Data Mining

Chapter 4: Support Vector Machine

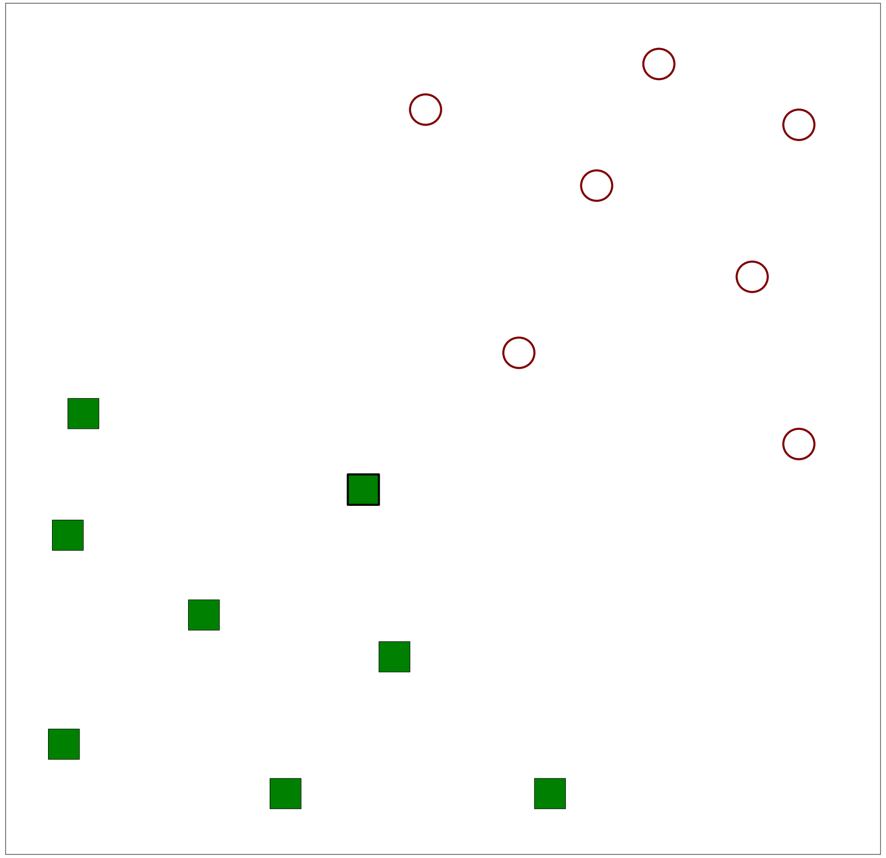
Yunming Ye, Baoquan Zhang
School of Computer Science
Harbin Institute of Technology, Shenzhen

Agenda

- Basic Idea of SVM
- Linear SVM
  - Hard-margin linear SVM
  - Soft-margin linear SVM
- Non-linear SVM
- SVM 编程实现

