机器学习

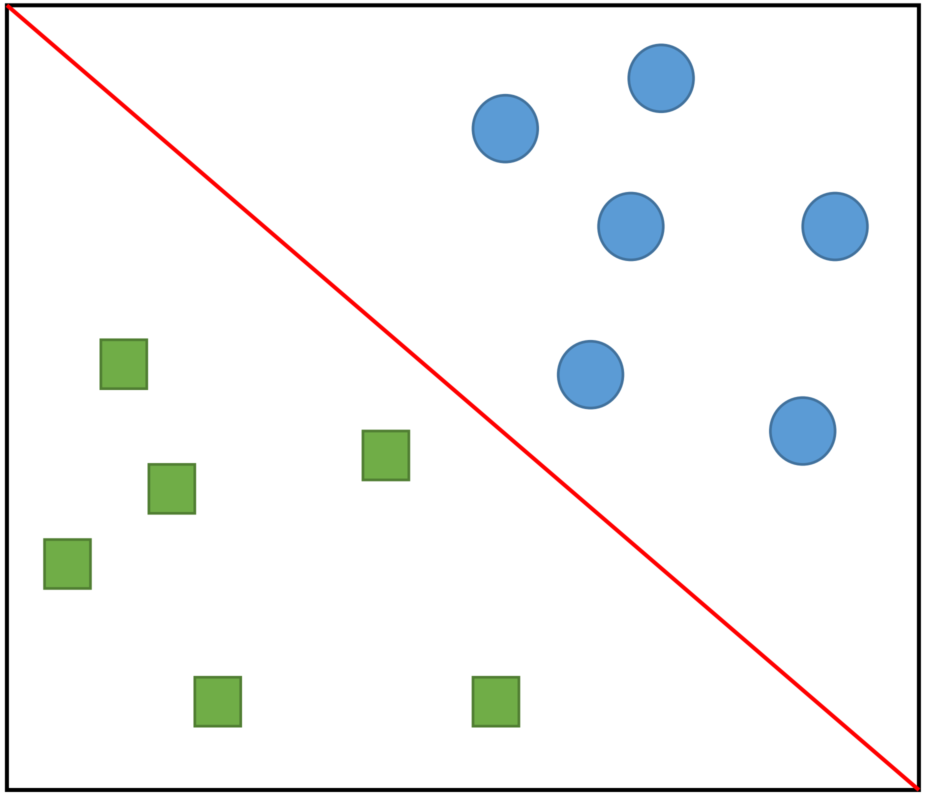
苏州大学计算机科学与技术学院
自然语言处理实验室
主讲：周夏冰
邮箱：zhouxiabing@suda.edu.cn

