第6章 支持向量机*

1. 线性可分支持向量机与硬间隔最大化
2. 线性支持向量机与软间隔最大化
3. 非线性支持向量机与核函数
4. 序列最小最优化算法

*参阅《统计学习方法》第7章

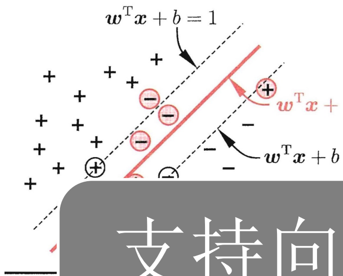
第6章 支持向量机*

1. 线性可分支持向量机与硬间隔最大化
2. 线性支持向量机与软间隔最大化
3. 非线性支持向量机与核函数
4. 序列最小最优化算法

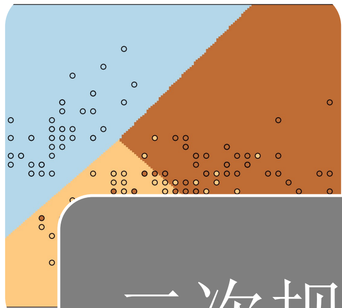
*参阅《统计学习方法》第7章

本章主要内容

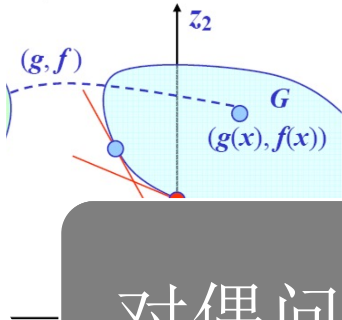
概念



支持向量机问题建模



二次规划求解



对偶问题求解

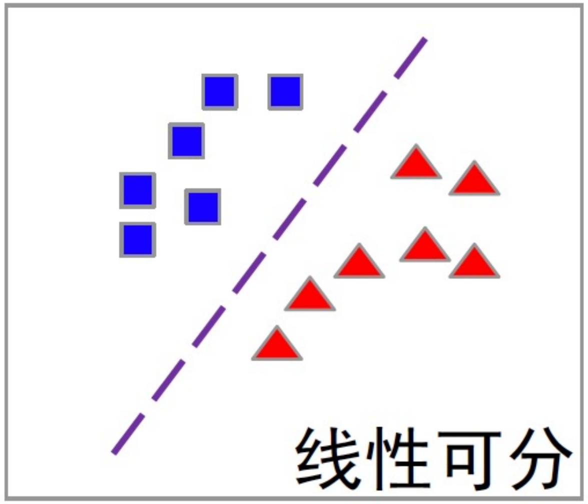
概念——支持向量机

- 支持向量机（support vector machine, SVM）

  SVM是一种二分类模型，其基本模型是定义在特征空间上的间隔最大的分类器

  - 线性可分支持向量机

    当训练数据线性可分时，通过硬间隔最大化（hard margin maximization），学习的一种线性的分类器



概念——支持向量机

- 支持向量机（support vector machine, SVM）

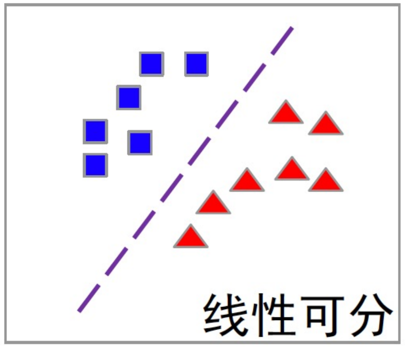
  SVM是一种二分类模型，其基本模型是定义在特征空间上的间隔最大的分类器

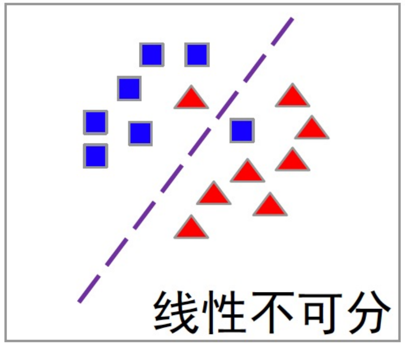
  - 线性可分支持向量机

    当训练数据线性可分时，通过硬间隔最大化（hard margin maximization），学习的一种线性的分类器

  - 线性支持向量机（linear SVM）

    当训练数据近似线性可分时，通过软间隔最大化（soft margin maximization），学习的一种线性的分类器





概念——支持向量机

- 支持向量机（support vector machine, SVM）

  SVM是一种二分类模型，其基本模型是定义在特征空间上的间隔最大的分类器

  - 线性可分支持向量机

    当训练数据线性可分时，通过硬间隔最大化（hard margin maximization），学习的一种线性的分类器

  - 线性支持向量机（linear SVM）

    当训练数据近似线性可分时，通过软间隔最大化（soft margin maximization），学习的一种线性的分类器

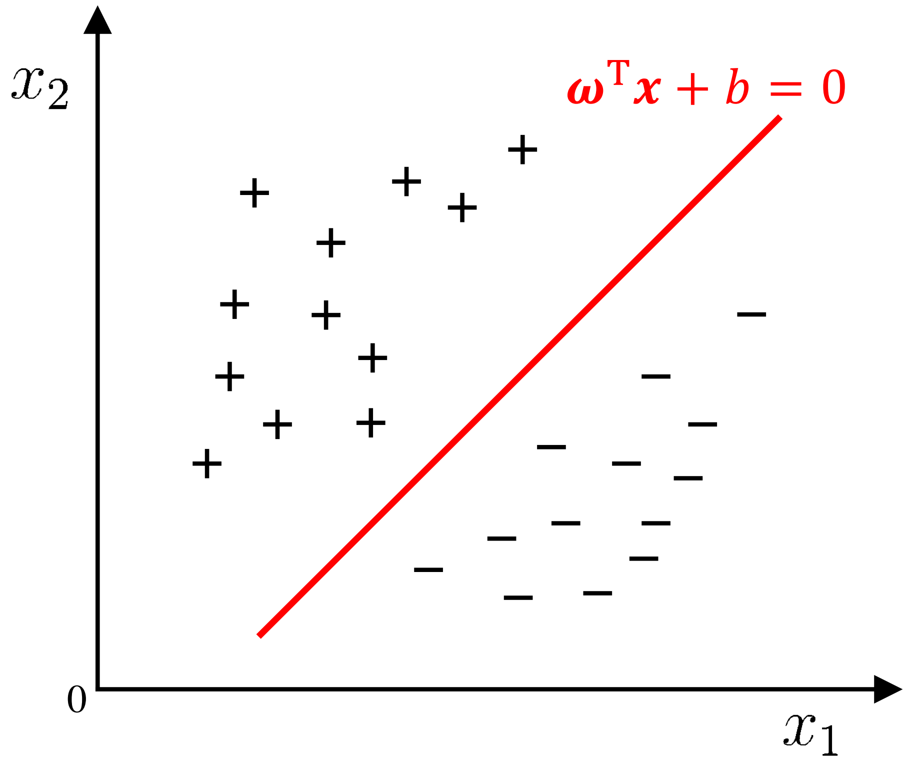
  - 非线性支持向量机（non-linear SVM）

    当训练数据线性不可分时，通过使用核技巧（kernel trick）及软间隔最大化，学习的一种非线性的分类器

概念——线性可分

◻ 支持向量机（support vector machine, SVM）

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



线性可分：

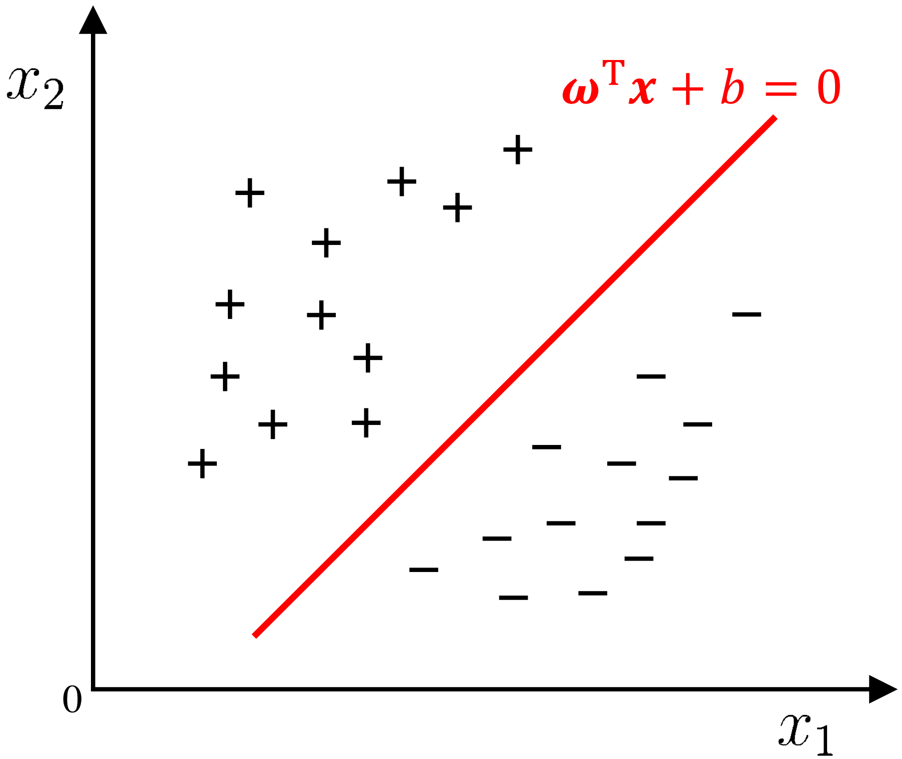
如果存在某个超平面$\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b=0$能够将数据集$D$的正实例点和负实例点完全正确地划分到超平面的两侧。对所有$y_i=+1$的实例$\boldsymbol{x}_i$，有$\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b>0$，对所有$y_i=-1$的实例$\boldsymbol{x}_i$，有$\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b<0$，则称$D$为线性可分数据集（linearly separable data set）。

$$y_i(\boldsymbol{\omega}^T x_i+b)\geq 0$$

概念——超平面

- 支持向量机（support vector machine, SVM）

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



超平面$(\boldsymbol{\omega},b)$：

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

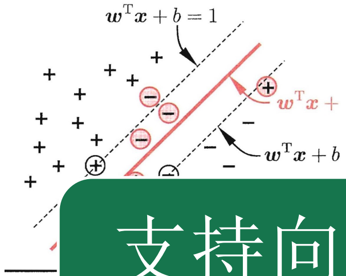
- $\boldsymbol{\omega}=(\omega_1;\omega_2;\cdots;\omega_d)$为法向量，决定了超平面的方向
- b为位移项，决定了超平面与原点之间的距离

