1. 概述

若两类训练样本是线性可分的，SVM 即找到一个超平面，使得两类最靠近的样本点的距离都达到最大，该平面就是最优分界面。最优边界上的样本称之为“支持向量”

对线性可分的两类问题 $\left\{\vec{x}^p, d^p\right\}_{p=i}^{N}, d^p \in \{-1, 1\}$，找到 $\vec{\omega}, b$ 满足

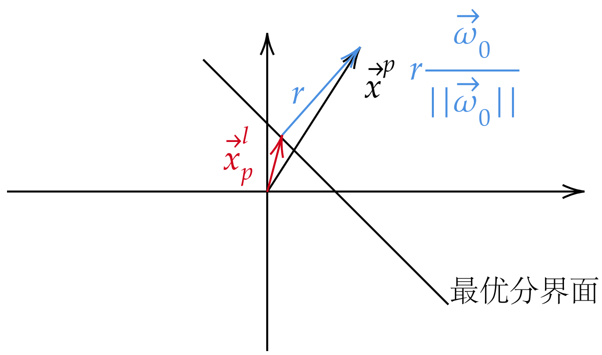
$$\begin{cases} \vec{w}^\top \vec{x}^p + b \geqslant 1 & d^p = 1 \\ \vec{\omega}^\top \vec{x}^p + b \leqslant -1 & d^p = -1 \end{cases}$$

2. 最优平面的支撑向量

2.1. 样本到最优分界面的距离

假设最优分界面已经找到，记为 $\vec{\omega}_0^\top x^l + b_0 = 0$，$\vec{\omega}_0$ 和 b_0 为最优分界面的法向量和偏差

令 $\vec{x}^p$ 为任一样本，且 $\vec{x}^p$ 在最优分界面的投影为 $\vec{x}_p^l$



则可知 $\vec{x}^p = \vec{x}_p^l + r\left(\frac{\vec{\omega}_0}{||\vec{\omega}_0||}\right)$　$r : \vec{x}^p$ 到最优分界面的距离

令 $g(\vec{x}) = \vec{\omega}_0^\top \vec{x} + b_0$，则可知

$$g\left(\vec{x}^p\right) = \vec{\omega}_0^\top\left(\vec{x}_p^l + r\left(\frac{\vec{\omega}_0}{||\vec{\omega}_0||}\right)\right) + b_0 = \vec{\omega}_0^\top \vec{x}_p^l + b_0 + r\frac{\vec{\omega}_0^\top \vec{\omega}_0}{||\vec{\omega}_0||} = r\frac{||\vec{\omega}_0||^2}{||\vec{\omega}_0||} = r||\vec{\omega}_0||$$

$\because$ $\vec{x}_p^l$ 在最优分界面上　$\therefore$ $\vec{\omega}_0^\top \vec{x}_p^l + b_0 = 0$

则可知 $r = \frac{g\left(\vec{x}^p\right)}{||\vec{\omega}_0||}$，即为支撑向量到最优分界面的距离

最优边界 $g\left(\vec{x}^p\right) = 1$ 和 $g\left(\vec{x}^p\right) = -1$ 之间距离为 $\rho = 2r = \dfrac{2}{||\vec{\omega}_0||}$

故可知要使 $\rho \to \max \Rightarrow ||\vec{\omega}_0|| \to \min \Rightarrow \sqrt{\vec{\omega}^\top \vec{\omega}} \to \min \Rightarrow \vec{\omega}^\top \vec{\omega} \to \min$

2.2. 寻找最优分界面

$\begin{cases} \vec{\omega}^\top \vec{x}^p + b \geqslant 1 & d^p = 1 \\ \vec{\omega}^\top \vec{x}^p + b \leqslant -1 & d^p = -1 \end{cases} \Rightarrow d^p\left(\vec{\omega}^\top \vec{x}^p + b\right) \geqslant 1$ （d^p 的取值是编码方式）

