机器学习

05 支持向量机

李祎

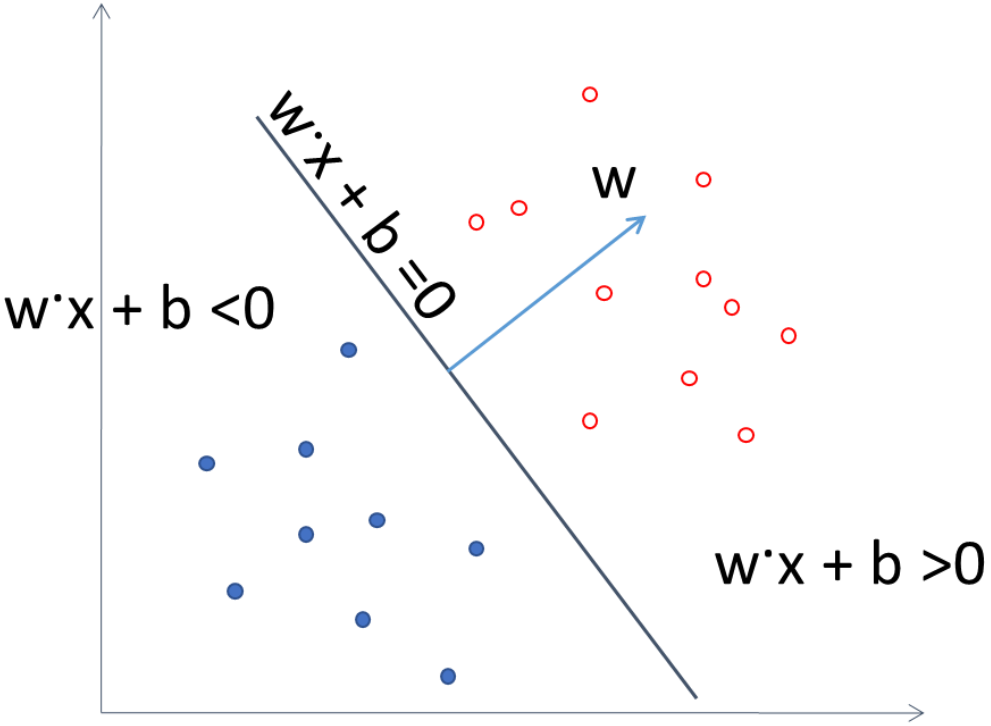
liyi@dlut.edu.cn

大连理工大学 人工智能学院
School of Artificial Intelligence, Dalian University of Technology

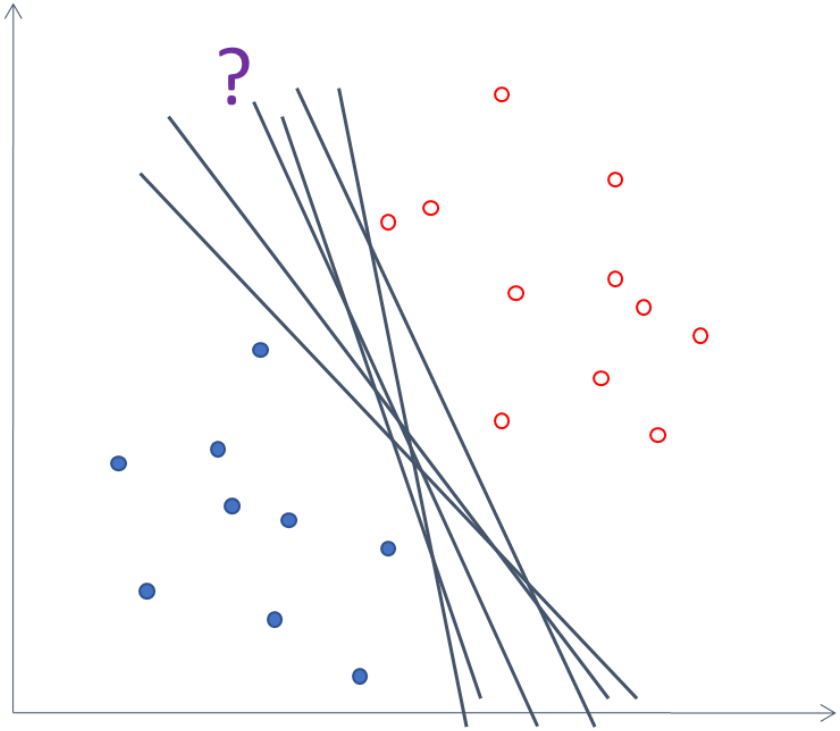
目录

线性可分

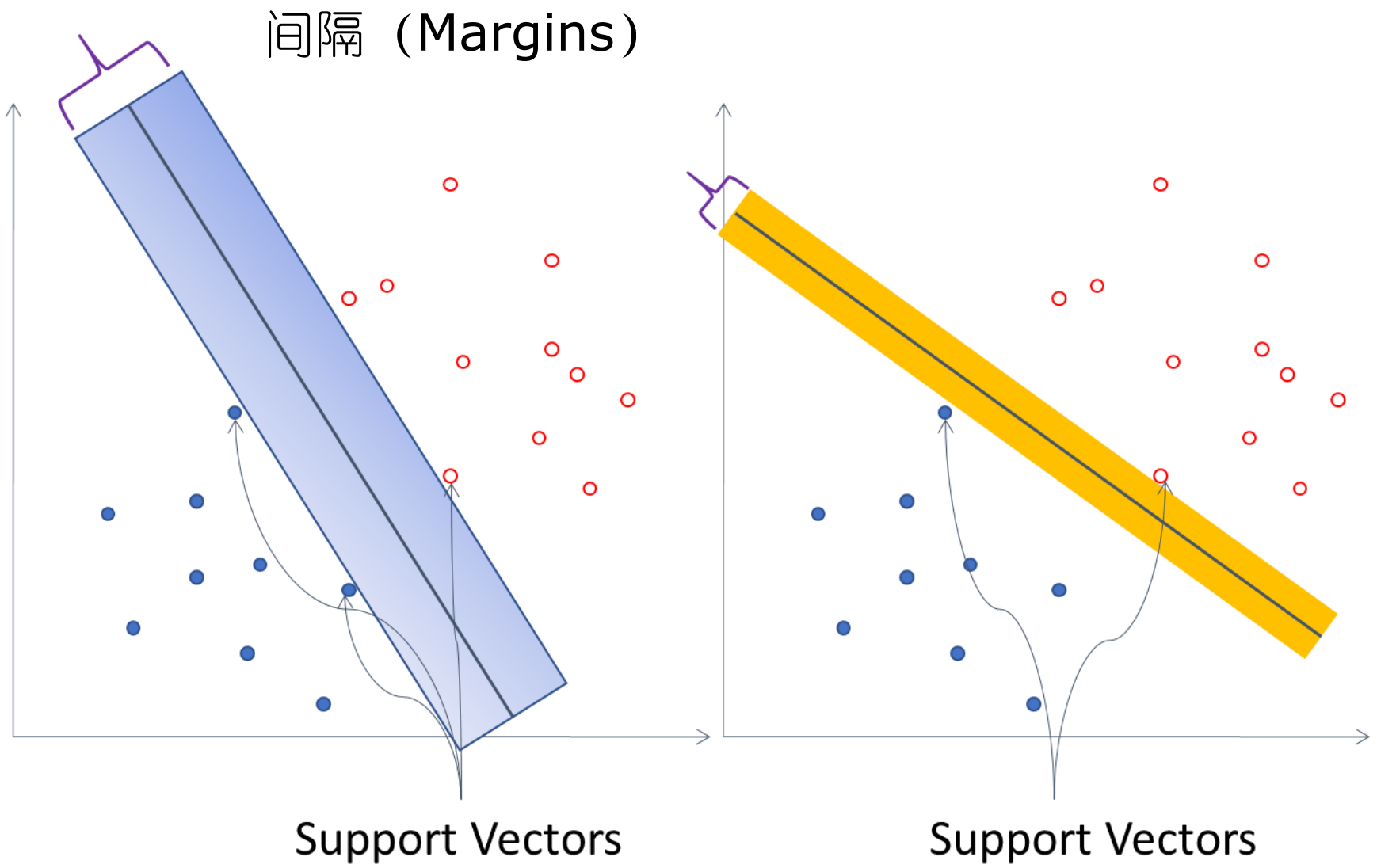
线性分类器



超平面选择

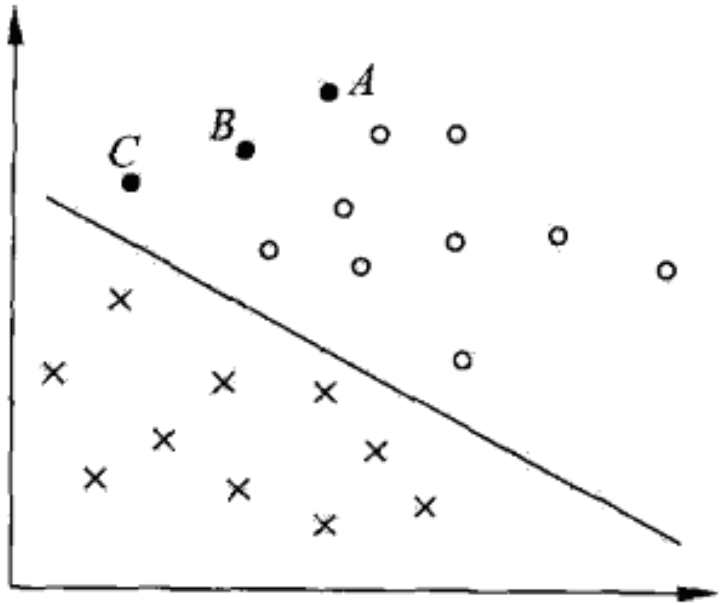


