支持向量机

W.J.Z

1 硬间隔支持向量机

硬间隔指超平面能完全正确将数据划分为两类，法向量一侧为正类，相反一侧为负类，所有样本点都能满足函数间隔大于等于 1 的约束条件。超平面公式为：

$$w \cdot x + b = 0 \tag{1}$$

w 为支持向量，b 为位移项。

空间任意一点到超平面的距离为：

$$\frac{|w \cdot x + b|}{\|w\|} \tag{2}$$

对于完全正确分类的样本，令：

$$\begin{cases} w \cdot x_i + b \geq +1, & y_i = +1 \\ w \cdot x_i + b \leq -1, & y_i = -1 \end{cases} \tag{3}$$

两个离超平面最近的两个异类点相减得

$$w \cdot (x_1 - x_2) = \|w\| \cdot \|x_1 - x_2\| \cdot \cos\theta = 2 \tag{4}$$

$w \quad (x_1 - x_2)$ 为向量相乘，θ 为两向量之间的夹角，而 $\|x_1 - x_2\| \cdot \cos\theta$ 就是我们所求的间隔。令 $d = \|x_1 - x_2\| \cdot \cos\theta$，则

$$d = \frac{2}{\|w\|} \tag{5}$$

原始问题可描述为：

$$\max_{w,b} \frac{2}{\|w\|} \tag{6}$$

$$s.t. \quad y_i\,(w \cdot x_i + b) \geq 1, \quad i = 1, 2, \ldots, m \tag{7}$$

最大化 $\frac{1}{\|w\|}$ 等价于最小化 $\|w\|^2$, 公式 6又可描述为

$$\min_{w,b} \frac{1}{2}\left\|w\right\|^2 \tag{8}$$

$$s.t. \quad y_i\left(w\cdot x_i+b\right)\geq 1, \quad i=1,2,\ldots,m$$

引入拉格朗日乘子进行对偶化：

$$L\left(w,b,\alpha\right)=\frac{1}{2}\left\|w\right\|^2-\sum_{i=1}^{N}\alpha_i y_i\left(w\cdot x_i+b\right)+\sum_{i=1}^{N}\alpha_i \tag{9}$$

对 w 和 b 求偏导并令其等于 0 得：

$$w=\sum_{i=1}^{N}\alpha_i y_i x_i \tag{10}$$

$$\sum_{i=1}^{N}\alpha_i y_i=0 \tag{11}$$

将公式 10代入公式 9和利用公式 11得：

$$\max_{\alpha} -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\left(x_i\cdot x_j\right)+\sum_{i=1}^{N}\alpha_i \tag{12}$$

$$s.t. \quad \sum_{i=1}^{N}\alpha_i y_i=0 \tag{13}$$

$$\alpha_i\geq 0, i=1,2,\ldots,N \tag{14}$$

将求极大转为求极小：

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\left(x_i\cdot x_j\right)+\sum_{i=1}^{N}\alpha_i \tag{15}$$

$$s.t. \quad \sum_{i=1}^{N}\alpha_i y_i=0 \tag{16}$$

$$\alpha_i\geq 0, i=1,2,\ldots,N \tag{17}$$

假设最优化问题公式 15– 公式 17对 α 的解为 a^*，则

$$w^*=\sum_{i=1}^{N}\alpha_i^* y_i x_i \tag{18}$$

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i \left(x_i \cdot x_j\right) \tag{19}$$

则分离超平面可以写成：

$$\sum_{i=1}^{N} \alpha_i^* y_j \left(x \cdot x_i\right) + b^* = 0 \tag{20}$$

Algorithm 1: 硬间隔支持向量机

Input: 训练集 T

Output: 分离超平面和分类决策函数

1 构造并求解约束最优化问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \left(x_i \cdot x_j\right) + \sum_{i=1}^{N} \alpha_i$$

$$s.t. \quad \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\alpha_i \geq 0, i = 1, 2, \ldots, N$$

求得最优解 a^*.

2 计算

$$w^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$$

并选择 a^* 的一个正分量 $a_j^* > 0$ 计算 $b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i \left(x_i \cdot x_j\right)$

3 求得分离超平面

$$w^* \cdot x + b^* = 0$$

分类决策函数：

$$f\left(x\right) = sign\left(w^* \cdot x + b^*\right)$$

---

2 软间隔支持向量机

软间隔是为了解决某些样本点不能满足函数间隔大于等于 1 的约束条件。约束条件和目标函数改变为：

$$y_i\left(w \cdot x_i + b\right) \geq 1 - \xi_i \tag{21}$$

$$\frac{1}{2}\left\|w\right\|^2 + C\sum_{i=1}^{N}\xi_i \tag{22}$$

ξ_i 为松弛变量，C 为惩罚函数，C 值大时对误分类的惩罚增大，C 值小时对误分类的惩罚减小。原始最优化问题的拉格朗日函数为：

$$L\left(w, b, \xi, \alpha, \mu\right) = \frac{1}{2}\left\|w\right\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\left(y_i\left(w \cdot x_i + b\right) - 1 + \xi_i\right) - \sum_{i=1}^{N}\mu_i\xi_i \tag{23}$$

