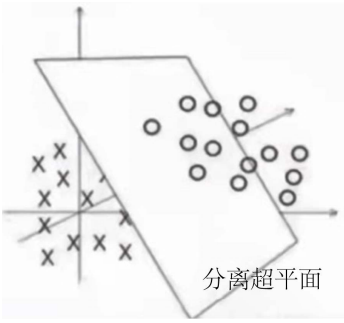

Figure 3. Data is mapped to high-dimensional space separable
图 3. 数据映射到高维空间可分

求解后可得到

$$f(x)=w^{\mathrm{T}}\varphi(x)+b=\sum_{i=1}^{m}\partial_i y_i k(x_i,x_j)+b$$

这里的函数$k(x_i,x_j)$为核函数。

以上就是支持向量机的基本思想，但在现实中，往往很难确定合适的核函数使训练数据线性可分，退一步讲，即使找到了这样的函数也很难判断是不是由于过拟合造成的。为了寻求一个更好的解决办法，国内外的学者做了很多的努力，在模型和算法上都取得了很大的进展，本文从这两方面分别进行梳理。

3. 模型

支持向量机的模型分为硬间隔和软间隔两种，由于现实中的应用大都是软间隔，所以在这里本文重点说软间隔的模型发展，硬间隔的模型在上述数据线性可分中已经给出，这里不再赘述。首先给出软间隔的基本模型

$$\min_{w\in R^n,b\in R}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{m}l\left(1-y_i\langle w,x_i\rangle+b\right)$$

其中l为损失函数，用来定义训练样本与真实值之间的误差；$C>0$为罚参数，用来惩罚那些误差大的样本。因此，软间隔的模型允许有分类误差。可以看出该模型有两项组成，第一项是正则项，用来衡量模型复杂度，防止过拟合；第二项是损失项用来衡量模型误差。所以我们总的目的是使两项都尽可能小，正则项小可以用很少的支持向量确定超平面，降低成本；损失项小可以使分类误差更小，准确率更高。

纵观支持向量机的发展，学者们都是在不断的改进损失项，也就是通过不断地引入新的损失函数来产生新的模型。这些新的损失函数大致分为凸函数和非凸函数两类。由于凸函数求解相对容易，所以早期的研究成果损失项主要集中在凸函数上。近年来，由于零范数和稀疏集的成果不断发展，尤其是变分性质的出现，如次微分，近端算子，切锥与法锥等为学者们提供了解决非凸非光滑问题的工具，所以近年的研究成果损失项以非凸函数为主。接下来本文就从凸和非凸两类函数对一些经典研究成果分别展开。

3.1. 损失函数为凸函数

1995 年 Vapnik 和 Cortes 首次提出 Hinge 损失函数，即 $l_{hinge}(t)=\max\{0,t\},\forall t\in R$，该函数只在 $t=0$ 这一点不可微，也是首次提出支持向量机的问题。提出该函数的目的是为了惩罚 $t\geq 0$ 的样本。考虑到该函数只能惩罚部分样本，2013 年 V. Jumutc 和 X. Huang 等人给出了 Pinball 损失函数[3]，其具体形式为 $l^{\tau}_{pinball}(t)=\max\{t,-\tau t\},0\leq\tau\leq 1$，这个函数的不可微点也只在零点，但与上一个函数不同的是这个函数可以惩罚所有样本，利用该函数实现了 Pegasos 算法新的优化目标。2008 年 L. Wang 和 J. Zhu 等人首次把 $l^{\tau}_{HH}(t)=\max\{0,t-\tau\}-\left(\max\{0,\tau/2-t^2/2\tau\}-t^2/2\right)$，$\tau>0$ 应用在 SVM 的分类当中[4]，该函数的特点是处处可微。利用该函数实现了基因自动选择以及分组效应—高度相关的基因往往被一起选择或删除。还有我们比较常见的平方损失函数 $l_{square}(t)=t^2$，但我们知道最小二乘支持向量机对数据中的异常值和噪声极为敏感，为克服这一问题后又引入了加权最小二乘支持向量机，然而为训练数据设置权重是一项工作量很大的任务，并且会极大的影响鲁棒性。因此在 2014 年 X. Yang 和 L. Tan 为了避免设置权重提出了一种基于截断最小二乘损失的新型鲁棒支持向量机，并借助其等效模型从理论上分析了鲁棒性高的原因[5]。2000 年 J. Friedman 和 T. Hastie 提出了逻辑回归损失函数[6]，由于其凸且光滑的性质，这一函数也成为了一个很经典的损失函数，后来还有交叉熵等函数都与之类似，应用的也非常多。

