SVM

等式约束优化

$\min_x f(x)$

$s.t.\quad h_i(x)=0, i=1,2,\dots,m$

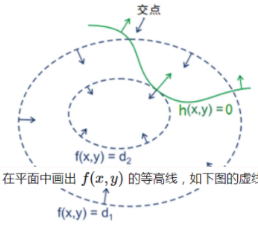
此时不一定能找到使得 $\nabla_x f(x)$ 为 0 的点，只需找到在可行域内使得 $f(x)$ 最小的值即可

常用的方法即为拉格朗日

$$\boxed{L(x,\alpha)=f(x)+\sum_{i=1}^{m}\alpha_i h_i(x)}$$

求解方法如下：
$$\begin{cases}\nabla_x L(x,\alpha)=0\\ \nabla_\alpha L(x,\alpha)=0\end{cases}$$

直观解释



这就是等式优化的拉格朗日乘子法的解释

目标函数是 $f(x,y)$，在平面中画出 $f(x,y)$ 的等高线，如下图的虚线所示，并只给出一个约束等式 $h(x,y)=0$

没交集肯定不是解,只有相交或者相切可能是解

但相交得到的一定不是最优值，因为相交意味着肯定还存在其它的等高线在该条等高线的内部或者外部，使得新的等高线与目标函数的交点的值更大或者更小，只有等高线与目标函数的曲线相切的时候，

因此给出结论：拉格朗日乘子法取得极值的必要条件是目标函数与约束函数相切，这时两者的法向量是平行的，即

所以有 $\nabla_x f(x)-\alpha\nabla_x h(x)=0$

不等式约束优化

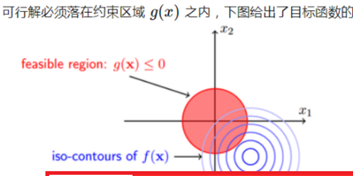
$\min_x f(x)$

$s.t.\quad g(x)\le 0$

对应的 Lagrangian 与图形分别如下所示：

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

可行解必须落在约束区域 $g(x)$ 之内，下图给出了目标函数的等高线与约束：



当可行解 x 落在 $g(x)=0$ 即边界上，此时等价于等式约束

当当约束区域包含目标函数原有的可行解时，对应 $g(x)<0$ 的情况，这时约束条件不起作用

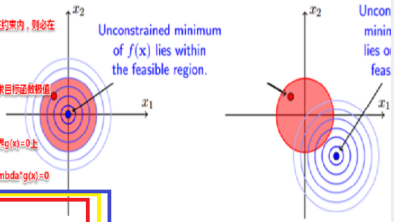
当当约束区域不包含目标函数原有的可行解时，此时加上约束后可行解落在边界 $g(x)=0$ 上。 **等式约束优化**

如果不受约束的可行解不在约束内，则必在边界g(x)=0上，

综述在g(x)<0上，则直接求目标函数极值即可，此时有lambda=0

不在g(x)<0中，则必在边界g(x)=0上

综述 必有KKT条件之一 lambda*g(x)=0



无论哪种， $\lambda g(x)=0$

$$\min_x f(x)$$
$$\boxed{\begin{aligned}s.t.\quad & h_i(x)=0,\ i=1,2,\dots,m\\ & g_j(x)\le 0,\ j=1,2,\dots,n\end{aligned}}$$

列出 Lagrangian 得到无约束优化问题：

$$\boxed{L(x,\alpha,\beta)=f(x)+\sum_{i=1}^{m}\alpha_i h_i(x)+\sum_{j=1}^{n}\beta_i g_i(x)}$$

经过之前的分析，便得知加上不等式约束后可行解 x 需要满足的就是以下的 KKT 条件：

$$\begin{aligned}\nabla_x L(x,\alpha,\beta)&=0 &(1)\\ \boxed{\beta_j g_j(x)=0,\ j=1,2,\dots,n} & &(2)\\ h_i(x)=0,\ i=1,2,\dots,m & &(3)\\ g_j(x)\le 0,\ j=1,2,\dots,n & &(4)\\ \boxed{\beta_j\ge 0,\ j=1,2,\dots,n} & &(5)\end{aligned}$$

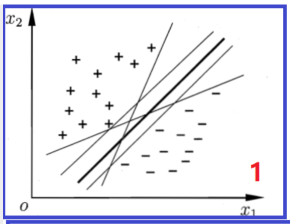
满足 KKT 条件后极小化 Lagrangian 即可得到在不等式约束条件下的可行解。 KKT 条件看起来很多，其实很好理解：

(1) ：拉格朗日取得可行解的必要条件；

(2) ：这就是以上分析的一个比较有意思的约束，称作松弛互补条件；

(3) ∼ (4) ：初始的约束条件；

(5) ：不等式约束的 Lagrange Multiplier 需满足的条件。



问题变成 要找到能满足式(6.3)中约束的参数 $\boldsymbol{w}$ 和 b，使得 γ 最大，即

$$\max_{\boldsymbol{w},b} \frac{2}{\|\boldsymbol{w}\|}$$
$$\text{s.t. } y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\geqslant 1,\quad i=1,2,\ldots,m.$$

4

为了最大化间隔，仅需最大化 $\|\boldsymbol{w}\|^{-1}$，这等价于最小化 $\|\boldsymbol{w}\|^2$.

