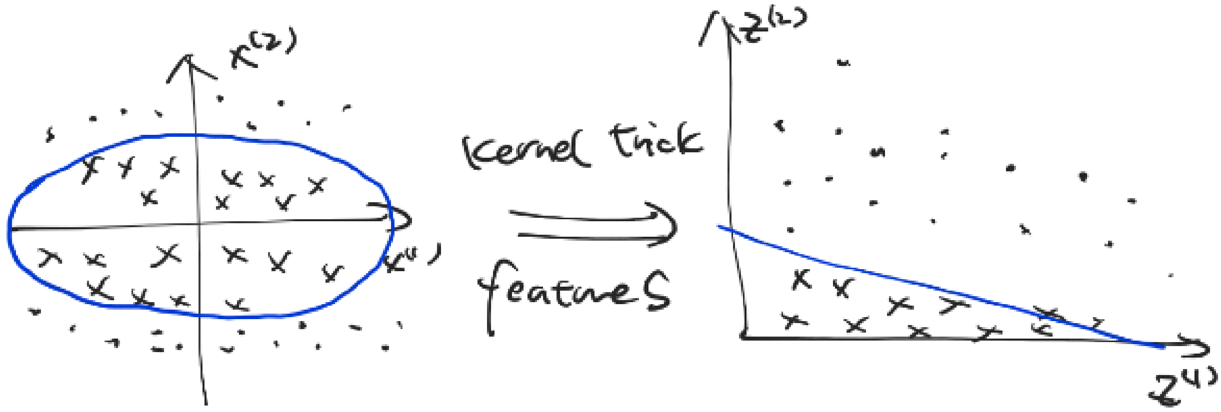
SVM

1. 核技巧 (kernel trick)



非线性 $X \subset R^2, x=(x^{(1)},x^{(2)})$　　线性 $Z \subset R^2, z=(z^{(1)},z^{(2)})$

$$z=\phi(x) = ((x^{(1)})^2, (x^{(2)})^2)$$

将椭圆 $w_1(x^{(1)})^2 + w_2(x^{(2)})^2 = 0$

⇒ 直线 $w_1(z^{(1)}) + w_2(z^{(2)}) = 0$

kernel function $K(x,y) =$

$$K(x,y) = \phi(x) \cdot \phi(y)$$

其中，ϕ 为 $X \to Z$ 的映射，$x, y \in X$，$\cdot$ 为内积

常用 kernel function：

① polynomial kernel function

$$K(x,y) = (x \cdot y + 1)^2$$

② Gaussian kernel function

$$K(x,y) = \exp\left(-\frac{\|x-y\|^2}{\ }\right)$$

$$K(x, z) = \exp\left(-\frac{\quad}{2\sigma^2}\right)$$

③ String kernel function

2. Sequential Minimal optimazation (SMO)

所求的最优化问题：

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

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

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

变量为 拉格朗日乘子 α_i，N个

SMO求解思路为：如果所有的 α_i 都满足 KKT，则为解

即：选择两个变量（如 α_1, α_2），固定其它变量（如 $\alpha_3, \cdots \alpha_N$）
构造一个二次规划，求解 (α_1, α_2)

注意，α_1、α_2 并不独立 $\alpha_1^* = -y_1\sum_{i=2}^{N}\alpha_i y_i$

SMO ⇒ ① 如何求解 α_1^*, α_2^*
② 如何选择 α_1, α_2

2.1 求解两个变量二次规划的方法

假设所选变量为 α_1, α_2，固定 $\alpha_i (i=3, \cdots N)$

$$\min_{\alpha_1, \alpha_2} W(\alpha_1, \alpha_2) = \frac{1}{2} K_{11} \alpha_1^2 + \frac{1}{2} K_{22} \alpha_2^2 + y_1 y_2 K_{12} \alpha_1 \alpha_2 - (\alpha_1 + \alpha_2) + y_1 \alpha_1 \sum_{i=3}^{N} y_i \alpha_i K_{i1} + y_2 \alpha_2 \sum_{i=3}^{N} y_i \alpha_i K_{i2}$$

$$\text{s.t.} \quad \alpha_1 y_1 + \alpha_2 y_2 = -\sum_{i=3}^{N} y_i \alpha_i = \varsigma$$