任意点$\boldsymbol{x}_i$到超平面$(\boldsymbol{\omega},b)$的距离为：

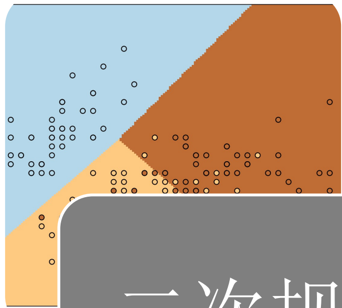
$$r_i=\frac{\left|\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b\right|}{\|\boldsymbol{\omega}\|}$$

本章主要内容

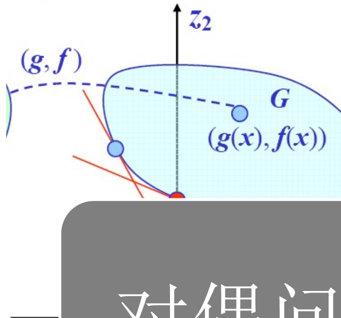
概念



支持向量机问题建模



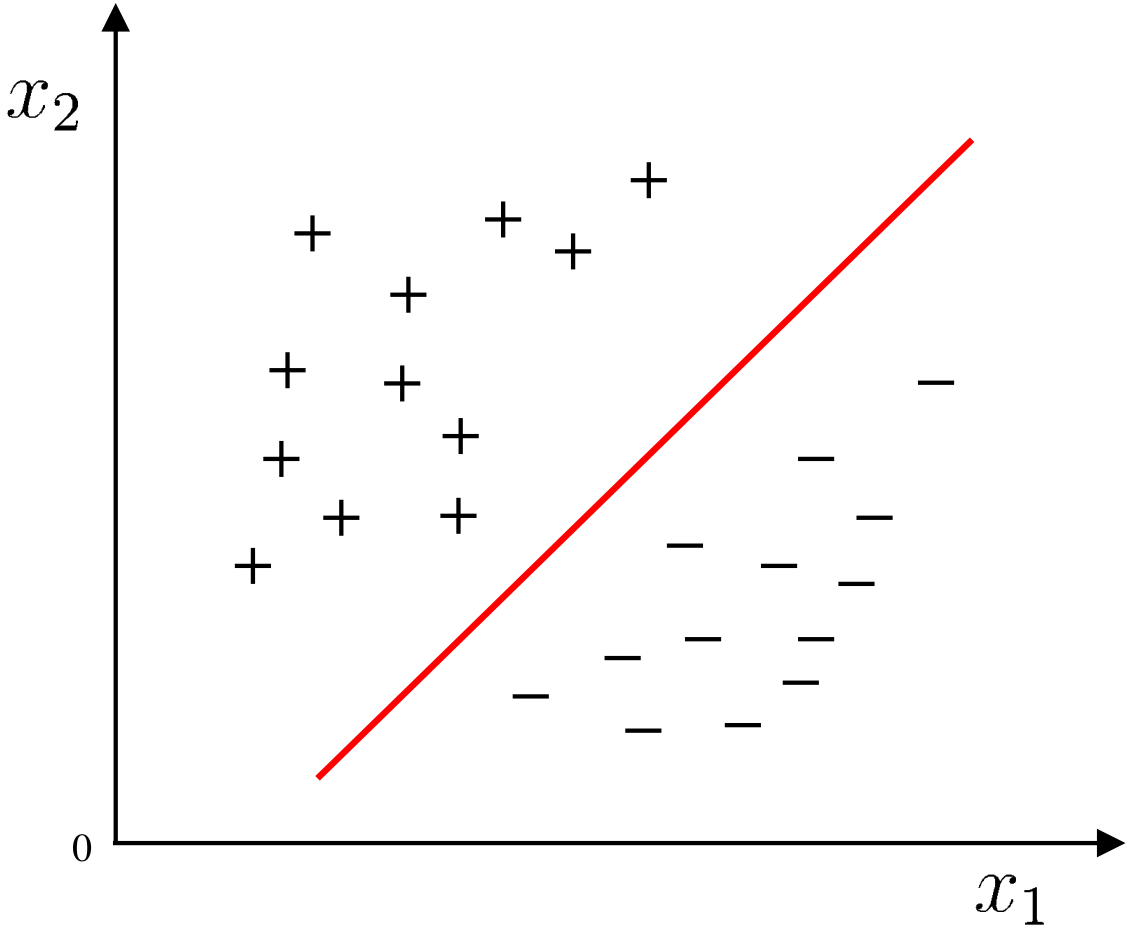
二次规划求解



对偶问题求解

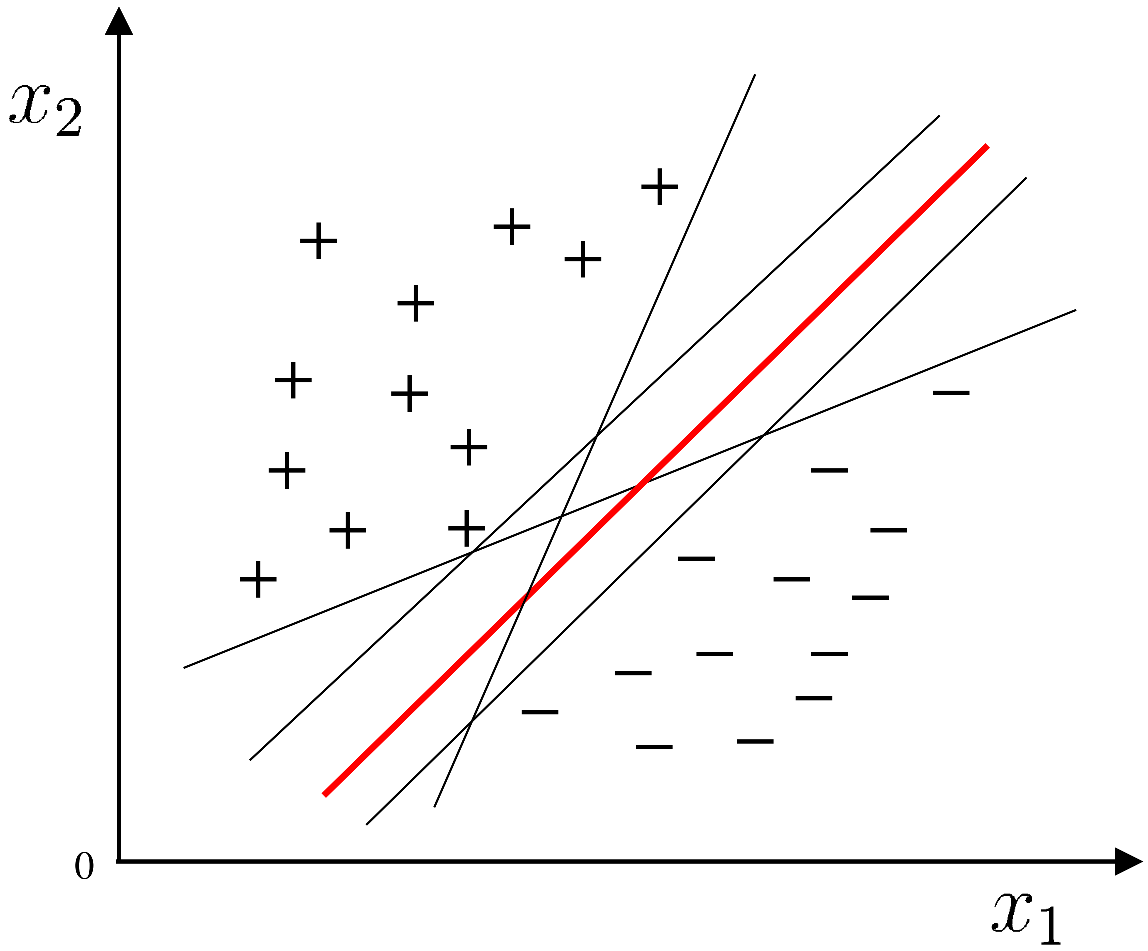
线性可分

在样本空间寻找一个超平面，将不同类型的样本分开

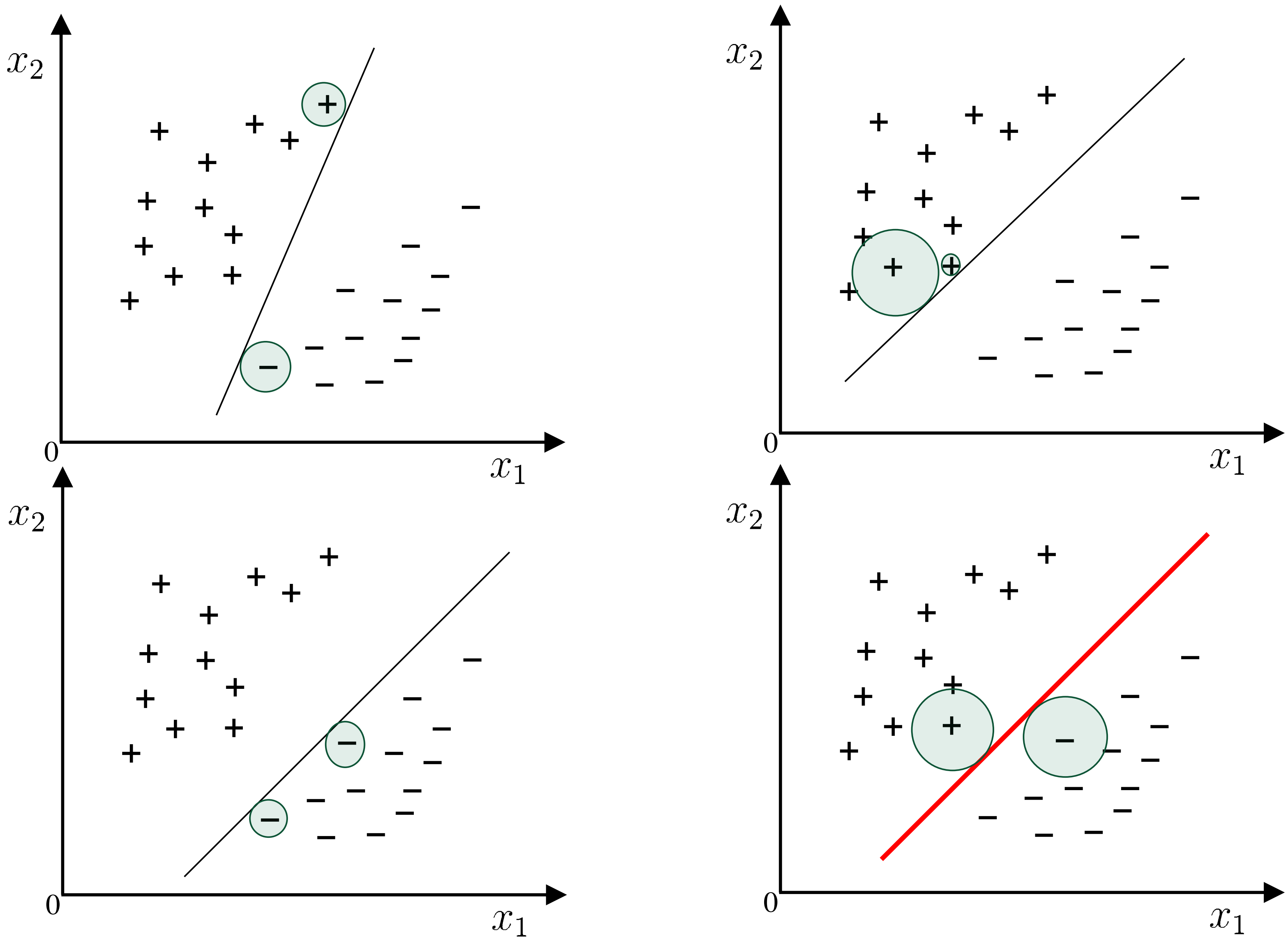


线性可分

将训练样本分开的超平面可能有很多，哪一个更好呢？

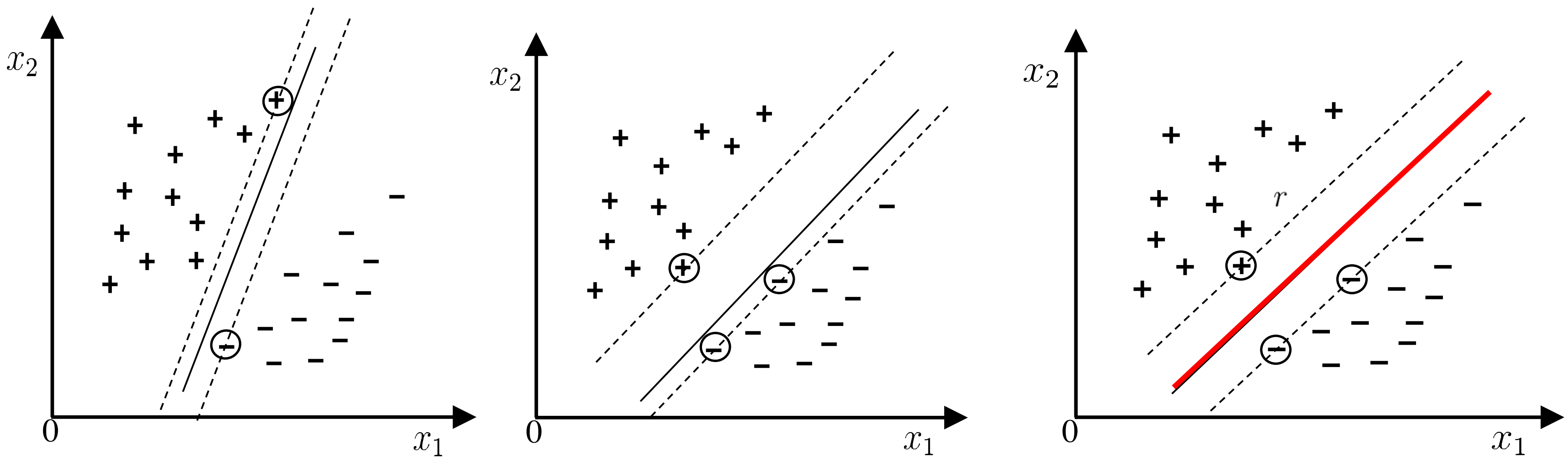


思考——哪条线是最好的呢?



