支持向量机

6.1 间隔与支持向量

鲁棒性：对未见示例的泛化能力（越强越好）。

$w^T x + b = 0$. $w = (w_1, w_2, \cdots, w_d)$ 为法向量，决定超平面方向。

b 位移项，决定超平面与原点距离。

$r = \frac{|w^T x + b|}{\|w\|}$ 样本空间任意点 x 到超平面距离。

范数。

若正确分类，$\begin{cases} w^T x_i + b \geq 1 & y_i = 1 \\ w^T x_i + b \leq -1 & y_i = -1 \end{cases}$

支持向量：距离最近的几个使不等式成立的样本点。两个异类支持向量到超平面距离之和为 $\rho = \frac{2}{\|w\|}$ ← 间隔

最大间隔：$\max_{w,b} \frac{2}{\|w\|}$ s.t. $y_i(w^T x_i + b) \geq 1$. $i = 1, 2, \cdots, n$.

也可 $\min_{w,b} \frac{1}{2}\|w\|^2$ s.t. $y_i(w^T x_i + b) \geq 1$, $i = 1, 2, \cdots, m$.

6.2 对偶问题

用拉格朗日乘子法转为对偶问题。

拉格朗日函数：$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{m} \alpha_i (1 - y_i(w^T x_i + b))$

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

令 $L(w, b, \alpha)$ 偏导为0. → $w = \sum_{i=1}^{m} \alpha_i y_i x_i$

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

带回原式，即 $\underline{\underline{\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j x_i^T x_j}}$ s.t. $\sum_{i=1}^{m} \alpha_i y_i = 0$

$\alpha_i \geq 0$.

解出 α 后，$f(x) = \sum_{i=1}^{m} \alpha_i y_i x_i^T x + b$ $i = 1, 2, \cdots, m$

最终模型只与支持向量有关. $\alpha_i = 0$. | $y_i f(x_i) = 1$

SMO算法求解.

先固定α_i以外所有参数, 求α_i极值.

先选取违背KKT条件程度最大的变量, 再选使目标函数值增长最快的变量.

优化选择: 两变量对应样本间隔最大.

$\alpha_i y_i + \alpha_j y_j = c \qquad \alpha_i \geq 0,\ \alpha_j \geq 0.$

$c = -\sum_{k \neq i,j} \alpha_k y_k$

$b = \frac{1}{|S|} \sum_{s \in S} \left(y_s - \sum_{i \in S} \alpha_i y_i x_i^T x_s \right).$

6.3. 核函数.

原样本线性不可分, 则映射到更高维可分.

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

$\min_{w,b} \frac{1}{2} \|w\|^2 \quad s.t.\ y_i(w^T \phi(x_i) + b) \geq 1. \quad i = 1, 2, \cdots, m.$

对偶. $\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \underbrace{\phi(x_i)^T \phi(x_j)}_{\downarrow\ K(x_i, x_j)}.$

$s.t. \sum_{i=1}^{m} \alpha_i y_i = 0. \quad \alpha_i \geq 0. \quad i = 1, 2, \cdots, m$

$f(x) = \sum_{i=1}^{m} \alpha_i y_i \underbrace{K(x_1, x_2)}_{\text{核函数}} + b$ 支持向量展式.

只要一个对称函数所对应的矩阵半正定, 即可作为核函数.

6.4 软间隔与正则化

防止过拟合，允许支持向量机在一些样本上出错。

即某些样本不满足 $y_i(w^Tx_i+b)\geq 1$，不满足约束样本尽量少。

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

≥ 0

$\ell_{0/1}(z)\begin{cases} 1 & \text{if}(z<0) \\ 0 \end{cases}$

替代损失函数：
- hinge 损失：$\ell_{hinge}(z)=\max(0,1-z)$.
- 指数损失：$\ell_{exp}(z)=\exp(-z)$.
- 对率损失：$\ell_{log}(z)=\log(1+\exp(-z))$.

