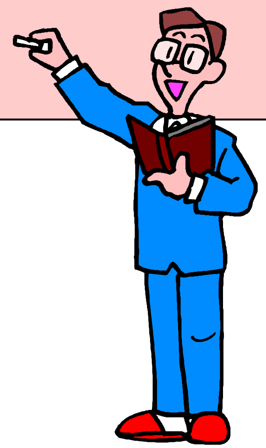
计算机模式识别与机器学习

—— 支持向量机SVM



主讲：图像处理与模式识别研究所
赵群飞

邮　　箱：zhaoqf@sjtu.edu.cn
办公室：电院 2-441
电　　话：13918191860

第7章　支持向量机(SVM)

- **本章学习目标**

✓理解大间隔原理

✓掌握基本的支持向量机分类模型

✓能够熟练运用拉格朗日对偶优化技术

✓掌握数据线性不可分情形下的分类模型，以及核方法的建模原理

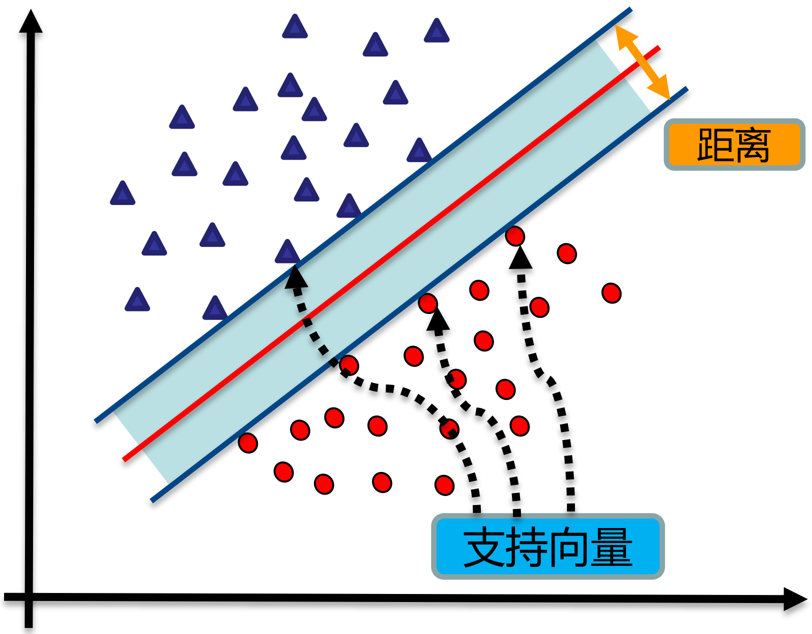
✓理解支持向量机回归的原理

✓了解支持向量机的模型扩展

目录

SVM（Support Vector Machine）的基本原理

- 支持向量机的理论基础于1964年被提出，上世纪90年代后得到快速发展并衍生出一系列改进和扩展算法。
- 支持向量机是一类按监督学习方式对数据进行二元分类的广义线性分类器，其决策边界是对学习样本求解的最大边距超平面（maximum-margin hyperplane）。

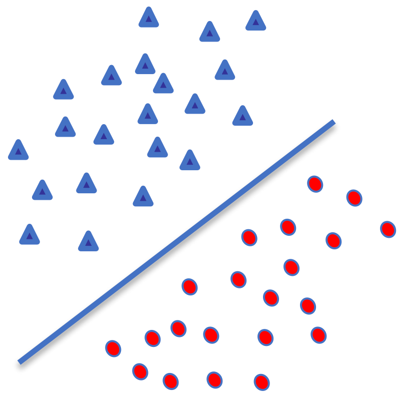


- 具体说，SVM可以看作是一个二类分类模型，其求解目标是在确定一个分类超平面使得间隔（所有样本与分类超平面之间距离的最小值）最大。
- 通过将SVM的原问题转化为对偶问题，SVM的学习核心从间隔最大化的学习问题转化为支持向量的学习问题。其中，支持向量指的是最终用于确定分类器参数的向量。
- 另一方面，基于对偶问题，可以明确地看出不同SVM的核心体现在核矩阵（或者对应核函数）的构造。
- 主要应用场景有图像分类、文本分类、面部识别和垃圾邮件检测等领域。