01

线性可分支持向量机

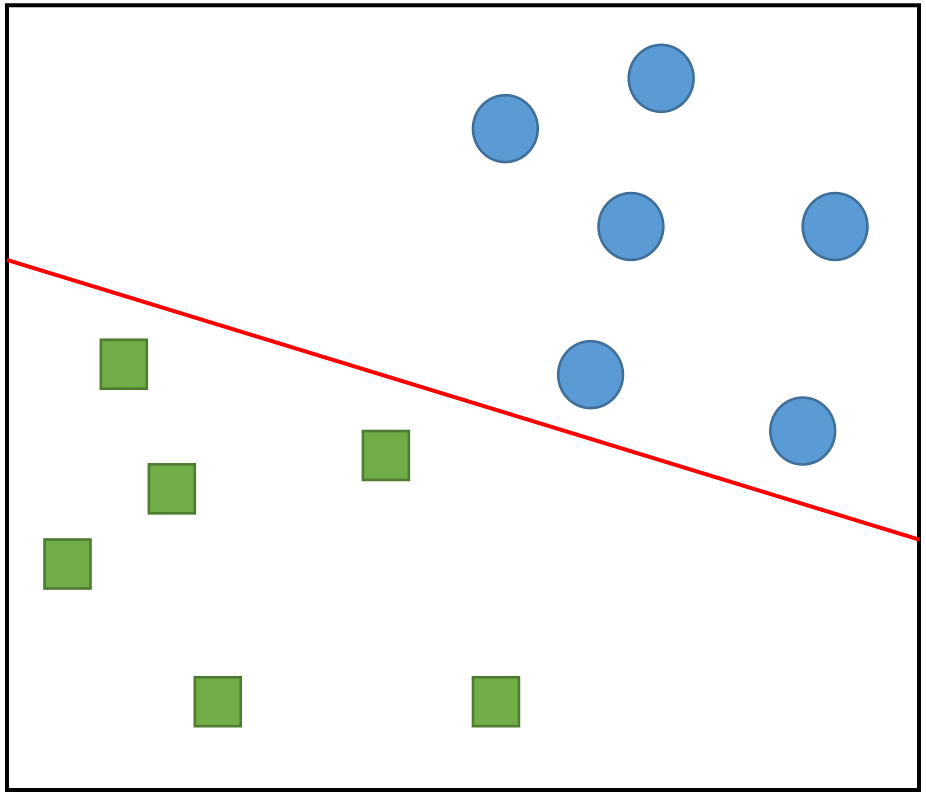
线性可分支持向量机

- 线性分类器：寻找一个超平面，将不同样本分开



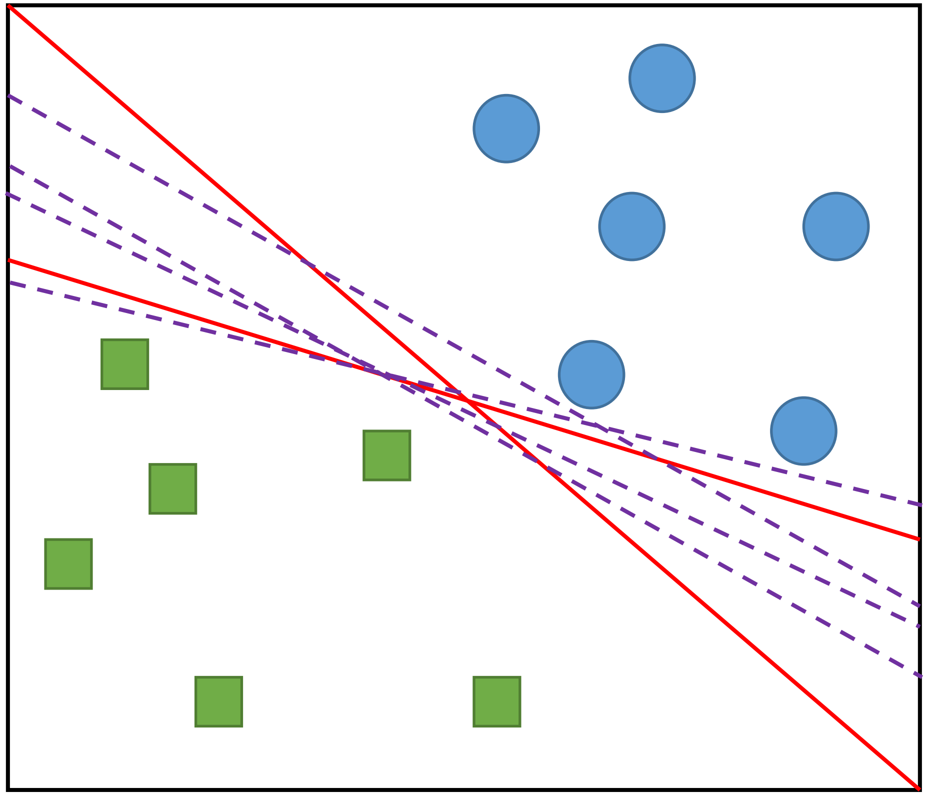
线性可分支持向量机

- 线性分类器：寻找一个超平面，将不同样本分开



线性可分支持向量机

- 线性分类器：寻找一个超平面，将不同样本分开

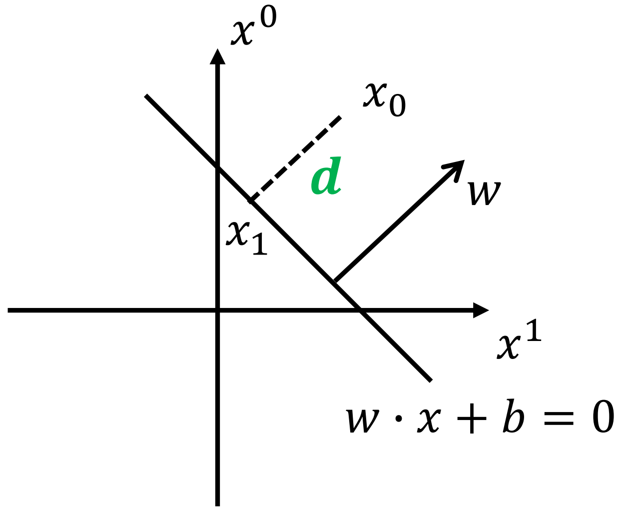


支持向量机（SVM）

- 支持向量机是90年代中期发展起来的基于统计学习理论的一种机器学习方法，通过寻求结构化风险最小来提高学习机的泛化能力，从而达到在统计样本量比较少的情况下，亦能获得良好统计规律的目的
- 二分类的模型
- 基本模型：在特征空间上的间隔最大的线性分类器，即SVM的学习策略便是间隔最大化，最终转化为一个凸二次规划问题的求解