其中 $\alpha_i \geq 0, \mu_i \geq 0$，对 w, b, ξ 求偏导并将其等于 0，得到的对偶问题为：

$$\underset{\alpha}{min}\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\left(x_i \cdot x_j\right) - \sum_{i=1}^{N}\alpha_i \tag{24}$$

$$s.t. \quad \sum_{i=1}^{N}\alpha_i y_i = 0 \tag{25}$$

$$0 \leq \alpha_i \leq C, i = 1, 2, \ldots, N \tag{26}$$

假设 α^* 是对偶问题的一个解，存在 α^* 的一个分量 α_j^*，使 $0 < \alpha_j^* < C$ 则：

$$w^* = \sum_{i=1}^{N}\alpha_i^* y_i x_i \tag{27}$$

$$b^* = y_j - \sum_{i=1}^{N}y_i\alpha_i^*\left(x_i \cdot x_j\right) \tag{28}$$

则分离超平面可以写成：

$$\sum_{i=1}^{N}\alpha_i^* y_j\left(x \cdot x_i\right) + b^* = 0 \tag{29}$$

分类决策函数可写成：

$$sign(\sum_{i=1}^{N}\alpha_i^* y_j\left(x \cdot x_i\right) + b^*) \tag{30}$$

Algorithm 2: 软间隔支持向量机

Input: 训练集 T

Output: 分离超平面和分类决策函数

1 选择惩罚参数 $C > 0$, 构造并求解凸二次规划问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \left(x_i \cdot x_j\right) - \sum_{i=1}^{N} \alpha_i$$

$$s.t. \quad \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$0 \le \alpha_i \le C, i = 1, 2, \ldots, N$$

求得最优解 α^*

2 计算 $w^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$, 选择 α^* 的一个分量 α_j^* 适合条件 $0 < \alpha_j^* < C$ 计算

$$b^* = y_j - \sum_{i=1}^{N} y_i \alpha_i^* \left(x_i \cdot x_j\right)$$

3 求得分离超平面

$$w^* \cdot x + b^* = 0$$

分类决策函数

$$\sum_{i=1}^{N} \alpha_i^* y_j \left(x \cdot x_i\right) + b^*$$

3 非线性支持向量机

求解非线性问题的方法之一为将非线性问题进行一个非线性变换，将非线性问题转换为线性问题，步骤为：首先使用一个变换将源空间的数据映射到新空间，然后在新空间里用线性分类学习方法从训练数据中学习分类模型，核技巧属于这种方法。

观察线性支持向量机的对偶问题，无论是目标函数还是决策函数，都只涉及输入实例和实例之间的内积，在多普问题的目标函数中的内积 $x_i \cdot x_j$ 可以用核函数 $K\left(x_i \; y_j\right)$ 来代替。

对偶问题为：

$$\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K\left(x_i\ y_j\right)-\sum_{i=1}^{N}\alpha_i \tag{31}$$

分类决策函数为：

$$sign(\sum_{i=1}^{N}\alpha_i^* y_j K\left(x_i\ y_j\right)+b^*) \tag{32}$$

Algorithm 3: 非线性支持向量机

Input: 训练集 T

Output: 分类决策函数

1 选择适当的核函数 K 和适当的参数 C，构造并求解最优化问题

$$\min_{\alpha}\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K\left(x_i,x_j\right)-\sum_{i=1}^{N}\alpha_i$$

$$s.t.\quad \sum_{i=1}^{N}\alpha_i y_i=0$$

$$0\le\alpha_i\le C, i=1,2,\ldots,N$$

求得最优解 α^*

2 选择 α^* 的一个正分量 α_j^* 适合条件 $0<\alpha_j^*<C$ 计算

$$b^*=y_j-\sum_{i=1}^{N}y_i\alpha_i^*\left(x_i\cdot x_j\right)$$

3 构造决策函数

$$sign(\sum_{i=1}^{N}\alpha_i^* y_j K\left(x_i\ y_j\right)+b^*)$$

4 * 当 $K(x,z)$ 是正定核实函数时，解释存在的

正定核的充要条件：设 $K:\chi\times\chi\to R$ 是对称函数，对任意 $x_i\epsilon\chi, i=1,2,\ldots,m$,$K(x,z)$ 对应的 Gram 矩阵是半正定矩阵。

定义：n维欧式空间中任意$k(k \leq n)$个向量$\alpha_1, \alpha_2, \ldots, \alpha_k$的内积所组成的矩阵

$$\Delta(\alpha_1, \alpha_2, \ldots, \alpha_k) = \begin{pmatrix} (\alpha_1, \alpha_1) & (\alpha_1, \alpha_2) & \ldots & (\alpha_1, \alpha_k) \\ (\alpha_2, \alpha_1) & (\alpha_2, \alpha_2) & \ldots & (\alpha_2, \alpha_k) \\ \ldots & \ldots & \ldots & \ldots \\ (\alpha_k, \alpha_1) & (\alpha_k, \alpha_2) & \ldots & (\alpha_k, \alpha_k) \end{pmatrix}$$

称为k个向量$\alpha_1, \alpha_2, \ldots, \alpha_k$的格拉姆矩阵（Gram矩阵），它的行列式称为Gram行列式。

http://blog.csdn.net/wangyang20170901