最佳分类超平面

将训练样本分开的超平面可能有很多，哪一个更好呢？

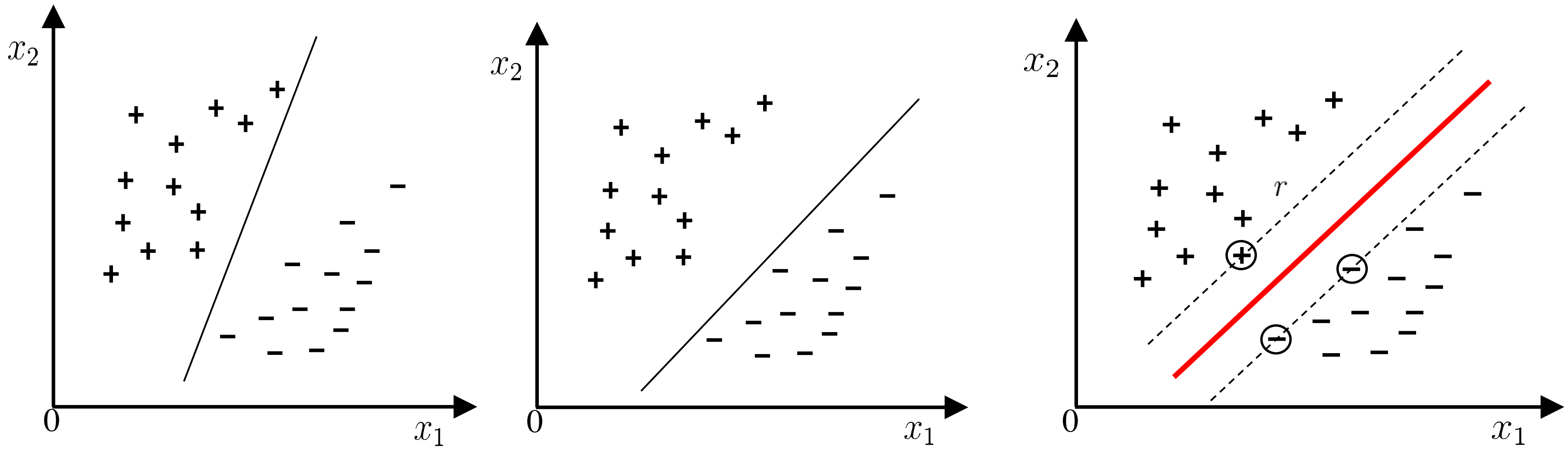


“正中间”的“容忍”性最好：

对接近两个类的分隔界的样本，红色的超平面受影响最小，

分类结果最鲁棒，对未见示例的泛化能力最强

最佳分类超平面



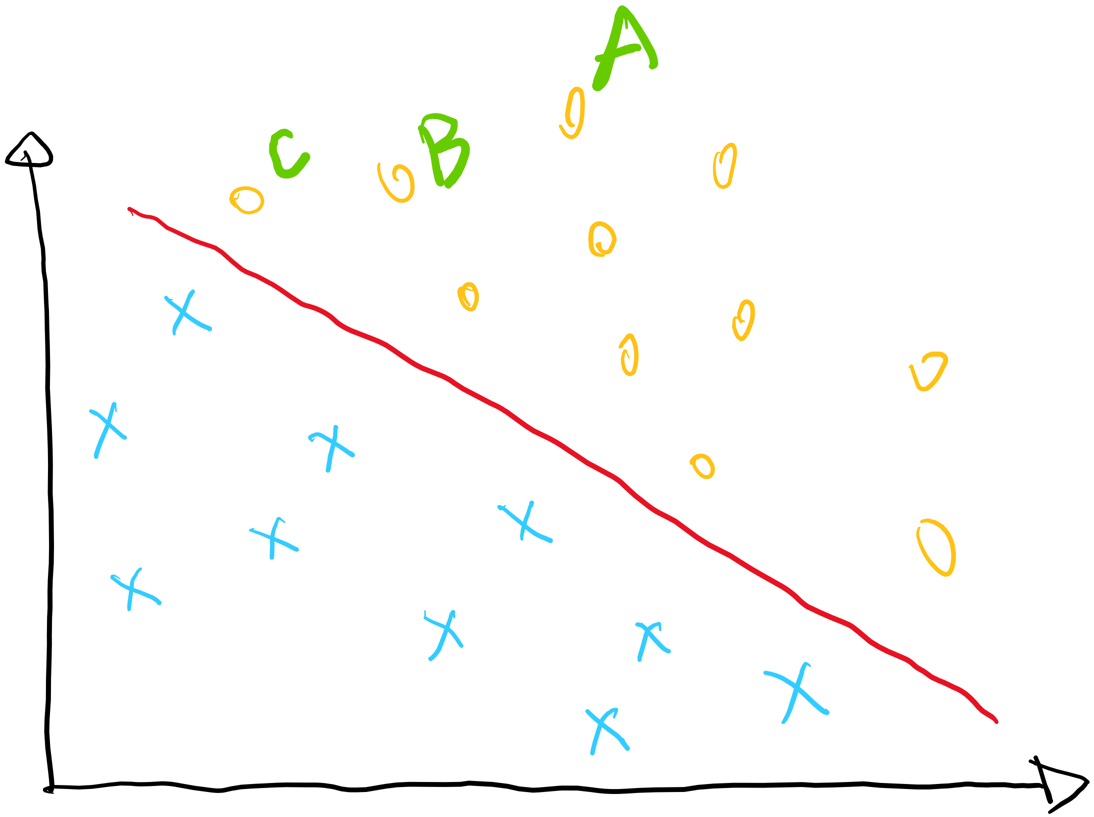
$$\max_{\boldsymbol{\omega},b} \quad \text{margin}(\boldsymbol{\omega})$$ 怎么计算？

$$\text{s.t.} \quad y_i(\boldsymbol{\omega}^T x_i + b) \geq 0, \text{for each } x_i$$

概念——函数间隔

◻ 线性可分支持向量机

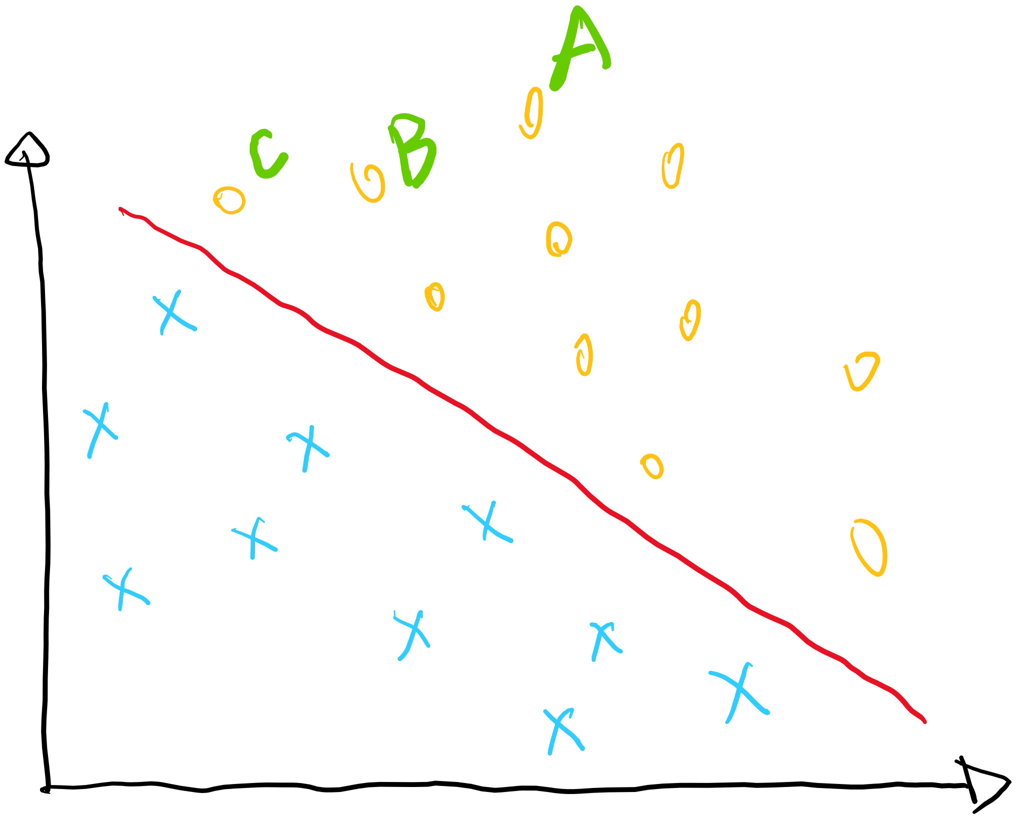
- ➢ $\left|\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b\right|$相对地表示到分离超平面的远近，表示分类的确信度



