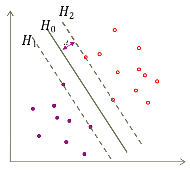
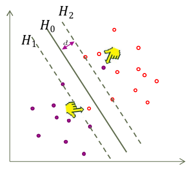
线性不可分

训练数据中有一些特异点(outlier)，不能满足函数间隔大于等于1的约束条件。





松弛变量与软间隔

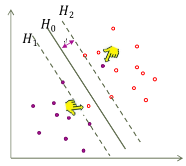
- 由于无法找到完美的分割，那么放松一定的限制
- 对每个样本点 (x_i, y_i) 引进一个松弛变量 $\xi_i \geq 0$
- 使得函数间隔加上松弛变量

大于等于1，约束条件变为：

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i$$

目标函数变为：$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$

C > 0 为惩罚参数



软间隔最大化

- 线性不可分的线性支持向量机的学习问题：

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$s.t. \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i, \ \ i = 1, 2, ..., N$$

$$\xi_i \geq 0, \ \ i = 1, 2, ..., N$$

- 凸二次规划问题
- 设该问题的解是w*, b*,可得到分离超平面和决策函数

$$w^* \cdot x + b^* = 0$$

$$f(x) = sign(w^* \cdot x + b^*)$$

软间隔最大化

- 原始问题：

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$s.t. \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i, \ \ i = 1, 2, ..., N$$

$$\xi_i \geq 0, \ \ i = 1, 2, ..., N$$

- 拉格朗日函数：

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

- 其中： $\alpha_i \geq 0, \mu_i \geq 0$

软间隔最大化

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

$$\alpha_i \geq 0, \mu_i \geq 0$$

- 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

$$\begin{cases} \nabla_w L(w,b,\xi,\alpha,\mu) = w - \sum_{i=1}^{N}\alpha_i y_i x_i = 0 \\ \nabla_b L(w,b,\xi,\alpha,\mu) = -\sum_{i=1}^{N}\alpha_i y_i = 0 \\ \nabla_{\xi_i} L(w,b,\xi,\alpha,\mu) = C - \alpha_i - \mu_i = 0 \end{cases}$$

得：

$$\begin{cases} w = \sum_{i=1}^{N}\alpha_i y_i x_i \\ \sum_{i=1}^{N}\alpha_i y_i = 0 \\ C - \alpha_i - \mu_i = 0 \end{cases}$$

软间隔最大化

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

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

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

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

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

软间隔最大化

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

- 再对 $\min_{w,b,\xi} L(w,b,\xi,\alpha,\mu)$ 求α的极大，得到对偶问题：

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

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

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

$$\alpha_i \geq 0 \qquad \Rightarrow \qquad 0 \leq \alpha_i \leq C$$

$$\mu_i \geq 0, \ i = 1,2,...,N$$

软间隔最大化

- 原始问题的对偶问题：

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

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

$$0 \leq \alpha_i \leq C, \ i = 1,2,...,N$$

- 定理：设 $\alpha^* = (\alpha_1^*, \alpha_2^*, ..., \alpha_l^*)^T$ 是对偶问题的一个解，若存在α^*的一个分量 $\alpha_j^*, 0 < \alpha_j^* < C$ ，则原始问题的解w*, b*

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

线性支持向量机学习算法

- 输入：线性不可分训练数据集 $T = \{(x_1,y_1),(x_2,y_2),...,(x_N,y_N)\}$

$$x_i \in \mathcal{X} = R^n, y_i \in \mathcal{Y} = \{-1,+1\}, i = 1,2,...,N$$

- 输出：分离超平面和分类决策函数

1、构造并求解约束最优化问题

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

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

$$0 \leq \alpha_i \leq C, \ i = 1,2,...,N$$

求得最优解：

$$\alpha^* = (\alpha_1^*, \alpha_2^*, ..., \alpha_l^*)^T$$

线性支持向量机学习算法

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

并选择α*，适合条件$0 < \alpha_j^* < C$ ，计算

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

3、求得分离超平面 $w^* \cdot x + b^* = 0$

分类决策函数$f(x) = sign(w^* \cdot x + b^*)$

支持向量

实例x_i到间隔边界的距离

$$y_i(w \cdot x_i + b) = 1$$

$$y_i(w \cdot x_i + b) = 1 - \xi_i$$

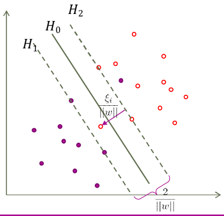
$$\Downarrow$$

$$\frac{\xi_i}{||w||}$$

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

$$\alpha_i \geq 0$$

$$\mu_i \geq 0, \ i = 1,2,...,N$$



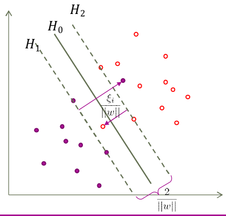
支持向量

若 $\alpha_i^* < C$，则$\xi_i = 0$
若 $\alpha_i^* = C$，$0 < \xi_i < 1$
若 $\alpha_i^* = C$，$\xi_i = 1$
若 $\alpha_i^* = C$，$\xi_i > 1$

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

$$\alpha_i \geq 0$$

$$\mu_i \geq 0, \ i = 1,2,...,N$$



合页损失(hinge loss)

•线性支持向量机原始最优化问题:

$$\min_{w,b,\xi}\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$s.t.\ \ y_i(w\cdot x_i+b)\geq 1-\xi_i,\ i=1,2,...,N$$

$$\xi_i\geq 0,\ i=1,2,...,N$$

•等价于：

$$\min_{w,b}\ \sum_{i=1}^{N}[1-y_i(w\cdot x_i+b)]_+ + \lambda\|w\|^2$$

•第一项：经验风险
•第二项：结构风险

合页损失(hinge loss)

•线性支持向量机原始最优化问题:

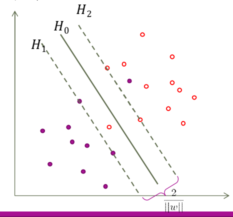
$$\min_{w,b,\xi}\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$s.t.\ \ y_i(w\cdot x_i+b)\geq 1-\xi_i,\ i=1,2,...,N$$

$$\xi_i\geq 0,\ i=1,2,...,N$$

$$\xi_i=[1-y_i(w\cdot x_i+b)]_+$$

$$\min_{w,b}\ \sum_{i=1}^{N}[1-y_i(w\cdot x_i+b)]_+ + \lambda\|w\|^2$$



合页损失(hinge loss)

$$L(y(w\cdot x+b))=[1-y(w\cdot +b)]_+$$

$$[z]_+=\begin{cases} z, & z>0\\ 0, & z\leq 0\end{cases}$$