支持向量机的前世今生

作者：张翀
班级：计算机试验班61
学号：2140506063

> 支持向量机是监督学习时代最优秀、最具代表性的分类算法，用于处理有标签数据的二分类问题。基于训练集D，支持向量机希望找到一个线性分割器（超平面）实现二分类，并具有一定的泛化能力。

支持向量机的原始模型

超平面的方程如$w^Tx+b=0$所示，样本中任意一点x到超平面的距离$r=\frac{|w^Tx+b|}{||w||}$.
在模型能对样本点进行正确分类时，不妨设立严格性要求，使如下条件满足：

$$\begin{cases} w^Tx+b\geq +1, y_i=+1 \\ w^Tx+b\leq -1, y_i=-1 \end{cases}$$

这一步实际上是对超平面系数的标准化。此时样本点与超平面的距离用间隔$\gamma=\frac{2}{||w||}$来衡量。因此，支持向量机可用如下的优化问题表示：

$$\underset{w,b}{max}\frac{2}{||w||}, s.t. y_i(w^Tx_i-b)\geq 1, 1\leq i\leq m$$

我们讨论等价的优化问题

$$\underset{w,b}{min}\frac{||w||^2}{2}, s.t. y_i(w^Tx_i-b)\geq 1, 1\leq i\leq m$$

间隔与支持向量

通过KKT方程求解上述问题：

$$\begin{cases} \alpha_i\geq 0 \\ y_if(x_i)-1\geq 0 \\ \alpha_i(y_if(x_i)-1)=0 \end{cases}$$

当且仅当$\alpha_i>0$时，$y_if(x_i)=1$.此时x_i是正好位于边缘的样本点，决定了当前的间隔。将此时的x_i称为支持向量。
实际上，支持向量指出了距离分界面最近、最难分类的样本点，只有这些点对计算间隔有影响。因此，建立的最终模型只与支持向量有关。然而，随参数w,b的变化，支持向量不一定是绝对固定的。

松弛约束

原始模型的局限性在于，实际情况下的样本点不可能完全用线性平面分隔开来。即使真的可以，我们也希望建立泛化能力更强的模型来保证更好的鲁棒性。因此，我们对先前的约束进行一定的松弛，使得更多分类出错的情况能进入可行解集。
具体而言，优化问题将错误样本个数加入目标函数，以期出错的样本个数最少，改动后的优化问题如下所示：

$$\underset{w,b}{min}\frac{||w||^2}{2}+C\sum_{i=1}^{m}l_0(y_i(w^Tx_i-b)-1), s.t. y_i(w^Tx_i-b)\geq 1, 1\leq 1\leq m$$

其中$l_0(x)=\begin{cases}1, z<0\\0, z\geq 0\end{cases}$
因为$l_0(x)$性质不好（不连续可导），软间隔支持向量机模型引入松弛变量，用软间隔$\xi_i=w^Tx_i-b$代替出错样本计数。这样不但优化了函数性态，还定量地说明了样本点偏离分类结果的程度，所以是一个更精细的模型。软间隔支持向量机模型如下所示：

$$\underset{w,b}{min}\frac{||w||^2}{2}+C\sum_{i=1}^{m}\xi_i, s.t. y_i(w^Tx_i-b)\geq 1-\xi_i, 1\leq i\leq m$$

核函数

样本点非线性可分的另外一个解决方法是利用映射函数$\phi(x):\mathbb{R}^n\rightarrow\mathbb{R}^k$将样本点映射到较高维度的空间，以期样本点在高维的特征空间上线性可分。事实上，这样的特征空间总是存在的。
映射过后的对偶问题形式如下：

$$\underset{\alpha}{max}\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_jy_iy_j\phi(x_i)^T\phi(x_j), s.t.\sum_{i=1}^{m}\alpha_iy_i=0, \alpha_i\geq 0$$

可以发现，$\phi(x)$的形式在计算中不会用到；反而是核函数$\kappa(x_i,x_j)=\phi(x_i)^T\phi(x_j)$比较重要。在使用对偶问题/KKT方程的方法求解高维特征对应的问题时，只要知道核函数$\kappa(x_i,x_j)=\phi(x_i)^T\phi(x_j)$的值，就可以对高维问题进行求解。常用的核函数可能由确切的$\phi(x)$指导，比如线性核，也可能由其他的核函数构造（满足性质即可）。

支持向量机的求解

SMO方法

SMO方法用于求解原问题的对偶问题，其问题模型如下：

$$\underset{\alpha}{max}\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_jy_iy_jx_i^Tx_j, s.t.\sum_{i=1}^{m}\alpha_iy_i=0, \alpha_i\geq 0$$

在未引入核函数，求解原始分类问题时，支持向量机的原始模型是一个二次规划问题，理论上可以直接按通用方法求解；但这样时间复杂度较高。由前文所述，分割平面的选取只和支持向量有关，可以利用这一点性质优化求解过程。具体而言，原始的二次规划问题转化为对偶问题，后者只和对偶变量有关；接下来使用使用SMO方法求解原对偶问题：求解时，先固定α_i之外的所有参数，再求α_i上的极值；每次取两个变量α_i和α_j进行优化，直至收敛。这样做使得每一步的计算等价于求单变量二次函数的最小值，大大减小了计算量。

支持向量回归和核方法

通过KKT条件直接求解，得到分析解：

$$\begin{cases} w = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i)\phi(x_i) \\ b = y_i + e - \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) x_i^T x \end{cases}$$

针对核方法，优化问题的最优解$h^*(x)$必然为核函数$\kappa(x, x_i)$的线性组合 。

参考内容

优化方法课件
《机器学习》（周志华著）, ISBN: 978-7-302-42328-7