概念——函数间隔

◘ 线性可分支持向量机

- $\left|\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right|$相对地表示到分离超平面的远近，表示分类的确信度



有A，B，C三点，表示3个实例，均在分离超平面的正类一侧，预测它们的类。

点A距分离超平面较远，若预测为正类，就比较确信预测是正确的；

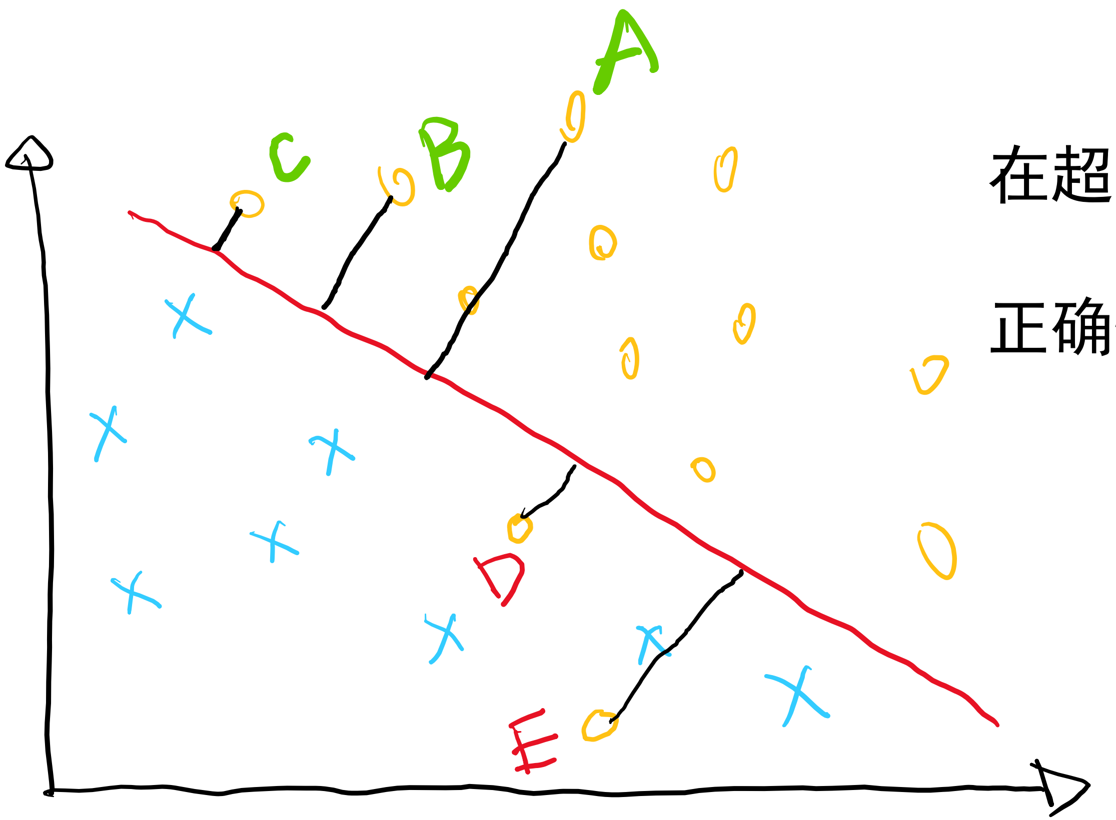
点C距分离超平面较近，若预测该点为正类就不那么确信；

点B介于点A与C之间，预测其为正类的确信度也在A与C之间。

概念——函数间隔

线性可分支持向量机

- $|\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b|$相对地表示到分离超平面的远近，表示分类的确信度
- $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b$的符号与类标记$y_i$的符号是否一致，表示分类的正确性



在超平面$w \cdot x + b = 0$确定的情况下，

正确性 $A > B > C > \underbrace{D > E}_{负数}$

概念——函数间隔

线性可分支持向量机

- $\left|\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right|$相对地表示到分离超平面的远近，表示分类的确信度
- $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b$的符号与类标记$\boldsymbol{y}_i$的符号是否一致，表示分类的正确性

函数间隔（functional margin）

$$\hat{r}_i = y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right)$$

几何间隔

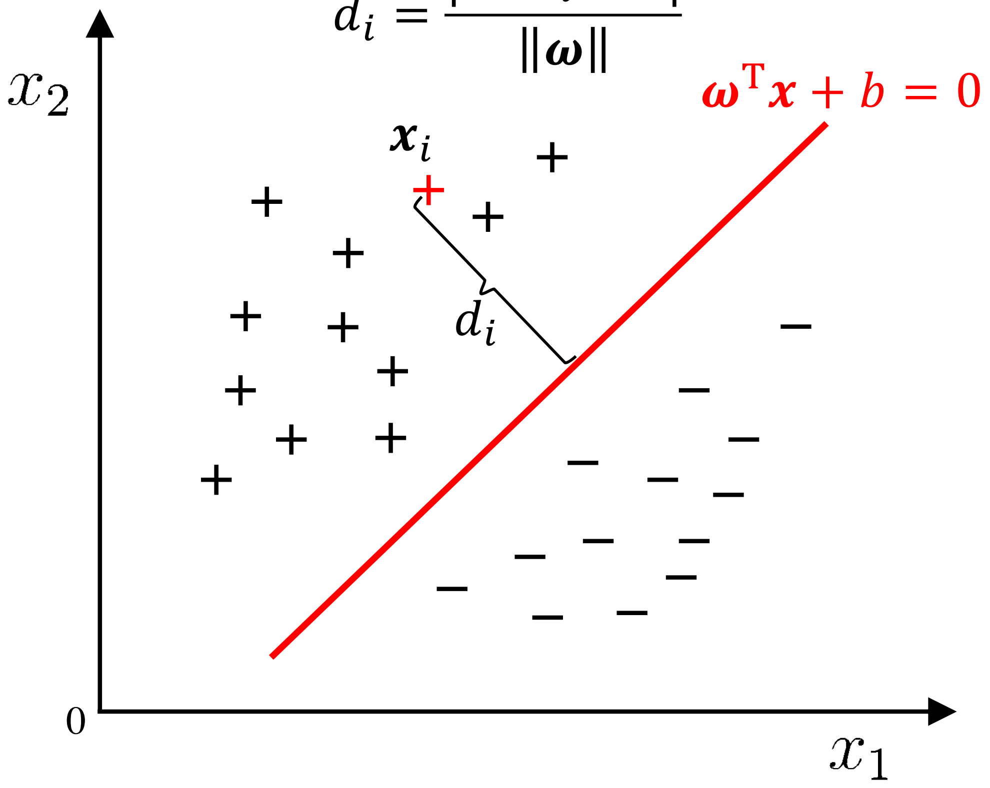
- 支持向量机（support vector machine, SVM）

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

模型：$f(\boldsymbol{x}_i)=\mathrm{sign}\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b\right)$，使得$f(\boldsymbol{x}_i)\simeq y_i$，其中$\boldsymbol{\omega}\in\mathbb{R}^d$

$$\mathrm{sign}(x)=\begin{cases}+1, x\geq 0\\ -1, x<0\end{cases}$$

$$d_i=\frac{|\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b|}{\|\boldsymbol{\omega}\|}$$



- 当样本点$(\boldsymbol{x}_i,y_i)$被超平面$(\boldsymbol{\omega},b)$正确分类时，点$\boldsymbol{x}_i$与超平面$(\boldsymbol{\omega},b)$的距离可以表示为

$$r_i=y_i\frac{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b}{\|\boldsymbol{\omega}\|}$$

几何间隔

◘ 几何间隔（geometric margin）

对于给定的训练数据集D和超平面$(\boldsymbol{\omega}, b)$：

- 定义超平面$(\boldsymbol{\omega}, b)$ 关于样本点$(\boldsymbol{x}_i, y_i)$的几何间隔为

$$r_i = \frac{y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right)}{\|\boldsymbol{\omega}\|} = \frac{\hat{r}_i}{\|\boldsymbol{\omega}\|}$$

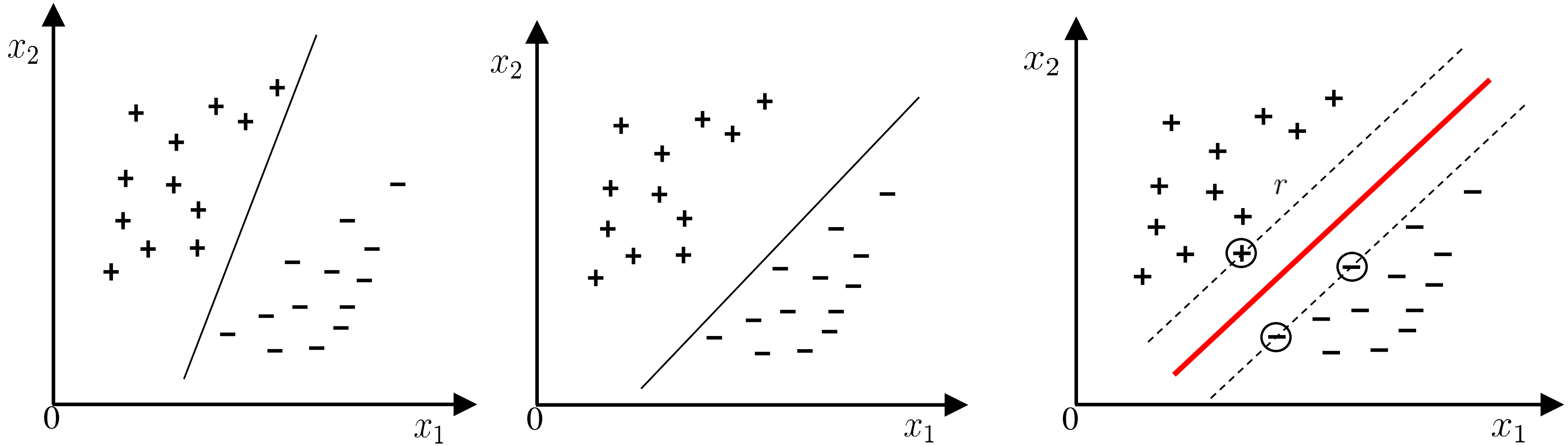
- 定义超平面$(\boldsymbol{\omega}, b)$ 关于训练数据集D的几何间隔为超平面$(\boldsymbol{\omega}, b)$关于D中所有样本点$(\boldsymbol{x}_i, y_i)$的几何间隔的最小值，即

$$r = \min_{i=1,\cdots,m} \frac{y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right)}{\|\boldsymbol{\omega}\|} = \frac{\hat{r}}{\|\boldsymbol{\omega}\|}$$