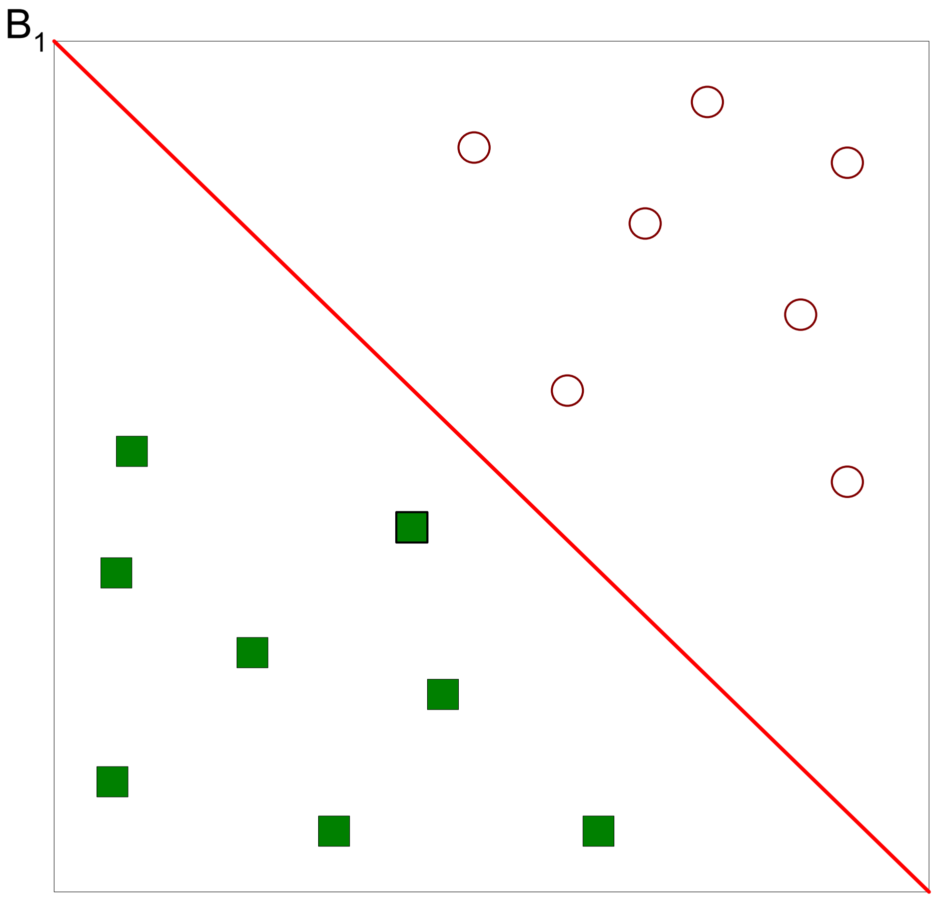
5.1 Basic Idea of SVM

Support Vector Machine

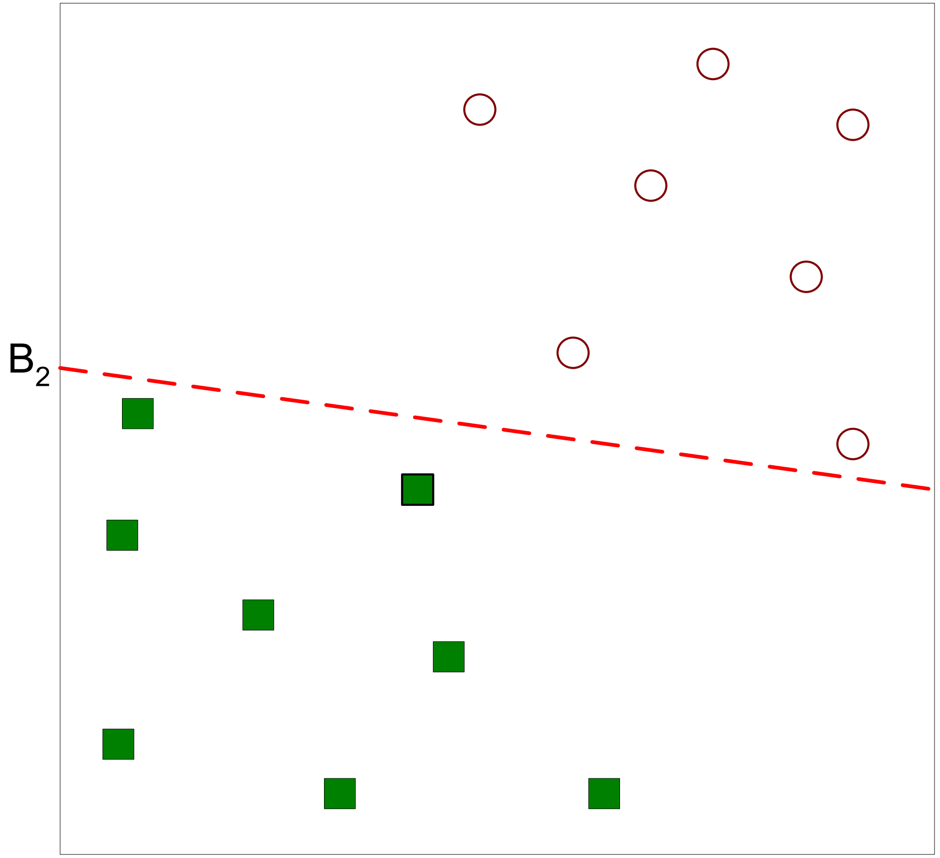


- Find a linear hyperplane (decision boundary) that will separate the data

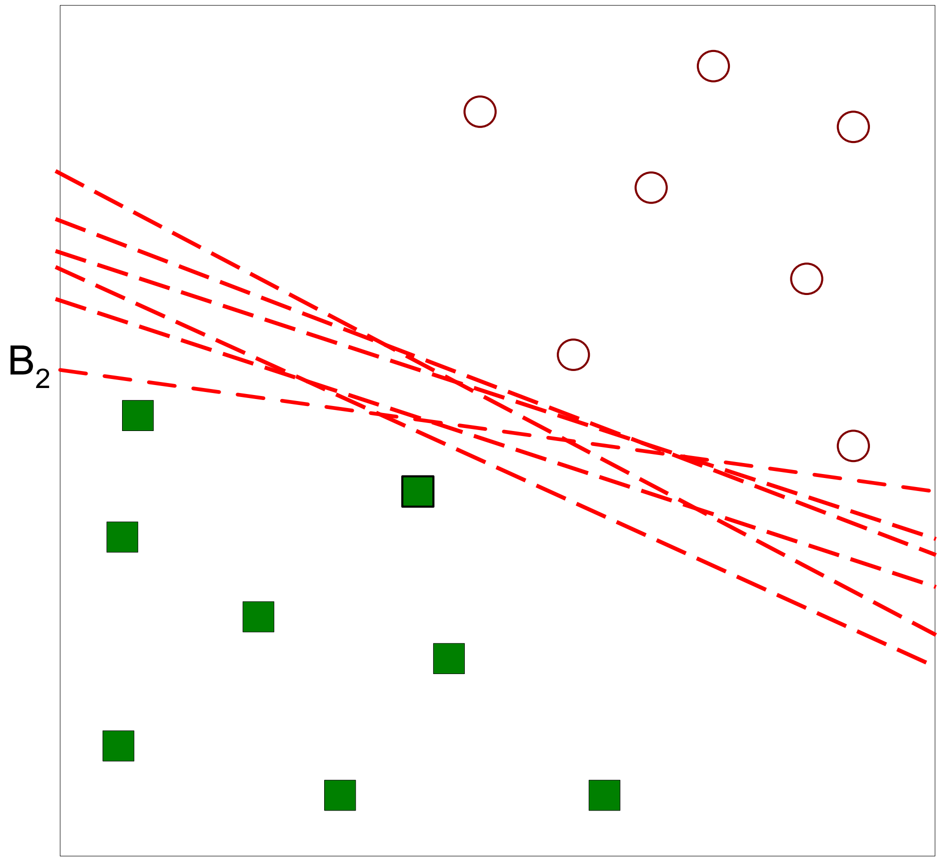
- One Possible Solution



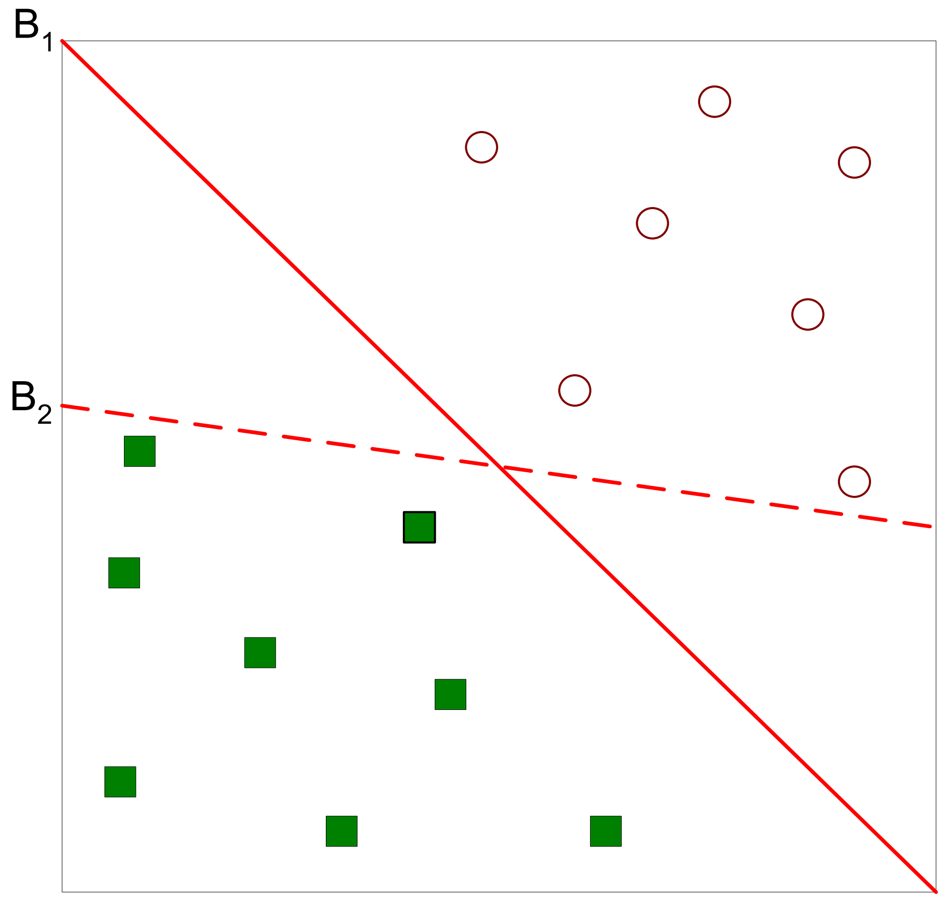
- Another possible solution



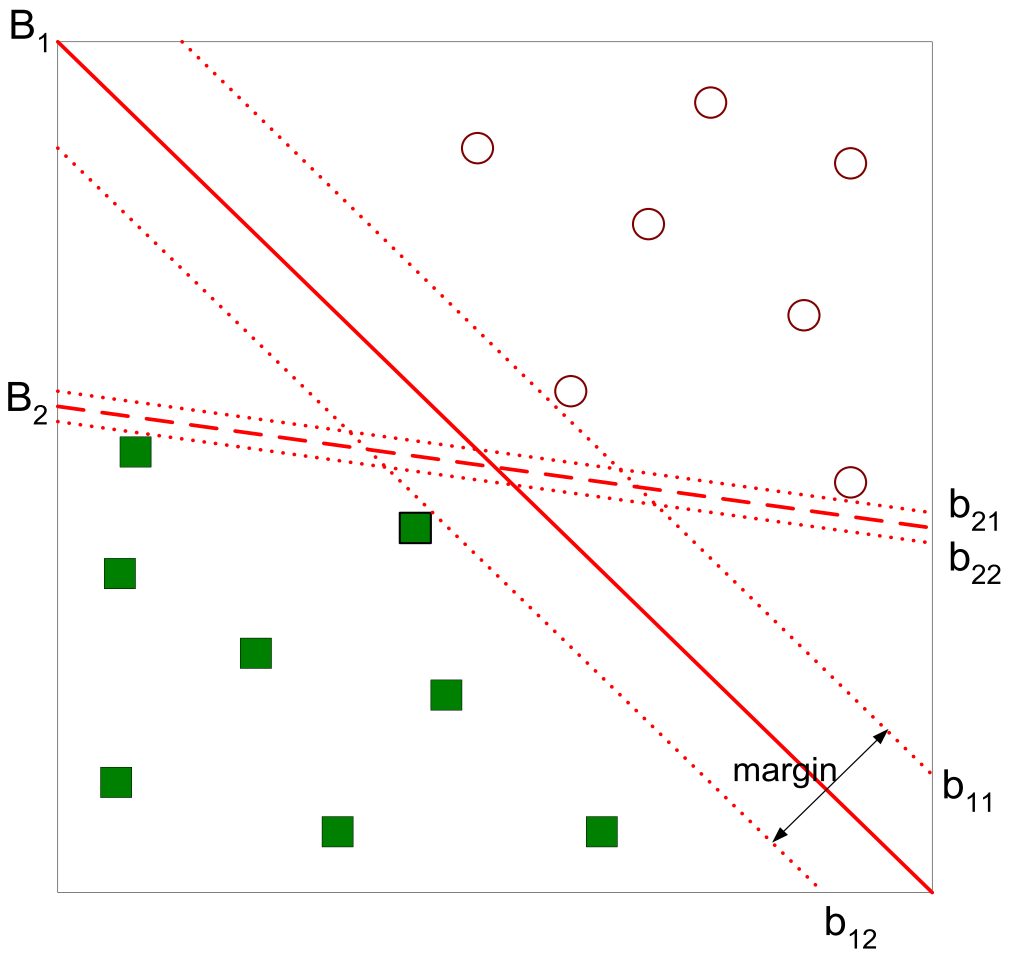
- Many possible solutions



- **Which one is better? B1 or B2?**
- **How do you define better?**

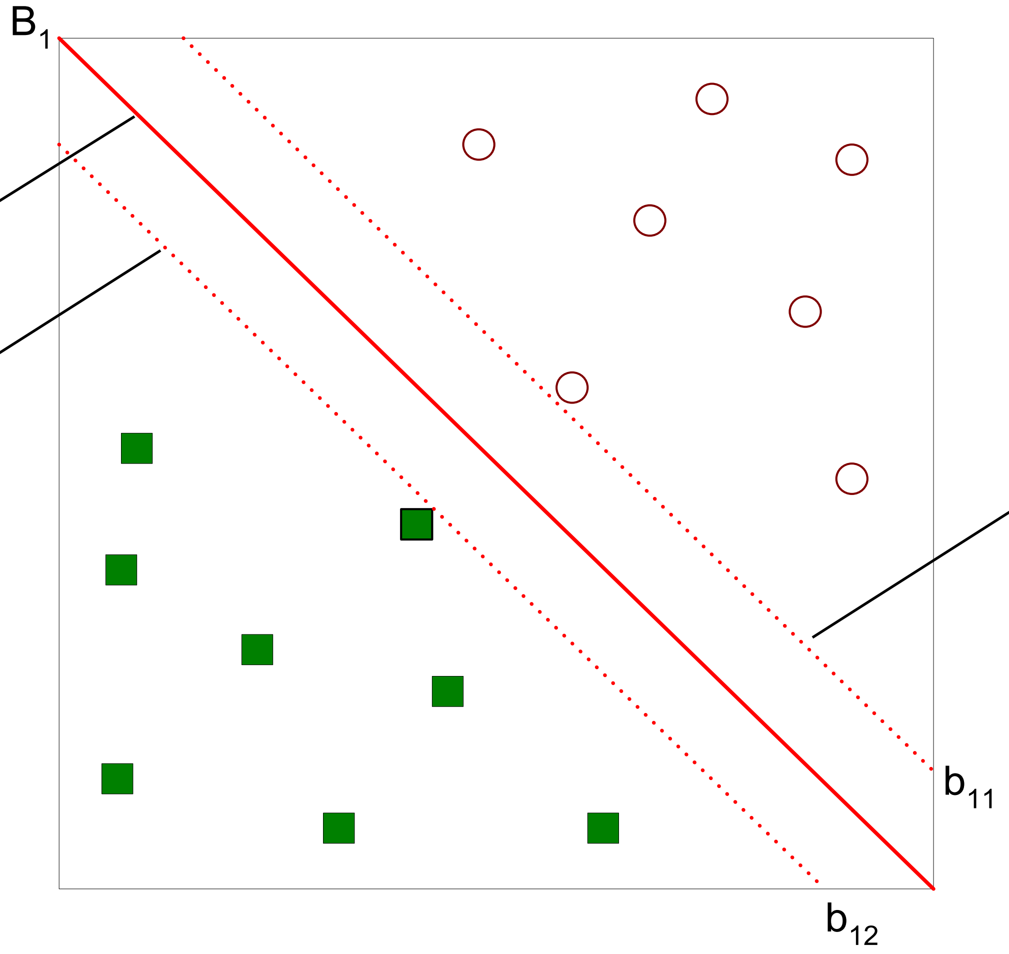


- Find hyperplane maximizes the margin => B1 is better than B2



More robust decision with larger margin!

Support Vector Machine



$$\vec{w} \bullet \vec{x} + b = 0$$

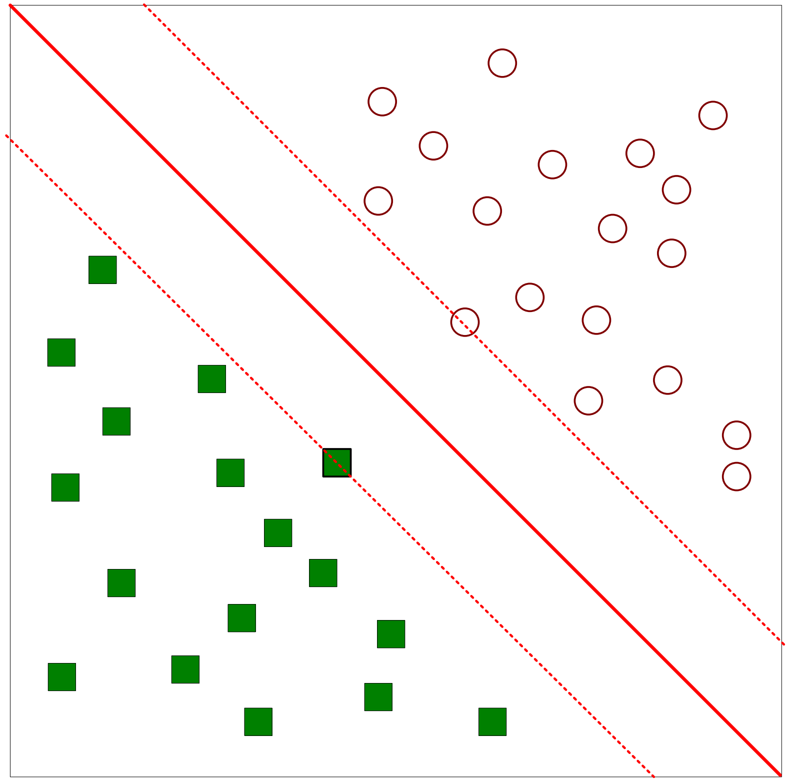
$$\vec{w} \bullet \vec{x} + b = -1$$

$$\vec{w} \bullet \vec{x} + b = +1$$

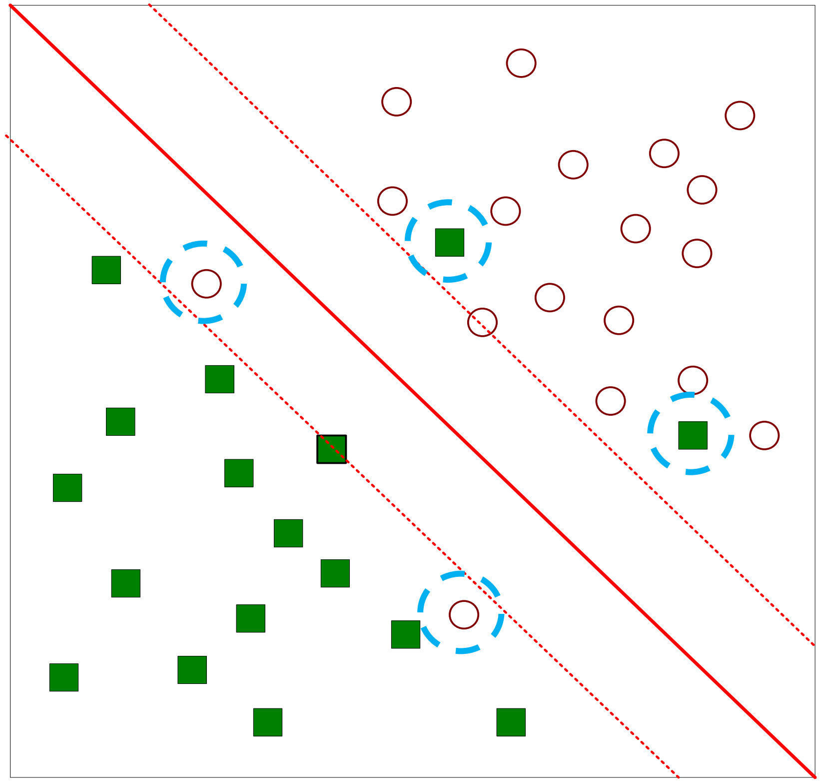
$$f(\vec{x}) = \begin{cases} 1 & \text{if } \vec{w} \bullet \vec{x} + b \geq 1 \\ -1 & \text{if } \vec{w} \bullet \vec{x} + b \leq -1 \end{cases}$$

$$\text{Margin} = \frac{2}{\|\vec{w}\|}$$

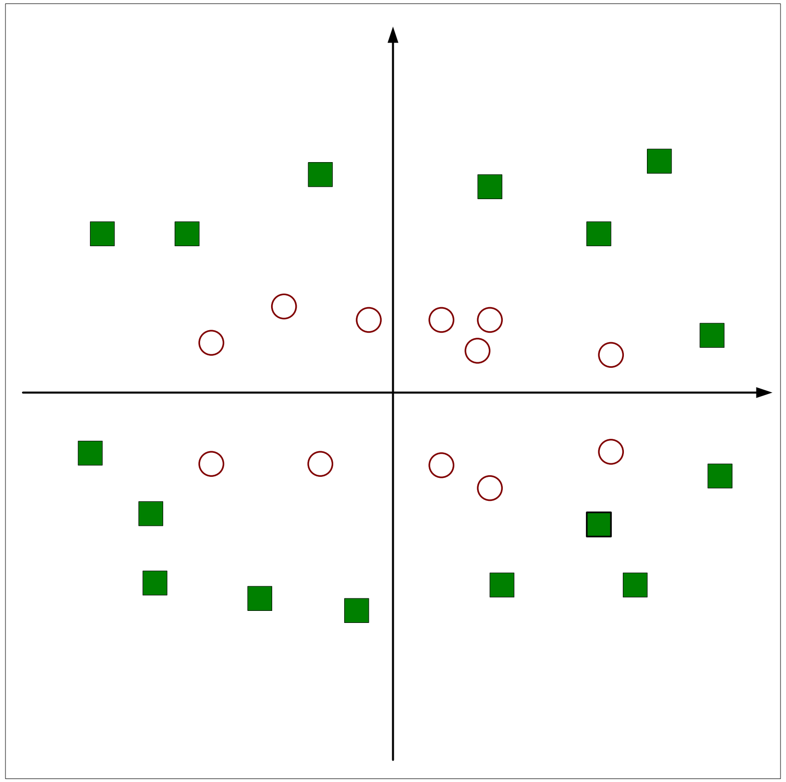
“可分性”不同的三类数据集



线性可分数据



“弱”线性不可分数据



“强”线性不可分数据

5.2 Hard-margin linear SVM
硬间隔线性支持向量机

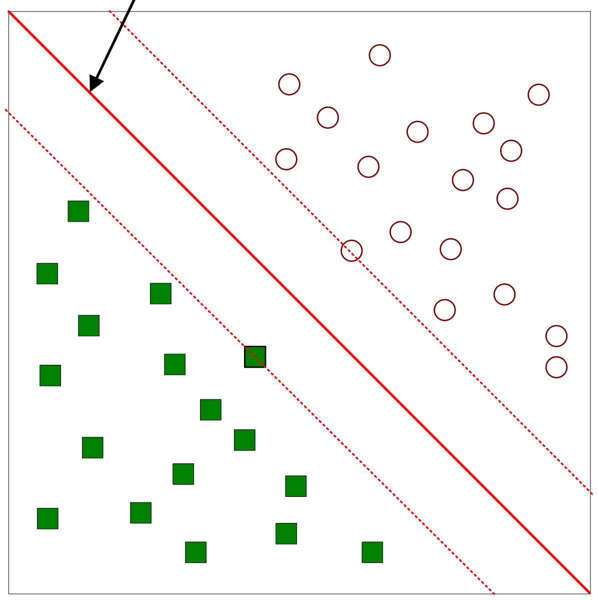
定义

- **超平面（hyperplane）**

$$\boldsymbol{w}^T\boldsymbol{x} + b = 0$$

$\boldsymbol{w} = (w_1, w_2, \dots, w_n)$代表超平面的法向量

b 代表原点到超平面的距离



- **线性可分数据集**
  - 数据集 $D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_m, y_m)\},\ \boldsymbol{x}_i \in \mathbb{R}^n,\ y_i\ \{\mathbf{-1}, \mathbf{1}\}$
  - 存在一个**超平面**能将D中的正负样本**严格地划分**到两侧
  - 这样的超平面称为：**分隔超平面**（separating hyperplane）

