Support Vector Machines

■ 学习的对偶问题

原始问题的对偶问题是

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

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

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

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

Support Vector Machines

对偶问题是拉格朗日函数的极大极小问题. 首先求 $L(w,b,\xi,\alpha,\mu)$ 对 w,b,ξ 的极小，由

$$\nabla_w L(w,b,\xi,\alpha,\mu) = w - \sum_{i=1}^{N}\alpha_i y_i x_i = 0$$

$$\nabla_b L(w,b,\xi,\alpha,\mu) = -\sum_{i=1}^{N}\alpha_i y_i = 0$$

$$\nabla_{\xi_i} L(w,b,\xi,\alpha,\mu) = C - \alpha_i - \mu_i = 0$$

得

$$w = \sum_{i=1}^{N}\alpha_i y_i x_i$$

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

$$C - \alpha_i - \mu_i = 0$$

$$\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j) + \sum_{i=1}^{N}\alpha_i$$

Support Vector Machines

再对 $\min\limits_{w,b,\xi} L(w,b,\xi,\alpha,\mu)$ 求 α 的极大，即得对偶问题：

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

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

$$C - \alpha_i - \mu_i = 0$$

$$\alpha_i \geqslant 0$$

$$\mu_i \geqslant 0, \quad i = 1,2,\cdots,N$$

将对偶最优化问题进行变换：消去 μ_i，从而只留下变量 α_i，
并将约束写成 $0 \leqslant \alpha_i \leqslant C$

再将对目标函数求极大转换为求极小，于是得到对偶问题

Support Vector Machines

定理 设$\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_N^*)^{\mathrm{T}}$是对偶问题的一个解，若存在$\alpha^*$的一个分量$\alpha_j^*$，$0 < \alpha_j^* < C$，则原始问题的解$w^*, b^*$可按下式求得：

$$w^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$$

$$b^* = y_j - \sum_{i=1}^{N} y_i \alpha_i^* (x_i \cdot x_j)$$

由此定理可知，分离超平面可以写成 $\sum_{i=1}^{N} \alpha_i^* y_i (x \cdot x_i) + b^* = 0$

分类决策函数可以写成 $f(x) = \mathrm{sign}\left(\sum_{i=1}^{N} \alpha_i^* y_i (x \cdot x_i) + b^* \right)$