如果超平面的参数$\boldsymbol{\omega}$和b成比例地改变，超平面和几何间隔都保持不变，而函数间隔会按此比例改变！

最佳分类超平面

策略：找到能够正确划分训练数据的超平面并且使得该超平面关于训练数据集的几何间隔最大



$$\max_{\boldsymbol{\omega},b} \quad r$$

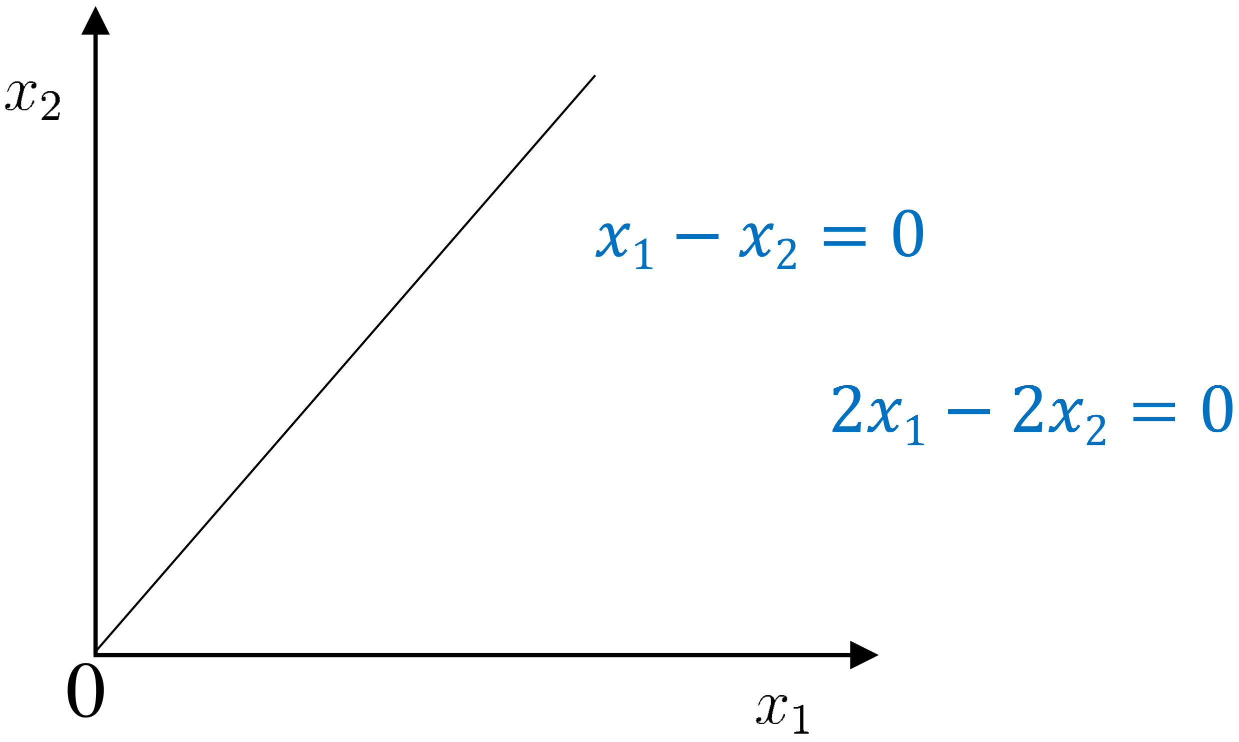
$$\text{s.t.} \quad y_i(\boldsymbol{\omega}^T x_i + b) \geq 0, \text{for each } x_i$$

$$r = \min_{i=1,\cdots,m} \frac{y_i\left(\boldsymbol{\omega}^{\mathrm{T}} \boldsymbol{x}_i + b\right)}{\|\boldsymbol{\omega}\|}$$

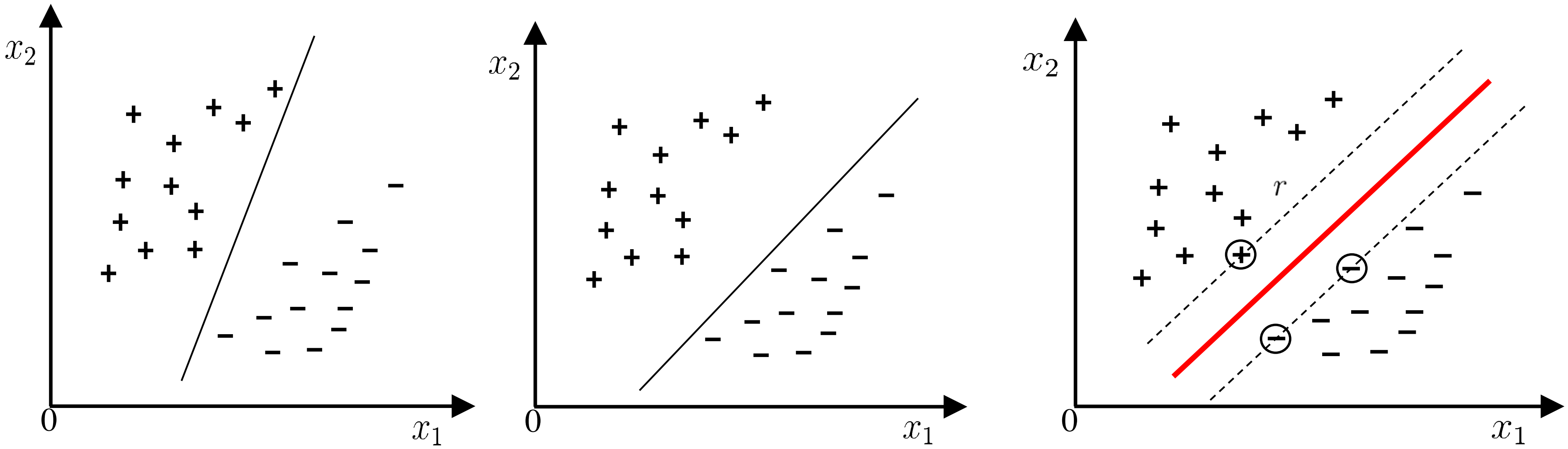
即为到分割线的最小距离 $r = \min_{i=1,\cdots,m} \mathrm{dis}(\boldsymbol{x_i}, \boldsymbol{\omega})$

超平面系数的放缩

- 对于任意一个超平面 （**w**，b）
- 与（a**w**, ab）为同一个平面，a>0



最佳分类超平面



超平面系数可以放缩：

$(\boldsymbol{\omega}, \ b)$ $(\mathbf{5}\boldsymbol{\omega}, \mathbf{5b})$ $(\mathbf{aw}, \mathbf{ab}), \mathbf{a} > \mathbf{0}$

$$\max_{\boldsymbol{\omega}, b} \quad r$$

$$\text{s.t.} \quad y_i(\boldsymbol{\omega}^T x_i + b) \geq 0, \text{for each } x_i$$

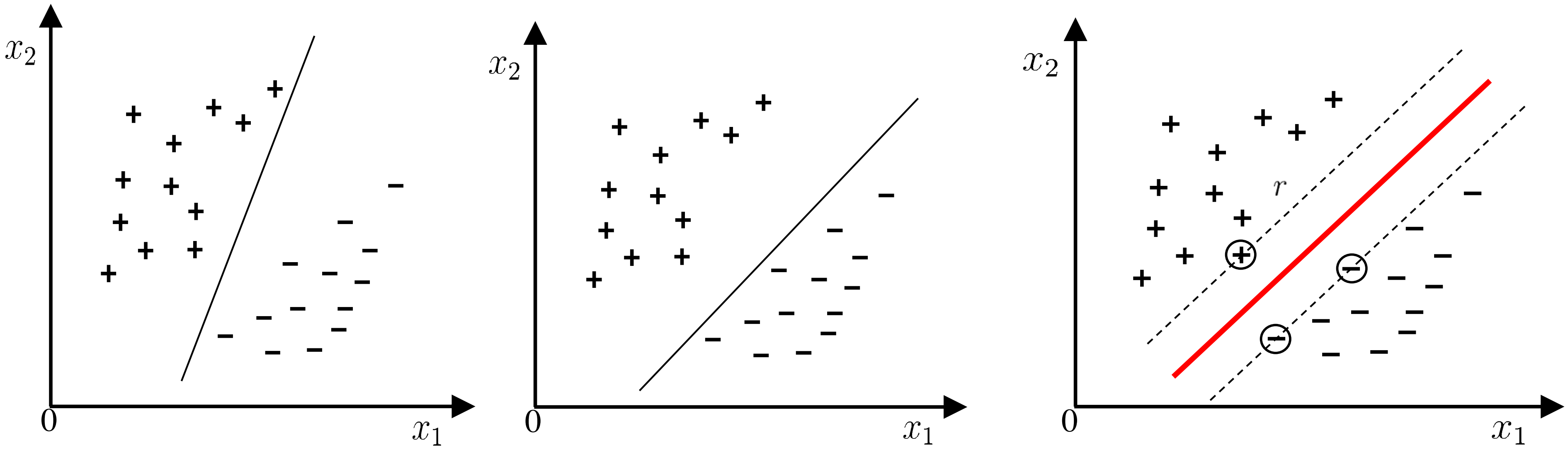
$$r = \min_{i=1,\cdots,m} \frac{y_i\left(\boldsymbol{\omega}^{\mathrm{T}} \boldsymbol{x}_i + b\right)}{\|\boldsymbol{\omega}\|} = \frac{\hat{r}}{\|\boldsymbol{\omega}\|}$$

可以放缩至：

$$|\boldsymbol{\omega}^T x_{\mathrm{i}} + b| = 1$$

此时超平面位置不变，
优化问题不变

最佳分类超平面



$$\max_{\boldsymbol{\omega},b} \quad r$$

$$\text{s.t.} \quad y_i(\boldsymbol{\omega}^T x_i + b) \geq \boxed{1} \quad \text{for each } x_i$$

$$r = \min_{i=1,\cdots,m} \frac{y_i\left(\boldsymbol{\omega}^{\mathrm{T}} \boldsymbol{x}_i + b\right)}{\|\boldsymbol{\omega}\|} = \frac{\boxed{1}}{\|\boldsymbol{\omega}\|}$$

