第6章 支持向量机

6.1 间隔与支持向量

划分超平面: $w^Tx+b=0$ (w,b)

$w=(w_1;w_2;\cdots w_d)$ 法向量,决定超平面的方向。

b: 位移项 决定超平面与原点的距离

样本空间点x到(w,b)的距离:

$$r=\frac{|w^Tx+b|}{||w||}$$

正确分类 (x_i,y_i):
$$\begin{cases} y_i=+1, & w^Tx+b>0 \\ y_i=-1, & w^Tx+b<0 \end{cases}$$

令
$$\begin{cases} w^Tx_i+b\geq+1, & y_i=+1 \\ w^Tx_i+b\leq-1, & y_i=-1 \end{cases}$$

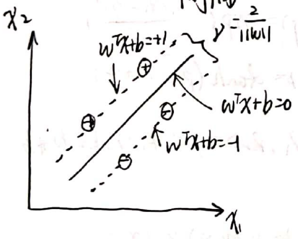
使上式等号成立的点(x_i,y_i)称为"支持向量"

两个异类支持向量到超平面的距离之和

$$\gamma=\frac{2}{||w||}$$

↓

间隔



找到"最大间隔"的划分超平面

$$\max_{w,b}\ \frac{2}{||w||}$$

s.t. $y_i(w^Tx_i+b)\geq1$ $i=1,2,\cdots m$

最大化 $\frac{2}{||w||}$ $\Leftrightarrow$ 最小化 $\frac{1}{2}||w||^2$

∴ 重写为

$$\max_{w,b}\ \frac{1}{2}||w||^2$$

s.t. $y_i(w^Tx_i+b)\geq1$ $i=1,2,\cdots m$

为二次规划,参数为w,b

6.2 对偶问题

① 更容易求解
② 自然引入核函数 ☆

使用拉格朗日乘子法:

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

其中 $\alpha=(\alpha_1,\alpha_2\cdots\alpha_m)$

$$\frac{\partial L}{\partial w}=w-\sum_{i=1}^{m}\alpha_iy_ix_i=0 \Rightarrow w=\sum_{i=1}^{m}\alpha_iy_ix_i \quad ②$$

$$\frac{\partial L}{\partial b}=-\sum_{i=1}^{m}\alpha_iy_i=0 \quad ③$$

②式带入①式:

$$L(w,b,\alpha)=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_jy_iy_j(x_i\cdot x_j)-\sum_{i=1}^{m}\alpha_iy_i\left(\left(\sum_{j=1}^{m}\alpha_jy_jx_j\right)\cdot x_i+b\right)+\sum_{i=1}^{m}\alpha_i$$
$$=-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_jy_iy_j(x_i\cdot x_j)+\sum_{i=1}^{m}\alpha_i$$

s.t. $\sum_{i=1}^{m}\alpha_iy_i=0$

$\alpha_i\geq0$ $i=1,2\cdots m$

解出α后,即得w,b

$$\begin{cases} w=\sum_{i=1}^{m}\alpha_iy_ix_i \\ b=y_j-\sum_{i=1}^{m}\alpha_iy_i(x_i\cdot x_j) \end{cases}$$

$\alpha_j>0$ 对应的(x_j,y_j)

α_i是拉格朗日乘子 对应(x_i,y_i)

因为①式不等式

∴ 满足KKT条件

$$\begin{cases} \alpha_i\geq0 \\ y_if(x_i)-1\geq0 \\ \alpha_i(y_if(x_i)-1)=0 \end{cases}$$

对任意样本(x_i,y_i), $\alpha_i=0$ 或 $y_if(x_i)=1$

①若$\alpha_i=0$ 则不会在w的求和中出现

②若$\alpha_i>0$ 则 $y_if(x_i)=1$,位于最大间隔边界上 即支持向量

SMO算法: 固定α_i之外参数,求α_i上的极值

∵ $\sum_{i=1}^{m}\alpha_iy_i=0$,固定$\alpha_i$之外的其他变量,$\alpha_i$可由其他导出。每次选择两变量、固定其他参数。

SMO：

①选取一对 α_i, α_j

② 固定 α_i, α_j 以外参数，求解获取更新后的 α_i, α_j

先选背KKT条件程度最大的变量 α_i

后选一两变量对应样本间隔最大的 α_j

6.3 核函数

原始空间 $\xrightarrow{\text{映射}}$ 更高维特征空间

（线性不可分） （线性可分）

注：若原始空间有限维，属性有限，则一定存在一个高维特征空间使样本可分

$\phi(x)$ 为 x 映射后的特征向量

$f(x) = w^T\phi(x)+b$ (w, b)

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

s.t $y_i(w^T\phi(x_i)+b) \geq 1$

对偶问题，

$$\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m \alpha_i\alpha_j y_i y_j (\phi^T(x_i)\cdot\phi(x_j))$$

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

$\alpha_i \geq 0$ $i=1,2,\dots m$

计算 $\phi^T(x_i)\cdot\phi(x_j)$，映射到特征空间后的内积

$\left\{\begin{array}{l}\text{维数很高}\\ \text{无穷维}\end{array}\right. \rightarrow$ 直接算 $\phi^T(x_i)\cdot\phi(x_j)$ ⊗

设 $k(x_i, x_j) = \langle\phi(x_i), \phi(x_j)\rangle = \phi^T(x_i)\phi(x_j)$ } 核技巧

即 x_i, x_j 在特征空间的内积等于在原始空间通过函数 $k(\cdot,\cdot)$ 计算

则 $\max \sum_{i=1}^m \alpha_i - \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m \alpha_i\alpha_j y_i y_j k(x_i, x_j)$