间隔与支持向量



函数间隔和几何间隔

- 点到分离超平面的远近 $|w \cdot x + b|$
- ➡ 表示分类预测的确信程度
- $w \cdot x + b$ 的符号与类标记y的符号是否一致
- ➡ 表示分类是否正确
- 所以：$y(w \cdot x + b)$
- 表示分类的正确性和确信度



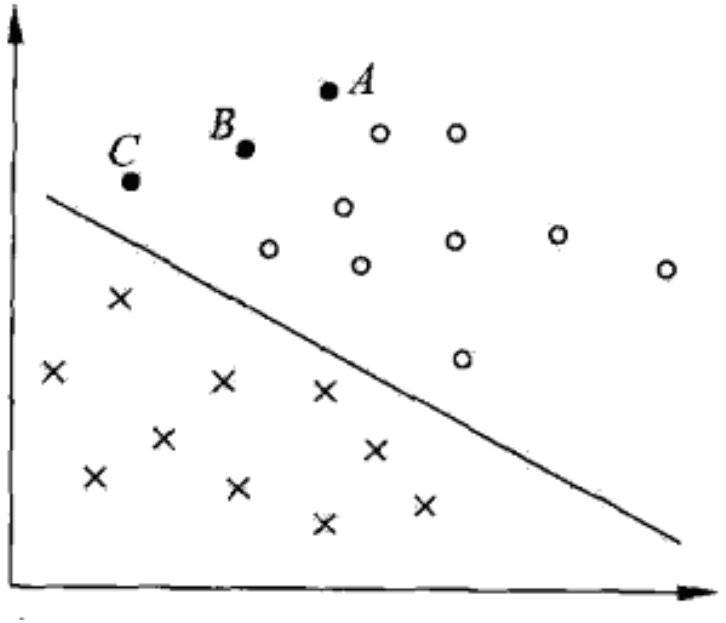
函数间隔和几何间隔

- 函数间隔
  - 样本点的函数间隔

$$\hat{\gamma}_i = y_i(w \cdot x_i + b)$$

  - 训练数据集的函数间隔