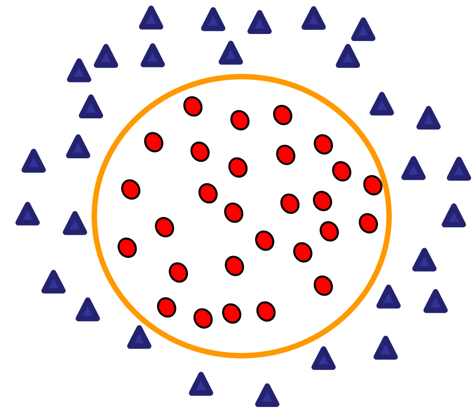
- SVM可以通过核方法进行非线性分类，是常见的核学习方法之一。基于精心构造（或通过多核学习得到）的核函数，可以有效地处理数据的非线性难题。
- 同时，通过核函数，可以在高维特征空间中，甚至无限维特征空间中实现分类问题。模型求解主要使用凸优化技术。
- 此外，SVM使用铰链（合页）损失函数（hinge loss）计算经验风险并在求解系统中加入了正则化项以优化结构风险，是一个具有稀疏性和稳健性的分类器。

硬间隔、软间隔和非线性 SVM

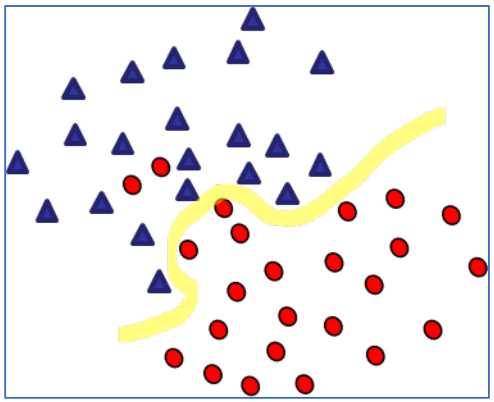
- 假如数据是完全的线性可分的，那么学习到的模型可以称为硬间隔支持向量机。
- 换个说法，硬间隔指的就是完全分类准确，不能存在分类错误的情况。软间隔，就是允许一定量的样本分类错误。



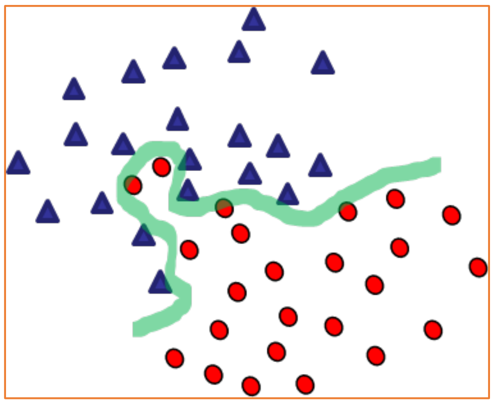
线性可分



线性不可分



软间隔



硬间隔

大间隔原理

以两分类问题为例，即假设给定了一些训练样本，每个样本只属于其中一个类别，目标是正确确定新样本所属的类别。假设训练数据是线性可分的，即可用超平面进行正确分类。希望*两类（正负）样本分得越开越好，也就是两类样本之间的几何间隔越大越好。*

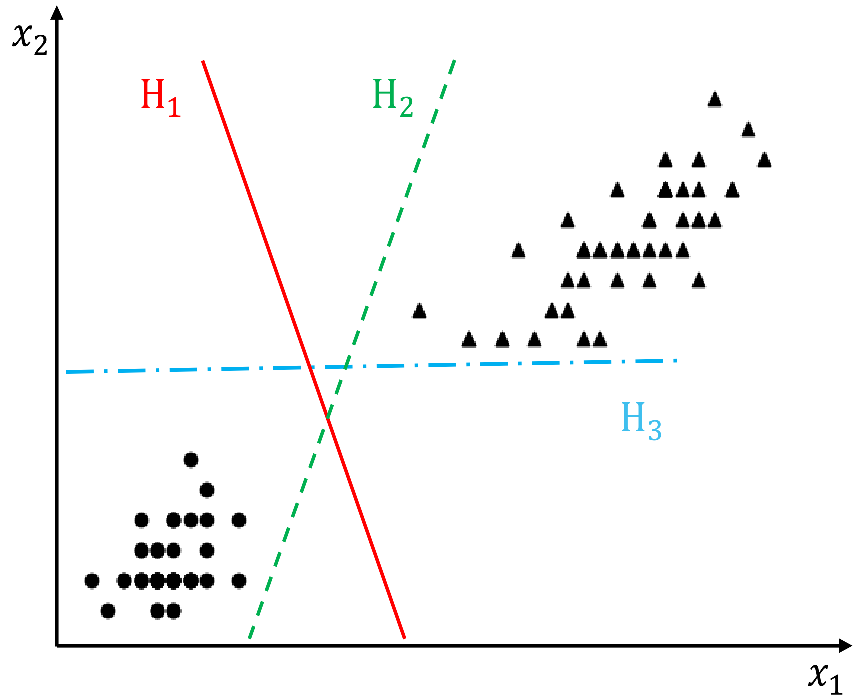


图5-1 大间隔分界面选择示意图
（H1是最大间隔分界面）

目录

基本分类模型

给定一个面向两类分类问题的线性可分训练集，其中包含N个样本$(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots, (\boldsymbol{x}_N, y_N)$。通常，令标签$y \in \{+1, -1\}$。需要学习到一个分类超平面，设对应的参数化表示为：