求最优分界面时应在约束条件 $d^p\left(\vec{\omega}^\top \vec{x}^p + b\right) \geqslant 1$ 下求 $\vec{\omega}^\top \vec{\omega}$ 的最小值，即使目标函数 J 最小，作拉格朗日函数得

$$\mathrm{J}(\vec{\omega}, b, \vec{\alpha}) = \frac{1}{2}\vec{\omega}^\top \vec{\omega} - \sum_{p=1}^{N} \alpha_p \left(d^p\left(\vec{\omega}^\top \vec{x}^p + b\right) - 1\right) \quad \vec{\alpha} = (\alpha_1, \cdots, \alpha_N)^\top$$

求导并令其等于 0 得

$$\frac{\partial \mathrm{J}}{\partial \vec{\omega}} = \vec{\omega} - \sum_{p=1}^{N} \alpha_p \left(d^p \vec{x}^p\right) = 0 \Rightarrow \vec{\omega} = \sum_{p=1}^{N} \alpha_p \left(d^p \vec{x}^p\right)$$

$$\frac{\partial \mathrm{J}}{\partial b} = \sum_{p=1}^{N} \alpha_p d^p = 0 \Rightarrow \sum_{p=1}^{N} \alpha_p d^p = 0$$

将上述结果代入 J 中得到新的目标函数

$$\mathrm{Q}(\vec{\alpha}) = \frac{1}{2}\sum_{p=1}^{N} \alpha_p d^p \vec{x}^{p\top} \cdot \sum_{q=1}^{N} \alpha_q d^q \vec{x}^q - \sum_{p=1}^{N} \alpha_p d^p \left(\sum_{q=1}^{N} \alpha_q d^q \vec{x}^{q\top}\right) \vec{x}^p - \sum_{p=1}^{N} \alpha_p d^p b + \sum_{p=1}^{N} \alpha_p$$

$\because \sum_{p=1}^{N} \alpha_p d^p = 0 \quad \therefore \sum_{p=1}^{N} \alpha_p d^p b = b \sum_{p=1}^{N} \alpha_p d^p = 0$

故化简得 $\mathrm{Q}(\vec{\alpha}) = \sum_{p=1}^{N} \alpha_p - \dfrac{1}{2} \sum_{p=1}^{N} \sum_{q=1}^{N} \alpha_p \alpha_q d^p d^q \vec{x}^{p\top} \vec{x}^q$

令 $\dfrac{\partial \mathrm{Q}(\vec{\alpha})}{\partial \alpha_p} = 0$，得到 N 个方程，进一步求得 α_p，得出 $\vec{\omega}_0 = \sum_{p=1}^{N} \alpha_p d^p \vec{x}^p$

当 $\alpha_p > 0$ 时，对应的样本为支撑向量

在 $d^s\left(\vec{\omega}_0^\top \vec{x}^s + b_0\right) = 1$ 中代入支撑向量，可求得 $b_0 = \dfrac{1}{d^s} - \vec{\omega}_0^\top \vec{x}^s$

2.3. 非线性支撑向量机

2.3.1. 概述

对于线性不可分的样本作非线性变换，变成线性可分，在映射的空间内采用前述的方法

映射后的最优分界面 $\vec{\omega}_0^\top \vec{\phi}(\vec{x}) + b_0 = 0 \quad \vec{\omega}_0 = \sum_{p=1}^{N} \alpha_p d^p \vec{\phi}\left(\vec{x}^p\right)$

$$g(\vec{x}) = \operatorname{sgn}\left(\sum_{p=1}^{N} \alpha_p d^p \vec{\phi}\left(\vec{x}^p\right)^\top \vec{\phi}(\vec{x}) + b_0\right) = \operatorname{sgn}\left(\sum_{p=1}^{N} \alpha_p d^p K\left(\vec{x}^p, \vec{x}\right) + b_0\right)$$

