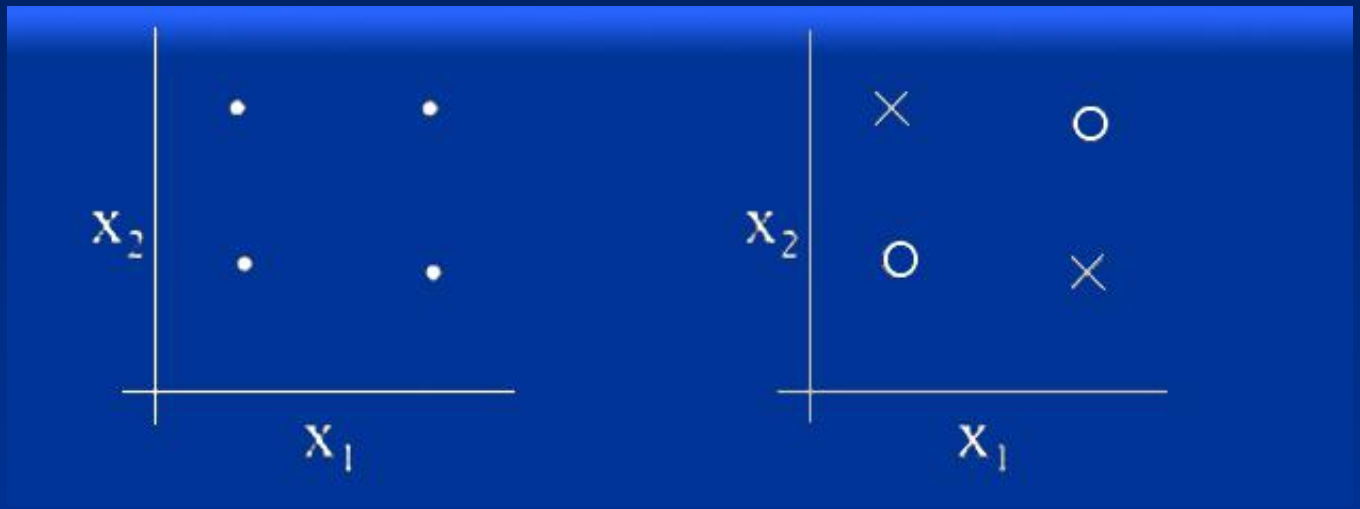
第2-1节　支持向量机基础

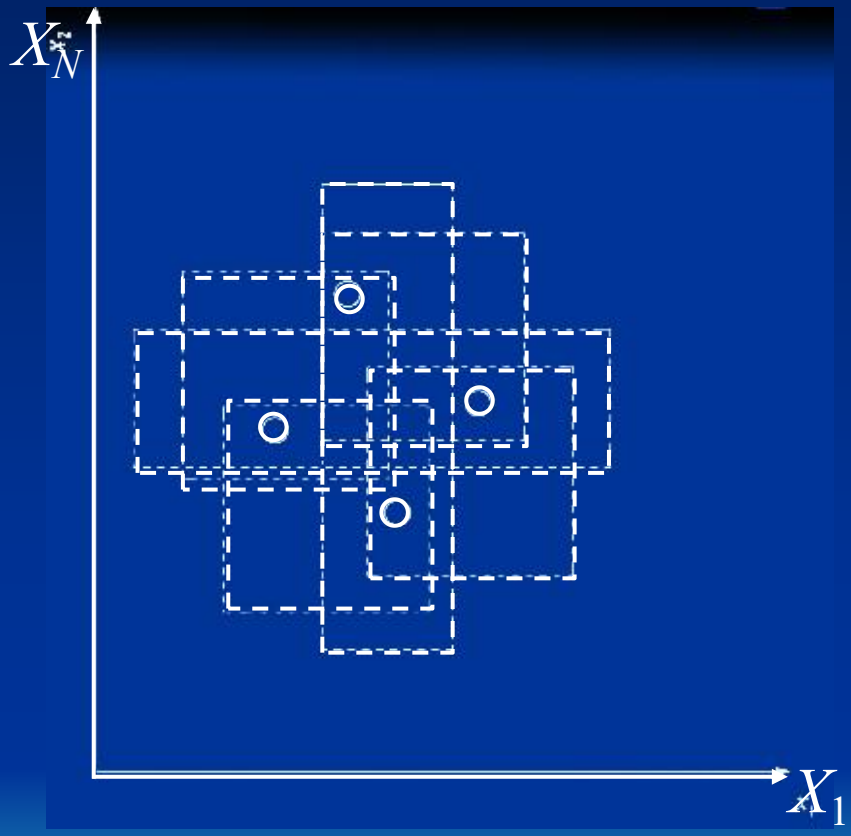
—— 结构风险最小化

1、基于VC维的分类器选择



- 无论4个点在一条直线上还是不在一条直线上（任意位置），直线分类器都不能进行$2^4=16$种类型的分类，因此VC维=3的直线分类器不能满足4个样本点的分类要求 。

