SVM的对偶算法

- 对于线性可分支持向量机的优化问题，原始问题：

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$
$$s.t. \quad y_i(w\cdot x_i+b)-1\geq 0, i=1,2,...,N$$

- 应用拉格朗日对偶性，通过求解对偶问题，得到原始问题的解。
- 优点：
  - 对偶问题往往容易解
  - 引入核函数，推广到非线性分类问题

SVM的对偶算法

- 定义拉格朗日函数

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

- 原问题：极小极大，对偶问题：极大极小

$$\min_{w,b}\theta_P(w,b)=\min_{w,b}\max_{\alpha:\alpha_i\geq 0}L(w,b,\alpha)$$

$$\Downarrow$$

$$\max_{\alpha:\alpha\geq 0}\min_{w,b}L(w,b,\alpha)$$

学习的对偶算法

- 先求$L(w,b,\alpha)$对w，b的极小，再求对α的极大
- 1、求：$\min_{w,b}L(w,b,\alpha)$ ，对w，b分别求偏导并令等于0

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

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

得：

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

$$\Updownarrow$$

$$\min_{w,b}L(w,b,a)$$

SVM的对偶算法

- 求 $\min_{w,b}L(w,b,\alpha)$ 对α的极大：

$$\max_{\alpha}\quad -\frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)+\sum_{i=1}^N\alpha_i$$
$$s.t.\quad \sum_{i=1}^N\alpha_i y_i=0$$
$$\alpha_i\geq 0,\ i=1,2,...,N$$

$$\Rightarrow$$

$$\min_{\alpha}\quad \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N\alpha_i\alpha_j y_i y_j(x_i\cdot x_j)-\sum_{i=1}^N\alpha_i$$
$$s.t.\quad \sum_{i=1}^N\alpha_i y_i=0$$
$$\alpha_i\geq 0,\ i=1,2,...,N$$

SVM的对偶算法

- 定理：设$\alpha^*=(\alpha_1^*,\alpha_2^*,...,\alpha_l^*)$是对偶最优问题的解，
  则存在下标j，使得$\alpha_j^*>0$，并可按下式求得原始问题的解。

$$\boxed{\begin{aligned} w^*&=\sum_{i=1}^N\alpha_i^* y_i x_i \\ b^*&=y_j-\sum_{i=1}^N\alpha_i^* y_i(x_i\cdot x_j)\end{aligned}}$$

证明：由

$$\nabla_w L(w^*,b^*,\alpha^*)=w^*-\sum_{i=1}^N\alpha_i^* y_i x_i=0$$
$$\nabla_b L(w^*,b^*,\alpha^*)=-\sum_{i=1}^N\alpha_i^* y_i=0$$
$$\alpha_i^*(y_i(w^*\cdot x_i+b^*)-1)=0, i=1,2,...,N$$
$$y_i(w^*\cdot x_i+b^*)-1\geq 0, i=1,2,...,N$$
$$\alpha_i^*\geq 0, i=1,2,...,N$$

得：$w^*=\sum_i\alpha_i^* y_i x_i$

SVM的对偶算法

- 定理：设$\alpha^*=(\alpha_1^*,\alpha_2^*,...,\alpha_l^*)^T$ 是对偶最优问题的解
- 则存在下标j，使得$\alpha_j^*>0$，并可按下式求得原始问题的解。
- 证明：由$w^*=\sum_i\alpha_i^* y_i x_i$ ，其中至少有一个$\alpha_j^*>0$

  反证法：

  假设：$\alpha^*=0$ ，则$w*$=0，

  但这不是原始优化问题的解，产生矛盾

  对此：j 有 $y_j(w^*\cdot x_j+b^*)-1=0$

  将式 $w^*=\sum_{i=1}^N\alpha_i^* y_i x_i$ 代入并注意到$y_j^2=1$

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

SVM的对偶算法

- 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：

$$f(x) = sign(\sum_{i=1}^{N}\alpha_i^* y_i (x \cdot x_i) + b^*)$$

- 这就是说，分类决策函数只依赖于输入x和训练样本输入的内积，上式称为线性可分支持向量机的对偶形式。

线性可分支持向量机学习算法

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

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

- 输出：最大间隔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

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

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

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

- 求得最优解：

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

线性可分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*的一个正分量 $\alpha_j^* > 0$，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

- 考虑原始优化问题和对偶优化问题，
- 将数据集中对应于$\alpha_j^* > 0$ 的样本(x_i, y_i) 的实例$x_i \in R^n$
- 称为支持向量

- 支持向量一定在分割边界上，由KKT互补条件：

$$\alpha_i^*(y_i(w^* \cdot x_i + b^*) - 1) = 0, i=1,2,...,N$$

- 对应于$\alpha_j^* > 0$ 的样本x_i

$$y_i(w^* \cdot x_i + b^*) - 1 = 0$$

- 或

$$w^* \cdot x_i + b^* = \pm 1$$