五. SVM

1. 决策面方程,

二维:

$$y = ax + b$$

$$x_2 = ax_1 + b$$

$$ax_1 - x_2 + b = 0$$

$$[a \ -1]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b = 0$$

$$\Downarrow$$

$$w^T x + \gamma = 0$$

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

w 为直线的法向量

多维:

$$w^T x + \gamma = 0$$

$$w = [w_1, w_2, \dots, w_n]^T$$

$$x = [x_1, x_2, \dots, x_n]^T$$

(2). 分类间隔

$$d = \frac{|w^T x + \gamma|}{\|w\|}$$

$$\|w\| = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$$

(3). 约束条件

对每个样本点加上类别标签 y_i

$$y_i = \begin{cases} 1 & \text{正样本} \\ -1 & \text{负样本} \end{cases}$$

$$\begin{cases} w^T x_i + \gamma > 0 & y_i = 1 \\ w^T x_i + \gamma < 0 & y_i = -1 \end{cases}$$

$$\begin{cases} \frac{w^T x_i + \gamma}{\|w\|} \geq d & \forall y_i = 1 \\ \frac{w^T x_i + \gamma}{\|w\|} \leq -d & \forall y_i = -1 \end{cases}$$

$$\begin{cases} w_d^T x_i + \gamma_d \geq 1 & \forall y_i = 1 \\ w_d^T x_i + \gamma_d \leq -1 & \forall y_i = -1 \end{cases}$$

$$w_d = \frac{w}{\|w\| d} \quad \gamma_d = \frac{\gamma}{\|w\| d}$$

$$y_i(w_d^T x_i + \gamma_d) \geq 1 \quad \forall x_i$$

$$\cancel{|w_d^T x_i + \gamma_d|}$$

$|w_d^T x_i + \gamma_d| = 1$ $\forall$ 支持向量上的样本点 x_i

$$d = \frac{|w^T x + \gamma|}{\|w\|} = \frac{1}{\|w\|}$$

求解 d 最大化 $\Rightarrow \min \|w\|$

$$\Downarrow$$

$$\min \frac{1}{2}\|w\|^2$$

$$\begin{cases} \min \frac{1}{2}\|w\|^2 \\ s.t. \quad y_i(w^T x_i + b) \geq 1, i = 1, 2, \dots, n \end{cases}$$

↓

subject to：服从某某条件

即构造一个函数，在可行区域内与原函数相同，可行外无穷大的最小化问题，

(4) 拉格朗日函数

将有约束的优化问题转换为无约束的优化问题

$$L(w,b,a) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} a_i\left(y_i(w^T x_i + b) - 1\right)$$

拉格朗日乘子

$$\theta(w) = \max_{a_i \ge 0} L(w,b,a)$$

当样本点不满足约束条件：$y_i(w^T x_i + b) < 1$

a_i 设为 $+\infty$

满足 $y_i(w^T x_i + b) \ge 1$

$\theta(w)$ 为原函数本身

$$\theta(w) = \begin{cases} \frac{1}{2}\|w\|^2 & x \in \text{可区域} \\ +\infty & x \in \text{非可行区域} \end{cases}$$

等价于带约束条件的原函数

$$\min_{w,b} \theta(w) = \min_{w,b} \max_{a_i \ge 0} L(w,b,a) = p^*$$ ← 表示最优解

拉格朗日对偶性:

$$\max_{a_i \geq 0} \min_{w,b} L(w,b,a) = d^*$$

$$d^* \leq p^*$$

何时 $d = p$:

① 是凸优化问题

② 满足KTT条件

先求 $L(w,b,a)$ 关于 w,b 的最小化.

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{n} a_i y_i x_i$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} a_i y_i = 0.$$

$$\begin{aligned} L(w,b,a) &= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} a_i \left(y_i (w^T x_i + b) - 1\right) \\ &= \frac{1}{2} w^T w - w^T \sum_{i=1}^{n} a_i y_i x_i - b \cancel{\sum_{i=1}^{n} a_i y_i} + \sum_{i=1}^{n} a_i \\ &= \sum_{i=1}^{n} a_i - \frac{1}{2} \left(\sum_{i=1}^{n} a_i y_i x_i\right)^T \cdot \sum_{i=1}^{n} a_i y_i x_i \\ &= \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_j x_i^T \cdot x_j \end{aligned}$$

$$\max_{a} \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_j x_i^T x_j$$

s.t. $a_i \geq 0, \; i = 1, \cdots, n$

$\sum_{i=1}^{n} a_i y_i = 0.$

KKT: $L(x,\lambda) = f(x) + \lambda g(x)$

$$\begin{cases} g(x) \le 0 \\ \lambda \ge 0 \\ \lambda g(x) = 0 . \end{cases}$$

$$W(a_2) = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j x_i^T x_j a_i a_j$$

$$= a_1 + a_2 + \sum_{i=3}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\left(\sum_{j=1}^{2} y_i y_j x_i^T x_j a_i a_j + \sum_{j=3}^{n} y_i y_j x_i^T x_j a_i a_j\right)$$

$$= a_1 + a_2 + \sum_{i=3}^{n} a_i - \frac{1}{2}\sum_{i=1}^{2}(\cdots) - \frac{1}{2}\sum_{i=3}^{n}(\cdots)$$

$$= a_1 + a_2 + \sum_{i=3}^{n} a_i - \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}(\cdots) - \sum_{i=1}^{2}\sum_{j=3}^{n}(\cdots) - \frac{1}{2}\sum_{i=3}^{n}\sum_{j=3}^{n}(\cdots)$$

$$= a_1 + a_2 - \frac{1}{2}a_1^2 x_1^T x_1 - \frac{1}{2}a_2^2 x_2^T x_2 - y_1 y_2 a_1 a_2 x_1^T x_2 - y_1 a_1 \sum_{j=3}^{n} a_j y_j x_1^T x_j - y_2 a_2 \sum_{j=3}^{n} a_j y_j x_2^T x_j + \sum_{i=3}^{n} a_i - \frac{1}{2}\sum_{i=3}^{n}\sum_{j=3}^{n}\cdots$$