轴平行的矩形能够打散二维空间的 4 个点：



- 轴平行的矩形分类器的VC维等于4。

VC维在支持向量机学习中的作用：

- 在小样本（样本数量少）数据的学习中，出现过拟合问题。因为，对于小样本，显然找到完全拟合样本的函数。

  为了解决小样本数据的过拟合问题，需要估计具有最小维（VC维）的分类器，以便正确分类。

推广性的界： $R(\omega) \leq Remp(\omega) + \Phi(n/h)$

当n/h值较小，即VC维h较大时，置信范围较大，解的推广性较差；n/h值较大，即VC维h较小时，置信范围较小，取得的解接近最优解。

推广性的界是针对最坏情况的，该条件在很多情况下是很松的，当VC维h较高时尤其如此。

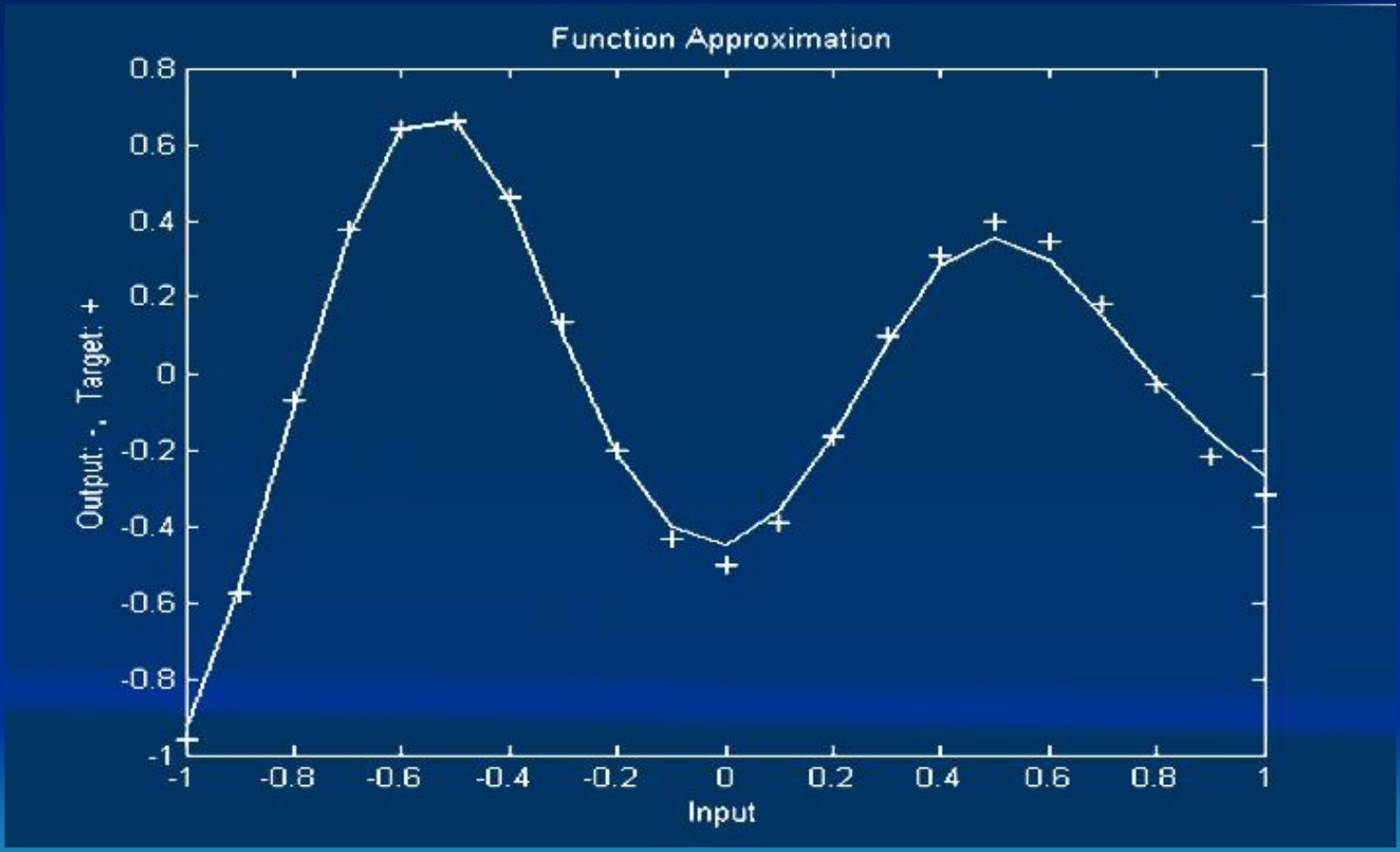
一般在样本较少的情况下，得到小的置信的范围采用线性分类器，原因是线性分类器的VC维较低。