模型学习问题

对于线性可分训练数据集 $D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\ldots,(\boldsymbol{x}_m,y_m)\},\boldsymbol{x}_i\in\mathbb{R}^n,y_i\ \{-1,1\}$

学习目标是找到一个分隔超平面 $\boldsymbol{w}^T\boldsymbol{x}+b=0$使其满足：

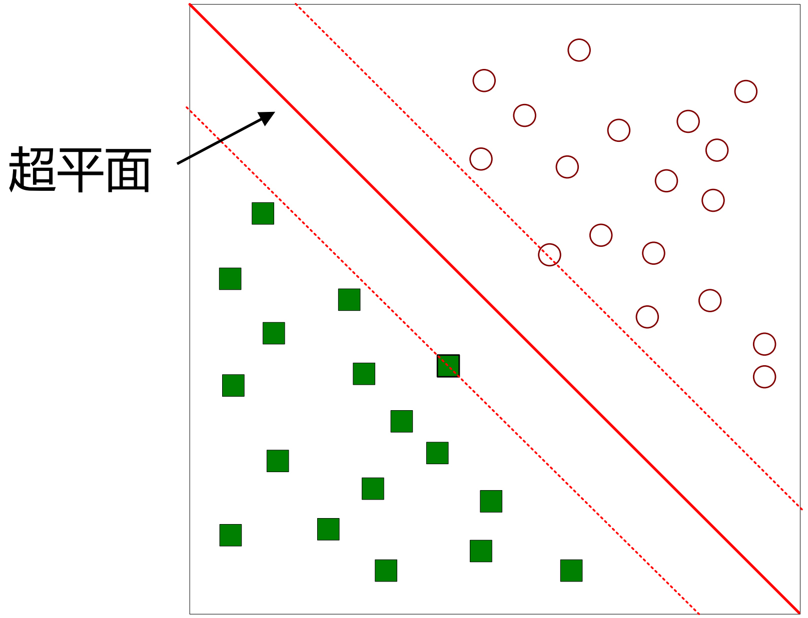
- 正确划分所有数据：

$$\left.\begin{array}{l}\boldsymbol{w}^T\boldsymbol{x}_i+b>0,\ \ y_i=\mathbf{1}\\ \boldsymbol{w}^T\boldsymbol{x}_i+b<0,\ \ y_i=\mathbf{-1}\end{array}\right\}\quad y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i+b\right)>0$$

- 间隔（margin）最大:

$$<\boldsymbol{w}^*,\boldsymbol{b}^*>=\max_{\boldsymbol{w},b}\ \text{margin}(\boldsymbol{w},b)$$

$$\text{margin}(\boldsymbol{w},b)=\min_{i=1,\ldots,m}\ \text{distance}(\boldsymbol{x}_i,<\boldsymbol{w},b>)$$

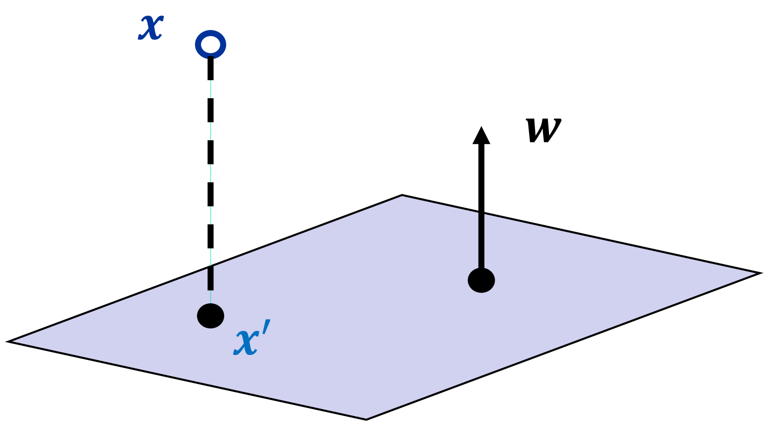


点到超平面的距离

- **定理** n维欧几里得空间中，点 $\boldsymbol{x} \in \mathbb{R}^n$ 到超平面 $\boldsymbol{w}^T\boldsymbol{x} + b = 0$的距离为：

$$d = \frac{|\boldsymbol{w}^T\boldsymbol{x} + b|}{\|\boldsymbol{w}\|}$$

$$\|\boldsymbol{w}\| = \sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}$$



模型学习问题（续）

- 对于线性可分的训练数据集$D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}, \boldsymbol{x}_i \in \mathbb{R}^n, y_i \; \{-1,1\}$；
- 超平面$\boldsymbol{w}^T\boldsymbol{x} + b = 0$ 的间隔可以定义为：

$$\text{margin}\,(\boldsymbol{w}, b) = \min_{i=1,2,\ldots,m} \frac{\left|\boldsymbol{w}^T\boldsymbol{x}_i + b\right|}{\|\boldsymbol{w}\|} = \min_{i=1,2,\ldots,m} \frac{1}{\|\boldsymbol{w}\|} y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)$$

原问题可改写：

$$< \boldsymbol{w}^*, \boldsymbol{b}^* > = \max_{\boldsymbol{w},b} \text{margin}(\boldsymbol{w}, b) \qquad y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) > 0$$

$$< \boldsymbol{w}^*, \boldsymbol{b}^* > = \max_{\boldsymbol{w},b} \quad \min_{i=1,2,\ldots,m} \frac{1}{\|\boldsymbol{w}\|} y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)$$

$$\text{s.t.} \quad y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) > 0, \qquad i = 1,2,\ldots,m$$

问题化简的目标

原问题：

$$\max_{\boldsymbol{w},b} \quad \min_{i=1,2,\dots,m} \quad \frac{1}{\|\boldsymbol{w}\|} y_i(\boldsymbol{w}^T \boldsymbol{x_i} + b)$$

$$\text{s.t.} \quad y_i\left(\boldsymbol{w}^T \boldsymbol{x_i} + b\right) > 0, \qquad i = 1,2,\dots,m$$

二次规划问题：

$$\min_{\boldsymbol{u}} \quad \frac{1}{2}\boldsymbol{u}^T \boldsymbol{Q} \boldsymbol{u} + \boldsymbol{p}^T \boldsymbol{u}$$

$$\text{s.t.} \quad \boldsymbol{a}_i^T \boldsymbol{u} \geq c_i, \quad i = 1,2,\dots,M$$

$$\boldsymbol{Q} \in \mathbb{R}^{N\times N}, \ \boldsymbol{u} \in \mathbb{R}^N, \ \boldsymbol{a_i} \in \mathbb{R}^N, \ \boldsymbol{p} \in \mathbb{R}^N, \ c_i \in \mathbb{R}$$

问题化简

原问题：

$$\max_{w,b} \boxed{\min_{i=1,2,\ldots,m} \frac{1}{\|\boldsymbol{w}\|} y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)}$$

$$\text{s.t.} \quad y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) > 0, \qquad i = 1,2,\ldots,m$$

- 按相同比例变化$\boldsymbol{w}$ 和 b 后，所得间隔不变，例如 $\boldsymbol{w}^T\boldsymbol{x} + b = 0$ 和 $6\boldsymbol{w}^T\boldsymbol{x} + 6b = 0$是同一超平面
- 不妨设：

$$\min_{i=1,2,\ldots,m} \ y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \ = 1$$

- 则原问题可以**化简**为以下有约束优化问题：

$$\max_{\boldsymbol{w},b} \ \frac{1}{\|\boldsymbol{w}\|}$$

$$\text{s.t.} \quad \min_{i=1,2,\ldots,m} \ y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \ = 1$$

问题： $\max_{\boldsymbol{w},b} \dfrac{1}{\|\boldsymbol{w}\|}$

$\text{s.t.} \quad \min_{i=1,2,\dots,m} y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \ = 1$

二次规划问题：

$$\min_{\boldsymbol{u}} \quad \frac{1}{2}\boldsymbol{u}^T\boldsymbol{Q}\boldsymbol{u} + \boldsymbol{p}^T\boldsymbol{u}$$

$$\text{s.t.} \quad \boldsymbol{a}_i^T\boldsymbol{u} \geq c_i, \quad i = 1,2,\dots,M$$

- 将条件改写：

$$\min_{i=1,2,\dots,m} y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \ = 1 \quad \Rightarrow \quad y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1, i = 1,2,\dots,m$$

- 但这样改写之后会不会导致求出来的$(\boldsymbol{w}, b)$并不能满足条件$\min_{i=1,2,\dots,m} y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \ = 1$ ？

- 可以证明并不会！

假设最终所求得的最佳$(\boldsymbol{w}, b)$只满足$y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) > 1$，而不满足$\min_{i=1,2,\dots,m} y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \ = 1$
那我们**假设此时**$\min_{i=1,2,\dots,m} y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \ = a > 1$，此时的margin $= \frac{\mathbf{1}}{\|\boldsymbol{w}\|}$**最大。**
但我们可以将$(\boldsymbol{w}, b)$缩小为$(\boldsymbol{w}/a, b/a)$，此时的margin $= \frac{\boldsymbol{a}}{\|\boldsymbol{w}\|}$ 比原来的大，这与**假设冲突。**
所以可以证明改写条件后并不会导致问题解的变化

问题：

$$\max_{\boldsymbol{w}\ b} \quad \frac{1}{\|\boldsymbol{w}\|}$$

$$\text{s.t.} \quad y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \geq 1, \qquad i = 1,2,\dots,m$$

- 可进一步转化为：

$$\max_{\boldsymbol{w},b} \frac{1}{\|\boldsymbol{w}\|} \quad \rightarrow \quad \min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\| \quad \rightarrow \quad \min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 \quad \rightarrow \quad \min_{\boldsymbol{w},b} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$$

$$\text{s.t.} \quad y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \geq 1, \qquad i = 1,2,\dots,m$$

最终可得：

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

$$\text{s.t.} \quad y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \geq 1, \qquad i = 1,2,\dots,m$$

问题求解

标准问题

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

$$\text{s.t.} \quad y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1, \qquad i = 1,2,\dots,m$$

凸二次规划问题

$$\min_{\boldsymbol{u}} \quad \frac{1}{2}\boldsymbol{u}^T\boldsymbol{Q}\boldsymbol{u} + \boldsymbol{p}^T\boldsymbol{u}$$

$$\text{s.t.} \quad \boldsymbol{a}_i^T\boldsymbol{u} \geq c_i, \qquad i = 1,2,\dots,M$$

$$\boldsymbol{Q} \in \mathbb{R}^{N\times N}, \boldsymbol{u} \in \mathbb{R}^N, \boldsymbol{a_i} \in \mathbb{R}^N, \boldsymbol{p} \in \mathbb{R}^N, c_i \in \mathbb{R}$$

- 标准问题是标准凸二次规划问题。代入凸二次规划形式，可得：

$$\boldsymbol{u} = \begin{bmatrix} b \\ \boldsymbol{w} \end{bmatrix}; \quad \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0}_n^T \\ \boldsymbol{0}_n & \boldsymbol{I}_{n\times n} \end{bmatrix}; \quad \boldsymbol{p} = \boldsymbol{0}_{n+1}; \quad \boldsymbol{a}_i^T = y_i[1 \quad \boldsymbol{x}_i^T]; \quad c_i = 1; \quad M = m$$

- 对于凸二次规划问题，可以使用如内点法、椭球法和梯度投影法等方法求解
- 在求出最优解$\boldsymbol{w}^*$和b^*之后，我们可以得到最终的分类决策函数为：

$$f(\boldsymbol{x}) = \text{sign}(\boldsymbol{w}^{*T}\boldsymbol{x} + b^*)$$

硬间隔线性支持向量机标准问题的学习算法

输入：线性可分数据集$D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}, \boldsymbol{x}_i \in \mathbb{R}^n, y_i \{-1,1\}$

输出：硬间隔最大化分离超平面和分类决策函数

- 第一步，构造并求解约束最优化问题

$$\min_{\boldsymbol{w},b} \quad \frac{1}{2}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}$$
$$\text{s.t.} \quad y_i\left(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x}_i + b\right) \geq 1, \qquad i = 1,2,\ldots,m$$

- 第二步，返回硬间隔最大化的分隔超平面$\boldsymbol{w}^{*\boldsymbol{T}}\boldsymbol{x} + b^* = 0$和分类决策函数$f(\boldsymbol{x}) = \text{sign}(\boldsymbol{w}^{*\boldsymbol{T}}\boldsymbol{x} + b^*)$

- 从线性可分数据集当中学习到**硬间隔（hard margin）最大化**的分离超平面及相应分类决策函数的过程称为**硬间隔线性支持向量机**，这里的**硬间隔**是指所有数据点都可以严格分开。
- 根据间隔最大化思想，硬间隔最大化分离超平面是唯一存在的。

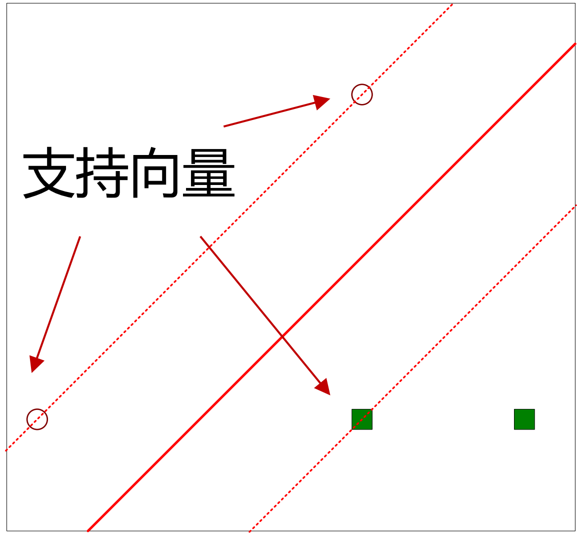
问题求解举例

标准问题：

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

$$\text{s.t.} \quad y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \geq 1, \qquad i = 1,2,\dots,m$$

使等号成立的点即为：**支持向量！**



$$\boldsymbol{X} = \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \\ 3 & 0 \end{bmatrix} \quad \boldsymbol{y} = \begin{bmatrix} +1 \\ +1 \\ -1 \\ -1 \end{bmatrix}$$