感知机学习策略

- 输入空间中任一点x_0到超平面S的距离



$$\frac{1}{\|w\|}|w \cdot x_0 + b|$$

间隔

- 函数间隔
  - 一个点距离分离超平面的远近可以表示分类预测的确信程度

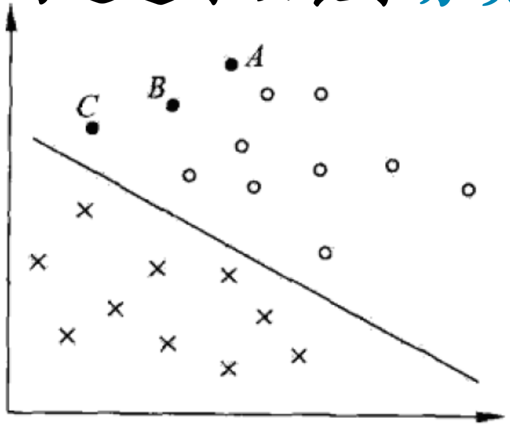


图 7.1　二类分类问题

  - $|wx+b|$相对表示点x到超平面的远近；
  - $wx+b$的符号与类标记y的符号是否一致能够表示分类是否正确
  - 所以，用$y(wx+b)$来表示分类的正确性及确信度

间隔

- **函数间隔**

定义 7.2（函数间隔） 对于给定的训练数据集 T 和超平面 (w,b)，定义超平面 (w,b) 关于样本点 (x_i,y_i) 的函数间隔为

$$\hat{\gamma}_i = y_i(w \cdot x_i + b) \tag{7.3}$$

定义超平面 (w,b) 关于训练数据集 T 的函数间隔为超平面 (w,b) 关于 T 中所有样本点 (x_i,y_i) 的函数间隔之最小值，即

$$\hat{\gamma} = \min_{i=1,\cdots,N} \hat{\gamma}_i \tag{7.4}$$

间隔

- 几何间隔

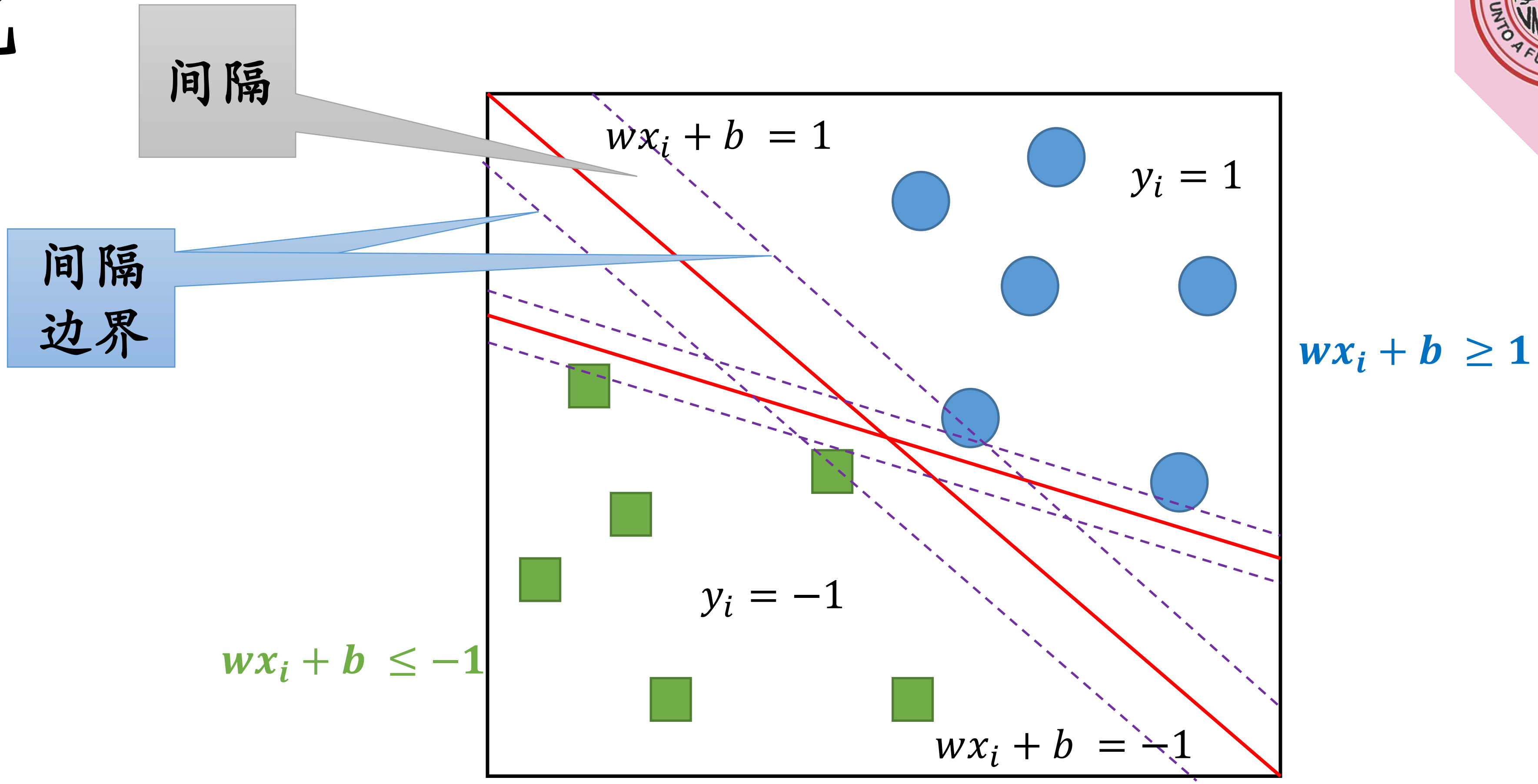
定义 7.3（几何间隔） 对于给定的训练数据集 T 和超平面 (w,b)，定义超平面 (w,b) 关于样本点 (x_i,y_i) 的几何间隔为