$$f(\boldsymbol{x}) = \boldsymbol{w}^\top \boldsymbol{x} + b = 0$$

其中，$\boldsymbol{w}$是超平面的法向量，标量b是偏差参数。

- 训练完成得到参数以后，分类函数的输出由以下符号函数给出：

$$h(\boldsymbol{w}) = \operatorname{sign}[\boldsymbol{w}^\top \boldsymbol{x} + b]$$

训练得到参数确定的分类函数：

$$h(\boldsymbol{w}) = \mathrm{sign}[\boldsymbol{w}^\top \boldsymbol{x} + b]$$

那么，对于一个新样本 $\boldsymbol{x}$ 输入

- 如果 $\boldsymbol{w}^\top \boldsymbol{x} + b \geqq 0$ 则分为+1类
- 如果 $\boldsymbol{w}^\top \boldsymbol{x} + b < 0$ 则分为−1类
- 如果 $\boldsymbol{w}^\top \boldsymbol{x} + b = 0$ 分类器无法判断，一般分为+1类。

- 任意一点$\boldsymbol{x}$到该超平面的距离可表示为：

$$d = \frac{|\boldsymbol{w}^\top \boldsymbol{x} + b|}{\|\boldsymbol{w}\|} = \frac{y(\boldsymbol{w}^\top \boldsymbol{x} + b)}{\|\boldsymbol{w}\|}$$

其中，$\|\boldsymbol{w}\| = \sqrt{w_1^2 + \cdots w_n^2}$

- 考虑到偏差参数b的灵活性，可认为距离d的大小与$\boldsymbol{w}$的长度无关，只与$\boldsymbol{w}$的方向有关系。因此可把参数向量$\boldsymbol{w}$的长度固定为1，即令$\|\boldsymbol{w}\| = 1$。

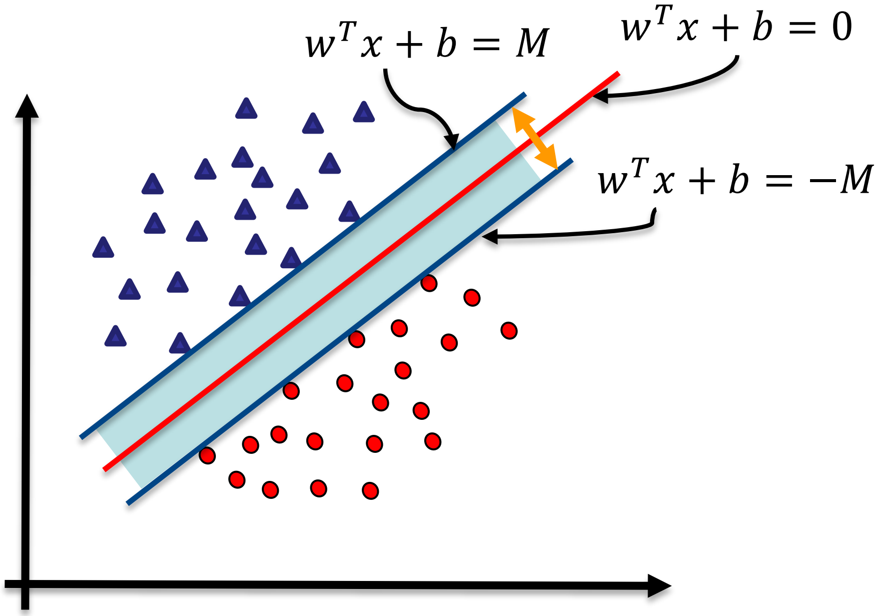
- 由于训练数据线性可分，我们希望能找到一个线性函数$f(\boldsymbol{x})$对所有的样本都满足 $y_i f(\boldsymbol{x}_i) > 0$，而且可以确定这样的函数一定存在。

- 那么，根据大间隔原理，分类模型的优化问题可表达为：

$$
\begin{aligned}
&\max_{\boldsymbol{w},b} \quad M \\
\text{s.t.} \quad & y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq M \\
& \|\boldsymbol{w}\| = 1
\end{aligned}
$$