构造有约束优化问题

$$\min_{w,b} \quad \frac{1}{2}(w_1^2 + w_2^2)$$

$$\text{s.t.} \quad \begin{aligned} b &\geq 1 \quad \text{(i)} \\ 2w_1 + 2w_2 + b &\geq 1 \quad \text{(ii)} \\ -2w_1 \qquad - b &\geq 1 \quad \text{(iii)} \\ -3w_1 \qquad - b &\geq 1 \quad \text{(iv)} \end{aligned}$$

$$\left\{\begin{matrix} \text{(i)} + \text{(iii)} & \Rightarrow & w_1 \leq -1 \\ \text{(ii)} + \text{(iii)} & \Rightarrow & w_2 \geq +1 \end{matrix}\right\} \Longrightarrow \frac{1}{2}(w_1^2 + w_2^2) \geq 1$$

进一步求解可得该最优化问题的解为：$w_1 = -1,\ w_2 = 1,\ b = 1$

故硬间隔最大化分割超平面为：$-x_1 + x_2 + 1 = 0$，可以利用函数 $\text{sign}(-x_1 + x_2 + 1)$ 来分类

存在问题

$$\min_{\boldsymbol{w},b} \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$$
$$\text{s.t.} \quad y_i(\boldsymbol{w}^T\boldsymbol{x_i} + b) \geq 1, \qquad i = 1,2,\dots,m$$

凸二次规划问题：

- $n+1$个变量 $(b \in \mathbb{R},\ \boldsymbol{w} \in \mathbb{R}^n)$
- m个约束条件

- 算法存在的问题
  - 计算复杂度依赖于参数$\boldsymbol{w}$的维度，也就是特征空间的维度n。
  - 因为特征工程或非线性映射的缘故，特征空间的维度往往非常大，特别是在特征维度大于样本数量 $(n > m)$时，这会大大增加计算量。
  - 解决方法：将原问题转换成**对偶问题**，可以让计算复杂度不依赖于特征空间维度

原问题：

- $n+1$个变量
- m个约束条件

对偶问题：

- m个变量
- m个约束条件

问题转换的基本思路

原问题：

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x})$$

$$\text{s.t.} \quad c_i(\boldsymbol{x}) \le 0, i = 1,2,\dots,k$$

$$h_j(\boldsymbol{x}) = 0, j = 1,2,\dots,l$$

引入广义**拉格朗日函数**：

$$\boldsymbol{L}(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta})$$

$$= f(\boldsymbol{x}) + \sum_{i=1}^{k}\alpha_i c_i(\boldsymbol{x}) + \sum_{\mathrm{j}=1}^{l}\beta_i h_i(\boldsymbol{x})$$

对偶问题：

$$\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \quad \min_{\boldsymbol{x}} \boldsymbol{L}(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta})$$

$$\text{s.t.} \quad \alpha_i \ge 0, i = 1,2,\dots,k$$

问题转换

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

$$\text{s.t.} \quad y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \geq 1, \qquad i = 1,2,\dots,m$$

利用拉格朗日乘子向量 $\boldsymbol{\lambda} = [\lambda_i], \lambda_i \geq 0, i = 1,2,\dots,m$

构建原问题的拉格朗日函数为：

$$\boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\lambda}) = \underbrace{\boxed{\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}}}_{\text{优化目标}} + \sum_{i=1}^{m}\lambda_i\underbrace{\boxed{\left(1 - y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)\right)}}_{\text{约束条件}}$$

利用拉格朗日函数，原问题就等价于以下min-max问题：

$$\min_{\boldsymbol{w},b}\left(\max_{\boldsymbol{\lambda}:\lambda_i\geq 0}\boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\lambda})\right) = \min_{\boldsymbol{w},b}\left(\underbrace{\boxed{\infty}}_{\text{不符合约束条件}}, \underbrace{\boxed{\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}}}_{\text{符合约束条件}}\right) = \underbrace{\min_{\boldsymbol{w},b}\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}}_{\text{符合约束条件}}$$

将约束条件隐藏在max中！

证明：

- 不符合约束条件的$(\boldsymbol{w},b)$，即存在某个i 使得$1 - y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) > 0$，则令对应的$\lambda_i \to +\infty$，其它$\lambda = 0$，则：

$$\max_{\boldsymbol{\lambda}:\lambda_i\geq 0}\boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\lambda}) \to +\infty$$

- 符合约束条件的$(\boldsymbol{w},b)$，即$\forall i, 1 - y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \leq 0$，则**令 $\boldsymbol{\lambda_i(1 - y_i(w^T x_i + b)) = 0}$** 可使$\boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\lambda})$ 最大，可得:

$$\max_{\boldsymbol{\lambda}:\lambda_i\geq 0}\boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$$

在$(\boldsymbol{w}, b)$给定的情况下，对于任何$\boldsymbol{\lambda}: \lambda_i \geq 0$，都有

$$\max_{\boldsymbol{\lambda}:\lambda_i \geq 0} \boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\lambda}) \geq \boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\lambda})$$

故对于任何$\boldsymbol{w}, b, \boldsymbol{\lambda}: \lambda_i \geq 0$，都有

$$\min_{\boldsymbol{w},b} \left(\max_{\boldsymbol{\lambda}:\lambda_i \geq 0} \boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\lambda}) \right) \geq \min_{\boldsymbol{w},b} \boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\lambda})$$

比任何的都大，那肯定比**最大**的也大

$$\underbrace{\min_{\boldsymbol{w},b} \left(\max_{\boldsymbol{\lambda}:\lambda_i \geq 0} \boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\lambda}) \right)}_{p^*} \geq \underbrace{\boxed{\max_{\boldsymbol{\lambda}:\lambda_i \geq 0} \left(\min_{\boldsymbol{w},b} \boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\lambda}) \right)}}_{d^*}$$

拉格朗日对偶问题

- 设原问题的最优值为p^*，对偶问题的最优值为d^*
  - 若$p^* \geq d^*$，则两个问题的关系是**弱**对偶性
  - 若$p^* = d^*$，则两个问题的关系是**强**对偶性
- 若原问题满足以下条件（Slater条件），则原问题与对偶问题有强对偶性
  - ✓ 原问题是凸问题
  - ✓ 原问题有解
  - ✓ 原问题的约束条件是线性约束条件

$$\min_{\boldsymbol{w},b} \left(\max_{\boldsymbol{\lambda}:\lambda_i \geq 0} \frac{1}{2} \boldsymbol{w}^T \boldsymbol{w} + \sum_{i=1}^{m} \lambda_i (1 - y_i(\boldsymbol{w}^T \boldsymbol{x}_i + b)) \right)$$

- SVM原问题满足上述条件，故$p^* = d^*$，**对偶问题的解就是原问题的解**，我们解决对偶问题即可

对偶问题简化

$$\max_{\boldsymbol{\lambda}:\lambda_i\geq 0}\left(\min_{\boldsymbol{w},b}\boxed{\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{i=1}^{m}\lambda_i(1-y_i(\boldsymbol{w}^T\boldsymbol{x}_i+b))}\right) \qquad \boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\lambda})$$

- $(\boldsymbol{w},b)$要满足$\boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\lambda})$最小，则对应梯度应为0，即：

➢ $\frac{\partial L(\boldsymbol{w},b,\boldsymbol{\lambda})}{\partial b}=0=-\sum_{i=1}^{m}\lambda_i y_i$，我们将$\sum_{i=1}^{m}\lambda_i y_i=0$ 加到对偶问题的条件中不会影响问题的最优解，故

$$\max_{\boldsymbol{\lambda}:\lambda_i\geq 0;\sum\lambda_i y_i=0}\left(\min_{\boldsymbol{w},b}\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{i=1}^{m}\lambda_i(1-y_i(\boldsymbol{w}^T\boldsymbol{x}_i))-\cancel{\sum_{i=1}^{m}\lambda_i y_i\cdot b}\right)$$

$$\Rightarrow \max_{\boldsymbol{\lambda}:\lambda_i\geq 0;\sum\lambda_i y_i=0}\left(\min_{\boldsymbol{w},b}\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{i=1}^{m}\lambda_i(1-y_i(\boldsymbol{w}^T\boldsymbol{x}_i))\right)$$

➢ $\frac{\partial L(\boldsymbol{w},b,\boldsymbol{\lambda})}{\partial \boldsymbol{w}}=\boldsymbol{0}=\boldsymbol{w}-\sum_{i=1}^{m}\lambda_i y_i\boldsymbol{x}_i$，同样将$\boldsymbol{w}=\sum_{i=1}^{m}\lambda_i y_i\boldsymbol{x}_i$ 加到对偶问题的条件中，

$$\max_{\boldsymbol{\lambda}:\lambda_i\geq 0;\sum\lambda_i y_i=0;\boldsymbol{w}=\sum\lambda_i y_i\boldsymbol{x}_i}\left(\min_{\boldsymbol{w},b}\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{i=1}^{m}\lambda_i(1-y_i(\boldsymbol{w}^T\boldsymbol{x}_i))\right)$$

$$\Rightarrow \max_{\boldsymbol{\lambda}:\lambda_i\geq 0;\sum\lambda_i y_i=0;\boldsymbol{w}=\sum\lambda_i y_i\boldsymbol{x}_i}\left(\min_{\boldsymbol{w},b}\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{i=1}^{m}\lambda_i-\boldsymbol{w}^T\boldsymbol{w}\right)$$

展开

$$\Rightarrow \max_{\boldsymbol{\lambda}:\lambda_i\geq 0;\sum\lambda_i y_i=0;\boldsymbol{w}=\sum\lambda_i y_i\boldsymbol{x}_i}\left(\min_{\boldsymbol{w},b}-\frac{1}{2}\left\|\sum_{i=1}^{m}\lambda_i y_i\boldsymbol{x}_i\right\|^2+\sum_{i=1}^{m}\lambda_i\right) \Rightarrow \boxed{\max_{\boldsymbol{\lambda}:\lambda_i\geq 0;\sum\lambda_i y_i=0;\boldsymbol{w}=\sum\lambda_i y_i\boldsymbol{x}_i}-\frac{1}{2}\left\|\sum_{i=1}^{m}\lambda_i y_i\boldsymbol{x}_i\right\|^2+\sum_{i=1}^{m}\lambda_i}$$

与$\boldsymbol{w},b$无关，可以去掉最小化

最终需求解的对偶问题

KKT条件

$$\max_{\lambda:\lambda_i\geq 0;\sum\lambda_i y_i=0;\boldsymbol{w}=\sum\lambda_i y_i \boldsymbol{x}_i} -\frac{1}{2}\left\|\sum_{i=1}^{m}\lambda_i y_i \boldsymbol{x}_i\right\|^2+\sum_{i=1}^{m}\lambda_i$$

总结一下，$(\boldsymbol{w}, b, \boldsymbol{\lambda})$ 是原问题与对偶问题最优解的充要条件是$(\boldsymbol{w}, b, \boldsymbol{\lambda})$满足下面的Karush-Kuhn-Tucker（KKT）条件：

- 满足原问题的约束条件：$y_i\left(\boldsymbol{w}^T \boldsymbol{x}_i + b\right) \geq 1,\ i = 1,2,\dots,m$
- 满足对偶问题约束条件：$\lambda_i \geq 0$
- 满足对偶问题优化条件：
  - $\frac{\partial L(\boldsymbol{w},b,\boldsymbol{\lambda})}{\partial b} = 0 = -\sum_{i=1}^{m}\lambda_i y_i$，即$\sum_{i=1}^{m}\lambda_i y_i = 0$
  - $\frac{\partial L(\boldsymbol{w},b,\boldsymbol{\lambda})}{\partial \boldsymbol{w}} = \boldsymbol{0} = \boldsymbol{w} - \sum_{i=1}^{m}\lambda_i y_i \boldsymbol{x}_i$，即$\boldsymbol{w} = \sum_{i=1}^{m}\lambda_i y_i \boldsymbol{x}_i$
- 满足原问题的优化条件：将原问题转换成 minmax问题时，$\lambda_i\left(1 - y_i\left(\boldsymbol{w}^T \boldsymbol{x}_i + b\right)\right) = 0$

对偶互补条件

求解对偶问题

$$\max_{\lambda:\lambda_i\geq 0;\sum\lambda_i y_i=0;\boldsymbol{w}=\sum\lambda_i y_i \boldsymbol{x}_i} -\frac{1}{2}\left\|\sum_{i=1}^{m}\lambda_i y_i \boldsymbol{x}_i\right\|^2+\sum_{i=1}^{m}\lambda_i$$

改写上述问题：最大化转换为最小化（取负号）

$$\min_{\boldsymbol{\lambda}} \quad \frac{1}{2}\left\|\sum_{i=1}^{m}\lambda_i y_i \boldsymbol{x}_i\right\|^2-\sum_{i=1}^{m}\lambda_i$$

展开

$$\min_{\boldsymbol{\lambda}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j-\sum_{i=1}^{m}\lambda_i$$

$$\text{s.t.} \quad \boxed{\sum_{i=1}^{m}\lambda_i y_i=0}$$

$$\lambda_i\geq 0, \qquad i=1,2,\dots,m$$

这仍是一个凸二次规划的问题

m个变量 ($\boldsymbol{\lambda}\in\mathbb{R}^m$)

m+1个约束条件

二次规划问题：

$$\min_{\boldsymbol{u}} \quad \frac{1}{2}\boldsymbol{u}^T\boldsymbol{Q}\boldsymbol{u}+\boldsymbol{p}^T\boldsymbol{u}$$

$$\text{s.t.} \quad \boldsymbol{a}_i^T\boldsymbol{u}\geq c_i, \qquad i=1,2,\dots,M$$

$\boldsymbol{Q}\in\mathbb{R}^{N\times N}$，$\boldsymbol{u}\in\mathbb{R}^N$，$\boldsymbol{a_i}\in\mathbb{R}^N$，$\boldsymbol{p}\in\mathbb{R}^N$，$c_i\in\mathbb{R}$

等号转换成两条不等式

$$\sum_{i=1}^{m}\lambda_i y_i\geq 0$$

$$-\sum_{i=1}^{m}\lambda_i y_i\geq 0$$

- 代入凸二次规划形式，可得：

$$\boldsymbol{u}=\boldsymbol{\lambda};\ \boldsymbol{Q}=\left[q_{ij}=y_i y_j\boldsymbol{x}_i^T\boldsymbol{x}_j\right]_{m\times m};\ \boldsymbol{p}=-\mathbf{1}_m;$$