因为：

$$|\boldsymbol{\omega}^T x_{\mathrm{i}} + b| = 1$$

此时：

$$y_i(\boldsymbol{\omega}^T x_i + b) \geq 1$$

$$\hat{r} = 1, \quad r = \frac{1}{\|\boldsymbol{\omega}\|}$$

间隔最大化

- 最大间隔划分超平面（几何间隔r最大的划分超平面）

  - 最大化$\frac{1}{\|\boldsymbol{\omega}\|}$和最小化$\frac{1}{2}\|\boldsymbol{\omega}\|^2$是等价的

$$
\begin{aligned}
&\max_{\boldsymbol{\omega},b} \quad \frac{1}{\|\boldsymbol{\omega}\|} \\
&\text{s.t.} \quad y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right) \geq 1, i = 1, \cdots, m
\end{aligned}
$$

$$\Updownarrow$$

$$
\begin{aligned}
&\min_{\boldsymbol{\omega},b} \quad \frac{1}{2}\|\boldsymbol{\omega}\|^2 \\
&\text{s.t.} \quad y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right) \geq 1, i = 1, \cdots, m
\end{aligned}
$$

间隔与线性可分支持向量机

□ 线性可分支持向量机

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

模型： $f(\boldsymbol{x}_i)=\mathrm{sign}\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b\right)$，使得$f(\boldsymbol{x}_i)\simeq y_i$，其中$\boldsymbol{\omega}\in\mathbb{R}^d$

函数间隔： $\hat{r}_i=y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b\right)$

几何间隔： $r_i=\dfrac{y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b\right)}{\|\boldsymbol{\omega}\|}=\dfrac{\hat{r}_i}{\|\boldsymbol{\omega}\|}$

策略： 找到能够正确划分训练数据的超平面并且使得该超平面关于训练数据集的几何间隔最大

$$\max_{\omega,b}\ \frac{2}{\|\boldsymbol{\omega}\|}$$

$$\text{s.t.(约束于)}\quad y_i(\boldsymbol{\omega}^T x_i+b)\geq 1,\qquad i=1,2,\ldots,m.$$

间隔与线性可分支持向量机

□ 线性可分支持向量机

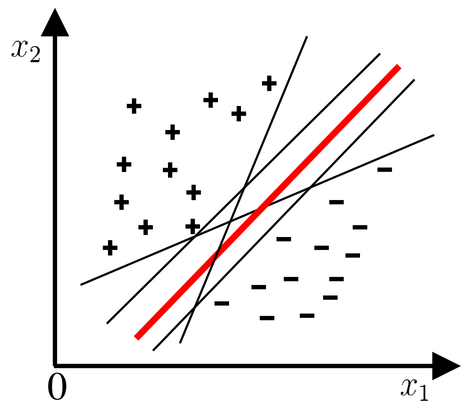
数据： $D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_m,y_m)\}$，$\boldsymbol{x}_i\in\mathbb{R}^d, y_i\in\{-1,1\}$

模型： $f(\boldsymbol{x}_i)=\mathrm{sign}\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b\right)$，使得$f(\boldsymbol{x}_i)\simeq y_i$，其中$\boldsymbol{\omega}\in\mathbb{R}^d$

函数间隔： $\hat{r}_i=y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b\right)$

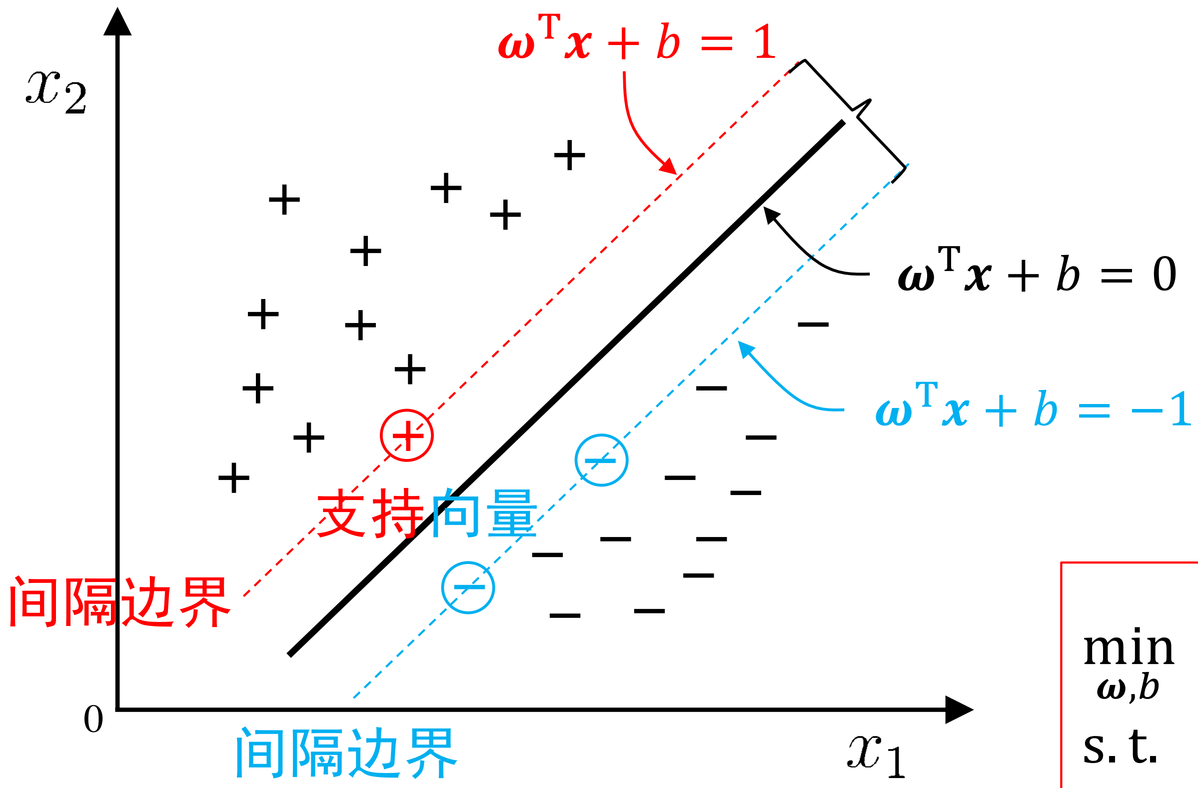
几何间隔： $r_i=\dfrac{y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b\right)}{\|\boldsymbol{\omega}\|}=\dfrac{\hat{r}_i}{\|\boldsymbol{\omega}\|}$

策略： 找到能够正确划分训练数据的超平面并且使得该超平面关于训练数据集的几何间隔最大



对于线性可分的训练数据集而言，
几何间隔最大的分离超平面是唯一的，
此时的间隔最大化又称为硬间隔最大化。

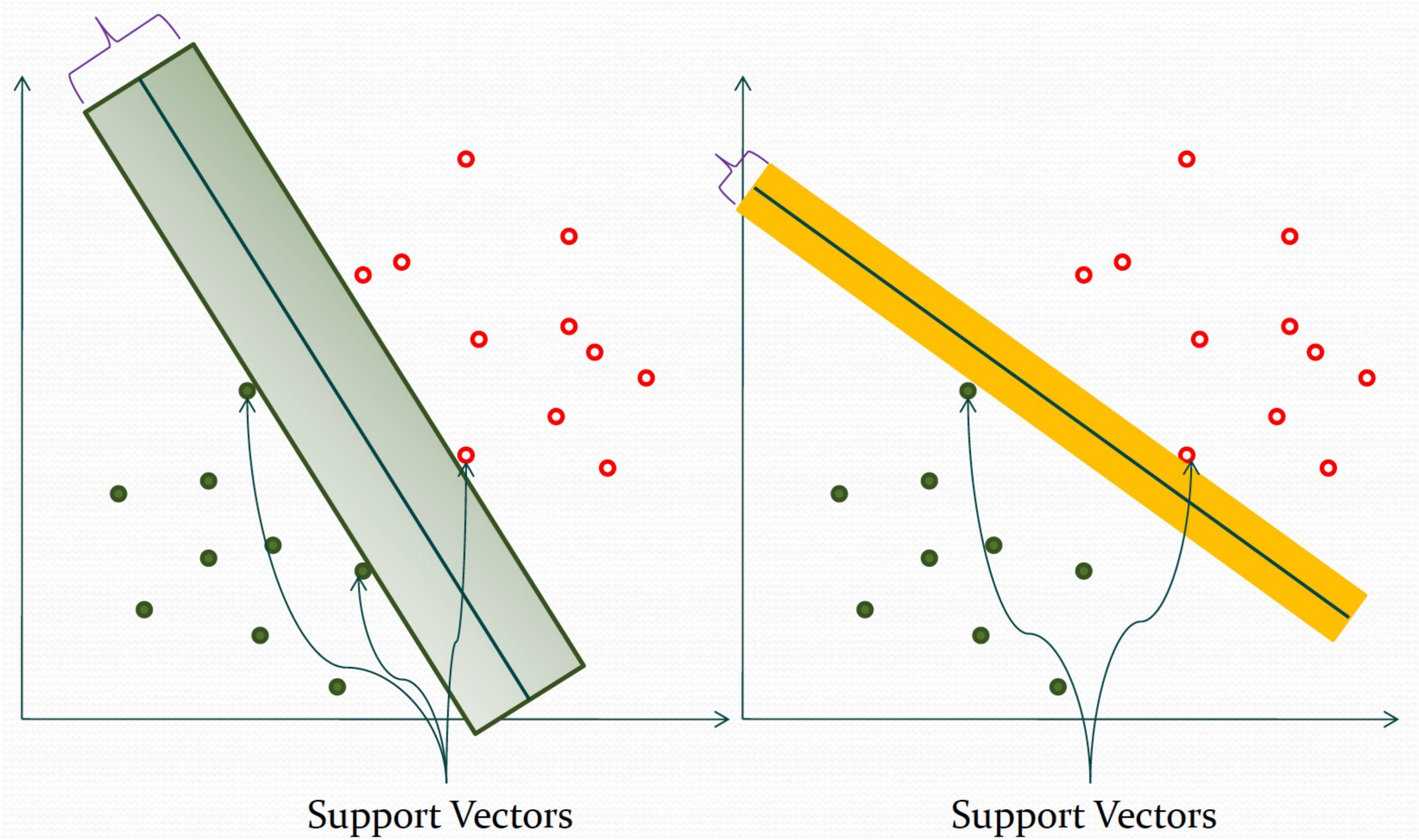
线性可分支持向量机



$$
\begin{aligned}
\min_{\boldsymbol{\omega},b} \quad & \frac{1}{2}\|\boldsymbol{\omega}\|^2 \\
\text{s.t.} \quad & y_i(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b)\geq 1, \\
& i=1,\cdots,m
\end{aligned}
$$

