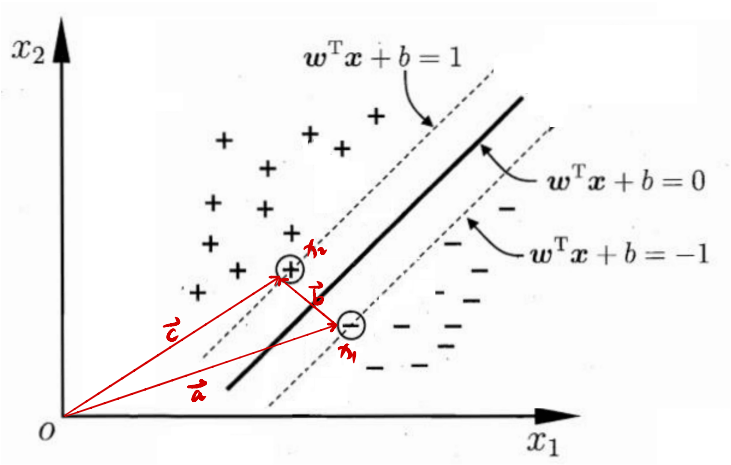
支持向量机（Support Vector Machine, SVM）

1. 线性可分的情况



相关知识：

$w^Tx+b=0$

↓

向量 w 便是这个直线方程的法向量

⊕、⊖ 两点（即向量 $\vec{c}$、$\vec{a}$）能称为"支持向量"（Support Vector）

$\Downarrow$

我们的目标，是找到能够区分 ⊕/⊖ 样本的最大间隔（margin）

↓

使分类的泛化能力最好

$\Downarrow$

构建一个超平面（hyperplane），

$$\begin{cases} w^Tx+b>0 & \text{if } y=1 \\ w^Tx+b<0 & \text{if } y=-1 \end{cases}$$

→ 若 $y(w^Tx+b)\geq 1$，则表示"分类正确"

↓

使数据点都在间隔外的区域

$\Downarrow$

假设我们在 $w^Tx+b=-1$ 上找到一个支持向量 x_1（即 $\vec{a}$），

那么其在 $w^Tx+b=1$ 上的关于分界面 $w^Tx+b=0$ 对称的点为 $x_1+\eta w$（即 $\vec{b}$），即：

↓

η 个单位法向量

$$w^T(x_1+\eta w)+b=1$$

$$\Rightarrow w^Tx_1+\eta w^Tw+b=1$$

$$\Rightarrow \underbrace{(w^Tx_1+b)}_{=-1}+\eta\|w\|^2=1$$

$$\Rightarrow \eta\|w\|^2=2$$

$$\Rightarrow \eta=\frac{2}{\|w\|^2}$$

$\Downarrow$

那么，我们的目标函数为：

$$\begin{cases} \max\limits_{w} \dfrac{2}{\|w\|^2} \\ \text{s.t. } y^{(i)}(w^Tx^{(i)}+b)\geq 1 \end{cases}$$

$$\Downarrow$$

目标函数求解复杂，我们：

Step 01 将其转化为拉格朗日函数形式

$$\begin{cases} \min\limits_{w} \frac{1}{2}\|w\|^2 \\ s.t.\ \ y^{(i)}(w^T x^{(i)}+b) \geq 1 \end{cases} \Rightarrow \underbrace{L(w,b,\pi)}_{\text{拉格朗日函数 (Lagrange Function)}} = \max_{\pi} \min_{w} \frac{1}{2}\|w\|^2 + \sum_{i=0}^{m} \pi^{(i)}\left[1-y^{(i)}(w^T x^{(i)}+b)\right]$$

Step 02 根据拉格朗日函数，求到其对偶函数（dual form）

$$\begin{cases} \dfrac{\partial L(w,b,\pi)}{\partial w} = 0 \Rightarrow w = \sum\limits_{i=1}^{m} \pi^{(i)} y^{(i)} x^{(i)} \\ \dfrac{\partial L(w,b,\pi)}{\partial b} = 0 \Rightarrow \sum\limits_{i=1}^{m} \pi^{(i)} y^{(i)} = 0 \end{cases}$$

$$\downarrow$$

将 w 值代入 $L(w,b,\pi)$，

$$L = \frac{1}{2}\underbrace{\|w\|^2}_{w^T w} + \sum_{i=0}^{m} \pi^{(i)}\left[1-y^{(i)}(w^T x^{(i)}+b)\right]$$