$\boldsymbol{e_i}\in\mathbb{R}^m$是第i行为1的单位向量

$$\boxed{\boldsymbol{a}_{\geq}=\boldsymbol{y}=[y_i]_m;c_{\geq}=0;\boldsymbol{a}_{\leq}=-\boldsymbol{y};c_{\leq}=0;}\ \boldsymbol{a_i}=\boldsymbol{e_i};c_i=0;$$

- 由于$\boldsymbol{Q}$可能比较大，实际应用中需利用专为SVM设计的方法解上述问题，得最优解 $\boldsymbol{\lambda}$

求解对偶问题

（KKT）条件：

- $y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1,\ i = 1,2,\dots,m$
- $\lambda_i \geq 0$
- $\sum_{i=1}^{m} \lambda_i y_i = 0$; $\boxed{\boldsymbol{w} = \sum_{i=1}^{m} \lambda_i y_i \boldsymbol{x}_i}$
- $\boxed{\lambda_i\left(1 - y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)\right) = 0}$（对偶互补条件）

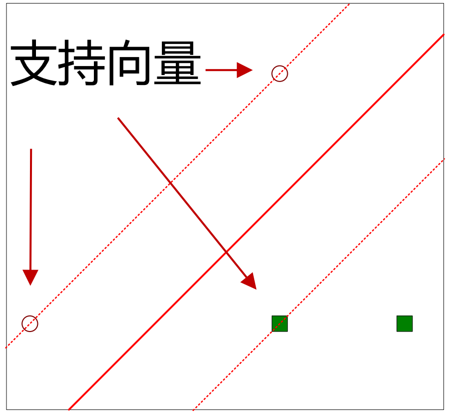
- 在求得最优解$\boldsymbol{\lambda}^* = [\lambda_i^*]$后
- 可以利用KKT条件求得原问题的最优解$\boldsymbol{w}^*$和b^*
  - $\boldsymbol{w}^* = \sum_{i=1}^{m} \lambda_i^* y_i \boldsymbol{x}_i$
  - 利用对偶互补条件，如果某个j 使得$\lambda_j^* > 0$，则$1 - y_j(\boldsymbol{w}^{*T}\boldsymbol{x}_j + b) = 0$，左右两边乘以$y_j$后可得：

当有多个$\lambda_j^* > 0$时，算得的b可能不一样

$$b^* = y_j - \boldsymbol{w}^{*T}\boldsymbol{x}_j$$

- 最后我们可以得到最终的分类决策函数的对偶形式为

$$f(x) = \mathrm{sign}(\boldsymbol{w}^{*T}\boldsymbol{x} + b^*)$$



- 支持向量
  - 对偶问题的解$\boldsymbol{\lambda}^* = [\lambda_i^*](i = 1,2,\dots,n)$中对应于分量$\lambda_i^* \neq 0$的样本点$(\boldsymbol{x}_i, y_i)$的实例$\boldsymbol{x}_i$称为**支持向量。**
  - 线性支持向量机中，支持向量可理解为经过间隔边界上的样本点，其作为向量支持着分类界限，决定决策超平面的参数取值。

硬间隔线性SVM对偶问题的学习算法

输入：线性可分数据集$D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\dots,(\boldsymbol{x}_m,y_m)\},\boldsymbol{x}_i\in\mathbb{R}^n,y_i\ \{-1,1\}$

输出：硬间隔最大化分离超平面和分类决策函数

- 第一步，构造带约束的凸二次规划问题，并求解得到最优解$\boldsymbol{\lambda}^*=[\lambda_i^*](i=1,2,\dots,m)$

$$\min_{\boldsymbol{\lambda}}\quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j-\sum_{i=1}^{m}\lambda_i$$

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

$$\lambda_i\geq 0,\qquad i=1,2,\dots,m$$

- 第二步，计算对偶问题对应的最优解$\boldsymbol{w}^*$，并任意选择$\boldsymbol{\lambda}^*$的一个正分量$\lambda_j^*>0$，以求解b^*

$$\boldsymbol{w}^*=\sum_{i=1}^{m}\lambda_i^* y_i\boldsymbol{x}_i$$

$$b^*=y_j-\boldsymbol{w}^{*T}\boldsymbol{x}_j$$

- 第三步，返回硬间隔最大化分离超平面$\boldsymbol{w}^{*T}\boldsymbol{x}+b^*=0$和分类决策函数$f(x)=\text{sign}(\boldsymbol{w}^{*T}\boldsymbol{x}+b^*)$

硬间隔线性SVM的两种学习算法对比

- 原问题学习算法

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

$$\text{s.t.} \quad y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \ge 1, \qquad i = 1,2,\dots,m$$

- n个变量
- m个约束条件

- 对偶问题学习算法

$$\min_{\boldsymbol{\lambda}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j - \sum_{i=1}^{m}\lambda_i$$

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

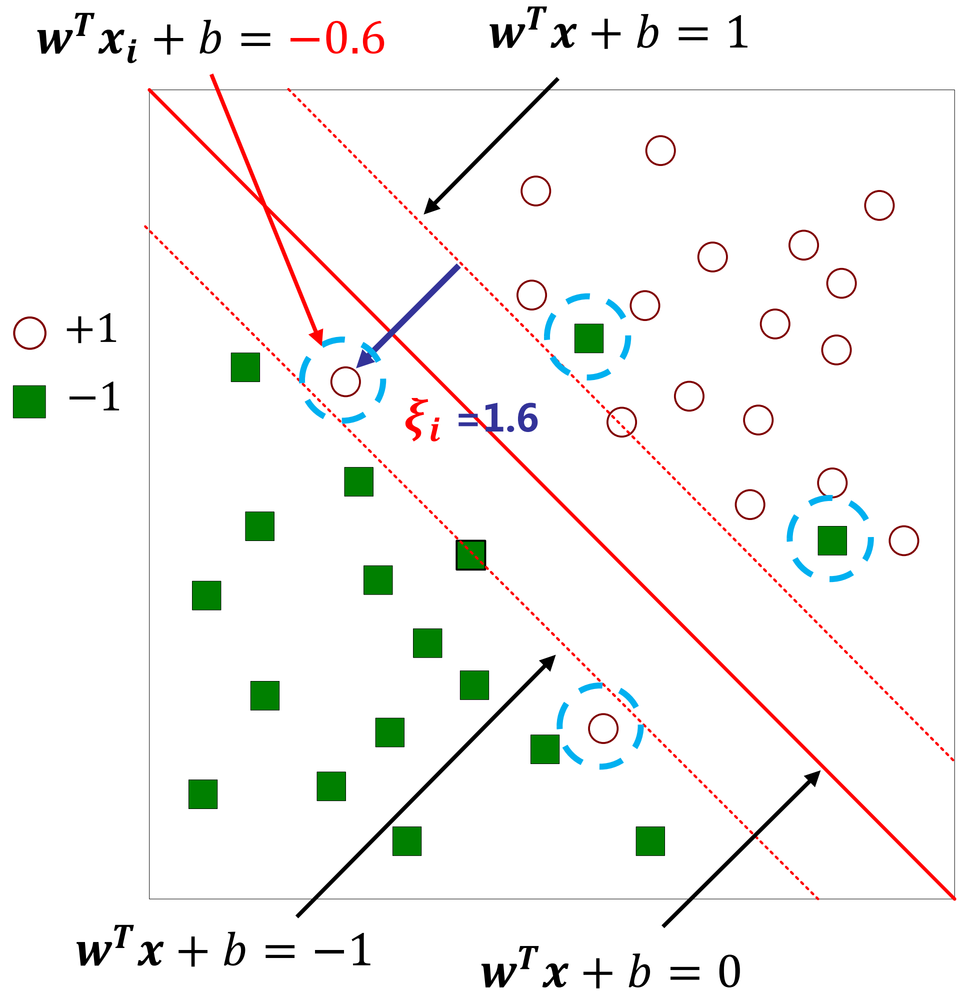
$$\lambda_i \ge 0, \qquad i = 1,2,\dots,m$$

- m个变量
- m+1个约束条件

- 当特征维度小于样本数量时($n < m$)，使用**原问题**算法较好
- 当特征维度大于样本数量时($n > m$)，使用**对偶问题**算法较好

5.3 Soft-margin linear SVM
软间隔线性支持向量机

软间隔线性SVM



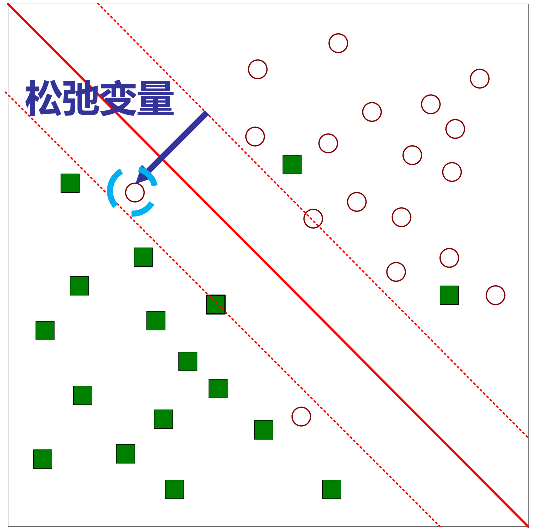
$$\min_{\boldsymbol{w},b} \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$$

$$\text{s.t.} \quad y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \geq 1, \qquad i = 1,2,\dots,m$$

- 数据异常点意味着其不满足约束条件 $y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \geq 1$
- 解决方案：引入**松弛变量** $\xi_i \geq 0$允许其违反约束条件：

$$y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \geq 1 - \xi_i$$

软间隔线性SVM的原问题



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

$$\text{s.t.} \quad y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1, \qquad i = 1,2,\dots,m$$

- 引入**松弛变量**$\xi_i \geq 0$允许数据点违反约束条件，需对松弛变量加以惩罚：

$$\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \underbrace{C\sum_{i=1}^{m}\xi_i}_{\text{惩罚项}}$$

$$\text{s.t.} \quad y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1 - \xi_i, \qquad i = 1,2,\dots,m$$

$$\xi_i \geq 0, \qquad i = 1,2,\dots,m$$

$C > 0$称为惩罚参数

- C值大时，对误分类的惩罚增大
- C值小时，对误分类的惩罚减小

上式属于凸二次规划问题：

- $n + m + 1$个变量 ($b \in \mathbb{R},\ \boldsymbol{w} \in \mathbb{R}^n, \boldsymbol{\xi} \in \mathbb{R}^m$)
- $2m$个约束条件

软间隔线性SVM

原问题 $\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{m}\xi_i$

s.t. $\quad y_i\left(\boldsymbol{w}^T\boldsymbol{x_i} + b\right) \ge 1 - \xi_i, \qquad i = 1,2,\dots,m$

$\qquad \xi_i \ge 0, \qquad i = 1,2,\dots,m$

利用拉格朗日乘子向量 $\boldsymbol{\lambda} = [\lambda_i], \lambda_i \ge 0, \boldsymbol{\beta} = [\beta_i], \beta_i \ge 0, i = 1,2,\dots,m$ ，构建原问题的拉格朗日函数为:

$$\boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \underbrace{\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{m}\xi_i}_{\text{优化目标}} + \sum_{i=1}^{m}\lambda_i\left(\underbrace{1 - \xi_i - y_i\left(\boldsymbol{w}^T\boldsymbol{x_i} + b\right)}_{\text{约束条件}}\right) + \sum_{i=1}^{m}\beta_i\left(\underbrace{-\xi_i}_{\text{约束条件}}\right)$$

利用拉格朗日函数，原问题就等价于以下min-max问题：

$$\min_{\boldsymbol{w},b,\boldsymbol{\xi}}\left(\max_{\boldsymbol{\lambda}:\lambda_i\ge 0,\boldsymbol{\beta}:\beta_i\ge 0}\boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta})\right) = \min_{\boldsymbol{w},b,\boldsymbol{\xi}}\Big(\underbrace{\infty}_{\text{不符合约束条件}}, \underbrace{\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{m}\xi_i}_{\text{符合约束条件}}\Big) = \underbrace{\min_{\boldsymbol{w},b,\boldsymbol{\xi}}\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{m}\xi_i}_{\text{符合约束条件}}$$

与原问题相同，将约束条件隐藏在max中

证明（min-max问题中对$(\boldsymbol{w}, b, \boldsymbol{\xi})$ **没约束**，所以要讨论）：

- 不符合约束条件的$(\boldsymbol{w}, b, \boldsymbol{\xi})$，即存在某个$i$ 使得$1 - \xi_i - y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) > 0$，则令对应的$\lambda_i \to \infty$，其他$\lambda, \beta = 0$，则

$$\max_{\boldsymbol{\lambda}:\lambda_i\ge 0,\boldsymbol{\beta}:\beta_i\ge 0}\boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta}) \to \infty$$

- 符合约束条件的$(\boldsymbol{w}, b, \boldsymbol{\xi})$，即$\forall i, 1 - \xi_i - y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \le 0$，$\xi_i \ge 0$则令$\boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta})$中的两个约束条件部分为0：

$$\max_{\boldsymbol{\lambda}:\lambda_i\ge 0,\boldsymbol{\beta}:\beta_i\ge 0}\boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{m}\xi_i$$

软间隔线性SVM**对偶问题**

在$(\boldsymbol{w}, b, \boldsymbol{\xi})$给定的情况下，对于任何$\lambda_i, \beta_i \geq \mathbf{0}$，都有

$$\max_{\boldsymbol{\lambda}:\lambda_i \geq 0, \boldsymbol{\beta}:\beta_i \geq 0} \boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta}) \geq \boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta})$$

故对于任何$\boldsymbol{w}, b, \boldsymbol{\xi}, \lambda_i \geq 0, \beta_i \geq 0$，都有

$$\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \left(\max_{\boldsymbol{\lambda}:\lambda_i \geq 0, \boldsymbol{\beta}:\beta_i \geq 0} \boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta}) \right) \geq \min_{\boldsymbol{w},b,\boldsymbol{\xi}} \boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta})$$

比任何的都大，那
肯定比**最大**的也大

$$\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \left(\max_{\boldsymbol{\lambda}:\lambda_i \geq 0, \boldsymbol{\beta}:\beta_i \geq 0} \boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta}) \right) \geq \boxed{\max_{\boldsymbol{\lambda}:\lambda_i \geq 0, \boldsymbol{\beta}:\beta_i \geq 0} \left(\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta}) \right)}$$