由于以上函数都具有凸性，所以对应的 SVM 模型解决起来也相对容易，但是，我们知道凸性往往会伴随着无界性，比如上述的这些函数可以发现几乎都是无界函数，而无界性常常会导致这些损失函数对数据中的异常值失去鲁棒性。为了克服这一问题，2003 年 F. Perez-Cruz 在文章[7]中对损失函数设置一个上界进行强制执行，使其在一定程度后停止增加，那么，凸函数就变成了非凸，从而解决了这个问题。

以上提到的就是一些常用的凸损失函数，经过多年的发展，不断克服新的问题，损失项为凸函数的理论已经趋于成熟，关于此类函数也有大量的成果，这里考虑到篇幅的限制，本文就列举到此。近年来，学者们更多的研究重心放在非凸函数方面，下面我们从非凸的损失函数来进行总结。

3.2. 损失函数为非凸函数

同样，由于已经研究了大量的非凸软间隔损失，在此只列举部分。

首先是 Ramp 损失函数，$l^{\mu}_{ramp}(t)=\max\{0,t\}-\max\{0,t-\mu\}$ 其中 $\mu\geq 0$。该函数在 $t=0$ 和 $t=\mu$ 这两个点都是不可微的，但在 0 到 μ 之间是有界的。这个函数的巧妙之处在于它不会去惩罚那些 $t<0$ 的数据样本，对于 $t>0$ 的那些样本又会分成两种情况，当 $0\leq t\leq\mu$ 时，数据样本会得到线性惩罚，而当 $t>\mu$ 时又会得到另一种特定的惩罚。正是因为这些巧妙的性质使得这个函数对一些异常值或噪声具有鲁棒性。2014 年 X. Huang 等人将该函数与支持向量机相结合，提出了一种通用的支持向量分类器[8]。为了克服传统模型过于简单而不能用于复杂的排序，很难将先验知识加入到模型中的困难，一种新的排序支持向量机被提出[9]，它利用 sigmoid 损失函数，其形式为 $l_{sigmoid}(t)=1/\left(1+\exp(-t)\right)$，在 0 到 1 之间光滑且有界，与上一个不同的是，它会惩罚所有样本。该函数用交叉检验进行训练，实际上该函数就是输入数据集的后验概率，可以用于处理一些后续工作。文章[10]中利用该函数得到了一个性能与原分类器相同的缩减支持向量分类器。2014 年 X. Shen 等发表的文章[11]中把 Truncated pinball 损失函数

$l_{Tpin}^{\tau,k}(t)=\max\{0,(1+\tau)t\}-(\max\{0,\tau(t+k)\}-\tau k)$，$0\le\tau\le1$和$k\ge0$与支持向量机相结合，在$t=0$和$t=-k$两点不可导且无界。对于固定的$k$，当$t<-k$时惩罚是固定的，除此之外惩罚都是线性的。最近，修教授团队提出了0/1损失函数，即当$t>0$时，$l_{0/1}(t)=1$，$t<0$时，$l_{0/1}(t)=0$。该函数在$t=0$不连续，由于它的取值是0或1，所以稀疏性和鲁棒性得到了很好的保证。学者们不仅推动了支持向量机的发展，同时还利用非凸非光滑的损失函数、次微分及近端算子的性质给出了优化模型的KKT条件、P-稳定点条件，这在稀疏优化领域也取得了很大的突破。

