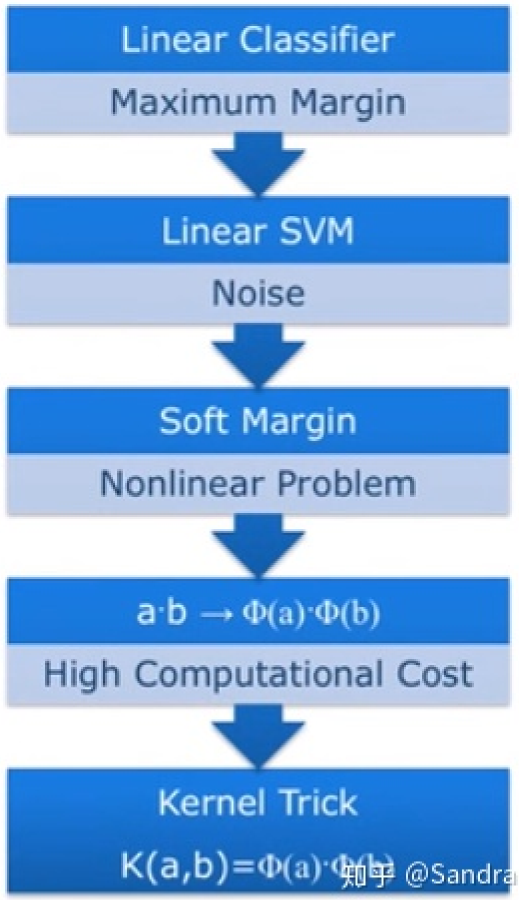
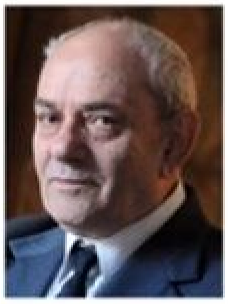
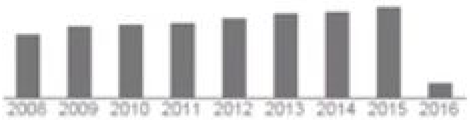
最初是线性分类器（Margin最大化）--> 线性分类器 --> Soft Margin（噪音）--> 高维映射使其线性可分（线性不可分） --> 提出Kernel（不去求高维的a,b，转为核函数）