$$\gamma_i = y_i\left(\frac{w}{\|w\|}\cdot x_i + \frac{b}{\|w\|}\right) \tag{7.5}$$

定义超平面 (w,b) 关于训练数据集 T 的几何间隔为超平面 (w,b) 关于 T 中所有样本点 (x_i,y_i) 的几何间隔之最小值，即

$$\gamma = \min_{i=1,\cdots,N} \gamma_i \tag{7.6}$$

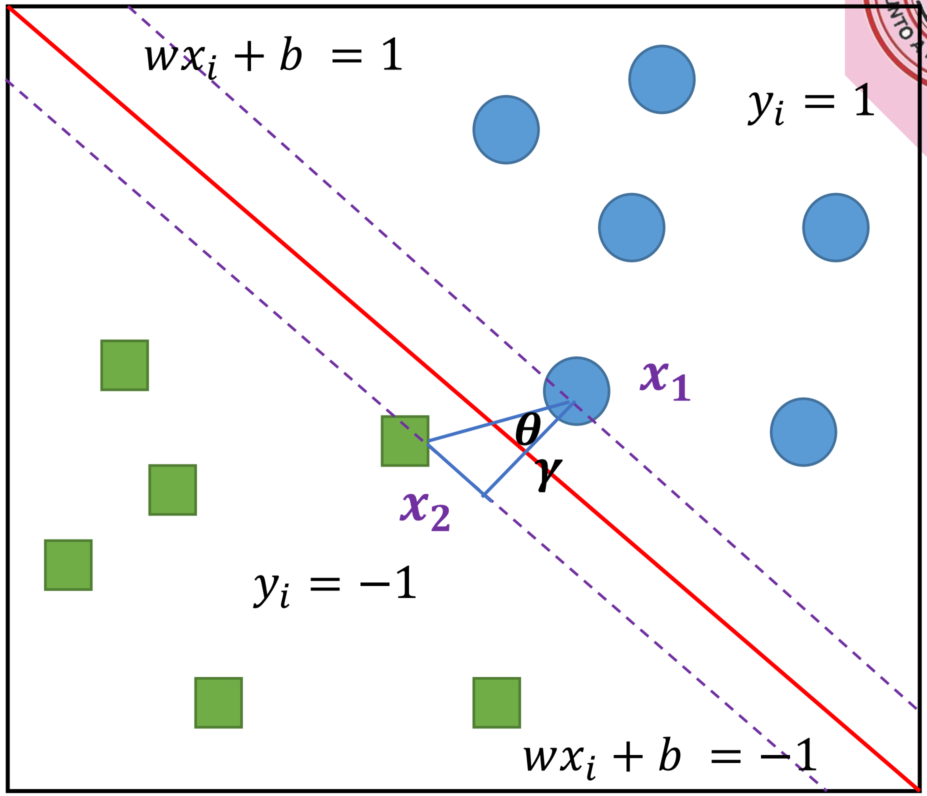
间隔最大化



- 支持向量在确定分离超平面中起着决定性作用（很少，但很重要），所以将这种分类模型称为支持向量机

线性可分支持向量机

- 间隔最大化：硬间隔最大化
  - 以充分大的确信度对训练数据进行分类
  - 泛化能力强；对新实例有很好的预测性



- $wx_1 + b = 1$; $wx_2 + b = -1$
- 两式子相减： $w \cdot \overrightarrow{x_1x_2} = 2 = \|w\| \boxed{\|x_1x_2\| cos\theta}$
- 等价于： $\|w\| \times \gamma = 2 \quad \gamma = \frac{2}{\|w\|}$

线性可分支持向量机

- 支持向量机学习的最优化问题：
  - $\min_{w,b} \frac{1}{2}\|w\|^2$
  - $s.t.\ \ y_i(wx_i + b) - 1 \geq 0, \quad i = 1, 2, \cdots, N$
- 最大间隔分离超平面的存在唯一性

求解最优化

- 线性可分支持向量机的**对偶算法**
  - **优点**：对偶问题往往更容易求解
    自然引入核函数
    样本的可解释性
- **对偶问题**
  - 对偶问题的对偶是原问题；
  - 无论原始问题是否是凸的，拉格朗日对偶可以转化为凸优化问题；
  - 对偶问题可以给出原始问题一个下界；
  - 当满足一定条件时，原始问题与对偶问题的解是完全等价的；

对偶问题

- 原问题

$$\min f(\theta)$$

$$s.t.\ h_i(\theta) = 0 \quad i = 1,2,\cdots,\mathrm{p}$$

$$g_j(\theta) \le 0 \quad j = 1,2,\cdots,q$$

- 拉格朗日函数

$$L(\theta,\alpha,\beta) = f(\theta) + \sum \alpha_i g_i(\theta) + \sum \beta_j h_j(\theta)$$