在上式中，两个超平面 $\boldsymbol{w}^\top\boldsymbol{x}+b=M$和$\boldsymbol{w}^\top\boldsymbol{x}+b=-M$ 之间的距离$2M$被称为**间隔**（margin）。每类数据到分界面的最近距离都是M，即：

$$\frac{y(\boldsymbol{w}^\top\boldsymbol{x}+b)}{||\boldsymbol{w}||}=M.$$



考虑到参数b的表达灵活性以及$\boldsymbol{w}$的长度并不影响分类结果，将上式的两边同时除以M，并采用新的表示，使其满足

$$y(\boldsymbol{w}^{\top}\boldsymbol{x}+b)\geq 1$$

两个超平面之间的距离也相应地发生变化。于是，优化问题转变为：

$$\max_{w,b}\ \frac{2}{\|\boldsymbol{w}\|}$$

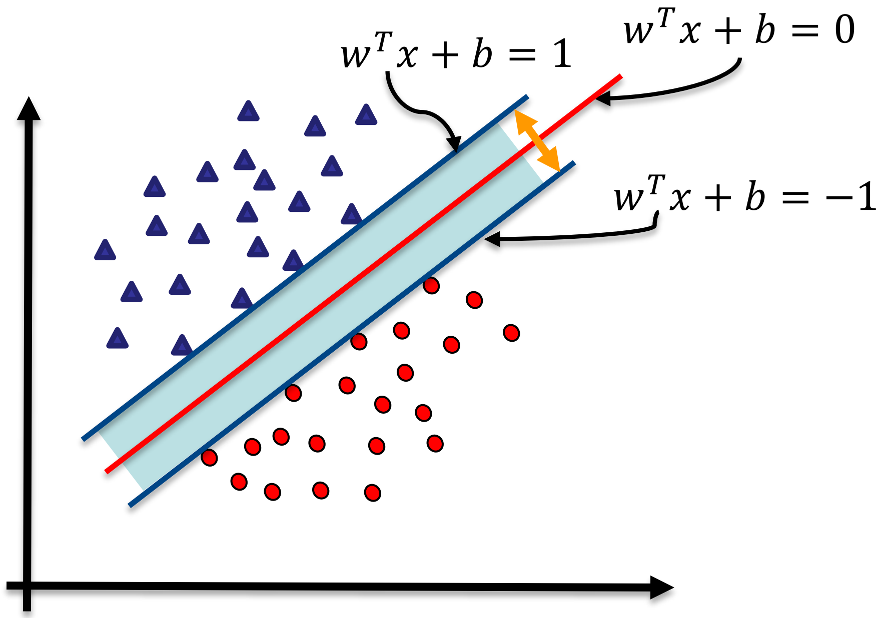
$$\text{s.t.}\quad y_i(\boldsymbol{w}^{\top}\boldsymbol{x}_i+b)\geq 1$$

做适当变换，可得到支持向量机用于线性可分问题的优化表达为：

$$\min_{w,b}\ \frac{1}{2}\|\boldsymbol{w}\|^2$$

$$\text{s.t.}\ \ y_i(\boldsymbol{w}^{\top}\boldsymbol{x}_i+b)\geq 1$$

$$(i=1,2,\dots,N)$$



目录

拉格朗日对偶优化

引入拉格朗日乘子向量$\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\cdots,\alpha_N]^{\mathrm{T}}$，便通过拉格朗日函数将约束条件融入到目标函数中，得到优化问题对应的拉格朗日函数为：

$$L(\boldsymbol{w},b,\boldsymbol{\alpha})=\frac{1}{2}\|\boldsymbol{w}\|^2-\sum_{i=1}^{N}\alpha_i\left(y_i(\boldsymbol{w}^{\top}\boldsymbol{x}_i+b)-1\right)$$

其中各乘子变量$\alpha_i(i=1,2,\cdots,N)$均为非负值。

拉格朗日对偶函数定义为：

$$g(\boldsymbol{\alpha})=\min_{\boldsymbol{w},b} L(\boldsymbol{w},b,\alpha)$$

可以证明，g(**α**)一定小于或等于原优化问题的最优值。

求解拉格朗日对偶优化问题：

$$\max_{\alpha_i \geq 0} \min_{\boldsymbol{w},b} L(\boldsymbol{w}, b, \boldsymbol{\alpha})$$

首先，固定$\boldsymbol{\alpha}$，关于$\boldsymbol{w}$和b最小化拉格朗日函数$L(\boldsymbol{w}, b, \boldsymbol{\alpha})$。

对$\boldsymbol{w}$和b求导，我们可以得出：

$$\frac{\partial L}{\partial \boldsymbol{w}} = \boldsymbol{0} \Rightarrow \boldsymbol{w} = \sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x}_i,$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0$$