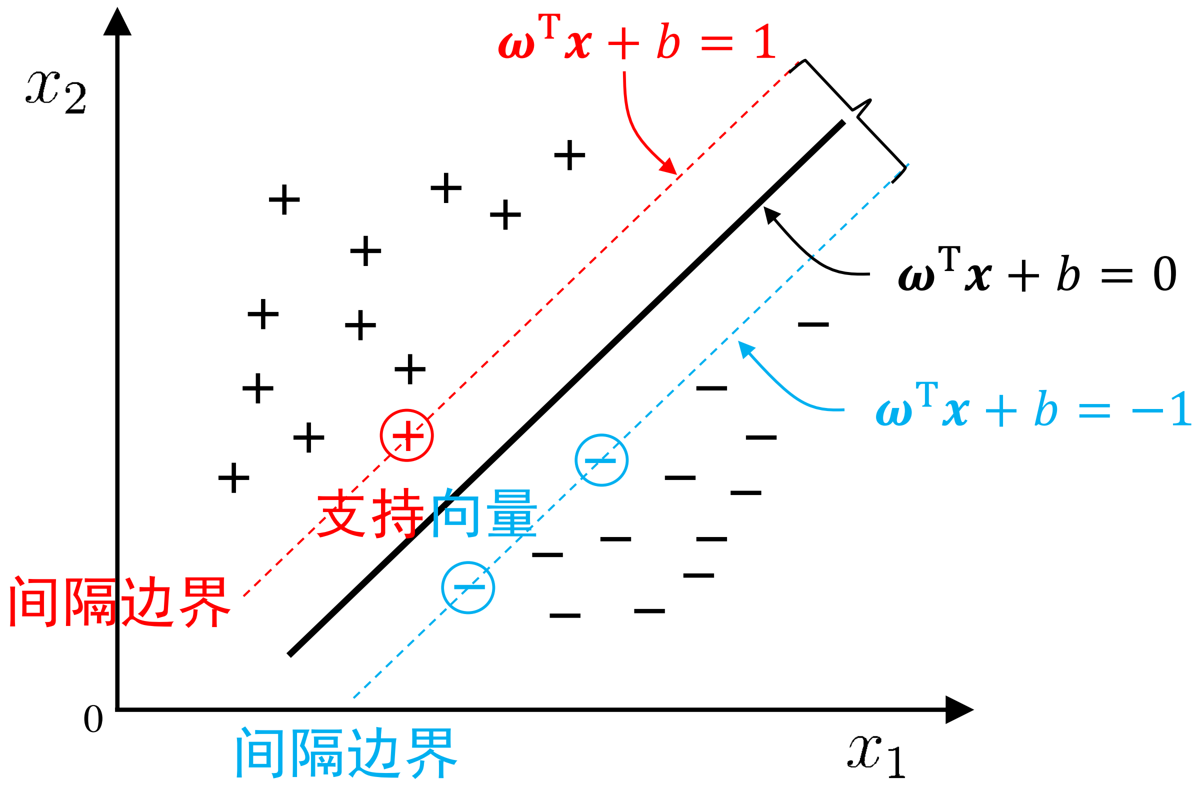
支持向量：在线性可分情况下，训练数据集的样本点中与分离超平面距离最近的样本点（使上述约束条件的等号成立）的示例称为支持向量(support vector)

线性可分支持向量机



支持向量：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==（使上述约束条件的等号成立）的示例称为支持向量(support vector)

线性可分支持向量机



支持向量的作用：在决定分离超平面时只有支持向量起作用，而其他实例点并不起作用。

由于支持向量在确定分离超平面中起决定性作用，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

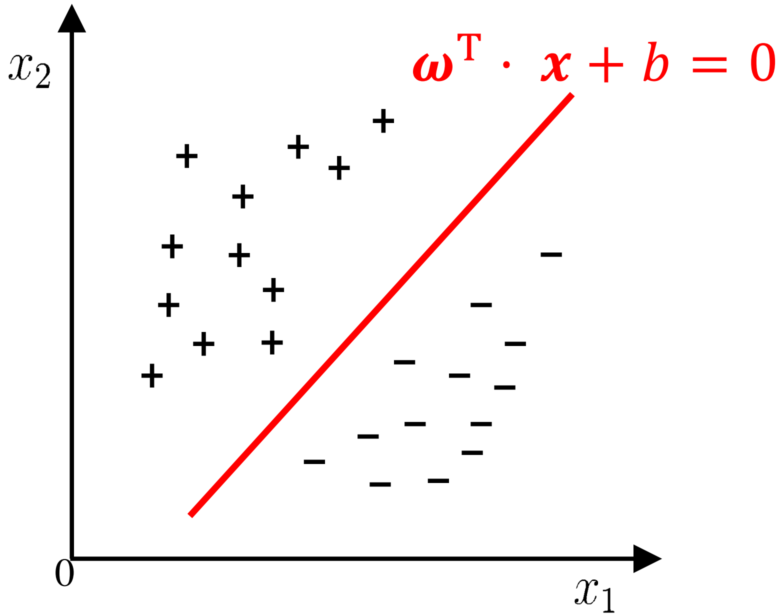
给定线性可分训练数据集，通过间隔最大化学习得到的分离超平面为：

$$\boldsymbol{\omega}^* \cdot \boldsymbol{x} + b^* = 0$$

以及相应的分类决策函数：

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

称为线性可分支持向量机。

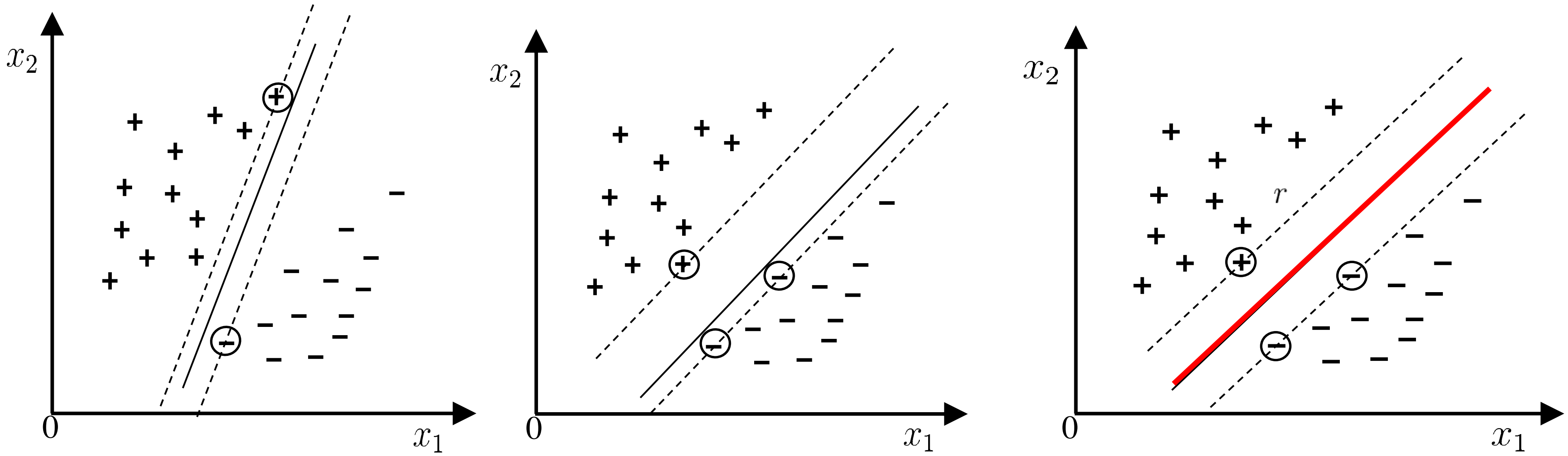


线性可分支持向量机

- 支持向量机（support vector machine, SVM）

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

  模型：$f(\boldsymbol{x}_i)=\mathrm{sign}\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b\right)$，使得$f(\boldsymbol{x}_i)\simeq y_i$，其中$\boldsymbol{\omega}\in\mathbb{R}^d$



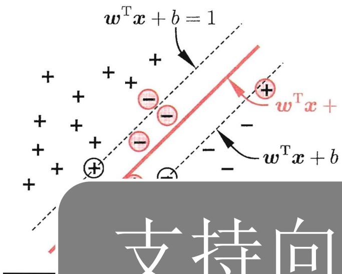
支持向量机需要找到一个分离超平面，

其所产生的分类结果最鲁棒，并对未见示例的泛化能力最强

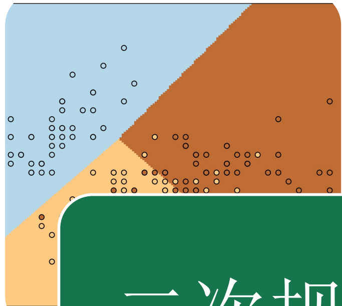
其基本模型是定义在特征空间上的间隔最大的二分类器

本章主要内容

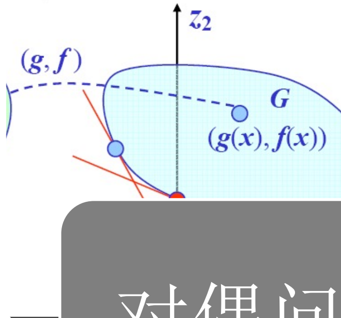
概念



支持向量机问题建模



二次规划求解



对偶问题求解

用二次规划来解

支持向量机问题

Find Optimal （b $\boldsymbol{\omega}$ ）

$$\min \quad \frac{1}{2}\|\boldsymbol{\omega}\|^2$$

s.t.

$$y_i(w^T x_i + b) \geq 1,\ i = 1,\ldots m$$

二次规划标准形式

Find Optimal x

$$\min \quad q(x) = \frac{1}{2}\left(x^T Q\, x \ + \ p^T x \right)$$

$$\text{s.t.} \quad a_i^T x \geq c_i \ , \ i = 1, \ldots m$$

$$x = \begin{bmatrix} b \\ \boldsymbol{\omega} \end{bmatrix} \quad Q = \begin{bmatrix} 0 & 0 \\ 0 & \mathrm{I} \end{bmatrix}$$

$$p = 0$$

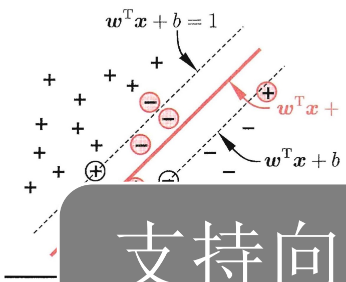
$$c_i = 1$$

$$a_i^T = \begin{pmatrix} y_i & y_i x_i^T \end{pmatrix}$$

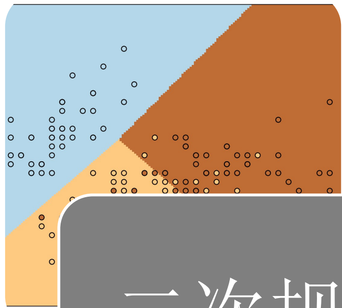
二次规划库函数(matlab)