- 原函数等价于： $f(\theta) = \max_{\alpha,\beta} L(\theta,\alpha,\beta)$

- 因此原问题： $\min f(\theta) = \min_{\theta} \max_{\alpha,\beta} L(\theta,\alpha,\beta)$

SVM

- $\min\limits_{w,b} \frac{1}{2}||w||^2$

- $s.t. \quad y_i(wx_i + b) - 1 \geq 0, \quad i = 1, 2, \cdots, N$

$$1 - y_i(wx_i + b) \leq 0$$

SVM的对偶问题

- 定义拉格朗日函数
  - $L = \frac{1}{2}||w||^2 + \sum_{i=1}^{N} \alpha_i(1 - y^i(w \cdot x^i + b)) \qquad \alpha_i \geq 0$
- 与原始问题等价

SVM的对偶问题

$$f = \min_{w,b}\left(\max_{\alpha} L(w,b,\alpha)\right)$$

$$f = \min_{w*,b*}\{g_{(w_1,b_1)}, g_{(w_2,b_2)}, g_{(w_3,b_3)}, \cdots\}$$

$$g_{(w_i,b_i)} = \max_{\alpha^*}\{L(w_i,b_i,\alpha_1), L(w_i,b_i,\alpha_2),\cdots\}$$

- 定义拉格朗日函数
  - $L = \frac{1}{2}||w||^2 + \sum_{i=1}^{N}\alpha_i(1 - y^i(w\cdot x^i + b))\quad \alpha_i \geq 0$
- 与原始问题等价
  - $\min_{b,w}(\max_{\alpha_i\geq 0} L(w,b,\alpha))$
  - 针对每一组(w,b)来调节α
  - $g(w,b) = \max_{\alpha_i\geq 0} L(w,b,\alpha)$ $\quad \min_{w,b} g(w,b)$
    - 无穷大
    - $\frac{1}{2}||w||^2$

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

➢ 如果(w,b)不好：在N个约束中至少有一条不满足，即$y^i(w\cdot x^i + b) < 1$，也就是$1 - y^i(w\cdot x^i + b) > 0$。此时，$\alpha_i$无穷大即可