- 表达式 $\boldsymbol{w}=\sum_{i=1}^{N}\alpha_i y_i \boldsymbol{x}_i$，体现了支持向量的思想。
- 由$L(\boldsymbol{w},b,\boldsymbol{\alpha})$可见，拉格朗日乘子与训练样本是一一对应的。
- 值为正的拉格朗日乘子所对应的训练样本被称为支持向量，因为它们支持了$\boldsymbol{w}$的计算。

将上面两式带入函数$L(\boldsymbol{w}, b, \boldsymbol{\alpha})$中，可以得到对偶函数为:

$$g(\boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{N} \alpha_i\, y_i \boldsymbol{w}^\top \boldsymbol{x}_i + \sum_{i=1}^{N} \alpha_i$$

$$= \sum_{i=1}^{N} \alpha_i + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\, \alpha_j y_i y_j \boldsymbol{x}_i^\top \boldsymbol{x}_j - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\, \alpha_j y_i y_j \boldsymbol{x}_i^\top \boldsymbol{x}_j$$

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\, \alpha_j y_i y_j \boldsymbol{x}_i^\top \boldsymbol{x}_j.$$

其次，求解对偶优化函数，计算对偶变量**α**的最优解。根据对偶公式得到对偶优化问题的具体表示为：

$$\max_{\alpha} \ \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j$$

$$\text{s.t.} \quad \alpha_i \geq 0$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

参数b可以通过如下互补松弛条件求得：

$$\alpha_i(y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) - 1) = 0, \quad i = 1,2,\ldots,N.$$

对于所有的支持向量，由于相应的对偶变量为正值，根据互补松弛条件可知，原优化问题中的不等式约束变成等式约束。

因此，根据任一个支持向量$\boldsymbol{x}_j$，可得出：

$$b = y_j - \boldsymbol{w}^\top \boldsymbol{x}_j$$

至此，对于待分类样本$\boldsymbol{x}$，支持向量机分类器表示为：

$$h(\boldsymbol{x}) = \mathrm{sign}[\sum_{i=1}^{N} y_i a_i \boldsymbol{x}^\top \boldsymbol{x}_i + b],$$

其中运用了变换为：

$$\boldsymbol{w}^\top \boldsymbol{x} = \sum_{i=1}^{N} y_i \alpha_i \, \boldsymbol{x}^\top \boldsymbol{x}_i.$$

目录

7.4 线性不可分数据的分类

- 对于大多数实际分类问题，训练数据通常都是线性不可分的，这时需要对基本分类模型做一定扩展，得到更为一般化的分类模型。
- 比如，可以选择坚持使用线性分类器模型，但允许一定的误分类，或选择使用非线性的分类器模型。

7.4.1 松弛变量

基本分类模型的约束条件为：

$$y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq 1.$$

引入松弛变量后约束条件为：

$$y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq 1 - \xi_i,$$

其中松弛变量$\xi_i (i = 1,2,\cdots,N)$体现了样本点$\boldsymbol{x}_i$允许偏离原间隔的量，而且满足条件$\xi_i \geq 0$。

在基本分类模型基础上最小化铰链损失，得到用于线性不可分问题的线性支持向量机分类器的优化问题如下：

$$\min_{w,b,\xi} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$\text{s.t.} \quad y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0 \quad (i = 1,2,\cdots,N)$$

其中$\boldsymbol{\xi}$是由ξ_i构成的向量，折衷参数C用于控制目标函数中大间隔和经验损失两项之间的权重。

引入非负的乘子变量α_i和β_i，可得拉格朗日函数表达式为：

$$L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \beta) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\left(y_i(\boldsymbol{w}^\top\boldsymbol{x}_i + b) - 1 + \xi_i\right) - \sum_{i=1}^{N}\beta_i\,\xi_i.$$

对上式关于$\boldsymbol{w}, b, \xi_i$求导，可以得到：

$$\frac{\partial L}{\partial \boldsymbol{w}} = \mathbf{0} \Rightarrow \boldsymbol{w} = \sum_{i=1}^{N}\alpha_i\,y_i\boldsymbol{x}_i$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N}\alpha_i\,y_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C - \alpha_i - \beta_i = 0$$

将这些表达式代入拉格朗日函数，可以得到对偶优化问题是：

$$\max_{\boldsymbol{\alpha}} \ \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j$$

$$\text{s.t.} \quad 0 \le \alpha_i \le C$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