拉格朗日对偶问题

原问题满足Slater条件，解决其**对偶问题**就可解决原问题：

- $(\boldsymbol{w}, b, \boldsymbol{\xi})$要满足$\boldsymbol{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta})$最小，则对$\xi$求导应为0，即：

因为$C - \lambda_i = \beta_i \geq 0$

- $\frac{\partial \boldsymbol{L}}{\partial \xi_i} = 0 = C - \lambda_i - \beta_i$，将$\beta_i = C - \lambda_i$，$\boxed{0 \leq \lambda_i \leq C}$加到条件中不会影响问题的最优解，而且可以借此去掉$\beta_i, \xi$：

$$\max_{\beta_i = C - \lambda_i; 0 \leq \lambda_i \leq C} \left(\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \frac{1}{2} \boldsymbol{w}^T \boldsymbol{w} + C \sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \lambda_i \left(1 - \xi_i - y_i(\boldsymbol{w}^T \boldsymbol{x}_i + b)\right) + \sum_{i=1}^{m} \beta_i(-\xi_i) \right)$$

$$\Rightarrow \max_{\beta_i = C - \lambda_i; 0 \leq \lambda_i \leq C} \left(\min_{\boldsymbol{w},b} \frac{1}{2} \boldsymbol{w}^T \boldsymbol{w} + \sum_{i=1}^{m} \lambda_i \left(1 - y_i(\boldsymbol{w}^T \boldsymbol{x}_i + b)\right) \right)$$

$C - \lambda_i - \beta_i = 0$

对偶问题简化

$$\max_{\beta_i=C-\lambda_i;0\le\lambda_i\le C}\left(\min_{\boldsymbol{w},b}\boxed{\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{i=1}^{m}\lambda_i(1-y_i(\boldsymbol{w}^T\boldsymbol{x}_i+b))}\right)\quad L(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\lambda},\boldsymbol{\beta})$$

- $(\boldsymbol{w},b,\boldsymbol{\xi})$要满足$L(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\lambda},\boldsymbol{\beta})$最小，则对$(\boldsymbol{w},b)$求导应为0，即：

➢ $\frac{\partial L(\boldsymbol{w},b,\boldsymbol{\lambda})}{\partial b}=0=-\sum_{i=1}^{m}\lambda_i y_i$，我们将$\sum_{i=1}^{m}\lambda_i y_i=0$ 加到对偶问题的条件中不会影响问题的最优解，故

$$\max_{\substack{\beta_i=C-\lambda_i;0\le\lambda_i\le C;\\ \sum\lambda_i y_i=0}}\left(\min_{\boldsymbol{w},b}\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{i=1}^{m}\lambda_i(1-y_i(\boldsymbol{w}^T\boldsymbol{x}_i))-\cancel{\sum_{i=1}^{m}\lambda_i y_i\cdot b}\right)\Rightarrow\max_{\substack{\beta_i=C-\lambda_i;0\le\lambda_i\le C;\\ \sum\lambda_i y_i=0}}\left(\min_{\boldsymbol{w},b}\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{i=1}^{m}\lambda_i(1-y_i(\boldsymbol{w}^T\boldsymbol{x}_i))\right)$$

➢ $\frac{\partial L(\boldsymbol{w},b,\boldsymbol{\lambda})}{\partial \boldsymbol{w}}=0=\boldsymbol{w}-\sum_{i=1}^{m}\lambda_i y_i\boldsymbol{x}_i$，同样将$\boldsymbol{w}=\sum_{i=1}^{m}\lambda_i y_i\boldsymbol{x}_i$ 加到对偶问题的条件中，

$$\max_{\substack{\beta_i=C-\lambda_i;0\le\lambda_i\le C;\\ \sum\lambda_i y_i=0;\boldsymbol{w}=\sum\lambda_i y_i\boldsymbol{x}_i}}\left(\min_{\boldsymbol{w},b}\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{i=1}^{m}\lambda_i(1-y_i(\boldsymbol{w}^T\boldsymbol{x}_i))\right)\Rightarrow\max_{\substack{\beta_i=C-\lambda_i;0\le\lambda_i\le C;\\ \sum\lambda_i y_i=0;\boldsymbol{w}=\sum\lambda_i y_i\boldsymbol{x}_i}}\left(\min_{\boldsymbol{w},b}\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}+\sum_{i=1}^{m}\lambda_i-\boldsymbol{w}^T\boldsymbol{w}\right)$$

展开

$$\Rightarrow\max_{\substack{\beta_i=C-\lambda_i;0\le\lambda_i\le C;\\ \sum\lambda_i y_i=0;\boldsymbol{w}=\sum\lambda_i y_i\boldsymbol{x}_i}}\left(\min_{\boldsymbol{w},b}-\frac{1}{2}\left\|\sum_{i=1}^{m}\lambda_i y_i\boldsymbol{x}_i\right\|^2+\sum_{i=1}^{m}\lambda_i\right)\Rightarrow\boxed{\max_{\substack{\beta_i=C-\lambda_i;0\le\lambda_i\le C;\\ \sum\lambda_i y_i=0;\boldsymbol{w}=\sum\lambda_i y_i\boldsymbol{x}_i}}-\frac{1}{2}\left\|\sum_{i=1}^{m}\lambda_i y_i\boldsymbol{x}_i\right\|^2+\sum_{i=1}^{m}\lambda_i}$$

与$\boldsymbol{w},b$无关，可以去掉内部的最小化

最终需求解的对偶问题

软间隔SVM KKT条件

总结一下，$(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta})$是原问题与对偶问题最优解的充要条件是$(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\beta})$满足下面的Karush-Kuhn-Tucker（KKT）条件：

- 满足原问题的约束条件： $y_i\left(\boldsymbol{w}^T\boldsymbol{x_i} + b\right) \geq 1 - \xi_i,\ \ \xi_i \geq 0,\quad i = 1,2,\dots,m$
- 满足对偶问题约束条件：$\lambda_i \geq 0,\ \ \beta_i \geq 0,\quad i = 1,2,\dots,m$
- 满足对偶问题优化条件：
  - $\frac{\partial \boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial \xi_i} = 0 = C - \lambda_i - \beta_i$，即$C - \lambda_i = \beta_i,\ \ 0 \leq \lambda_i \leq C,\ \ i = 1,2,\dots,m$
  - $\frac{\partial \boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial b} = 0 = -\sum_{i=1}^{m}\lambda_i y_i$，即$\sum_{i=1}^{m}\lambda_i y_i = 0$
  - $\frac{\partial \boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial \boldsymbol{w}} = 0 = \boldsymbol{w} - \sum_{i=1}^{m}\lambda_i y_i \boldsymbol{x}_i$，即$\boldsymbol{w} = \sum_{i=1}^{m}\lambda_i y_i \boldsymbol{x}_i$
- 满足原问题的优化条件：
  - 将原问题转换成 minmax问题时，$\lambda_i\left(1 - \xi_i - y_i\left(\boldsymbol{w}^T\boldsymbol{x_i} + b\right)\right) = 0,\ \beta_i\xi_i = 0,\ \ i = 1,2,\dots,m$

求解对偶问题

$$\max_{\substack{\beta_i=C-\lambda_i;0\le\lambda_i\le C;\\ \sum\lambda_i y_i=0;\boldsymbol{w}=\sum\lambda_i y_i \boldsymbol{x}_i}} -\frac{1}{2}\left\|\sum_{i=1}^{m}\lambda_i y_i \boldsymbol{x}_i\right\|^2+\sum_{i=1}^{m}\lambda_i$$

改写上述问题：最大化转换为最小化（取负号），条件下移后可得：

$$\min_{\boldsymbol{\lambda}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j-\sum_{i=1}^{m}\lambda_i$$

$$\mathrm{s.t.} \quad \sum_{i=1}^{m}\lambda_i y_i=0$$

$$0\le\lambda_i\le C, \qquad i=1,2,\dots,m$$

代入凸二次规划形式，可得：

$$\boldsymbol{u}=\boldsymbol{\lambda};\ \boldsymbol{Q}=\left[q_{ij}=y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j\right]_{\boldsymbol{m}\times\boldsymbol{m}};\ \boldsymbol{p}=-\boldsymbol{1}_{\boldsymbol{m}};$$

$\boldsymbol{e}_i\in\mathbb{R}^m$是第i行为1的单位向量

$$\boxed{\boldsymbol{a}_{\ge}=\boldsymbol{y}=[y_i]_m;c_{\ge}=0;\boldsymbol{a}_{\le}=-\boldsymbol{y};c_{\le}=0;}\ \boxed{\boldsymbol{a}_i=-\boldsymbol{e}_i;c_i=-C;\ \boldsymbol{a}_{i-}=\boldsymbol{e}_i;c_{i-}=0;}$$

对应约束条件$\sum_{i=1}^{m}\lambda_i y_i=0$

对应约束条件$0\le\lambda_i\le C$

求解对偶问题

（KKT）条件：

- $y_i(\boldsymbol{w}^T\boldsymbol{x_i}+b)\geq 1-\xi_i$，$\xi_i\geq 0$
- $\lambda_i\geq 0$，$\beta_i\geq 0$
- $\boxed{C-\lambda_i=\beta_i}$，$0\leq\lambda_i\leq C$
- $\sum_{i=1}^m\lambda_i y_i=0$；$\boxed{\boldsymbol{w}=\sum_{i=1}^m\lambda_i y_i\boldsymbol{x}_i}$
- $\boxed{\lambda_i\left(1-\xi_i-y_i(\boldsymbol{w}^T\boldsymbol{x_i}+b)\right)=0}$，$\boxed{\beta_i\xi_i=0}$

- 在求得最优解$\boldsymbol{\lambda}^*=[\lambda_i^*]$后
- 可以利用KKT条件求得原问题的最优解$\boldsymbol{w}^*$和b^*
  - $\boldsymbol{w}^*=\sum_{i=1}^m\lambda_i^* y_i\boldsymbol{x_i}$
  - 如果某个j 使得$0<\lambda_{\boldsymbol{j}}^*<C$，则$\beta_j=C-\lambda_j>0$。由 $\beta_j\xi_j=0$可知 $\xi_j=0$
  - 故$1-y_j(\boldsymbol{w}^T\boldsymbol{x}_j+b)=0$，可得：

$$b^*=y_j-\boldsymbol{w}^{*\boldsymbol{T}}\boldsymbol{x}_j$$

- 软间隔最大化分类决策函数的对偶求解结果式为

$$f(x)=\mathrm{sign}(\boldsymbol{w}^{*\boldsymbol{T}}\boldsymbol{x}+b^*)$$

软间隔线性SVM对偶问题的学习算法

输入：数据集$D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\dots,(\boldsymbol{x}_m,y_m)\},\boldsymbol{x}_i\in\mathbb{R}^n,y_i\ \{-1,1\}$

输出：**软间隔**最大化分离超平面和分类决策函数

- 第一步，选择惩罚参数C，构造带约束的凸二次规划问题，并求解得到最优解$\boldsymbol{\lambda}^*=[\lambda_i^*](i=1,2,\dots,m)$

$$\min_{\boldsymbol{\lambda}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j-\sum_{i=1}^{m}\lambda_i$$

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

$$0\le\lambda_i\le C,\qquad i=1,2,\dots,m$$

其中，暗含以下条件：

$$\boldsymbol{w}^*=\sum_{i=1}^{m}\lambda_i^* y_i\boldsymbol{x}_i$$

$$\beta_i=C-\lambda_i,\qquad i=1,2,\dots,m$$

- 第二步，任意选择$\boldsymbol{\lambda}^*$的一个正分量$0<\lambda_j^*<C$，计算对偶问题对应的最优解$\boldsymbol{w}^*$和b^*

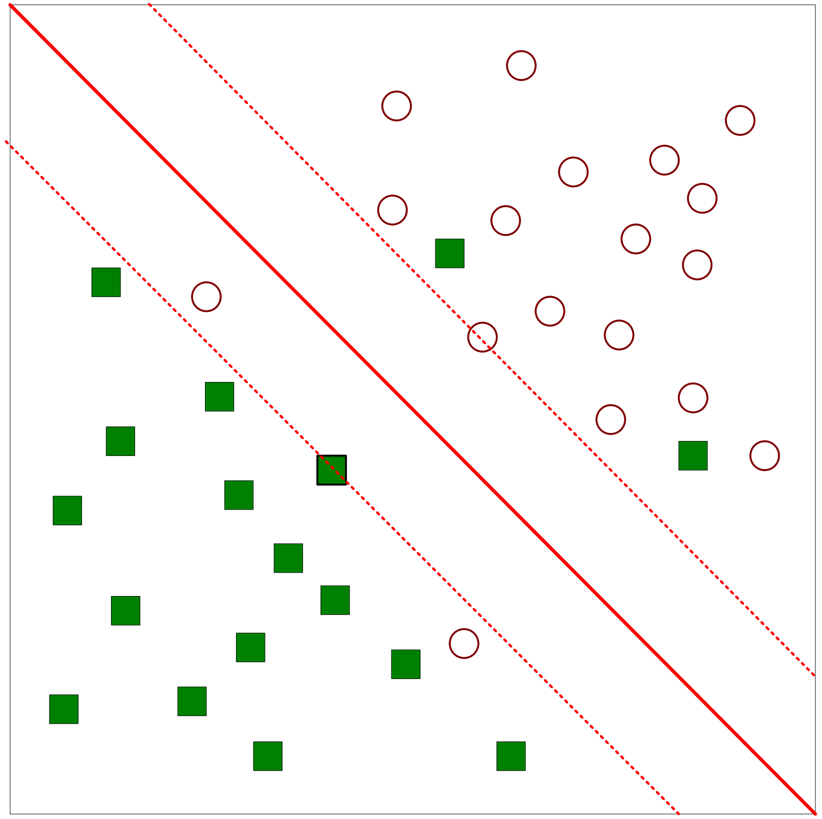
$$\boldsymbol{w}^*=\sum_{i=1}^{m}\lambda_i^* y_i\boldsymbol{x}_i$$