发明人：俄国数学家， 万普尼克



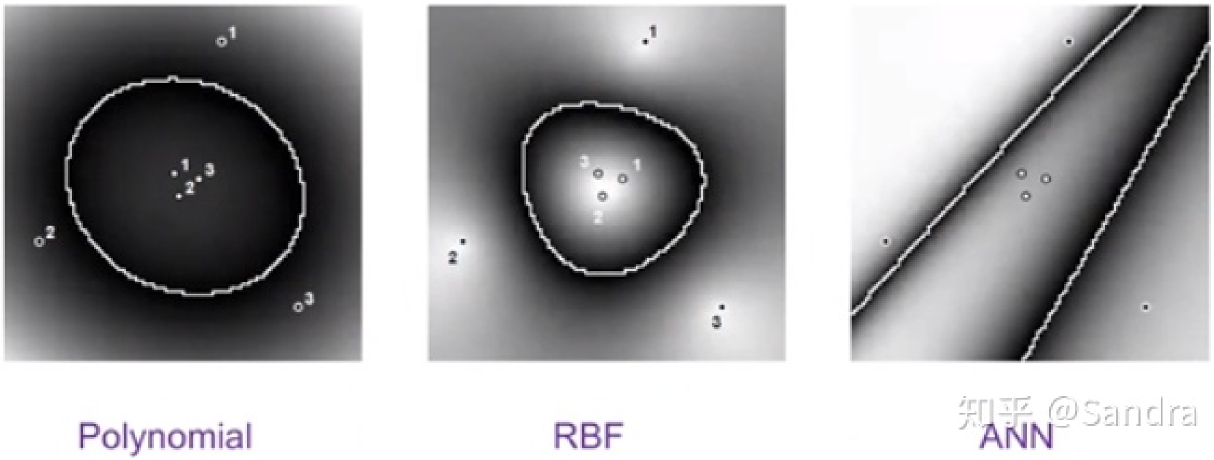


引用指数	总计	2011 年至今
引用	164864	74389
h 指数	109	72
i10 指数	371	270

A training algorithm for optimal margin classifiers

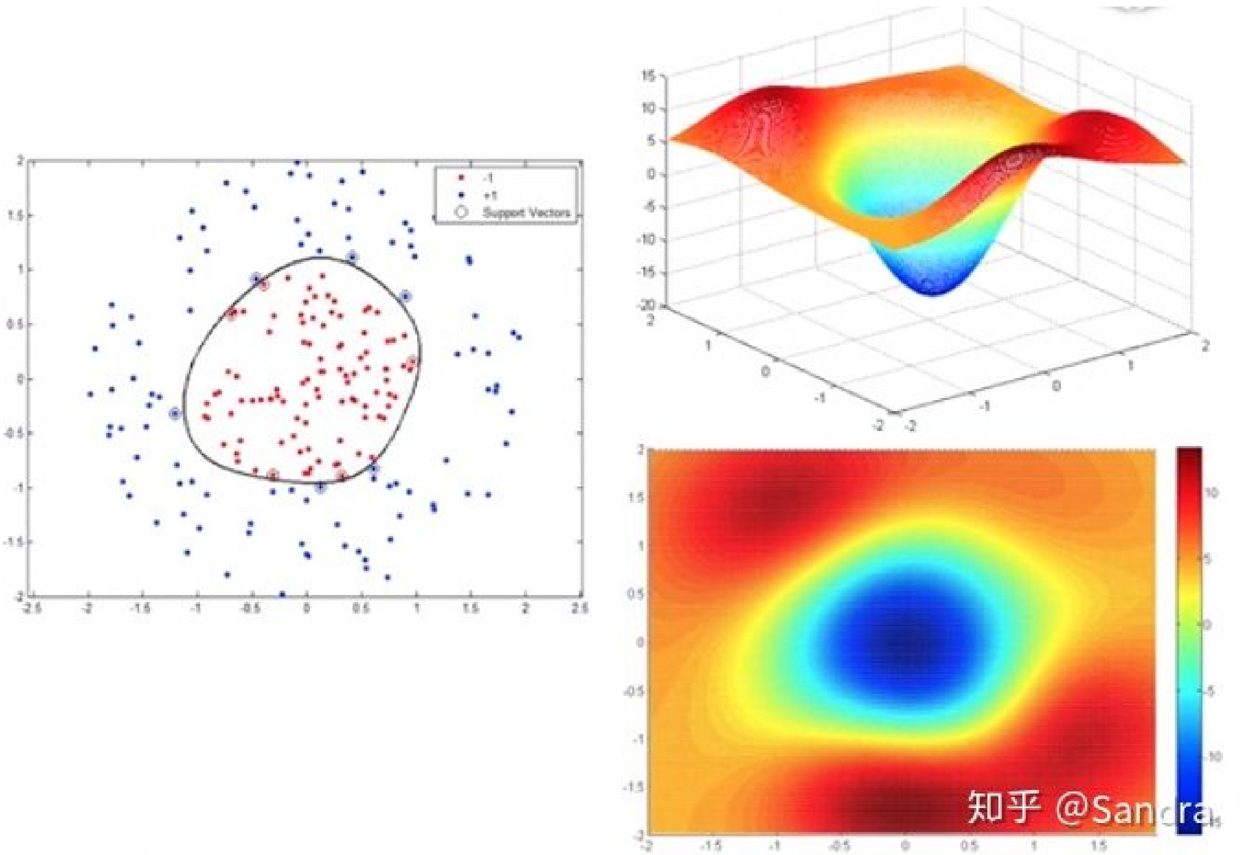
作者	Bernhard E Boser, Isabelle M Guyon, Vladimir N Vapnik
发表日期	1992/7/1
研讨会论文	Proceedings of the fifth annual workshop on Computational learning theory
页码范围	144-152
出版商	ACM
简介	Abstract A training algorithm that maximizes the margin between the training patterns and the decision boundary is presented. The technique is applicable to a wide variety of the classification functions, including Perceptrons, polynomials, and Radial Basis Functions. The effective number of parameters is adjusted automatically to match the complexity of the problem. The solution is expressed as a linear combination of supporting patterns. These are the subset of training patterns that are closest to the decision boundary. Bounds on the ...
引用总数	被引用次数：7251