$$\hat{\gamma} = \min_{i=1,\cdots,N} \hat{\gamma}_i$$



  - 表示分类预测的

    正确性和确信度

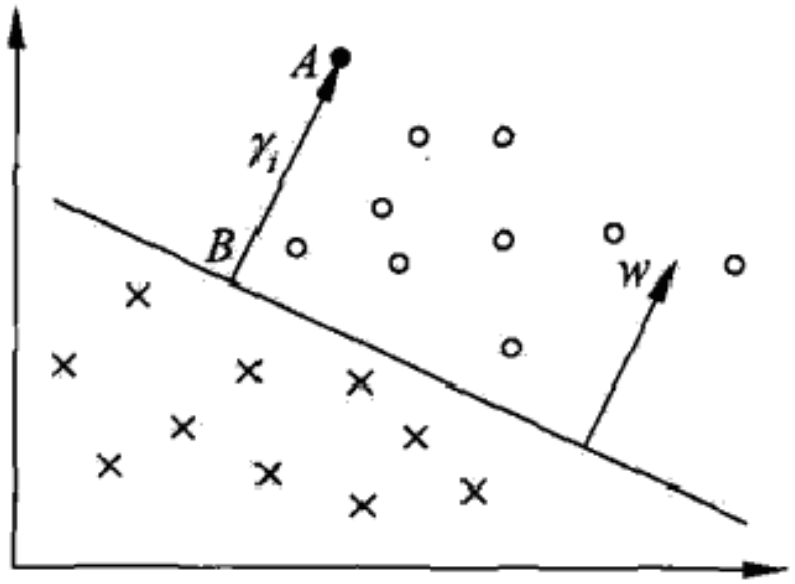
  - 当成比例改变w和b，超平面不变，但函数间隔会变化。

函数间隔和几何间隔

- 几何间隔
  - 样本点的几何间隔：正例和负例

$$\gamma_i = \left(\frac{w}{\|w\|} \cdot x_i + \frac{b}{\|w\|}\right) \qquad \gamma_i = -\left(\frac{w}{\|w\|} \cdot x_i + \frac{b}{\|w\|}\right)$$