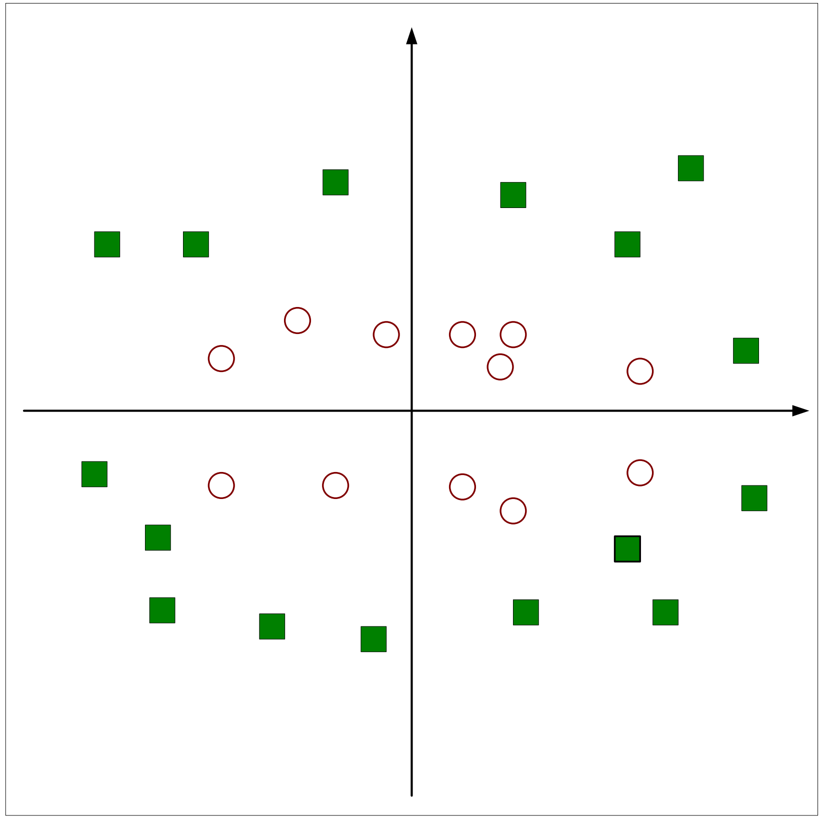
$$b^*=y_j-\boldsymbol{w}^{*T}\boldsymbol{x}_j$$

- 第三步，返回软间隔最大化分离超平面$\boldsymbol{w}^{*T}\boldsymbol{x}+b^*=0$和分类决策函数$f(x)=\mathrm{sign}(\boldsymbol{w}^{*T}\boldsymbol{x}+b^*)$

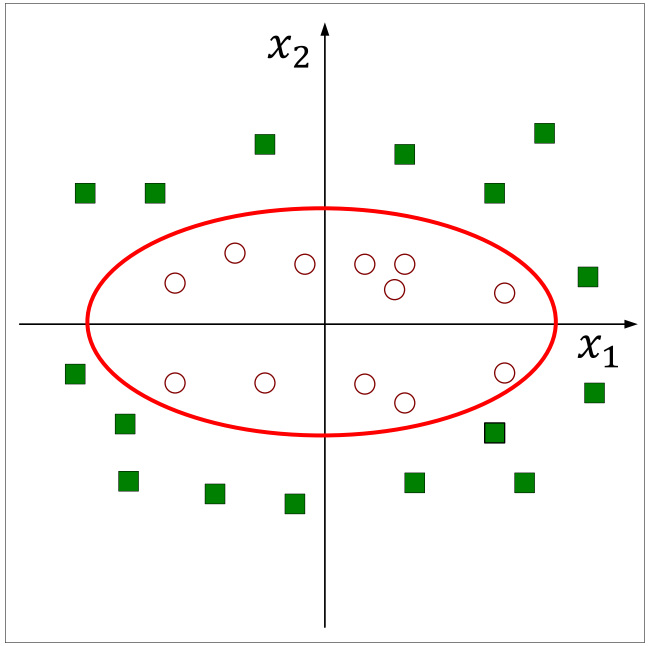
5.4 Non-linear SVM
非线性支持向量机

线性不可分数据集

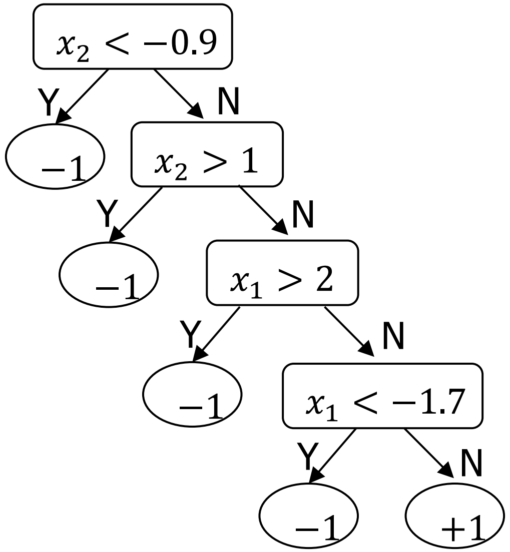


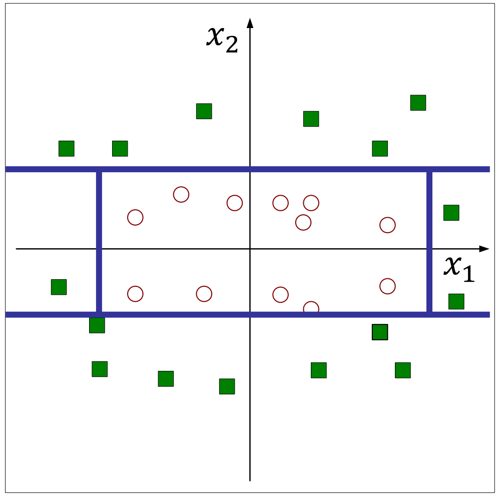


非线性数据分类的两个思路



非线性分类器

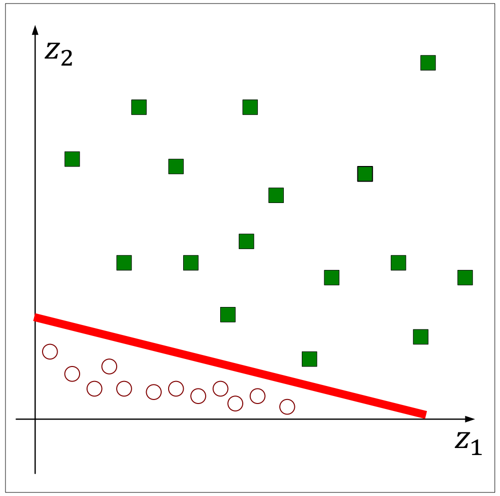




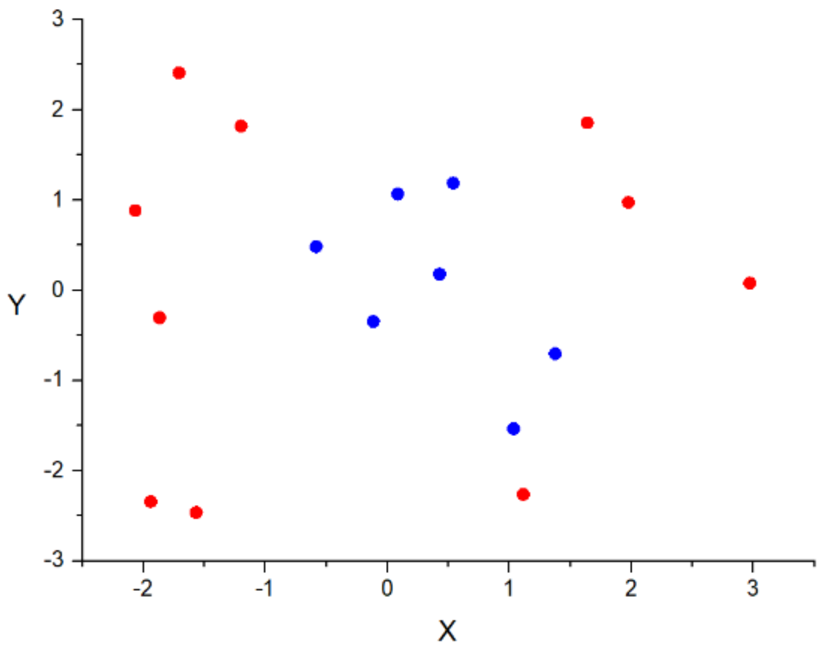
非线性变换

$$\begin{cases} z_1 = x_1^2 \\ z_2 = x_2^2 \end{cases}$$

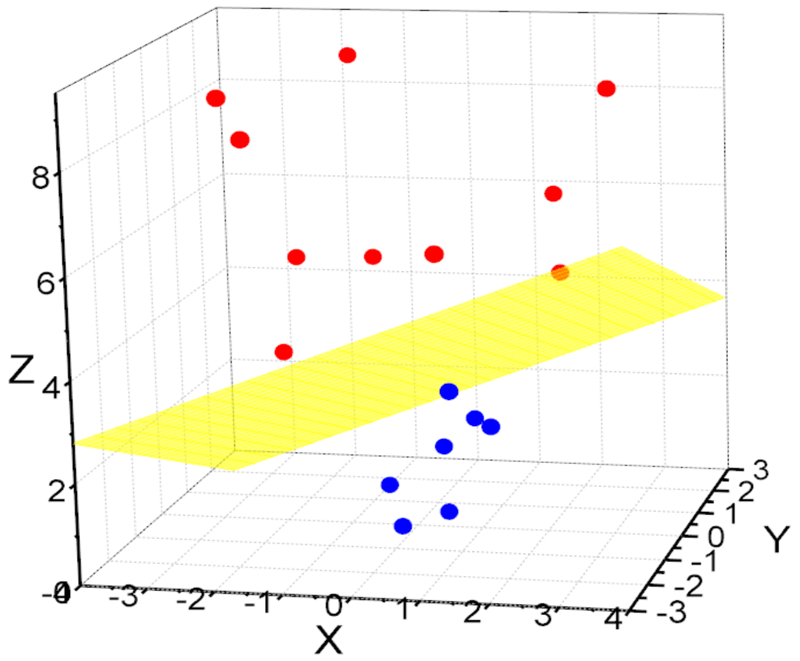
线性分类器求解



非线性SVM的思想



$$\begin{cases} z_1 = x_1 \\ z_2 = x_2 \\ z_3 = x_1^2 + x_2^2 \end{cases}$$



非线性支持向量机的问题定义

$$\min_{\boldsymbol{\lambda}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \phi(\boldsymbol{x}_i)^{\boldsymbol{T}}\phi(\boldsymbol{x}_j) - \sum_{i=1}^{m}\lambda_i$$

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

$$0 \le \lambda_i \le C, \qquad i = 1,2,\ldots,m$$

- **问题1**：需要显式计算任意两个实例非线性变换后的内积$\phi(\boldsymbol{x}_i)^{\boldsymbol{T}}\phi(\boldsymbol{x}_j)$
- **问题2**：需要显式定义映射函数表达式和特征空间，当特征空间**无限维**时，无法定义

快速计算内积的思路

- 以二阶多项式变换为例，$\boldsymbol{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$，其映射函数为：

n维→ $n^2 + 1$维

$$\phi_2(x) = (\underbrace{1}_{\text{零阶项}}, \underbrace{x_1, x_2, \dots, x_n}_{\text{一阶项}}, \underbrace{x_1^2, x_1x_2, \dots, x_1x_n, x_2x_1, x_2^2, \dots, x_2x_n, \dots, x_nx_1, x_nx_2, \dots, x_n^2}_{\text{二阶项}})$$

- 计算$\boldsymbol{x} = (x_1, x_2, \dots, x_n), \boldsymbol{x}' = (x_1', x_2', \dots, x_n') \in \mathbb{R}^n$二阶多项式变换后的内积$\phi_2(\boldsymbol{x})^T\phi_2(\boldsymbol{x}')$：

$$\begin{aligned}\phi_2(\boldsymbol{x})^T\phi_2(\boldsymbol{x}') &= 1 + \sum_{i=1}^{n} x_i x_i' + \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j x_i' x_j' \\ &= 1 + \sum_{i=1}^{n} x_i x_i' + \sum_{i=1}^{n} x_i x_i' \sum_{j=1}^{n} x_j x_j' \\ &= 1 + \boldsymbol{x}^T\boldsymbol{x}' + (\boldsymbol{x}^T\boldsymbol{x}')(\boldsymbol{x}^T\boldsymbol{x}')\end{aligned}$$

- 核函数思想：

$$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{\phi}(\boldsymbol{x})^{\boldsymbol{T}}\boldsymbol{\phi}(\boldsymbol{x}')$$

比如对于二阶多项式变换：

$$K(\boldsymbol{x}, \boldsymbol{x}') = 1 + \boldsymbol{x}^T\boldsymbol{x}' + (\boldsymbol{x}^T\boldsymbol{x}')(\boldsymbol{x}^T\boldsymbol{x}')$$

核函数的定义与判定

定义：设$\mathcal{X}$是输入空间（欧氏空间$\mathbb{R}^n$或者离散集合），$\mathcal{H}$为特征空间（希尔伯特空间），若存在一个从$\mathcal{X}$到$\mathcal{H}$的**映射**$\phi(\boldsymbol{x}): \mathcal{X} \to \mathcal{H}$，使得对于所有元素$\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{X}$，函数$K(\boldsymbol{x}_i, \boldsymbol{x}_j)$满足条件

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i) \cdot \phi(\boldsymbol{x}_j)$$

其中，$\phi(\boldsymbol{x}_i) \cdot \phi(\boldsymbol{x}_j)$是二者内积，则称$K(\boldsymbol{x}_i, \boldsymbol{x}_j)$为**核函数**，$\phi(\boldsymbol{x})$为核函数的基函数。

- **Mercer定理**：设$\mathcal{X}$是输入空间，$K(\cdot,\cdot)$是定义在$\mathcal{X} \times \mathcal{X}$上的函数，则$K(\cdot,\cdot)$是核函数的**充要条件**是：
  - $K(\cdot,\cdot)$是**对称**函数
  - 对于任意$\boldsymbol{x}_i \in \mathcal{X}, i = 1,2,\ldots,\mathrm{m}$，$K(\cdot,\cdot)$对应的Gram矩阵 $\boldsymbol{K}$ 是**半正定**矩阵。

$$\boldsymbol{K} = \left[K(\boldsymbol{x}_i, \boldsymbol{x}_j)\right]_{m \times m}$$

无限维核函数

- 假设$\boldsymbol{x} = (x)$ ，即其特征维度为1，可将其映射到无限维空间，即$\phi(\boldsymbol{x})$为**无限维：**

$$
\begin{aligned}
K(\boldsymbol{x}, \boldsymbol{x}') &= \exp(-(x - x')^2) \\
&= \exp(-(x)^2)\exp(-(x')^2)\exp(2xx')
\end{aligned}
$$