核函数 $K\left(\vec{x}^p, \vec{x}\right) = \vec{\phi}\left(\vec{x}^p\right)^\top \vec{\phi}(\vec{x}) = \sum_j \phi_j\left(\vec{x}^p\right)^\top \phi_j\left(\vec{x}^p\right)$

- $K\left(\vec{x}^p, \vec{x}\right) = K\left(\vec{x}, \vec{x}^p\right)$
- 最优分界面 $\sum_{p=1}^{N} \alpha_p d^p K\left(\vec{x}^p, \vec{x}\right) + b_0 = 0$

常用的核函数如下：

- $K\left(\vec{x}, \vec{x}^p\right) = \exp\left\{-\frac{\left|\left|\vec{x} - \vec{x}^p\right|\right|}{2\sigma^2}\right\}$
- $K\left(\vec{x}, \vec{x}^p\right) = \left(\vec{x}^\top \vec{x}^p + 1\right)^M$
- sigmoid 函数

2.3.2. 算法

已知输入样本 $\vec{x}^p, p = 1, 2, \cdots, N$ 和对应的输出 $d^p \in \{-1, 1\}$，非线性变换 $\vec{\phi}\left(\vec{x}^p\right)$

- 在约束条件 $\sum_{p=1}^{\alpha} \alpha_p d^p = 0$ 及 $\alpha_p \geqslant 0$ 的条件下，使目标函数

$$\mathrm{Q}(\vec{\alpha}) = \sum_{p=1}^{N} \alpha_p - \frac{1}{2} \sum_{p=1}^{N} \sum_{q=1}^{N} \alpha_p \alpha_q d^p d^q K\left(\vec{x}^p, \vec{x}^q\right)$$

由 $\frac{\partial Q(\vec{\alpha})}{\partial \vec{\alpha}} = 0$，求得 $\alpha_1, \alpha_2, \cdots, \alpha_N$

- 计算 $\vec{\omega}_0 = \sum_{p=1}^{N} \alpha_p d^p \vec{\phi}\left(\vec{x}^p\right) \quad b_0 = \frac{1}{d^s} - \vec{\omega}_0^\top \vec{\phi}\left(\vec{x}^s\right)$
- 对测试集样本，求 $K\left(\vec{x}, \vec{x}^p\right)$，计算 $g(\vec{x}) = \operatorname{sgn}\left(\sum_{p=1}^{N} d^p \alpha_p K\left(\vec{x}^p, \vec{x}\right) + b_0\right)$

  若 $\begin{cases} g(\vec{x}) = 1 & \vec{x} \in C_1 \\ g(\vec{x}) = -1 & \vec{x} \in C_2 \end{cases}$

例：SVM 求解异或问题

$$\vec{x} = \begin{cases} (-1,-1)^\top & d = -1 \\ (-1,1)^\top & d = 1 \\ (1,-1)^\top & d = 1 \\ (1,1)^\top & d = -1 \end{cases} \quad \text{核函数 } K\left(\vec{x}, \vec{x}^p\right) = \left(1 + \vec{x}^\top \vec{x}^p\right)^2$$

其中 $\vec{x}^p = \left(x_1^p, x_2^p\right)^\top \quad \vec{x} = (x_1, x_2)^\top \quad p = 1,2,3,4$

则可知 $K\left(\vec{x}, \vec{x}^p\right) = 1 + \left(x_1^p\right)^2 x_1^2 + 2x_1^p x_2^p x_1 x_2 + \left(x_2^p\right)^2 x_2^2 + 2x_1^p x_1 + 2x_2^p x_2$

$\vec{\phi}(\vec{x}) = \left(1, x_1^2, \sqrt{2}x_1x_2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2\right)^\top \quad \vec{\phi}\left(\vec{x}^p\right) = \left(1, \left(x_1^p\right)^2, \sqrt{2}x_1^p x_2^p, \left(x_2^p\right)^2, \sqrt{2}x_1^p, \sqrt{2}x_2^p\right)^\top$