过学习

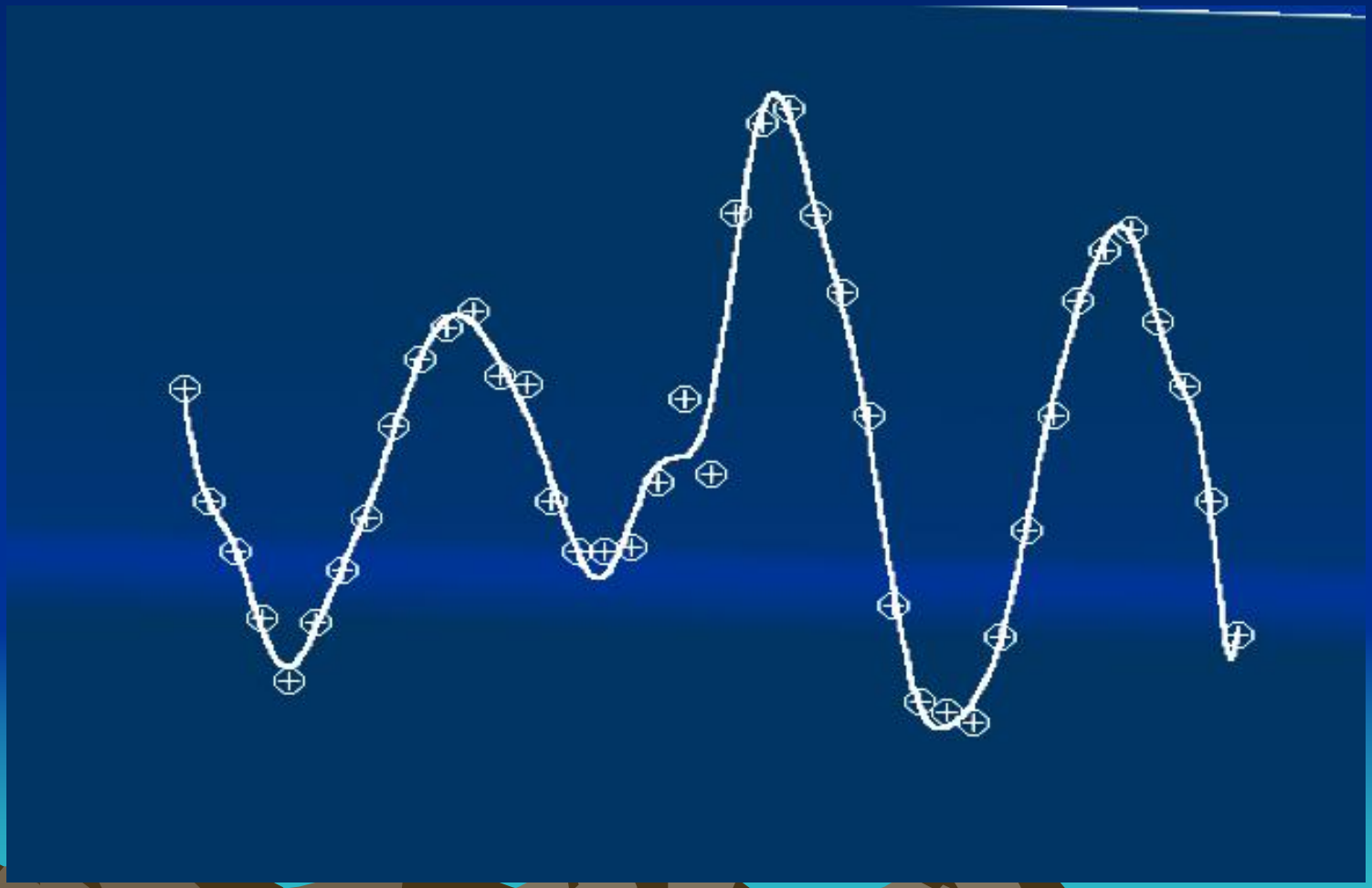
一般的学习方法(如神经网络)是基于 $Remp(\omega)$ 最小，满足对已有训练数据的最佳拟和。理论上可以通过增加算法(如神经网络)的规模使得$Remp(\omega)$ 不断降低至0。

但是，这样使得算法(神经网络)的复杂度增加，VC维h增加，从而$\Phi(n/h)$增大，致实际风险 $R(\omega)$ 增加，这就是学习算法的过拟合(Overfitting)。