➢ 如果(w,b)好：满足所有约束，即$y^i(w\cdot x^i + b) \geq 1$，也就是$1 - y^i(w\cdot x^i + b) < 0$。此时，$\alpha_i$最好取0，那么最大化的结果就是$\frac{1}{2}||w||^2$

对偶问题

- 定义对偶函数(dual function)
  - $\boldsymbol{D}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \min_{\boldsymbol{\theta}} \boldsymbol{L}(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta})$
- 对偶问题：$\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \min_{\boldsymbol{\theta}} \boldsymbol{L}(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta})$

- 假设原问题解$\boldsymbol{p}^*$, 对偶问题解$\boldsymbol{d}^*$
- 有$\boldsymbol{d}^* \le \boldsymbol{p}^*$
- 对偶问题是原问题解的一个下界

$$D(\alpha,\beta) = \min_{\theta^*} L(\theta^*,\alpha,\beta) \le L\left(\hat{\theta},\alpha,\beta\right) \le \max_{\alpha^*,\beta^*} L(\theta,\alpha^*,\beta^*) = f(\theta)$$

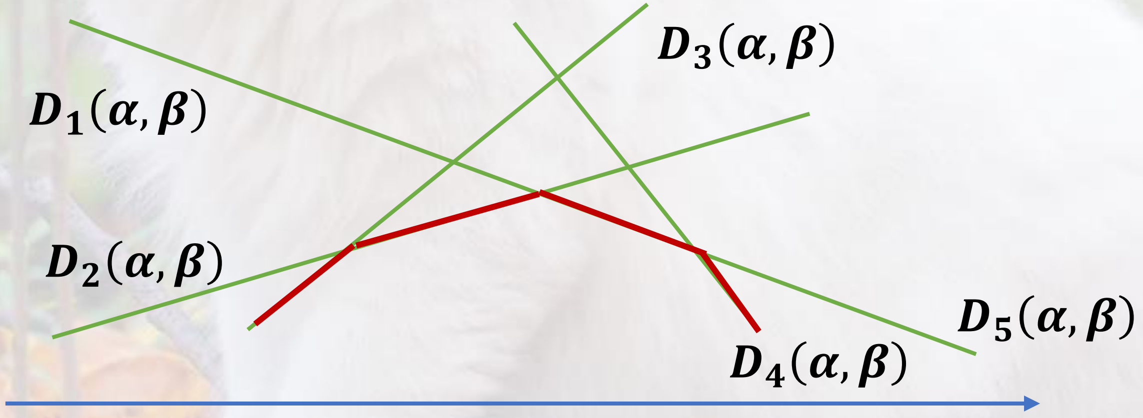
对偶函数一定凸（凹）函数

$$D(\alpha,\beta)=\min_{\theta} L(\theta,\alpha,\beta)$$

$L(x,\alpha,\beta)$看成一个函数集合，每个元素由$L(x_i,\alpha,\beta)$组成

$$D(\alpha,\beta)=inf\{D_1(\alpha,\beta),D_2(\alpha,\beta),\cdots,D_\infty(\alpha,\beta)\}$$

$$L(x,\alpha,\beta)=f(x)+\sum\alpha_i g_i(x)+\sum\beta_j h_j(x)$$ 关于α,β的仿射函数



求解最优化

- **拉格朗日乘子法（Lagrange Multiplier）**
  - 不等式情况

$$\min f(\theta)$$
$$s.t.\ h_i(\theta)=0 \quad i=1,2,\cdots,\mathrm{p}$$
$$g_j(\theta)\le 0 \quad j=1,2,\cdots,q$$

$$L(\theta,\alpha,\beta)=f(\theta)+\sum \alpha_i g_i(\theta)+\sum \beta_j h_j(\theta)$$