$$\gamma_i = y_i\left(\frac{w}{\|w\|} \cdot x_i + \frac{b}{\|w\|}\right)$$



点到平面的距离：
https://www.jianshu.com/p/2e3c0c583e85

函数间隔和几何间隔

- 几何间隔
  - 对于给定的训练数据集T和超平面(w, b)

$$\gamma_i = y_i\left(\frac{w}{\|w\|}\cdot x_i + \frac{b}{\|w\|}\right)$$

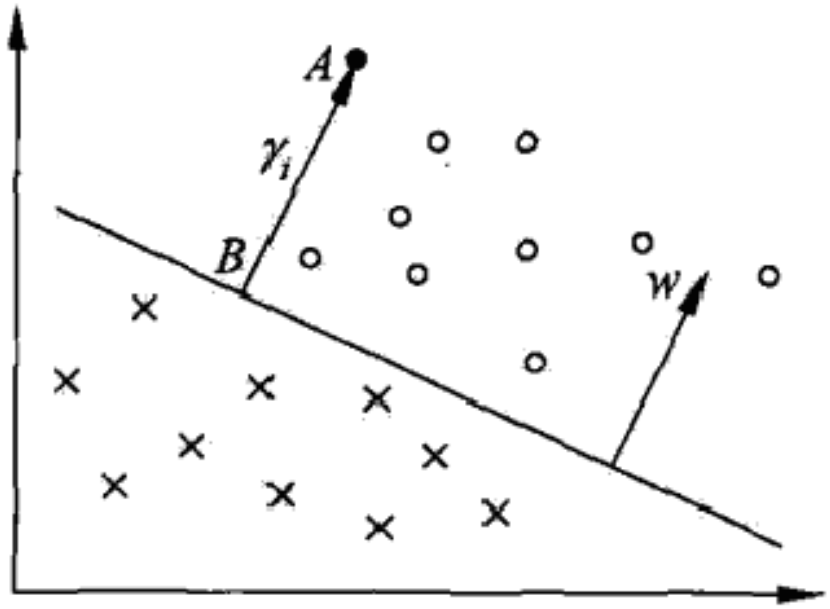
  - 训练数据集的几何间隔

$$\gamma = \min_{i=1,\cdots,N} \gamma_i$$

  - 即

$$\gamma_i = \frac{\hat{\gamma}_i}{\|w\|}$$

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



间隔最大化

- 最大间隔分类超平面

$$\max_{w,b} \quad \gamma$$

$$\text{s.t.} \quad y_i\left(\frac{w}{\|w\|}\cdot x_i+\frac{b}{\|w\|}\right)\geqslant\gamma, \quad i=1,2,\cdots,N$$

- 根据几何间隔和函数间隔的关系

$$\max_{w,b} \quad \frac{\hat{\gamma}}{\|w\|}$$

$$\text{s.t.} \quad y_i(w\cdot x_i+b)\geqslant\hat{\gamma}, \quad i=1,2,\cdots,N$$

- 考虑
  - 可以通过函数间隔的比例缩放，取 $\hat{\gamma}=1$
  - 最大化 $\frac{1}{\|w\|}$ 和最小化 $\frac{1}{2}\|w\|^2$ 等价

SVM基本型

- 线性可分支持向量机学习的最优化问题

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

$$\text{s.t.} \quad y_i(w \cdot x_i + b) - 1 \geqslant 0, \quad i = 1, 2, \cdots, N$$

- 问：为什么要写成 $\frac{1}{2}\|w\|^2$ 形式？
- 答：这样很符合一类已经很成熟的优化问题：凸二次规划 (convex quadratic programming)