上面就是近年一些经典的研究成果，当然，还有很多优秀的文章值得我们去学习，在此，由于作者个人能力有限，就列举这么多，读者可根据自身兴趣深入挖掘学习。接下来从算法的角度进行陈述。

4. 算法

提到支持向量机的算法求解，常规的思路是求解一个有约束的二次规划问题，如果该二次规划问题的规模比较小，我们可以利用现有的优化算法，如牛顿法、内点法等一些比较成熟且经典的方法来进行求解，但随着大数据时代的到来，问题的规模日渐庞大，此时传统的优化方法捉襟见肘，往往会出现速度慢、效率低等问题，于是国内外的学者进行了多种针对目前形势的算法研究。下面从计算角度和数学优化角度对一些算法进行列举。

早期的算法主要集中在计算方面，文献[12]在1999年最早提出了SVM增量训练算法，即每次只选一小批常规二次算法能处理的数据作为增量，保留原样本中的支持向量和新增样本混合训练，直到训练样本用完。此类算法是机器学习系统在处理新增样本时，能够只对原学习结果中与新样本有关的部分进行增加修改或删除操作，与之无关的部分则不被改变。此算法的一个突出特点是支持向量机的学习不是一次离线进行的，而是一个数据逐一加入反复优化的过程。文章[13]中最早提出了分解算法，其原理是将一个大型的二次规划问题分解成一系列小的二次规划子问题进行迭代求解，在每次迭代中，使用拉格朗日乘子分量的一个子集作为训练集，用传统优化算法求解一个二次规划的子问题，该算法主要适用于求解大规模问题。值得注意的是文献[14]提出的SMO算法是分解算法的一个特例，它的训练集中只含有2个样本，但是对于两个样本的二次规划问题能够有解析解，较适合样本稀疏的问题。1992年Boser B E等人提出Chunking算法[15]，它是通过删除矩阵中对应Lagrange乘数为零的行和列来进行，对于给定的样本，该算法的目标是运用某种迭代方式逐步排除非支持向量，这样做的好处在于可以很大程度上降低训练过程对存储器容量的要求。2013年V. Jumutc和X. Huang等人提出Pegasos算法利用Pinball损失的特性和强凸优化问题的理论实现了较快的收敛速度和较低的计算和内存成本[3]。

近年来，随着优化理论的发展，诸多学者从优化角度对算法进行设计改进，把SVM的算法设计与稀疏优化相结合，从而提高计算精度。2015年Bai等人针对稀疏支持向量机设计了l_1-l_2正则稀疏法[16]，此算法不仅有更高的分类准确率而且有更稀疏的分类器。2019年Shao等人将特征选择和支持向量机相结合，首次在分类的同时进行特征选择，为解决结合后的模型，提出了缩放交替方向乘数的算法[17]，因为对样本和特征同时选择，所以该算法的测试速度更快，泛化能力也更强。

以上就是本文列举的一些经典算法，整体来看，支持向量机的算法目标都是朝着提高计算精度、降低存储容量以及算法复杂度的方向来进行，这也是今后我们要持续努力的方向。

5. 总结

本文从数学角度综述了支持向量机的模型与算法方面的成果，主要介绍了一些经典的文章。经过多年的发展，该领域积累了大量的研究成果，但还有大量的问题需要去探讨，比如在处理大规模数据的问题时，还是会面临计算速度等问题，如何更精准，更稀疏，与多种领域深入结合，都是将来我们需要努

力的方向。希望本文能够为读者带来一些启发，推动该领域的发展。

参考文献

[1] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297.