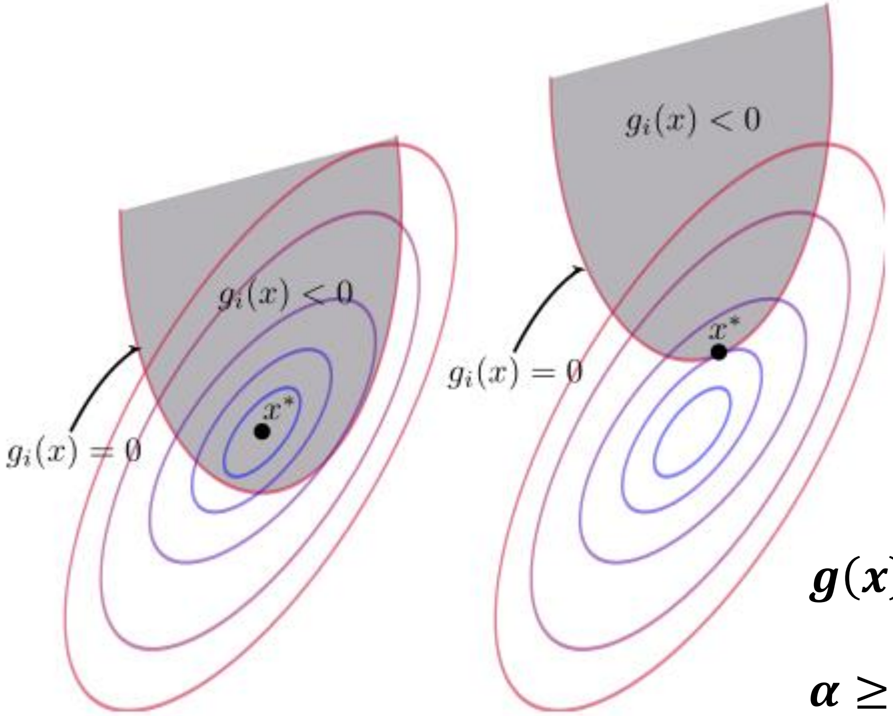
  - KKT条件（Karush Kuhn Tucker）

1. $\frac{\partial L}{\partial \theta}=0$
2. $\alpha_j g_j=0$
3. $h_i=0$
4. $g_j(\theta)\le 0$
5. $\alpha_j \ge 0$

$$L(x,\alpha) = f(x) + \sum \alpha_i g_i(x)$$

不等式情况

$$\min_{x} \ f(x)$$
$$s.t. \ \ g(x) \le 0$$



$$\alpha g(x) = 0$$

$$\frac{\partial L}{\partial x} = 0$$
$$\alpha_j g_j = 0$$
$$\alpha_j \ge 0$$

$$g(x) = 0$$

$$\alpha \ge 0$$

$$\alpha = 0$$

SVM的对偶问题

- $\min\limits_{bw} L = \frac{1}{2}||w||^2 + \sum_{i=1}^{N} \alpha_i (1 - y^i (w \cdot x^i + b))$

$$\frac{\partial L}{\partial w_j} = w_j - \sum_{i=1}^{N} \alpha_i y^i x_j^i = 0 \quad \boldsymbol{w_j = \sum_{i=1}^{N} \alpha_i y^i x_j^i}$$

$$\frac{\partial L}{\partial b} = \boldsymbol{\sum_{i=1}^{N} \alpha_i y^i = 0}$$

- 带入

$$\max_{\alpha} L = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \, y^i y^j (\boldsymbol{x}^i \cdot \boldsymbol{x}^j) + \sum_{i=1}^{N} \alpha_i$$

$$\sum_{i=1}^{N} \alpha_i y^i = 0 \qquad \alpha_i \geq 0$$

解的等价

原函数：$f(\boldsymbol{\theta}) = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \boldsymbol{L}(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta})$

对偶函数：$\boldsymbol{D}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \min_{\boldsymbol{\theta}} \boldsymbol{L}(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta})$

$f(\boldsymbol{\theta}^*) = \boldsymbol{D}(\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*)$

$$\boldsymbol{D}(\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*) = \min_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) + \sum \boldsymbol{\alpha}^* \boldsymbol{g}(\boldsymbol{\theta}) + \sum \boldsymbol{\beta}^* \boldsymbol{h}(\boldsymbol{\theta})$$

$$\leq f(\boldsymbol{\theta}^*) + \sum \boldsymbol{\alpha}^* \boldsymbol{g}(\boldsymbol{\theta}^*) + \boxed{0}$$

$$\leq f(\boldsymbol{\theta}^*)$$

$\boldsymbol{\alpha}^* \geq 0$

$\boldsymbol{g}(\boldsymbol{x}^*) \leq 0$

等号成立

$\boldsymbol{\theta}^*$是一个极大值点

$\nabla_{\boldsymbol{\theta}^*} \boldsymbol{L}(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta}) = 0$

等号成立

$\boldsymbol{\alpha}_j^* \boldsymbol{g}_j(\boldsymbol{\theta}^*)$都是非正

$\boldsymbol{\alpha}_j^* \boldsymbol{g}_j(\boldsymbol{\theta}^*) = 0$

满足KKT条件

- $\frac{\partial L}{\partial \theta} = 0$
- $\boldsymbol{\alpha}_j \boldsymbol{g}_j = 0$
- $\boldsymbol{h}_i = 0$
- $\boldsymbol{g}_j(\boldsymbol{\theta}) \leq 0$
- $\boldsymbol{\alpha}_j \geq 0$