$$= \frac{1}{2}\Big(\underbrace{\sum_{i=0}^{m} \pi^{(i)} y^{(i)} x^{(i)T}}_{\text{标量 标量 向量}}\Big)\Big(\sum_{j=0}^{m} \pi^{(j)} y^{(j)} x^{(j)}\Big) + \sum_{i=0}^{m} \pi^{(i)}\Big[1-y^{(i)}\Big(\sum_{j=0}^{m} \pi^{(j)} y^{(j)} x^{(j)T} x^{(i)} + b\Big)\Big]$$

$$= \frac{1}{2}\Big(\sum_{i=0}^{m}\sum_{j=0}^{m} \pi^{(i)}\pi^{(j)} y^{(i)} y^{(j)} x^{(i)T} x^{(j)}\Big) + \sum_{i=0}^{m} \pi^{(i)} - \sum_{i=0}^{m}\sum_{j=0}^{m} \pi^{(i)}\pi^{(j)} y^{(i)} y^{(j)} x^{(i)T} x^{(j)} + \overbrace{\boxed{\sum_{i=0}^{m} \pi^{(i)} y^{(i)}}}^{0} b$$

$$\begin{cases} = \boxed{\sum\limits_{i=0}^{m} \pi^{(i)} - \frac{1}{2}\sum\limits_{i=0}^{m}\sum\limits_{j=0}^{m} \pi^{(i)}\pi^{(j)} y^{(i)} y^{(j)} x^{(i)T} x^{(j)}} \\ s.t. \begin{cases} \pi^{(i)} \geq 0 & (\text{KKT条件之一}) \\ \sum\limits_{i=1}^{m} \pi^{(i)} y^{(i)} = 0 & (\text{之前 } \frac{\partial L}{\partial b} \text{ 推导出的结果}) \end{cases} \end{cases}$$

→ 称为原式的对偶式

↓

将对 w, b, π 的求解转化为对 π 的求解
简化了计算

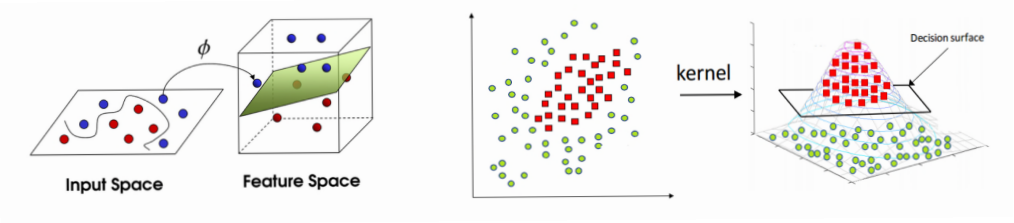
在KKT条件下，
对偶式能够提供与原式一样的解

Step 03 使用对偶函数替代原式进行求解，即：

$$\begin{cases} \max\limits_{\pi} \sum\limits_{i=0}^{m} \pi^{(i)} - \frac{1}{2}\sum\limits_{i=0}^{m}\sum\limits_{j=0}^{m} \pi^{(i)}\pi^{(j)} y^{(i)} y^{(j)} x^{(i)T} x^{(j)} \\ s.t.\ \ \pi^{(i)} \geq 0 \qquad i=1,2,3,\cdots,m \\ \qquad \sum\limits_{i=1}^{m} \pi^{(i)} y^{(i)} = 0 \end{cases}$$

这是一个二次规划问题，可使用SMO算法进行求解

2.线性不可分的情况



即：核函数（Kernel Function）

将数据根据某种规则投射到高维空间中，使其线性可分。

$$\begin{cases} \max\limits_{\eta} \sum\limits_{i=0}^{m} \eta^{(i)} - \frac{1}{2}\sum\limits_{i=0}^{m}\sum\limits_{j=0}^{m} \eta^{(i)}\eta^{(j)} y^{(i)} y^{(j)} \underbrace{x^{(i)T} x^{(j)}}_{K(x_i,\, x_j)} \\ \text{s.t.} \quad \eta^{(i)} \geq 0 \qquad i=1,2,3,\cdots,m \\ \qquad \sum\limits_{i=1}^{m} \eta^{(i)} y^{(i)} = 0 \end{cases}$$

这是线性核

也可以用其他核进行替换

名称	表达式	参数
线性核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j$	
多项式核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j)^d$	$d \geqslant 1$ 为多项式的次数
高斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\lVert \boldsymbol{x}_i - \boldsymbol{x}_j \rVert^2}{2\sigma^2}\right)$	$\sigma > 0$ 为高斯核的带宽(width)
拉普拉斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\lVert \boldsymbol{x}_i - \boldsymbol{x}_j \rVert}{\sigma}\right)$	$\sigma > 0$

（常用核函数）

$\Downarrow$

现在，一般使用人工神经网络来实现非线性分类，并且效果更好

※ 以上内容是基于我对SVM的理解所写，可能有误。