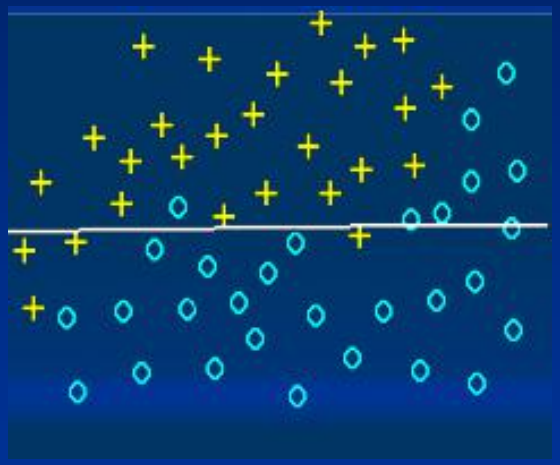
可以用三角函数拟合任意点：



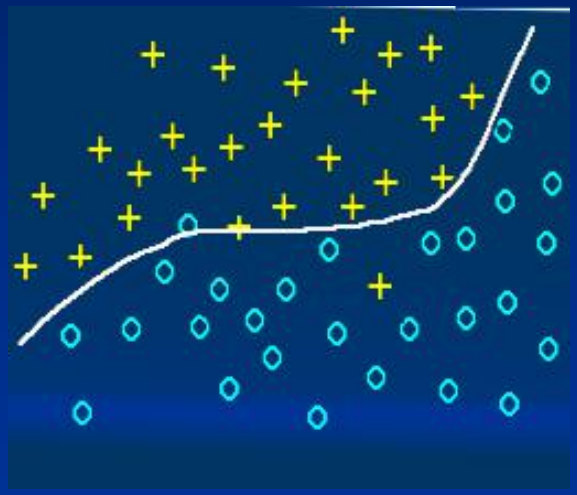
另一个过拟合实例：



其它过学习例子：



underfitting



good fit

overfitting

说明：经验风险 $Remp(\omega)$ 值小，并不代表真实风险 $R(\omega)$ 值也小。

学习机器的泛化能力

- 学习机器对未来输出进行正确预测的能力称作推广能力（也称为“泛化能力”）。
- 在某些情况下，训练误差过小反而导致推广能力的下降，这就是过学习问题。
- 神经网络的过学习问题是经验风险最小化原则失败的一个典型例子。

- 因此，选用过于复杂的分类器或神经网络，往往得不到好的学习效果。

例如：用 sin*x* 函数拟合任意点的例子，因为 sin*x* 函数的VC维为无穷大，因此虽然经验风险达到了0，但小样本数据学习的真实风险却很大，不具有较好的推广能力。