线性可分SVM学习算法

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

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

- 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

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

$$\text{s.t.} \quad y_i(w\cdot x_i+b)-1\geqslant 0, \quad i=1,2,\cdots,N$$

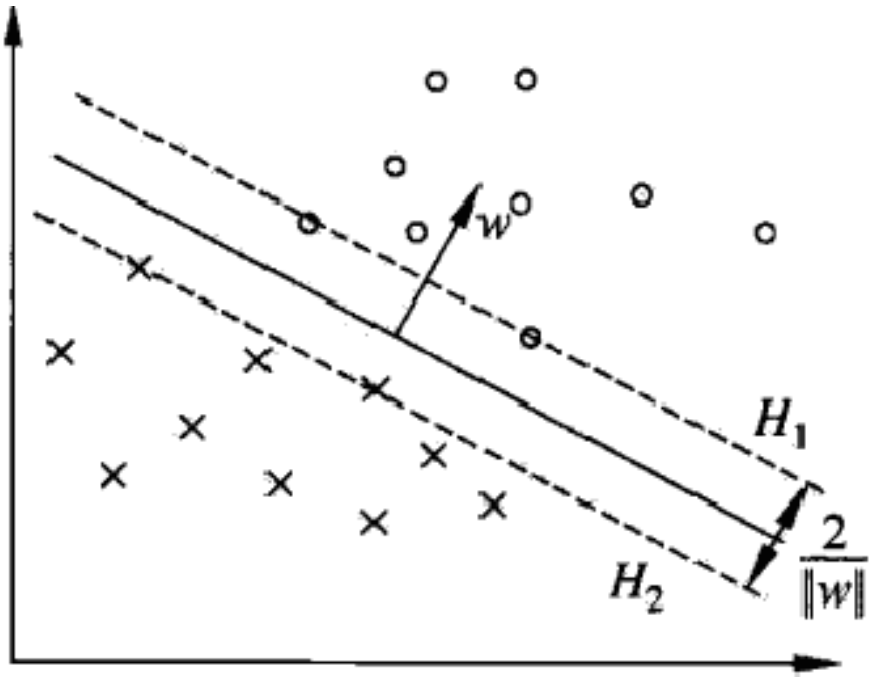
求得w*和b*

2、得到分离超平面 $w^*\cdot x+b^*=0$

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

支持向量和间隔边界

- 在线性可分情况下，训练数据集的样本点中与分离超平面距离最近的样本点的实例称为支持向量(support vector)；
- 支持向量是使约束条件式等号成立的点，即 $y_i(w \cdot x_i + b) - 1 = 0$
- 正例： $H_1 : w \cdot x + b = 1$
- 负例： $H_2 : w \cdot x + b = -1$
- H_1与H_2平行，并且没有点落在它们中间，H_1与H_2之间的距离称为间隔，H_1和H_2称为间隔边界。



线性可分SVM

硬间隔最大化

拉格朗日对偶

- 对于线性可分支持向量机的优化问题，原始问题：

1

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

- 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

拉格朗日对偶

- 对每条约束引入拉格朗日乘子 $\alpha_i \geq 0$ ，构建拉格朗日对偶函数

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

- 根据拉格朗日对偶性，原始问题的对偶问题为 $\max_{\alpha}\min_{w,b}L(w,b,\alpha)$
- 先求 $L(w,b,\alpha)$ 对w，b的极小，再求对a的极大

1. 令 $L(w,b,\alpha)$ 对w，b的偏导为0，可得

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

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

拉格朗日对偶

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

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

代入

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

可得

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

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

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

拉格朗日对偶

- 2、再求 $\min_{w,b} L(w,b,\alpha)$ 对a的极大，即是对偶问题：

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

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

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

2

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

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

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

拉格朗日对偶

- 定理：设 $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_l^*)^T$ 是对偶最优问题 2 的解，则存在下标j，使得 $\alpha_j^* > 0$，并可按下式求得原始问题 1 的解。

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

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

- 证明思路：KKT条件+反证法

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

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SVM学习算法

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

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

- 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

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

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

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

求得最优解：$\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_N^*)^{\mathrm{T}}$

线性可分SVM学习算法

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

并选择a*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 $f(x) = \text{sign}(w^* \cdot x + b^*)$

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$x_j \in \mathbf{R}^n$ 称为支持向量
- 支持向量一定在分割边界上，由KKT互补条件：

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

对应于 $\alpha_j^* > 0$ 的样本 x_j 有

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

或 $w^* \cdot x_i + b^* = \pm 1$

支持向量机解的稀疏性：训练完成后，大部分的训练样本都不需保留，最终模型仅与支持向量有关.

求解方法 - SMO

- 基本思路：不断执行如下两个步骤直至收敛.
  - 第一步：选取一对需更新的变量α_i和α_j.
  - 第二步：固定α_i和α_j以外的参数，求解对偶问题更新α_i和α_j.

- 仅考虑 α_i和α_j 时，对偶问题的约束变为

$$\alpha_i y_i + \alpha_j y_j = -\sum_{k \neq i,j} \alpha_k y_k, \quad \alpha_i \geq 0, \quad \alpha_j \geq 0.$$

用一个变量表示另一个变量，回代入对偶问题可得一个单变量的二次规划，该问题具有闭式解.

- 偏移项b：通过支持向量来确定.