另一个重要贡献是机器学习中，模型复杂度风险方面的：提出VC Dimension



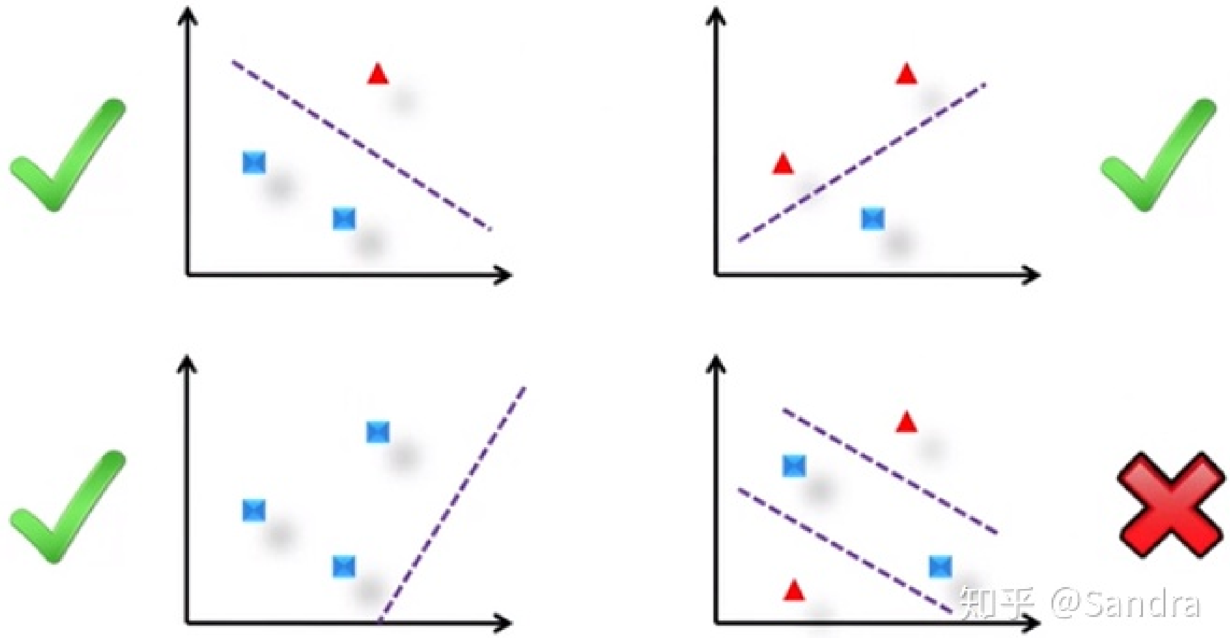
文献中的截图

线圈其实是一个等高面（Decision boundary），在根据Support Vectors计算出g(x)以后，令其等于0，将其取出

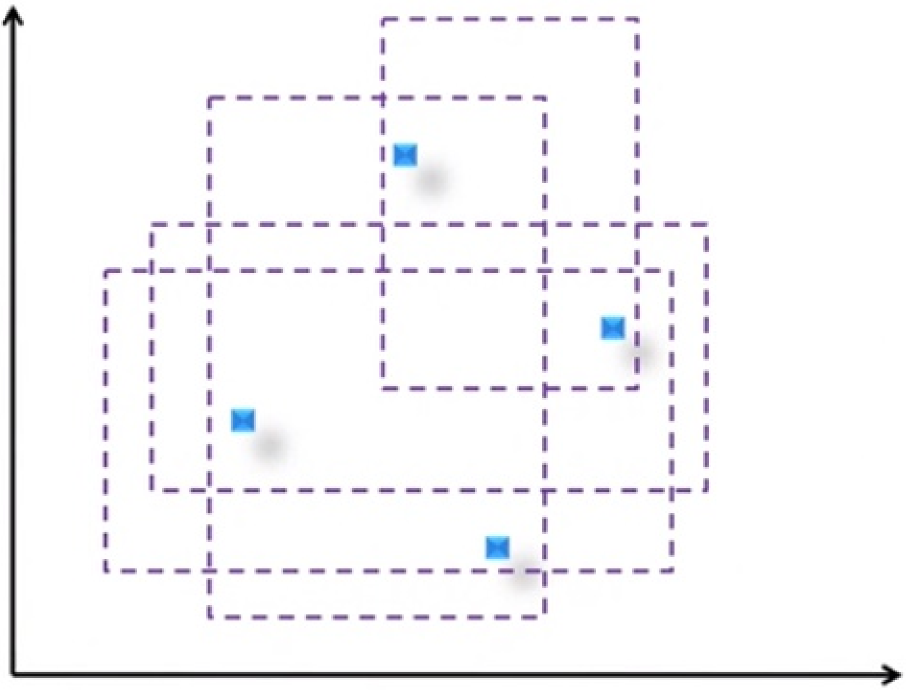


VC Dimension：

Model Capacity：



无论三个点怎样，这一条线都可以把他们分开；四个点一条线就无法分开



A rectangle can classify (shatter) a certain set of 4 points, regardless of their labels.

四个点无论怎样都可以用矩形框分开

- The VC dimension of a model M is *h* if there *exists* a set of (up to) *h* points that can be *shattered* by M.
- The *h* value of a hyperplane in $\mathbf{R}^n$ is n+1.
- The *h* value of DT is roughly the number of internal nodes.
- The *h* value of SVM depends on the kernel function in use.
- VC dimension is pessimistic: arbitrary assignment of labels.
- Real data sets: points with same labels tend to be close to each other.

存在H个点，这几个点无论怎么分布（打标签），这个模型M都能将其分开，那么就可以说这个模型的VC Dimension=H

超平面，在N维空间中VC Dimension就是N+1

决策树的中间节点部分决定了他的VC Dimension，数越高，VC Dimension越多