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2$$
$$\text{s.t. } y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\geqslant 1,\quad i=1,2,\ldots,m.$$

5

$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b=0$，

$\boldsymbol{w}=(w_1;w_2;\ldots;w_d)$ 为法向量，决定了超平面的方向；

b 为位移项，决定了超平面与原点之间的距离.

样本空间中任意点 $\boldsymbol{x}$ 到超平面 $(\boldsymbol{w},b)$ 的距离可写为

$$r=\frac{|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b|}{\|\boldsymbol{w}\|}.$$

2

SVM问题提出

分类正确则有：若 $y_i=+1$，则有 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b>0$；即
若 $y_i=-1$，则有 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b<0$.

$$\begin{cases}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b\geqslant +1, & y_i=+1;\\ \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b\leqslant -1, & y_i=-1.\end{cases}$$

距离超平面最近的这几个训练样本点使得上式成立，它们被称为“支持向量”(support vecto

两个异类支持向量到超平面的距离之和为 $\gamma=\frac{2}{\|\boldsymbol{w}\|}$ 它被称为“间隔”(margin)

3

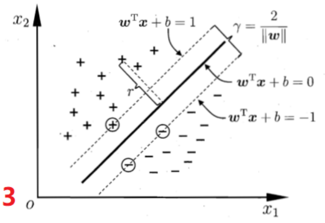


图 6.2 支持向量与间隔

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2$$
$$\text{s.t. } y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\geqslant 1,\quad i=1,2,\ldots,m.$$

SVM基本型

使用拉格朗日乘子法可得到其“对偶问题”(dual problem)

$$L(\boldsymbol{w},b,\boldsymbol{\alpha})=\frac{1}{2}\|\boldsymbol{w}\|^2+\sum_{i=1}^{m}\alpha_i\left(1-y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\right),$$

其实根据前面的拉格朗日乘子法和KKT条件的介绍，很容易写出KKT条件

$$\begin{cases}\alpha_i\geqslant 0;\\ y_if(\boldsymbol{x}_i)-1\geqslant 0;\\ \alpha_i\left(y_if(\boldsymbol{x}_i)-1\right)=0.\end{cases}$$

理解这里的KKT条件 对任意训练样本 $(\boldsymbol{x}_i,y_i)$，总有 $\alpha_i=0$ 或 $y_if(\boldsymbol{x}_i)=1$.

若 $\alpha_i=0$，则该样本不会对 $f(\boldsymbol{x})=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b=\sum_{i=1}^{m}\alpha_iy_i\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}+b$ 产生任何影响，因为没有贡献

若 $\alpha_i>0$，则必有 $y_if(\boldsymbol{x}_i)=1$，所对应的样本点位于最大间隔边界上，是一个支持向量

这显示出支持向量机的一个非常重要的性质：
训练完成后大部分样本点不需要保留，最终的模型只与支持向量有关

$$L(w,b,\alpha)=\frac{1}{2}\|w\|^2+\sum_{i=1}^{m}\alpha_i(1-y_i(w^Tx_i+b))$$

还有乘子没有求出

$$=\frac{1}{2}\|w\|^2+\sum_{i=1}^{m}\alpha_i-b\sum_{i=1}^{m}\alpha_iy_i-\sum_{i=1}^{m}\alpha_iy_iw^Tx_i$$

$$\frac{\partial L}{\partial w}=0\Rightarrow \boxed{w=\sum_{i=1}^{m}\alpha_iy_ix_i}$$

$$\frac{\partial L}{\partial b}=0\Rightarrow \boxed{0=\sum_{i=1}^{m}\alpha_iy_i}$$

$$f(\boldsymbol{x})=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b=\sum_{i=1}^{m}\alpha_iy_i\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}+b$$

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_jy_iy_j\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$$
$$\text{s.t. } \sum_{i=1}^{m}\alpha_iy_i=0,$$
$$\alpha_i\geqslant 0,\quad i=1,2,\ldots,m.$$

最终需要求解的对偶问题的化简形式

这是一个二次规划问题，可使用通用的二次规划算法来求解；

更高效的算法

SMO (Sequential Minimal Optimization)

SMO怎么求乘子alpha，不提，自己有需要再去查

因为 $y_s\left(\sum_{i\in S}\alpha_iy_ix_i^Tx_s+b\right)=1$

$$\Rightarrow y_s^2\left(\sum_{i\in S}\alpha_iy_ix_i^Tx_s+b\right)=y_s$$

$$\Rightarrow \sum_{i\in S}\alpha_iy_ix_i^Tx_s+b=y_s$$

$$\Rightarrow b=y_s-\sum_{i\in S}\alpha_iy_ix_i^Tx_s$$

更为鲁棒性的做法求均值

$$b=\frac{1}{|S|}\sum_{s\in S}\left(y_s-\sum_{i\in S}\alpha_iy_ix_i^Tx_s\right)$$

其中 $S=\{i\mid\alpha_i>0,\ i=1,2,\ldots,m\}$ 为所有支持向量的下标集.