参数b可以通过如下互补松弛条件求得：

$$\alpha_i(y_i(\boldsymbol{w}^{\top}\boldsymbol{x}_i + b) - 1 + \xi_i) = 0, \quad (i = 1,2,\dots,N)$$

$$\beta_i \xi_i = 0, \quad (i = 1,2,\dots,N)$$

其中，通过选择满足$0 < \alpha_j < C$的支持向量$\boldsymbol{x}_j$，可得:

$$b = y_j - \boldsymbol{w}^\top \boldsymbol{x}_j$$

最终的决策函数与基本分类模型的决策函数具有相同的表达式, 即:

$$h(\boldsymbol{x}) = \mathrm{sign}[\sum_{i=1}^{N} y_i a_i \boldsymbol{x}^\top \boldsymbol{x}_i + b]$$

7.4.2 核方法

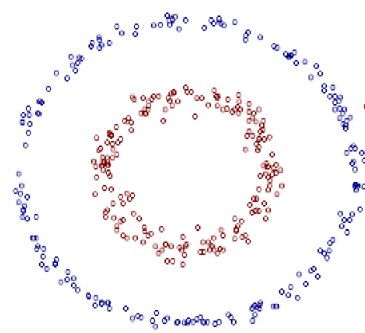
设x和z来自空间Γ（不一定是线性空间），满足下式的函数κ被称为**核函数**：

$$\kappa(x,z)=\langle\phi(x),\phi(z)\rangle$$

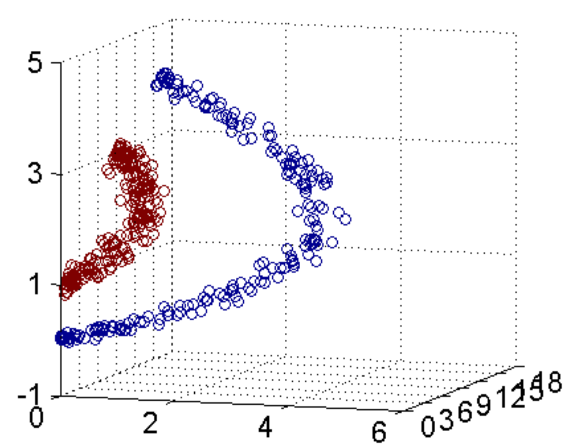
其中ϕ是从空间Γ到希尔伯特空间F的映射，即：

$$\phi: x\in\Gamma\mapsto\phi(x)\in F$$

空间F通常被称为**特征空间**。



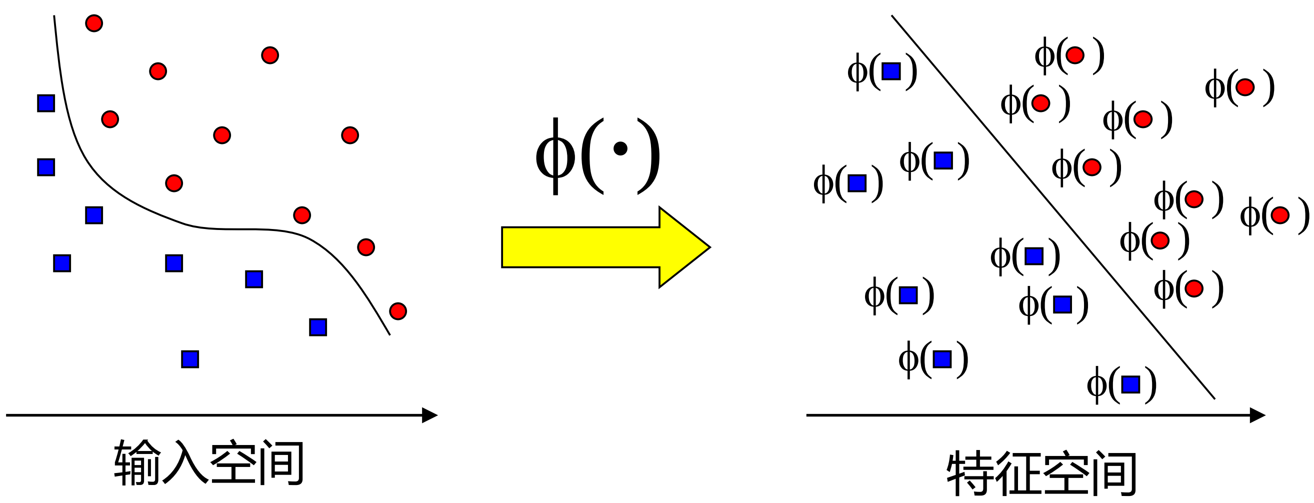
- 可以将样本从原始空间映射到一个更高维的特征空间中，使得样本在新的空间中线性可分。