- 指示函数可表示为：

$$f(x,\omega)=\Phi(\sin(\omega x))$$

它的VC维无穷大，对于直线上的下列点：

$$x_1=10^{-1},\ \ldots,\ x_l=10^{-l}$$

把上述数据分为由标签(取0或1的值)序列

$$\delta_1,\ \ldots,\ \delta_l$$

确定的两类，只要选择参数

$$\omega=\pi\left(\sum_{i=1}^{l}(1-\delta_i)\cdot 10^i+1\right)$$

即可，如图2.1所示。

- 这个例子反映了这样一个事实：即只要选择适当的参数ω，我们可以对[-1, +1]内的任意数目数据点用sin (ωx) 来进行分类。

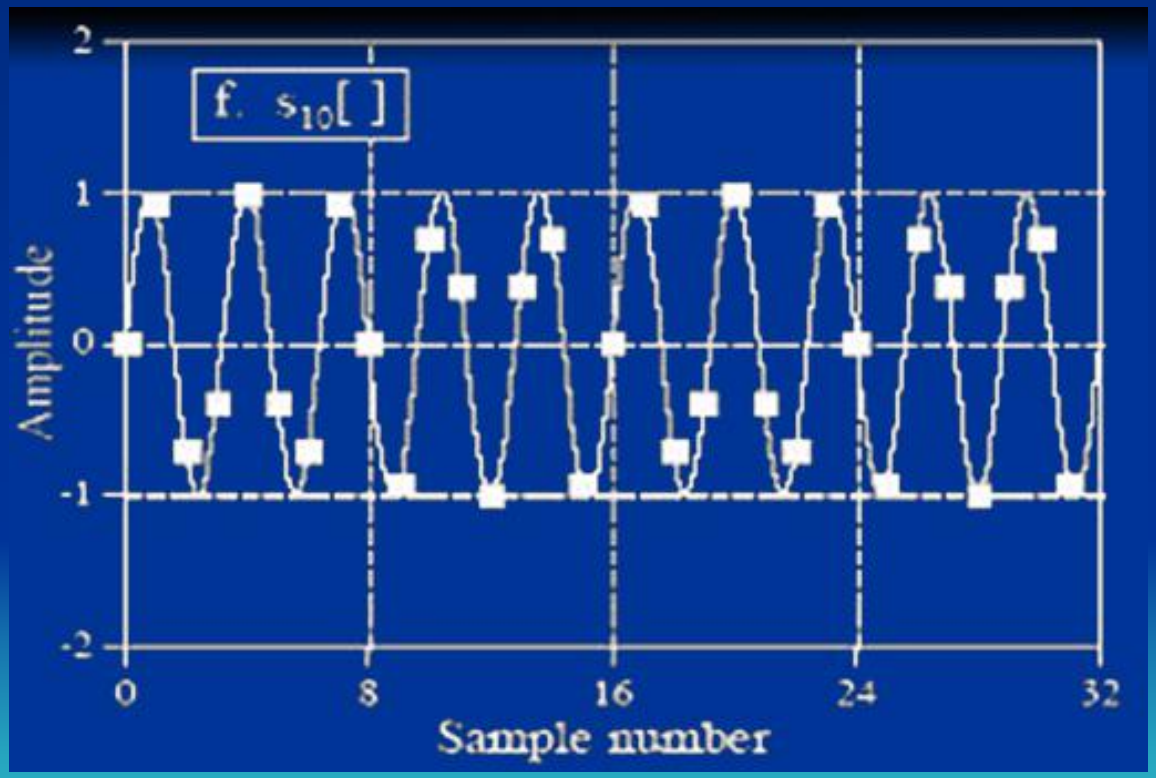


图2.1 用sin(ωx)拟合[-1, +1]内的数据点

小样本学习的推广能力：

- 在有限样本情况下，
  - 经验风险最小并不一定意味着期望风险最小；
  - 学习机器的复杂性不但与所研究的系统有关，而且要和有限的学习样本相适应；
  - 学习精度和推广性之间是一对不可调和的矛盾，采用复杂的学习机器虽然容易使得学习误差更小，却往往丧失推广性；
  - 传统的解决办法（例如：采用正则化、模型选择、噪声干扰等方法以控制学习机器的复杂度）缺乏坚实的理论基础。

- ***K*-近邻算法的VC维和收敛情况：**

***K*-近邻分类器的VC维是无穷大的，因为它对于任何训练集，总能找到一个分类器对其中任何样本都分类正确。

虽然这种算法拥有较好的分类效果，但它没有遵循ERM原则，VC维的讨论不适用于它。

结论：

- 传统机器学习方法中普遍采用的经验风险最小化原则在样本数目有限的时候是不合理的，只考虑到了经验风险。
- 正确的是：应该同时考虑最小化经验风险和置信范围。

2、结构风险最小化

经验风险最小化原则在样本数目有限时是不合理的，因为这时需要同时最小化经验风险和置信范围。

在传统方法中，我们选择学习模型和算法的过程就是优化置信范围的过程。如果选择的模型比较适合现有的训练样本 (相当于 h/n 值适当)，则可以得到比较好的学习效果。

学习机器的设计原则：

- 在设计分类器时，不但要使经验风险最小化，还要使VC维尽量小，从而缩小置信范围，使期望风险最小。

- 寻找反映学习机器的能力的更好参数，从而得到更好的界是今后的重要研究方向之一。

- 例如，在神经网络中，需要根据问题和样本的具体情况来选择不同的网络结构(对应不同的VC维)，然后再进行经验风险最小化。
- 因为缺乏对分类器 函数的认识，选择哪种分类器往往是依赖先验知识进行的，造成了神经网络等方法对使用者“技巧”的过分依赖。

- 虽然很多问题并不是线性可分的，但当样本数量有限时，我们使用线性分类器往往能得到不错的分类效果。这是因为线性分类器的VC维比较低，有利于在样本较少的情况下得到小的置信范围。