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

引入松弛变量，$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i$

软间隔支持向量机。

使用拉格朗日乘子法，获得拉格朗日函数

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

令偏导为0，
$$w=\sum_{i=1}^{m}\alpha_iy_ix_i$$
$$0=\sum_{i=1}^{m}\alpha_iy_i$$
$$C=\alpha_i+\mu_i$$

得到对偶问题 $\max_{\alpha} \sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_jy_iy_jx_i^Tx_j$

s.t. $\sum_{i=1}^{m}\alpha_iy_i=0$. $0\leq\alpha_i\leq C$. $i=1,2,\cdots,m$.

KKT 条件要求
$$\begin{cases} \alpha_i\geq 0, \quad \mu_i\geq 0 \\ y_if(x_i)-1+\xi_i\geq 0 \\ \alpha_i(y_if(x_i)-1+\xi_i)=0 \end{cases}$$

$$\xi_i \geq 0, \quad \mu_i \xi_i = 0.$$

损失函数一般形式：$\min_f \underbrace{\Omega(f)}_{\text{结构风险}} + \underbrace{C\sum_{i=1}^{m} l(f(x_i), y_i)}_{\text{经验风险}}$.

6.5 支持向量回归

$D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)\}$. 回归模型使 $f(x)$ 与 y 接近.

支持向量回归认为样本落入 2ϵ 间隔带为正确.

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} l_\epsilon(f(x_i)-y_i)$$

$$l_\epsilon(z)=\begin{cases} 0 & \text{if } |z|\leq\epsilon. \\ |z|-\epsilon & \text{otherwise} \end{cases}$$

引入松弛变量.

$$\min_{w,b,\xi_i,\hat{\xi}_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i+\hat{\xi}_i). \quad s.t.\ f(x_i)-y_i \leq \epsilon+\xi_i$$

$$y_i - f(x_i) \leq \epsilon + \hat{\xi}_i$$

$$\xi_i \geq 0,\ \hat{\xi}_i \geq 0,\ i=1,2,\cdots,m.$$

拉格朗日函数：

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

令偏导为0，得 $w=\sum_{i=1}^{m}(\hat{\alpha}_i-\alpha_i)x_i$

$$0=\sum_{i=1}^{m}(\hat{\alpha}_i-\alpha_i)$$

$$C=\alpha_i+\mu_i$$

$$C=\hat{\alpha}_i+\hat{\mu}_i$$

得到SVR对偶问题

$$\max_{\alpha,\hat{\alpha}} \sum_{i=1}^{m} y_i(\hat{\alpha}_i-\alpha_i) - \epsilon(\hat{\alpha}_i+\alpha_i) - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\hat{\alpha}_i-\alpha_i)(\hat{\alpha}_j-\alpha_j)x_i^Tx_j$$

$$s.t.\ \sum_{i=1}^{m}(\hat{\alpha}_i-\alpha_i)=0 \qquad 0\leq\alpha_i,\hat{\alpha}_i\leq C.$$

KKT, $\begin{cases} \alpha_i(f(x_i)-y_i-\epsilon-\xi_i)=0 \\ \hat{\alpha}_i(y_i-f(x_i)-\epsilon-\hat{\xi}_i)=0 \\ \alpha_i\hat{\alpha}_i=0, \quad \xi_i\hat{\xi}_i=0 \\ (C-\alpha_i)\xi_i=0. \quad (C-\hat{\alpha}_i)\hat{\xi}_i=0, \end{cases}$

SVR解. $f(x)=\sum_{i=1}^{m}(\hat{\alpha}_i-\alpha_i)x_i^Tx+b$. or. $f(x)=\sum_{i=1}^{m}(\hat{\alpha}_i-\alpha_i)K(x,x_i)+b$

$b=y_i+\epsilon-\sum_{i=1}^{m}(\hat{\alpha}_i-\alpha_i)x_i^Tx.$ $K(x_i,x_j)=\phi(x_i)^T\phi(x_j)$

6.6. 核方法.

表示定理. 对于单调递增函数和任意非负损失函数. 优化问题总有解.

向非线性拓展 → 核线性判别分析. KLDA