➢ 用核函数来替换原来的内积，即通过一个非线性转换后的两个样本间的内积。具体地，存在一个从输入空间到特征空间的映射函数$\phi(\boldsymbol{x})$，对于任意空间输入的$\boldsymbol{x}_i, \boldsymbol{x}_j$有

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^\top \phi(\boldsymbol{x}_j),$$

构建一个核函数（正定核）。



设映射函数为$\phi(\boldsymbol{x})$，那么$\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j)=\phi(\boldsymbol{x}_i)^\top\phi(\boldsymbol{x}_j)$，对偶问题的优化目标变为：

$$\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\,\alpha_j y_i y_j\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j).$$

原来的$\boldsymbol{w}^\top\boldsymbol{x}$变为$\boldsymbol{w}^\top\phi(\boldsymbol{x})$，同时考虑到参数$\boldsymbol{w}$可以由对偶变量和输入数据表达，可以得到：

$$\boldsymbol{w}^\top\phi(\boldsymbol{x})=\sum_{i=1}^{N}\alpha_i\,y_i\kappa(\boldsymbol{x}_i,\boldsymbol{x}).$$

用核函数$K(x,z)$代替线性支持向量机学习的对偶问题中的内积，求解得到非线性支持向量机

$$f(x)=\operatorname{sign}\left(\sum_{i=1}^{N}\alpha_i^*\,y_iK(x,x_i)+b^*\right)$$

常见的基本核函数

- 线性核函数：

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^{\top} \boldsymbol{x}_j$$

- 多项式核函数：

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^{\top} \boldsymbol{x}_j + 1)^d$$

其中参数d是多项式次数。

- 高斯核函数：

$$\boldsymbol{x}_i^{\top} \boldsymbol{x}_j = \exp\left\{-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}{2\sigma^2}\right\}$$

也称为径向基函数核，其中参数σ和多项式核函数中的参数d一样，通常需要通过模型选择来确定具体取值。

目录

支持向量机回归（SVR）

- 最大化间隔的思想同样适用于回归问题。在回归问题中，训练数据的标签变量是实数值，不再是离散的有限个类别。
- 为了得到决策函数的稀疏表达，引入如下的ϵ不敏感损失函数：

$$|y-f(\boldsymbol{x})|_{\varepsilon}=\max\{0,|y-f(\boldsymbol{x})|-\varepsilon\}$$

其中$f(\boldsymbol{x})$是回归函数，参数$\epsilon \geq 0$。

- 由于ϵ不能设置得过大，所以不敏感损失的值往往不为零。

因此，支持向量回归的优化问题表示为：

$$\min_{\boldsymbol{w},b} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}|y_i - f(\boldsymbol{x}_i)|_\varepsilon$$

其中非负参数C反应了函数复杂性与经验损失之间的折衷，对样本的预测输出函数为：

$$f(\boldsymbol{x}) = \boldsymbol{w}^\top \boldsymbol{x} + b$$

引入两组松弛变量 ξ 和 ξ^*，使得每个样本在间隔带两侧的松弛程度可以不同，则支持向$\|\boldsymbol{w}\|^2$量回归的优化函数等价为：

$$
\begin{aligned}
\min_{w,b,\xi,\xi^*} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\
\text{s.t.} \quad & y_i - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i - b \le \varepsilon + \xi_i \\
& \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b - y_i \le \varepsilon + \xi_i^* \\
& \xi_i \ge 0 \\
& \xi_i^* \ge 0, \quad i = 1,2,\ldots,N
\end{aligned}
$$

上式对应的拉格朗日函数为:

$$
\begin{aligned}
L(\boldsymbol{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\
- \sum_{i=1}^{N}\alpha_i\left(\epsilon + \xi_i - y_i + \boldsymbol{w}^\top\boldsymbol{x}_i + b\right) \\
- \sum_{i=1}^{N}\alpha_i^*\left(\epsilon + \xi_i^* + y_i - \boldsymbol{w}^\top\boldsymbol{x}_i - b\right)
\end{aligned}
$$

其中，$\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$是非负的乘子变量。

对拉格朗日函数关于w, b, ξ_i, ξ_i^*进行求导，可以得到:

$$\frac{\partial L}{\partial \boldsymbol{w}} = \mathbf{0} \Rightarrow \boldsymbol{w} = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)\, \boldsymbol{x}_i$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0$$

$$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C - \alpha_i - \eta_i = 0$$

$$\frac{\partial L}{\partial \xi_i^*} = 0 \Rightarrow C - \alpha_i^* - \eta_i^* = 0$$

将上述结果带入拉格朗日函数中，得到对偶优化问题:

$$min_{\alpha,\alpha^*} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\, x_i^T x_j + \varepsilon \sum_{i=1}^{N}(\alpha_i + \alpha_i^*) - \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*)$$

$$\text{s.t.} \quad \sum_{i=1}^{N}(\alpha_i - \alpha_i^*), \quad \alpha_i, \alpha_i^* \in [0, C]$$