[2] Wang, H.F., Zheng, B.C., Yoon, S.W., *et al*. (2018) A Support Vector Machine-Based Ensemble Algorithm for Breast Cancer Diagnosis. *European Journal of Operational Research*, **267**, 687-699. https://doi.org/10.1016/j.ejor.2017.12.001

[3] Jumutc, V., Huang, X. and Suykens, A.J. (2013) Fixed-Size Pegasos for Hinge and Pinball Loss SVM. *The* 2013 *International Joint Conference on Neural Networks*, Dallas, 4-9 August 2013, 1-7. https://doi.org/10.1109/IJCNN.2013.6706864

[4] Wang, L., Zhu, J. and Zou, H. (2008) Hybrid Huberized Support Vector Machines for Microarray Classification and Gene Selection. *Bioinformatics*, **24**, 412-419.

[5] Yang, X.W., Tan, L.J. and He, L.F. (2014) A Robust Least Squares Support Vector Machine for Regression and Classification with Noise. *Neurocomputing*, **140**, 41-52. https://doi.org/10.1016/j.neucom.2014.03.037

[6] Friedman, J., Hastie, T. and Tibshirani, R. (2000) Additive Logistic Regression: A Statistical View of Boosting (with Discussion and a Rejoinder by the Authors). *The Annals of Statistics*, **28**, 337-407. https://doi.org/10.1214/aos/1016218223

[7] Pérez-Cruz, F., Navia-Vázquez, A., Figueiras-Vidal, A.R., *et al*. (2003) Empirical Risk Minimization for Support Vector Classifiers. *IEEE Transactions on Neural Networks*, **14**, 296-303. https://doi.org/10.1109/TNN.2003.809399

[8] Huang, X., Shi, L. and Suykens, J.A.K. (2014) Ramp Loss Linear Programming Support Vector Machine. *The Journal of Machine Learning Research*, **15**, 2185-2211.

[9] Thuy, N.T.T., Vien, N.A., Viet, N.H., *et al*. (2009) Probabilistic Ranking Support Vector Machine. In: Yu, W., He, H. and Zhang, N., Eds., *International Symposium on Neural Networks*, Springer, Berlin, 345-353. https://doi.org/10.1007/978-3-642-01510-6_40

[10] Pérez-Cruz, F., Navia-Vazquez, A., Alarcón-Diana, P.L., *et al*. (2000) Support Vector Classifier with Hyperbolic Tangent Penalty Function. 2000 *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, 5-9 June 2000, 3458-3461. https://doi.org/10.1109/ICASSP.2000.860145

[11] Shen, X., Niu, L.F., Qi, Z.Q., *et al*. (2017) Support Vector Machine Classifier with Truncated Pinball Loss. *Pattern Recognition*, **68**, 199-210. https://doi.org/10.1016/j.patcog.2017.03.011

[12] Syed, N.A., Liu, H. and Sung, K.K. (1999) Handling Concept Drifts in Incremental Learning with Support Vector Machines. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, 15-18 August 1999, 317-321. https://doi.org/10.1145/312129.312267

[13] Osuna, E., Freund, R. and Girosi, F. (1997) An Improved Training Algorithm for Support Vector Machines. *Proceedings of the* 1997 *IEEE Signal Processing Society Workshop*, Amelia Island, 24-26 September 1997, 276-285. https://doi.org/10.1109/NNSP.1997.622408

[14] Hearst, M.A., Dumais, S.T., Osuna, E., *et al*. (1998) Support Vector Machines. *IEEE Intelligent Systems and Their Applications*, **13**, 18-28. https://doi.org/10.1109/5254.708428

[15] Boser, B.E., Guyon, I.M. and Vapnik, V.N. (1992) A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, 27-29 July 1992, 144-152. https://doi.org/10.1145/130385.130401

[16] Bai, Y.Q., Zhu, Z.Y. and Yan, W.L. (2015) Sparse Proximal Support Vector Machine with a Specialized Interior-Point Method. *Journal of the Operations Research Society of China*, **3**, 1-15.

[17] Shao, Y.H., Li, C.N., Huang, L.W., *et al*. (2019) Joint Sample and Feature Selection via Sparse Primal and Dual LSSVM. *Knowledge-Based Systems*, **185**, Article ID: 104915. https://doi.org/10.1016/j.knosys.2019.104915