泰勒展开

$$
\begin{aligned}
&= \exp(-(x)^2)\exp(-(x')^2)\left(\sum_{i=0}^{\infty}\frac{(2xx')^i}{i!}\right) \\
&= \sum_{i=0}^{\infty}\left(\exp(-(x)^2)\exp(-(x')^2)\sqrt{\frac{2^i}{i!}}\sqrt{\frac{2^i}{i!}}(x)^i(x')^i\right) \\
&= \phi(\boldsymbol{x})^T\phi(\boldsymbol{x}')
\end{aligned}
$$

- 上述特征变换后的特征空间就是无限维的，$\phi(\boldsymbol{x}) = \exp(-x^2)\cdot\left(1, \sqrt{\frac{2}{1!}}x, \sqrt{\frac{2^2}{2!}}x^2, \ldots\right)$
- 上述核函数就是**高斯核函数**，其有更一般的形式（$\gamma > 0$）：

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\gamma\|\boldsymbol{x} - \boldsymbol{x}'\|^2)$$

核函数

- 常用核函数

  - 线性核函数：$K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \boldsymbol{x}_i^T \boldsymbol{x}_j$

  - 多项式核函数：$K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = (\boldsymbol{x}_i^T \boldsymbol{x}_j + 1)^p$

  - 高斯核函数：$K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \exp(-\frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|}{2\sigma^2})$

非线性支持向量机的求解

$$\min_{\boldsymbol{\lambda}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \phi(\boldsymbol{x}_i)^{\boldsymbol{T}}\phi(\boldsymbol{x}_j) - \sum_{i=1}^{m}\lambda_i$$

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

$$0 \le \lambda_i \le C, \qquad i = 1,2,\dots,m$$

线性SVM求解对偶问题回顾

（KKT）条件：

- $y_i(\boldsymbol{w}^T\boldsymbol{x_i}+b)\geq 1-\xi_i,\ \ \xi_i\geq 0$
- $\lambda_i\geq 0,\ \ \beta_i\geq 0$
- $\boxed{C-\lambda_i=\beta_i},\ \ 0\leq\lambda_i\leq C$
- $\sum_{i=1}^m\lambda_i y_i=0;\ \boxed{\boldsymbol{w}=\sum_{i=1}^m\lambda_i y_i\boldsymbol{x}_i}$
- $\boxed{\lambda_i\left(1-\xi_i-y_i(\boldsymbol{w}^T\boldsymbol{x_i}+b)\right)=0},\ \boxed{\beta_i\xi_i=0}$

- 在求得最优解$\boldsymbol{\lambda}^*=[\lambda_i^*]$后
- 可以利用KKT条件求得问题的最优解$\boldsymbol{w}^*$和b^*
  - $\boldsymbol{w}^*=\sum_{i=1}^m\lambda_i^* y_i\boldsymbol{x_i}$
  - 如果某个j 使得$0<\lambda_j^*<C$，则$\beta_j=C-\lambda_j>0$。由 $\beta_j\xi_j=0$可知 $\xi_j=0$
  - 故$1-y_j(\boldsymbol{w}^T\boldsymbol{x}_j+b)=0$，可得：

$$b^*=y_j-\boldsymbol{w}^{*T}\boldsymbol{x}_j$$

- 软间隔最大化分类决策函数的对偶求解结果式为

$$f(x)=\operatorname{sign}(\boldsymbol{w}^{*T}\boldsymbol{x}+b^*)$$

$$\boxed{\boldsymbol{w}^*=\sum_{i=1}^m\lambda_i^* y_i\emptyset(\boldsymbol{x_i})}$$

$$\boxed{b^*=y_j-\boldsymbol{w}^{*T}\emptyset(\boldsymbol{x}_j)}$$

$$b^* = y_s - \boldsymbol{w}^{*\boldsymbol{T}}\phi(\boldsymbol{x}_s)$$

$$= y_s - \left(\sum_{i=1}^{m} \lambda_i^* y_i \phi(\boldsymbol{x}_i)\right)^{\boldsymbol{T}} \phi(\boldsymbol{x}_s)$$

$$= y_s - \sum_{i=1}^{m} \lambda_i^* y_i K(\boldsymbol{x}_i, \boldsymbol{x}_s)$$

KKT :

- $y_i\left(\boldsymbol{w}^{\boldsymbol{T}}\emptyset(\boldsymbol{x}_i) + b\right) \geq 1 - \xi_i, \ \ \xi_i \geq 0$
- $\lambda_i \geq 0, \ \ \beta_i \geq 0$
- $C - \lambda_i = \beta_i, \ 0 \leq \lambda_i \leq C$
- $\sum_{i=1}^{m} \lambda_i y_i = 0; \ \ \boldsymbol{w} = \sum_{i=1}^{m} \lambda_i y_i \emptyset(\boldsymbol{x}_i)$
- $\lambda_i\left(1 - \xi_i - y_i\left(\boldsymbol{w}^{\boldsymbol{T}}\emptyset(\boldsymbol{x}_i) + b\right)\right) = 0, \ \beta_i\xi_i = 0$

$$\boldsymbol{w}^{*\boldsymbol{T}}\phi(\boldsymbol{x}) = \left(\sum_{i=1}^{m} \lambda_i^* y_i \phi(\boldsymbol{x}_i)\right)^{\boldsymbol{T}}\phi(\boldsymbol{x}) = \sum_{i=1}^{m} \lambda_i^* y_i K(\boldsymbol{x}_i, \boldsymbol{x})$$

非线性支持向量机训练阶段

输入：数据集$D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\dots,(\boldsymbol{x}_m,y_m)\},\boldsymbol{x}_i\in\mathbb{R}^n,y_i\ \{-1,1\}$

输出：分类决策函数

- 第一步，选取适当的**核函数**$K(\cdot,\cdot)$和惩罚参数C，构造带约束的凸二次规划问题并求解

$$\min_{\boldsymbol{\lambda}}\quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j K(\boldsymbol{x}_i,\boldsymbol{x}_j)-\sum_{i=1}^{m}\lambda_i$$

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

$$0\le\lambda_i\le C,\qquad i=1,2,\dots,m$$

求解得到最优解$\boldsymbol{\lambda}^*=[\lambda_i^*](i=1,2,\dots,m)$

- 第二步，任意选择$\boldsymbol{\lambda}^*$的一个正分量$0<\lambda_s^*<C$，计算b^*

$$b^*=y_s-\boldsymbol{w}^{*\boldsymbol{T}}\phi(\boldsymbol{x}_s)=y_s-\left(\sum_{i=1}^{m}\lambda_i^* y_i\phi(\boldsymbol{x}_i)\right)^{\boldsymbol{T}}\phi(\boldsymbol{x}_s)=y_s-\sum_{i=1}^{m}\lambda_i^* y_i K(\boldsymbol{x}_i,\boldsymbol{x}_s)$$

- $\boldsymbol{w}^*$暂时不计算，因为$\boldsymbol{w}^*=\sum_{i=1}^{m}\lambda_i^* y_i\phi(\boldsymbol{x}_i)$，其中$\phi(\boldsymbol{x}_i)$可能维度很高，无法计算。

非线性支持向量机测试阶段

输入：数据集$D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\dots,(\boldsymbol{x}_m,y_m)\},\boldsymbol{x}_i\in\mathbb{R}^n,y_i\ \{-1,1\}$

输出：分类决策函数

- 利用训练阶段的b^*

$$b^* = y_s-\boldsymbol{w}^{*\boldsymbol{T}}\phi(\boldsymbol{x}_s)=y_s-\left(\sum_{i=1}^{m}\lambda_i^* y_i\phi(\boldsymbol{x}_i)\right)^{\boldsymbol{T}}\phi(\boldsymbol{x}_s)=y_s-\sum_{i=1}^{m}\lambda_i^* y_i K(\boldsymbol{x}_i,\boldsymbol{x}_s)$$

以及 $\boldsymbol{w}^{*\boldsymbol{T}}\phi(\boldsymbol{x})=\left(\sum_{i=1}^{m}\lambda_i^* y_i\phi(\boldsymbol{x}_i)\right)^{\boldsymbol{T}}\phi(\boldsymbol{x})=\sum_{i=1}^{m}\lambda_i^* y_i K(\boldsymbol{x}_i,\boldsymbol{x})$

- 返回分类决策函数$f(\boldsymbol{x})=\mathrm{sign}\left(\sum_{i=1}^{m}\lambda_i^* y_i K(\boldsymbol{x}_i,\boldsymbol{x})+b^*\right)$

5.5 SVM 编程实现

利用 sklearn库实现 SVM

sklearn是开源的Python机器学习库，提供了大量用于数据挖掘、分析的工具，**SVM**可以直接调用该库的接口**sklearn.svm**实现。

class sklearn.svm.SVC()

以下列出较为重要的参数，其余参数可参考官网介绍

参数	描述
C	**惩罚项，**必须为正，默认为1。
kernel	**核函数类型**，默认为rbf，线性应选择 'linear'
degree	**多项式核函数的次数**，当kernel设置为 'poly' 时才有效。默认为3
gamma	**核函数参数，**针对不同核函数，有不同设置

https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

利用 sklearn库实现线性SVM

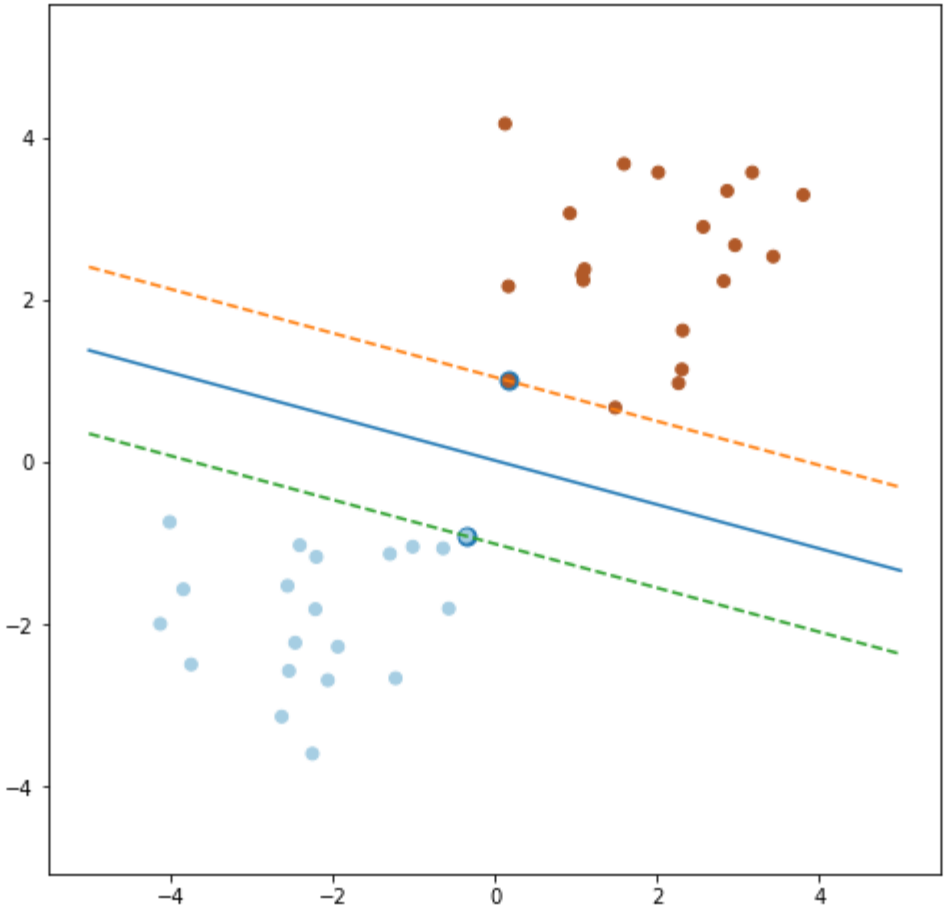
```
# 引入必要的库
from sklearn.svm import SVC
import numpy as np
import matplotlib.pyplot as plt

# 使用正态分布产生随机数据
x = np.r_[np.random.randn(20, 2) - [2, 2], np.random.randn(20, 2) + [2, 2]]
y = [0] * 20 + [1] * 20

# 先调用SVC类，设置好参数，再使用fit函数求解，使用predict函数预测
clf = SVC(kernel='linear')
clf.fit(x, y)
clf.predict([[2,2],[-2,-2]])

# 获取参数w,b,以及支持向量
w, b, sv = clf.coef_[0], clf.intercept_[0], clf.support_vectors_

# 绘图部分
x1 = np.linspace(-5, 5)
x2 = -(w[0] * x1 + b) / w[1]
x2up = -(w[0] * x1 + b - 1) / w[1]
x2down = -(w[0] * x1 + b + 1) / w[1]
plt.figure(figsize=(8, 8))
plt.plot(x1, x2)
plt.plot(x1, x2up, linestyle="--")
plt.plot(x1, x2down, linestyle="--")
plt.scatter(sv[:, 0], sv[:, 1], s=80)
plt.scatter(x[:, 0], x[:, 1], c=y, cmap=plt.cm.Paired)
plt.axis('equal')
plt.show()
```



利用 sklearn库实现非线性SVM

```
# 引入必要的库
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import make_moons
from sklearn.svm import SVC

# 使用make_moons函数产生交叉半圆形随机数据
X, y = make_moons(n_samples=100, noise=0.15, random_state=42)

# 先调用SVC类，设置好参数，此处使用三次多项式核函数再使用fit函数求解
clf = SVC(kernel='poly', degree=3, coef0=1, C=5)
clf.fit(X, y)

# 使用predict函数预测，并绘制等高线
x0s = np.linspace(-1.5, 2.5, 100)
x1s = np.linspace(-1, 1.5, 100)
x0, x1 = np.meshgrid(x0s, x1s)
X_pred = np.c_[x0.ravel(), x1.ravel()]
y_pred = clf.predict(X_pred).reshape(x0.shape)
plt.contourf(x0, x1, y_pred, cmap=plt.cm.brg, alpha=0.1)

# 绘制数据点
plt.scatter(X[:, 0], X[:, 1], c=y, cmap=plt.cm.Paired)
plt.axis('equal')
plt.show()
```