故可知

$$[K] = \begin{bmatrix} K\left(\vec{x}^1, \vec{x}^1\right) & K\left(\vec{x}^1, \vec{x}^2\right) & K\left(\vec{x}^1, \vec{x}^3\right) & K\left(\vec{x}^1, \vec{x}^4\right) \\ K\left(\vec{x}^2, \vec{x}^1\right) & K\left(\vec{x}^2, \vec{x}^2\right) & K\left(\vec{x}^2, \vec{x}^3\right) & K\left(\vec{x}^2, \vec{x}^4\right) \\ K\left(\vec{x}^3, \vec{x}^1\right) & K\left(\vec{x}^3, \vec{x}^2\right) & K\left(\vec{x}^3, \vec{x}^3\right) & K\left(\vec{x}^3, \vec{x}^4\right) \\ K\left(\vec{x}^4, \vec{x}^1\right) & K\left(\vec{x}^4, \vec{x}^2\right) & K\left(\vec{x}^4, \vec{x}^3\right) & K\left(\vec{x}^4, \vec{x}^4\right) \end{bmatrix} = \begin{bmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{bmatrix}$$

$$\mathrm{Q}(\vec{\alpha}) = \sum_{p=1}^{N} \alpha_p - \frac{1}{2} \sum_{p=1}^{N} \sum_{q=1}^{N} \alpha_p \alpha_q d^p d^q K\left(\vec{x}^p, \vec{x}^q\right)$$

$$\frac{\partial Q}{\partial \alpha_i} = 0 \;\; i = 1,2,3,4 \;\; \Rightarrow \begin{cases} 9\alpha_1 - \alpha_2 - \alpha_3 - \alpha_4 = 1 \\ -\alpha_1 + 9\alpha_2 + \alpha_3 - \alpha_4 = 1 \\ -\alpha_1 + \alpha_2 + 9\alpha_3 - \alpha_4 = 1 \\ \alpha_1 - \alpha_2 - \alpha_3 + 9\alpha_4 = 1 \end{cases} \quad \Rightarrow \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \frac{1}{8}$$

$$\vec{\omega}_0 = \sum_{p=1}^{N} \alpha_p d^p \vec{\phi}\left(\vec{x}^p\right) = \left(0, 0, -\frac{1}{\sqrt{2}}, 0, 0\right)^\top$$

$b_0 = \frac{1}{d^s} - \vec{\omega}_0^\top \vec{\phi}\left(\vec{x}^s\right)$，取 $\vec{x}_1$ 计算 b_0 得 $b_0 = 0$

最优分界面

$$\vec{\omega}_0^\top \vec{\phi}(\vec{x}) + 0 = 0 \Rightarrow \begin{bmatrix} 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \end{bmatrix} = 0 \Rightarrow -x_1x_2 = 0$$

最终的 $g(\vec{x}) = \text{sgn}\{-x_1, x_2\}$

注：$\vec{\omega}_0 \vec{\phi}(\vec{x}) + b_0 = \left(\sum_{p=1}^{N} \alpha_p d^p \vec{\phi}\left(\vec{x}^p\right)\right)^\top \vec{\phi}(\vec{x}) + b_0 = \left(\sum_{p=1}^{N} \alpha_p d^p\right) K\left(\vec{x}^p, \vec{x}\right) + b_0$

$b_0 = \frac{1}{d^s} - \vec{\omega}_0^\top \vec{\phi}\left(\vec{x}^s\right) = \frac{1}{d^s} - \left(\sum_{p=1}^{N} \alpha_p d^p \vec{\phi}\left(\vec{x}^p\right)^\top\right) \vec{\phi}\left(\vec{x}^s\right) = \frac{1}{d^s} - \left(\sum_{p=1}^{N} \alpha_p d^p\right) K\left(\vec{x}^s, \vec{x}^p\right)$

由此可以看出，只需要知道核函数即可计算结果，无需知道具体的线性变换函数 $\vec{\phi}$