SVM对偶问题解

- 根据KKT条件：

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

$$\nabla_b L(w^*,b^*,\alpha^*) = -\sum_{i=1}^{N} \alpha_i^* y_i = 0 \qquad b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j)$$

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

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

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

对于正分量$\alpha_j > 0$

$y_j\left(w^* x_j + b^*\right) = 1$

SVM对偶问题

- **SVM目标函数：**
  - $\max_{\alpha} L = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\, y^i y^j (\boldsymbol{x}^i \cdot \boldsymbol{x}^j) + \sum_{i=1}^{N}\alpha_i$
  - $\min_{\alpha} L = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\, y^i y^j (\boldsymbol{x}^i \cdot \boldsymbol{x}^j) - \sum_{i=1}^{N}\alpha_i$
  - $\sum_{i=1}^{N}\alpha_i y^i = 0$
  - $\alpha_i \geq 0$

例7.2

$$\min_{\alpha} L = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j y^i y^j (\boldsymbol{x}^i \cdot \boldsymbol{x}^j) - \sum_{i=1}^{N} \alpha_i$$
$$\sum_{i=1}^{N} \alpha_i y^i = 0$$
$$\alpha_i \geq 0$$

- **正实例$x_1 = (3, 3)^T$, $x_2 = (4, 3)^T$, 负实例$x_3 = (1, 1)^T$**

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

$$= \alpha_1^2 y_1^2 (3,3)\cdot(3,3) + 2\alpha_1\alpha_2 y_1 y_2 (3,3)\cdot(4,3) + 2\alpha_1\alpha_3 y_1 y_3 (3,3)\cdot(1,1) + 2\alpha_2\alpha_3 y_2 y_3 (4,3)\cdot(1,1) + \alpha_2^2 y_2^2 (4,3)\cdot(4,3) + \alpha_3^2 y_3^2 (1,1)\cdot(1,1)$$

$$= \frac{1}{2}(18\alpha_1^2 + 25\alpha_2^2 + 2\alpha_3^2 + 42\alpha_1\alpha_2 - 12\alpha_1\alpha_3 - 14\alpha_2\alpha_3) - \alpha_1 - \alpha_2 - \alpha_3$$

$$\text{s.t.} \quad \alpha_1 + \alpha_2 - \alpha_3 = 0$$

$$\alpha_i \geqslant 0, \quad i = 1, 2, 3$$

$$\alpha_3 = \alpha_1 + \alpha_2 \longrightarrow s(\alpha_1, \alpha_2) = 4\alpha_1^2 + \frac{13}{2}\alpha_2^2 + 10\alpha_1\alpha_2 - 2\alpha_1 - 2\alpha_2$$

例7.2

$$\frac{\partial s}{\partial \alpha_1} = 8\alpha_1 + 10\alpha_2 - 2 = 0$$

$$\frac{\partial s}{\partial \alpha_2} = 13\alpha_2 + 10\alpha_1 - 2 = 0$$

- $s(\alpha_1, \alpha_2) = 4\alpha_1^2 + \frac{13}{2}\alpha_2^2 + 10\alpha_1\alpha_2 - 2\alpha_1 - 2\alpha_2$
- 1. 两个均不为0，可得：$\alpha_1 = \frac{3}{2}, \alpha_2 = -1$ ✗
- 2. $\alpha_1 = 0, \alpha_2 = 0$ ✗

$$\frac{\partial s}{\partial \alpha_2} = 13\alpha_2 - 2 = 0$$

- 3. $\alpha_1 = 0, \alpha_2 \neq 0$，可得：$\alpha_2 = \frac{2}{13}$　$s = -\frac{2}{13}$
- 4. $\alpha_2 = 0, \alpha_1 \neq 0$，可得：$\alpha_1 = \frac{1}{4}$　$s = -\frac{1}{4}$ ✓　$\alpha_3 = \frac{1}{4}$

$$\frac{\partial s}{\partial \alpha_1} = 8\alpha_1 - 2 = 0$$

$$\alpha_i(1 - y_i(wx_i + b)) = 0$$

$$\Downarrow$$

$$y_i(wx_i + b) = 1$$

$$\Downarrow$$

α_1, α_3为支持向量