2.1 结构风险最小化方法

根据推广性的界公式，采用结构风险最小化的思维策略来解决小样本问题。它的过程分为下列两个：

1) 首先，把分类器函数集$S=\{f(x,\omega),\omega\in\Omega\}$分解成一个子集序列；

$$S_1\subset S_2\subset S_3\subset\ldots\ldots\subset S$$

令各个子集能够依照VC维的大小排列，即：

$$h_1<h_2<h_3<\ldots\ldots<h_k<\ldots\ldots$$

这样得到的同一个子集中，置信范围相同。

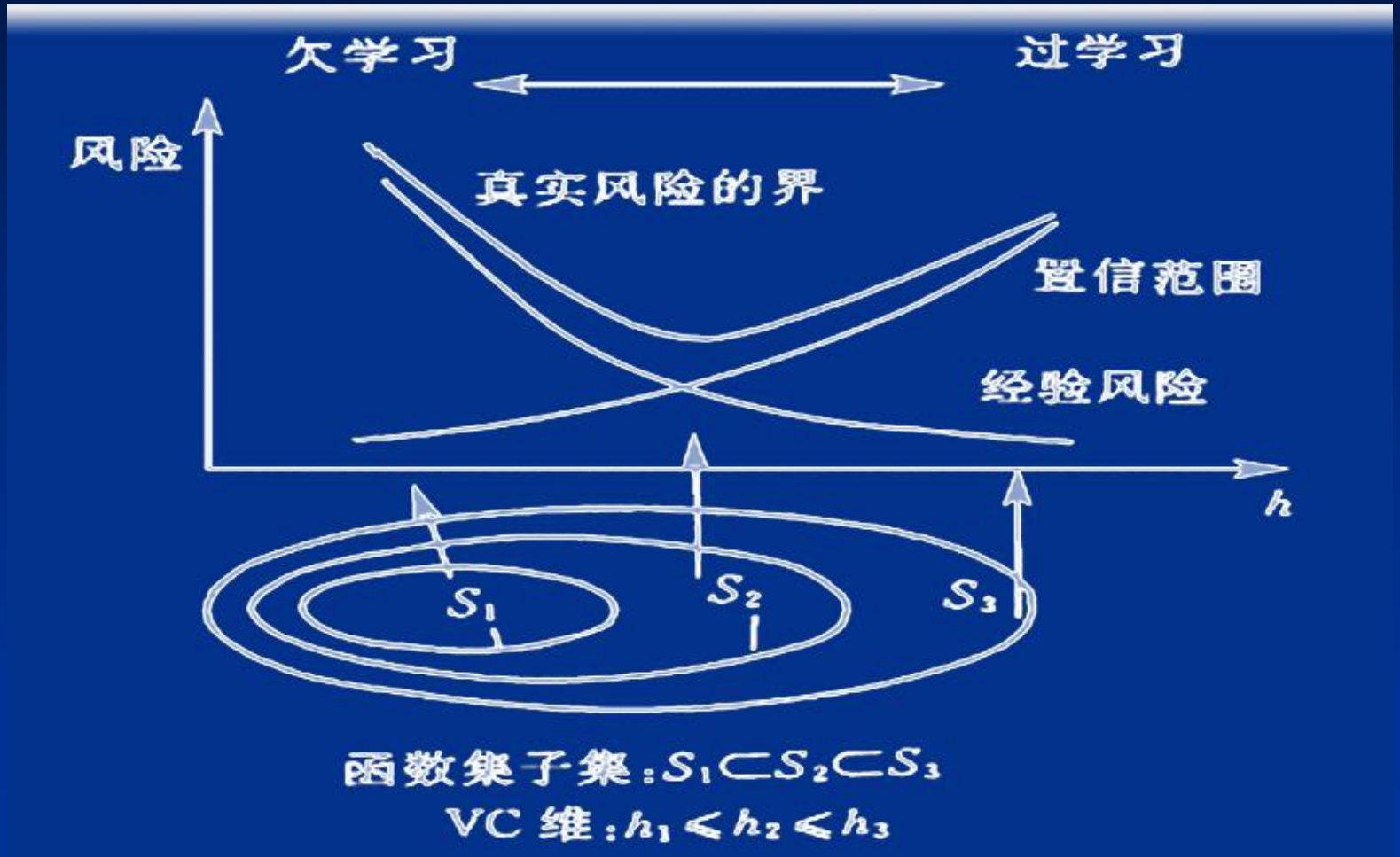
Φ

2) 在每一个子集中寻找最小经验风险

选择最小经验风险与置信范围之和最小的子集，则可以达到期望风险的最小，在这个子集中使经验风险最小的函数就是要求的最优函数。

上面的思想称为有序风险最小化，也称为结构风险最小化 (Structural Risk Minimization)，简称SRM原则。

结构风险最小化方法的示意图见图2.2。



2.2 结构风险最小化示意图

2.2 结构风险最小化原则下分类器的设计

1) 选择一个适当的函数子集，对问题有最优的分类能力。

这一步相当于模型选择，通过对分类器的推广性的界进行估计得到。

2) 从该子集中选择一个函数，令经验风险最小。

从上可知：结构风险最小化需要逐一计算子集，同时要恰当划分子集。目前尚未有一般化的函数子集结构构造方法。

2.3 实现SRM原则的两种思路：

1) 选择使最小经验风险和置信范围之和最小的子集。显然这种方法比较费时，当子集数目很大甚至是无穷时不可行。

   – 例如神经网络的学习，首先通过固定每个不同的网络拓扑结构，来固定学习机器的置信范围；然后在每个结构下最小化经验风险。

2) 设计函数集的某种结构使每个子集中都能取得最小的经验风险 (如使训练误差为0)，然后只需选择选择适当的子集使置信范围最小，则这个子集中使经验风险最小的函数就是最优函数。