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

$\alpha_i \geq 0$

则 $f(x) = w^T\phi(x)+b$

$= \sum_{i=1}^m \alpha_i y_i \phi^T(x_i)\cdot\phi(x)+b$

$= \underbrace{\sum_{i=1}^m \alpha_i y_i k(x, x_i)}_{\text{支持向量展开}} + b$

通过训练样本的核函数展开

核函数：令 $\mathcal{X}$ 是输入空间，$k(\cdot,\cdot)$ 是定义在 $\mathcal{X}\times\mathcal{X}$ 上的对称函数，则 k 是核函数当且仅当 $\forall D=\{x_1\dots x_m\}$ 核矩阵 K 总是半正定

$$K = \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_m) \\ \vdots & & \vdots \\ k(x_m,x_1) & \cdots & k(x_m,x_m)\end{bmatrix}$$

$\left\{\begin{array}{l}\text{对称函数}\\ \text{核矩阵半正定}\end{array}\right. \Rightarrow$ 核函数

常用：

① 线性核 $k(x_i,x_j) = x_i^T x_j$

② 多项式核 $k(x_i,x_j) = (x_i^T x_j)^d$

③ 高斯核 $k(x_i,x_j) = \exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right)$

④ 拉普拉斯核 $k(x_i,x_j) = \exp\left(-\frac{\|x_i-x_j\|}{\sigma}\right)$

⑤ sigmoid核 $k(x_i,x_j) = \tanh(\beta x_i^T x_j+\theta)$

且 1) k_1, k_2 为核 $\Rightarrow \gamma_1, \gamma_2 > 0$，$\gamma_1 k_1 + \gamma_2 k_2$ 为核

2) k_1, k_2 为核 $\Rightarrow$

$k_1 \otimes k_2(x,z) = k_1(x,z)k_2(x,z)$

也是核

3) k_1 是核 $\forall g(x)$

$k(x,z) = g(x)k_1(x,z)g(z)$

也是核

6.4

6.4 软间隔与正则化

硬间隔：所有样本均满足 $y_i(w^Tx_i+b)\geq 1$

软间隔：允许某些样本不满足

但数量应尽可能少

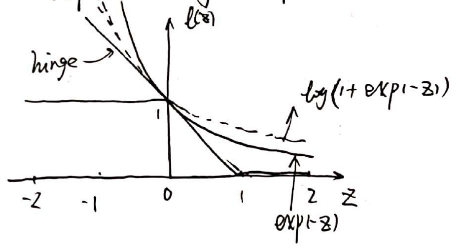
$$\min_{w,b}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{m}\ell_{0/1}(y_i(w^Tx_i+b)-1)\quad C>0$$

$$\ell_{0/1}(z)=\begin{cases}1 & \text{if } z<0\\ 0 & \text{otherwise}\end{cases}$$ 非凸 非连续 0/1损失

hinge损失 $\max(0,1-z)$

指数损失 $\exp(-z)$

对率损失 $\log(1+\exp(-z))$



采用hinge损失 $\ell_{hinge}=\max(0,1-z)$;

$$\min_{w,b}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{m}\max(0,1-y_i(w^Tx_i+b))$$

引入松弛变量，$\xi_i\geq 0$ 则

$$\min\frac{1}{2}\|w\|^2+C\sum_{i=1}^{m}\xi_i$$

$$\text{s.t. } y_i(w^Tx_i+b)\geq 1-\xi_i$$

$$\xi_i\geq 0\quad i=1,2,\cdots,m$$

每个样本都有个松弛变量

拉格朗日乘法

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

$\alpha_i\geq 0,\ \mu_i\geq 0$ 拉格朗日乘子

$$\frac{\partial L}{\partial w}=w-\sum_{i=1}^{m}\alpha_i y_i x_i=0$$

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

$$\frac{\partial L}{\partial \xi}\Rightarrow C=\alpha_i+\mu_i$$

对偶问题

$$\max_{\alpha}\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j$$

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

$$0\leq\alpha_i\leq C\quad i=1,2,\cdots,m$$

差别在于对偶变量：约束 $\begin{cases}0\leq\alpha_i\leq C\\ 0\leq\alpha_i\end{cases}$

KKT条件

$$\begin{cases}\alpha_i\geq 0\quad \mu_i\geq 0\\ y_i f(x_i)-1+\xi_i\geq 0\\ \alpha_i(y_i f(x_i)-1+\xi_i)=0\\ \xi_i\geq 0,\ \mu_i\xi_i=0\end{cases}$$

∴ $\alpha_i=0$ 或 $y_i f(x_i)=1-\xi_i$ (x_i,y_i)

若 $\alpha_i=0$，则没有作用

若 $\alpha_i\neq 0$ ① $\alpha_i>0,\alpha_i<C\Rightarrow\mu_i>0\Rightarrow\xi_i=0$

则该样本在最大间隔边界上

② $\alpha_i=C\Rightarrow\mu_i=0$ $\begin{cases}\xi_i\leq 1 & \text{间隔内部}\\ \xi_i>1 & \text{错误分类}\end{cases}$

0/1损失替换成对率损失 ⇒ 逻辑回归

更一般 第一项，描述间隔

第二项：误差

$$\min_f\ \underbrace{\Omega(f)}_{\text{结构风险, f的性质}}+\underbrace{C\sum_{i=1}^{m}\ell(f(x_i),y_i)}_{\text{经验风险, 模型与数据的契合}}$$

6.5 支持向量回归(SVR)

$$\min\frac{1}{2}\|w\|^2+C\sum_{i=1}^{m}\ell_\epsilon(f(x_i)-y_i)$$

$$\ell_\epsilon(z)=\begin{cases}0 & \text{if } |z|<\epsilon\\ |z|-\epsilon\end{cases}$$