定义: $f(x_i) = \sum_{j=1}^{n} a_j y_j x_i^T x_j + b$

$$v_i = \sum_{j=3}^{n} a_j y_j x_i^T x_j = f(x_i) - \sum_{j=1}^{2} a_j y_j x_i^T x_j - b .$$

$$W(a_2) = a_1 + a_2 - \frac{1}{2}a_1^2 x_1^T x_1 - \frac{1}{2}a_2^2 x_2^T x_2 - y_1 y_2 a_1 a_2 x_1^T x_2$$
$$- y_1 a_1 v_1 - y_2 a_2 v_2 + \text{constant}.$$

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

$\therefore a_1 y_1 + a_2 y_2 = B$

$a_1 = \gamma - s a_2$ $\quad \gamma = B y_1$, $s = y_1 y_2$

把 a_1 代入

$$W(a_2) = \gamma - s a_2 + a_2 - \frac{1}{2}(\gamma - s a_2)^2 x_1^T x_1 - \frac{1}{2} a_2^2 x_2^T x_2 -$$
$$s(\gamma - s a_2) a_2 x_1^T x_2 - y_1(\gamma - s a_2) v_1 - y_2 a_2 v_2 + \text{constant}$$

$$\frac{\partial W(a_2)}{\partial a_2} = -s + 1 + \gamma s x_1^T x_1 - a_2 x_1^T x_1 - a_2 x_2^T x_2 - \gamma s x_1^T x_2$$
$$+ 2 a_2 x_1^T x_2 + y_2 v_1 - y_2 v_2 = 0.$$

将 $s = y_1 y_2$ 代入

$$a_2^{new} = \frac{y_2(y_2 - y_1 + y_1 \gamma (x_1^T x_1 - x_1^T x_2) + v_1 - v_2)}{x_1^T x_1 + x_2^T x_2 - 2 x_1^T x_2}$$

令 $E_i = f(x_i) - y_i$

$\eta = x_1^T x_1 + x_2^T x_2 - 2 x_1^T x_2$

$$a_2^{new} = a_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$$

$$\because \begin{cases} a_1^{old} = \gamma - s a_2^{old} \\ a_1^{new} = \gamma - s a_2^{new} \end{cases}$$

$$\therefore a_1^{new} = a_1^{old} + y_1 y_2 (a_2^{old} - a_2^{new})$$

$$y_1(w^T x_1 + b) = 1$$

$$\sum_{i=1}^{n} a_i y_i x_i^T x_1 + b = y_1$$

$$b_1^{new} = \underbrace{y_1 - \sum_{i=3}^{n} a_i y_i x_i^T x_1}_{\parallel} - a_1^{new} y_1 x_1^T x_1 - a_2^{new} y_2 x_2^T x_1$$

$$-E_1 + a_1^{old} y_1 x_1^T x_1 + a_2^{old} y_2 x_2^T x_1 + b^{old}$$

$$b_1^{new} = b^{old} - E_1 - y_1(a_1^{new} - a_1^{old}) x_1^T x_1 - y_2(a_2^{new} - a_2^{old}) x_2^T x_1$$

同理 b_2^{new}

SMO

① 计算误差

$$E_i = f(x_i) - y_i$$
$$= \sum_{j=1}^{n} a_j y_i x_i^T x_j + b - y_i$$

② 计算上下边界 L, H

$$L = \max(0, a_j^{old} - a_i^{old}), H = \min(C, C + a_j^{old} - a_i^{old}) \text{ if } y_i \neq y_j$$
$$L = \max(0, a_j^{old} + a_i^{old} - C), H = \min(C, a_j^{old} + a_i^{old}) \text{ if } y_j = y_i$$

③ 计算最优修改量

$$\eta = x_i^T x_i + x_j^T x_j - 2 x_i^T x_j$$

④ 更新 a_j

$$a_j^{new} = a_j^{old} + \frac{y_i (E_i - E_j)}{\eta}$$

⑤ 根据取值范围修剪 a_j

$$a_j^{new} = \begin{cases} H & a_j^{new} \geq H \\ a_j^{new} & L \leq a_j^{new} \leq H \\ L & a_j^{new} \leq L \end{cases}$$

⑥ 更新 a_i

$$a_i^{new} = a_i^{old} + y_i y_j (a_j^{old} - a_j^{new})$$

⑦ 更新 b_1, b_2

$$b_1^{new} = b^{old} - E_i - y_i(a_i^{new} - a_i^{old}) x_i^T x_i - y_i (a_j^{new} - a_j^{old}) x_j^T x_i$$

$$b_2^{new} = b^{old} - E_j - y_i (a_i^{new} - a_i^{old}) x_i^T x_j - y_i (a_j^{new} - a_j^{old}) x_j^T x_j$$

⑧ 更新 b.

$$b = \begin{cases} b_1 & 0 < a_i < C \\ b_2 & 0 < a_j < C \\ \frac{b_1 + b_2}{2} & \text{otherwise} \end{cases}$$