- 求得最优的α_i, α_i^*后，支持向量机回归的预测函数为:

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)\, \boldsymbol{x}^\top \boldsymbol{x}_i + b.$$

其中，偏置项b通过互补松弛条件求得。

- 若$0 < \alpha_i < C$，则必有 $\xi_i = 0$，进而运用与α_i对应的样本输入和标签得出b的表达式为:

$$b = y_i - \epsilon - \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)\, \boldsymbol{x}^\top \boldsymbol{x}_i$$

在实际中，为了具有更好的鲁棒性，可以选取多个满足条件 $0 < \alpha_i < C$ 的样本求解 b，取得平均值。

目录

SVM总结与模型扩展

支持向量机分类的基本思想是在特征空间中学习一个超平面分类器，具体内容可以概括为：

1. 用$w^{\top}x+b$ 定义分类函数；
2. 求 $\boldsymbol{w}$, b;
3. 为了寻找最大间隔，引出$\frac{1/2}{||w||^2}$；
4. 引入拉格朗日因子，化为对拉格朗日乘子α的求解；
5. 求解过程中，处理一系列最优化与凸二次规划的问题；
6. 求解α时可以使用快速学习方法SMO；
7. 处理非线性情况时，使用核函数；
8. 核函数在低维计算，等效了高维表现，避免了直接映射到高维计算时可能出现的维度爆炸现象；

SVM使用的普遍准则：（n为特征数，m为训练样本数。）

1) 如果相较于m而言，n要大许多，即训练集数据量不够支持我们训练一个复杂的非线性模型，我们选用逻辑回归模型或者不带核函数的支持向量机。

2) 如果n较小，而且m大小中等，例如n在 1-1000 之间，而m在10-10000之间，使用高斯核函数的支持向量机。

3) 如果n较小，而m较大，例如n在1-1000之间，而m大于50000，则使用支持向量机会非常慢，解决方案是创造、增加更多的特征，然后使用逻辑回归或不带核函数的支持向量机。

支持向量机的优点：

- 对于线性不可分的情况可以通过核函数，映射到高维特征空间实现线性可分。
- SVM学习问题可以表示为凸优化问题，因此可以利用已知的有效算法发现目标函数的全局最小值。而其他分类方法（如基于规则的分类器和人工神经网络）都采用一种基于贪心学习的策略来搜索假设空间，这种方法一般只能获得局部最优解。
- 小集群分类效果好。

支持向量机的缺点：

- SVM仅仅只限于一个二类分类问题，对于多分类问题解决效果并不好。
- 仅局限于小集群样本，对于观测样本太多时，效率较低。
- 寻求合适的核函数相对困难。

支持向量机模型的扩展：

- 双平面支持向量机的思想是使得一个超平面离一类样本近并且离另一类样本有一定的距离。与支持向量机求解单个二次优化问题不同，它求解两个二次优化问题，而且每个优化问题涉及的样本数量少于支持向量机中的样本数量。

- 针对多视图数据建模的需要，人们也提出了多视图支持向量机及多视图双平面支持向量机。

- 此外，支持向量机还被扩展到半监督学习、主动学习和多任务学习等多个场景。在监督分类方面，前面正文介绍的支持向量机面向两类分类问题，后来支持向量机还被扩展到了多类分类问题。

参考文献

1. 机器学习实战（七）支持向量机（SVM）
   https://blog.csdn.net/qs17809259715/article/details/97761963

2. 支持向量机原理篇之手撕线性SVM
   https://blog.csdn.net/c406495762/article/details/78072313

3. 机器学习之支持向量机SVM及代码示
   https://blog.csdn.net/cxmscb/article/details/56277984?utm_medium=distribute.pc_relevant_download.none-task-blog-2~default~BlogCommendFromBaidu~default-1.nonecase&depth_1-utm_source=distribute.pc_relevant_download.none-task-blog-2~default~BlogCommendFromBaidu~default-1.nonecas

4. 支持向量机（原理公式推导+python实战代码分析）
   https://blog.csdn.net/z_feng12489/article/details/83030077?utm_medium=distribute.pc_relevant_download.none-task-blog-baidujs-1.nonecase&depth_1-utm_source=distribute.pc_relevant_download.none-task-blog-baidujs-1.nonecase

谢 谢！

本课件制作过程中，多处引用了国内外同行的网页、教材、以及课件PPT的内容或图片，没有随处标注，特此说明，并在此向各位作者表示感谢！