数据挖掘的规律性！通常都假设数据是有分布性的，不会任意打标签。但VC Dimension是一个保守的估计，是随意分布的。但实际不会这么复杂。

$$P\left(E_{test} < E_{train} + \sqrt{\frac{h\left(\log\left(2N / h\right) + 1\right) - \log\left(\eta / 4\right)}{N}}\right) = 1 - \eta$$

N: Number of training samples

之前一直在讲模型的训练误差，万普尼克提出：训练误差加上一个数，得到测试误差不超过的范围（bond）。N是训练样本个数（N远远大于h），

η

是置信度，其等于0.1的时候说明有90%的confidence我的测试误差不超过这个范围。h增加（模型VC Dimension增加），它的bond会增加，在实际应用中效果比较差的概率就会增加，风险更高。所以为什么当两个决策树都能解决同一个问题时（达到相同的训练效果），要用那个简单的。

补充阅读材料：

- N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, 2000.
- C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition”. *Data Mining and Knowledge Discovery*, vol. 2, pp. 121-167, 1998.
- H. Lodhi et al., “Text Classification Using String Kernels”. *The Journal of Machine Learning Research*, vol. 2, pp. 419-444, 2002.
- Online Resources
  - http://www.kernel-machines.org/
  - http://www.support-vector-machines.org/
  - http://www.tristanfletcher.co.uk/SVM%20Explained.pdf
  - http://www.csie.ntu.edu.tw/~cjlin/libsvm/