$$0 \le \alpha_i \le C, \quad i = 1, 2$$

初始可行解 $\alpha_1^{old}, \alpha_2^{old}$；最优解 $\alpha_1^{new}, \alpha_2^{new}$；
未进行 $0 \le \alpha_i \le C$ 约束解 α_1^{unc}；α_2^{unc}；

$$L \le \alpha_2^{new} \le H$$

if $y_1 \ne y_2$，(如 $y_1 = -1, y_2 = +1$)：

$$L = \max(0, \alpha_2^{old} - \alpha_1^{old}),$$
$$H = \min(C, C + \alpha_2^{old} - \alpha_1^{old})$$

if $y_1 = y_2$，(如 $y_1 = y_2 = 1 / -1$)：

$$L = \max(0, \alpha_2^{old} + \alpha_1^{old} - C)$$
$$H = \min(C, \alpha_2^{old} - \alpha_1^{old})$$

令 $$g(x) = \sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b$$

$$E_i = g(x_i) - y_i = \left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b \right) - y_i$$
$$i = 1, 2$$

则未经约束方向 $0 \le \alpha_i \le C$ 的解为：

$$\alpha_2^{unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$$

其中 $\eta = K_{11} + K_{22} - 2K_{12} = \|\Phi(x_1) - \Phi(x_2)\|^2$

约束后 α_2 的解为：

$$\alpha_2^{new} = \begin{cases} H, & \alpha_2^{unc} > H \\ \alpha_2^{unc}, & L \le \alpha_2^{unc} \le H \\ L, & \alpha_2^{unc} < L \end{cases}$$

$$\alpha_1^{new} = \alpha_1^{old} + y_1 y_2 (\alpha_2^{old} - \alpha_2^{new})$$

2.2. 如何选择这两个变量 ⇒ 启发式方法

(1) 1st 变量 ⇒ 违反KKT最严重的样本点 (x_i, y_i) ⇒ α_i

a.k.a. 外层循环

KKT Condition：

$$\begin{cases} \alpha_i = 0 \iff y_i g(x_i) \ge 1 & \Rightarrow \text{边界内} \\ 0 < \alpha_i < C \iff y_i g(x_i) = 1 & \Rightarrow \text{边界上 (SV)} \\ \alpha_i = C \iff y_i g(x_i) \le 1 & \Rightarrow \text{边界之间} \end{cases}$$

(2) 2nd 变量 ⇒ 使 1st α 有足够大的变化

a.k.a. 内层循环

$$\alpha_2^{new} \propto |E_1 - E_2| \begin{cases} \max(E_i), & \text{if } E_1 < 0 \\ \min(E_i), & \text{if } E_1 > 0 \end{cases}$$

将所有E_i存在一列表中.

(3) 计算 b 和 E_i

if $0 < \alpha_i < C \Rightarrow y_i g(x_i) = 1$

$$\Rightarrow y_i \left[\sum_{j=1}^{N} \alpha_j y_j K(x_j, x_i) + b \right] = 1$$

$$\Rightarrow \sum_{i=1}^{N} a_i y_i K_{i1} + b = y_1 \quad (i=1)$$

$$\Rightarrow b_1^{new} = y_1 - \sum_{i=3}^{N} \alpha_i y_i K_{i1} - \alpha_1^{new} y_1 K_{11} - \alpha_2^{new} y_2 K_{21}$$

$$E_i^{new} = \sum_{S} y_j \alpha_j K(x_i, x_j) + b^{new} - y_i$$

其中，S是所有支持向量x_j的集合

$\boxed{\text{SMO算法：}}$

Input：$T = \{(x_1, y_1), \cdots, (x_N, y_N)\}$，精度 ε

Output：α

① $\alpha^{(0)} = 0 \quad k = 0$

② 选取 $\alpha_1^{(k)}, \alpha_2^{(k)}$, $\Rightarrow \begin{cases} \alpha_1^{(k+1)} = \alpha_1^{new} \\ \alpha_2^{(k+1)} = \alpha_2^{new} \end{cases}$

update b, E_i

③ ε精度是否满足：$\sum_{i=1}^{N} \alpha_i y_i = 0$

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

$$y_i \cdot g(x_i) = \begin{cases} \ge 1, & \{x_i \mid \alpha_i = 0\} \\ = 1, & \{x_i \mid 0 < \alpha_i < C\} \\ \le 1, & \{x_i \mid \alpha_i = C\} \end{cases}$$

满足，则输出 $\hat{\alpha} = \alpha^{(k+1)}$

否则 转到 ②

上次修改：22:05