- 支持向量机方法实际上就是这种思想的具体实现。

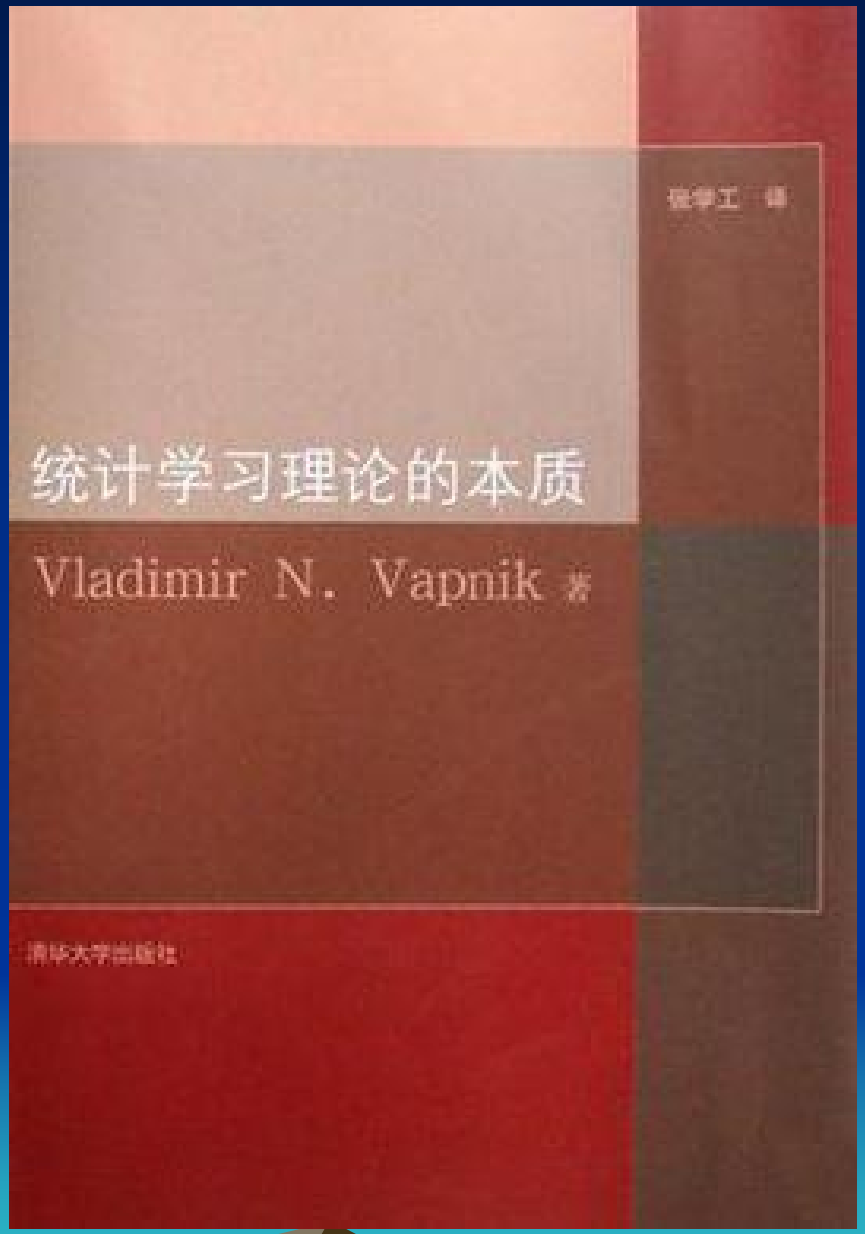
参考文献:

1. 统计学习理论的本质

Vapnik著，

张学工译，

清华大学出版社



参考文献:

2. 人工神经网络与模拟进化计算 第二版

阎平凡，张长水 编著，清华大学出版社

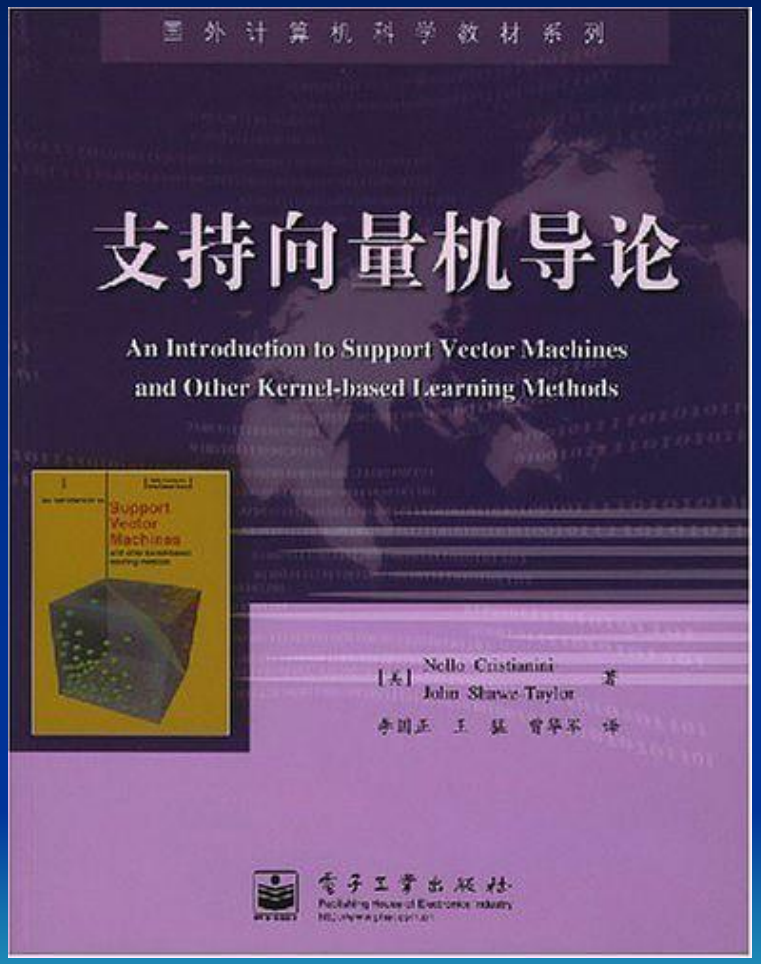
参考文献:

3. 支持向量机导论

克里斯特安尼 著，

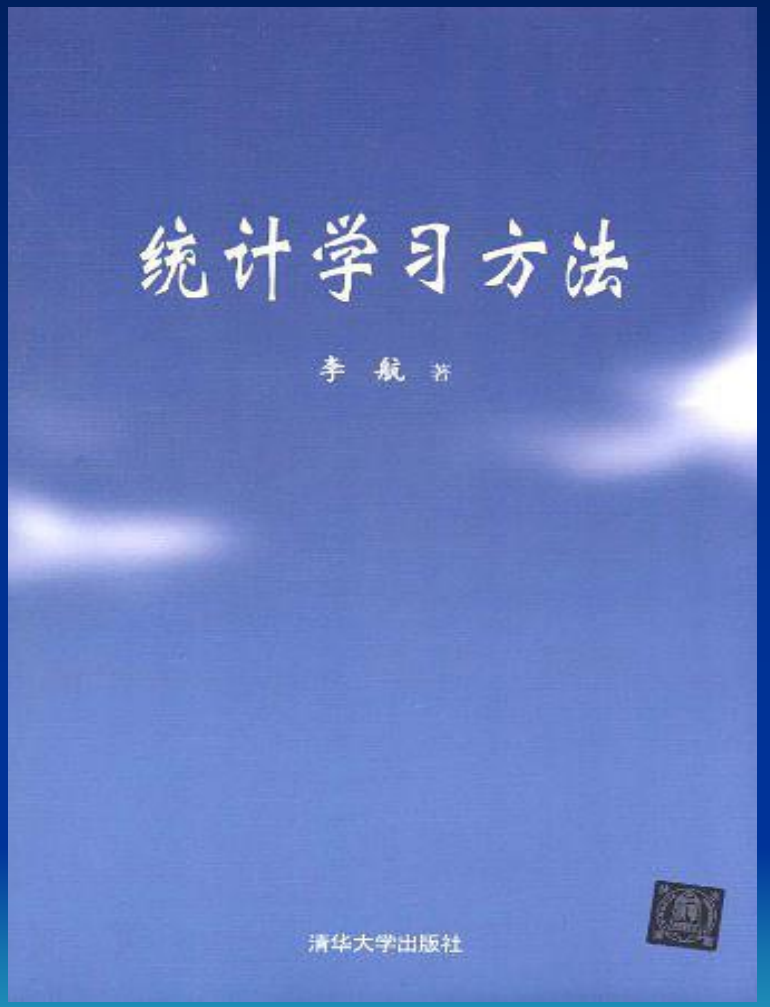
李国正等 译，

电子工业出版社



参考文献:

4. 统计学习方法



李航 著，

2012年 第1版，

清华大学出版社