本章主要内容

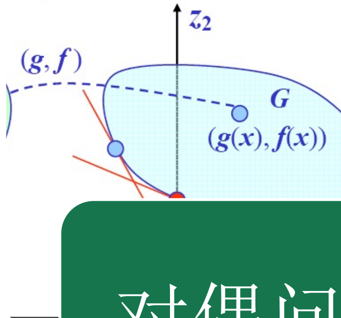
概念



支持向量机问题建模



二次规划求解



对偶问题求解

对偶问题(最优化理论)

https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

拉格朗日对偶性

$$
\begin{aligned}
\min_{\boldsymbol{\omega},b} \quad & \frac{1}{2}\|\boldsymbol{\omega}\|^2 \\
\text{s.t.} \quad & y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right) \geq 1, \\
& i = 1, \cdots, m
\end{aligned}
$$

在约束最优化问题中，常常利用拉格朗日对偶性(Lagrange duality) 将原始问题转换为对偶问题，对偶问题往往要比原始问题的求解要简单，因而通过解对偶问题得到原始问题的解。

拉格朗日乘数法

- 对于下面的优化问题

$$\min f(\mathbf{x})$$
$$h_i(\mathbf{x}) = 0, i = 1, \ldots, p$$

- 构造拉格朗日乘子函数，引入乘子变量 $\lambda_i, i = 1, \ldots, p$

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{p} \lambda_i h_i(\mathbf{x})$$

- 最优解的必要条件是对所有变量的偏导数均为0

$$\nabla_{\mathbf{x}} f + \sum_{i=1}^{p} \lambda_i \nabla_{\mathbf{x}} h_i = \mathbf{0}$$
$$h_i(\mathbf{x}) = 0$$

- 第二组方程就是原来的等式约束

拉格朗日对偶

原问题（Primal Problem）

- 带有等式和不等式约束的优化问题

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$
$$s.t.\, g_i(\boldsymbol{x}) \leq 0, i = 1, \dots k$$
$$h_i(\boldsymbol{x}) = 0, i = 1, \dots p$$

- 拉格朗日乘子函数

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = f(\boldsymbol{x}) + \sum_{i=1}^{k} \alpha_i g_i(\boldsymbol{x}) + \sum_{i=1}^{p} \beta_i h_i(\boldsymbol{x})$$

- 对偶问题（Dual Problem）

$$\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \boxed{\theta(\boldsymbol{\alpha}, \boldsymbol{\beta})} = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \boxed{\min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})}$$
$$s.t.\, \alpha_i \geq 0, i = 1, \dots k$$

- Lower bounds on optimal value $\theta(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) \leq f(\boldsymbol{x}^*)$

Lower bounds on optimal value

原问题

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$s.t.\, g_i(\boldsymbol{x}) \le 0, i = 1, \dots k$$

$$h_i(\boldsymbol{x}) = 0, i = 1, \dots p$$

对偶问题

$$\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \theta(\boldsymbol{\alpha},\boldsymbol{\beta}) = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{\beta})$$

$$s.t.\, \alpha_i \ge 0, i = 1, \dots k$$

$$\theta(\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*) \le f(\boldsymbol{x}^*)$$

- 假设 $\boldsymbol{x}^*$ 为原问题的一个解

$$\sum_{i=1}^{k} \alpha_i g_i(\boldsymbol{x}^*) + \sum_{i=1}^{p} \beta_i h_i(\boldsymbol{x}^*) \le 0$$

- 因此 $\mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\alpha},\boldsymbol{\beta}) = f(\boldsymbol{x}^*) + \sum_{i=1}^{k} \alpha_i g_i(\boldsymbol{x}^*) + \sum_{i=1}^{p} \beta_i h_i(\boldsymbol{x}^*) \le f(\boldsymbol{x}^*)$
- 假设 $\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*$ 是对偶问题的解

$$\theta(\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*) = \min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x},\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*) \le \mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*) \le f(\boldsymbol{x}^*)$$

Duality Gap

弱对偶定理

- 如果原问题和对偶问题都存在最优解，则对偶问题的最优值不大于原问题的最优值

$$G = f(\boldsymbol{x}^*) - \theta(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) \geq \mathbf{0}$$

强对偶定理

- 如果原问题与对偶问题的最优值相等，则称为强对偶

$$G = f(\boldsymbol{x}^*) - \theta(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = \mathbf{0}$$

强对偶

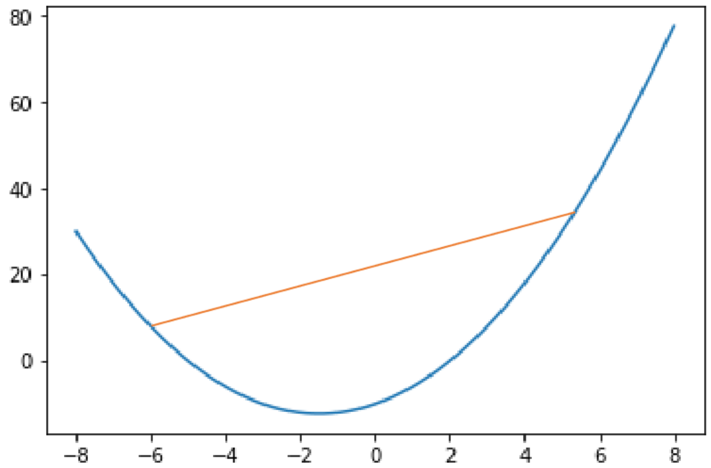
$$\min f(\boldsymbol{x})$$
$$s.t.\ g_i(\boldsymbol{x}) \le 0, i = 1, \dots k$$
$$h_i(\boldsymbol{x}) = 0, i = 1, \dots p$$

当原问题中

- $f(\boldsymbol{x})$， $g_i(\boldsymbol{x})$ 均为凸函数
- $h_i(\boldsymbol{x})$ 为线性函数
- 满足KKT条件的点也是原问题的最优解（充分条件）

在 $f(x)$ 的定义域内有两点 x,y，如果对于任意的 $0\le\theta\le1$ 都有

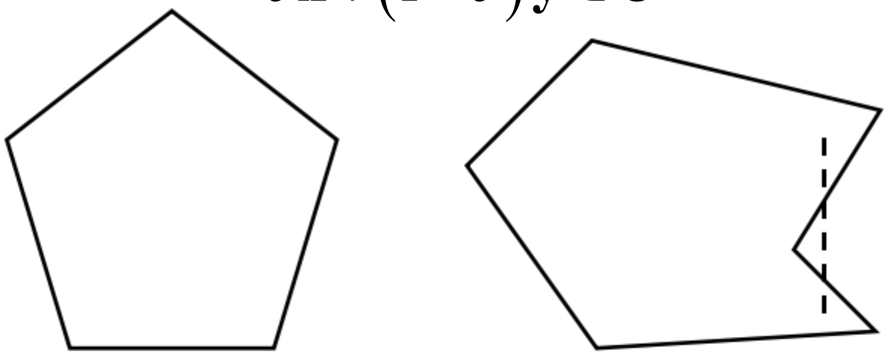
$$f\left(\theta x+(1-\theta)y\right)\le\theta f(x)+(1-\theta)f(y)$$



凸函数

对于n维空间中的点集C，如果对该集合中的任意两点$\mathbf{x}$和$\mathbf{y}$，以及实数$0\le\theta\le1$，都有

$$\theta\mathbf{x}+(1-\theta)\mathbf{y}\in C$$



凸集和非凸集

KKT条件

KKT条件（Karush-Kuhn-Tucker）

- 用于求解带有等式和不等式约束的优化问题，是拉格朗日乘数法的推广

对于下面的最优化问题

$$\min f(\boldsymbol{x})$$

$$s.t.\ g_i(\boldsymbol{x}) \le 0, i = 1, \dots k$$

$$h_i(\boldsymbol{x}) = 0, i = 1, \dots p$$

与拉格朗日乘数法相比，有不等式约束

取得极值的必要条件是

$$\nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = 0 \qquad \alpha_i \ge 0, i = 1, \dots k$$

Dual Feasible

$$g_i(\boldsymbol{x}) \le 0, i = 1, \dots k$$

$$h_i(\boldsymbol{x}) = 0, i = 1, \dots p$$

Primal Feasible

$$\alpha_i g_i(\boldsymbol{x}) = \mathbf{0}, i = 1, \dots k$$

Complementary Slack

Now Back to SVM Problem

支持向量机对偶问题

$$\text{Min} \quad \frac{1}{2}\|\boldsymbol{\omega}\|^2$$

s.t. $\mathbf{1} - \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) \le 0$ i = 1,....m

1：构造拉格朗日乘子函数

$$\mathcal{L}(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 + \sum_{i=1}^{m} \alpha_i (\mathbf{1} - \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b))$$

2：对偶问题

$$\max_{\boldsymbol{\alpha}}(\min_{b,\boldsymbol{\omega}} L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}))$$

3：求导 w,b

$$\max_{\boldsymbol{\alpha}}(\min_{b,\boldsymbol{\omega}} \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{m} \alpha_i (y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) - 1) \ )$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\omega}, b, \boldsymbol{\alpha})}{\partial \boldsymbol{\omega}} = \boldsymbol{\omega} - \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\omega}, b, \boldsymbol{\alpha})}{\partial \boldsymbol{b}} = - \sum_{i=1}^{m} \alpha_i y_i$$

$$\boldsymbol{\omega} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i$$

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

$$\frac{\partial \|\boldsymbol{\omega}\|^2}{\partial \boldsymbol{\omega}} = 2\boldsymbol{\omega}$$

$$\frac{\partial \boldsymbol{\omega}^T \boldsymbol{x}}{\partial \boldsymbol{\omega}} = \boldsymbol{x}$$

对偶问题

$$\max_{\alpha}\left(\min_{b,w}\frac{1}{2}\|\boldsymbol{\omega}\|^2-\sum_{i=1}^{m}\alpha_i\left(y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i+b)-1\right)\right)$$

- 4：回代

$$\boldsymbol{\omega}=\sum_{i=1}^{m}\alpha_i y_i\boldsymbol{x}_i \qquad \sum_{i=1}^{m}\alpha_i y_i=0$$

$$\max_{\alpha}\left(\min_{b,\boldsymbol{\omega}}\frac{1}{2}\|\boldsymbol{\omega}\|^2-\sum_{i=1}^{m}\alpha_i\left(y_i\boldsymbol{\omega}^T\boldsymbol{x}_i-1\right)-\underbrace{\sum_{i=1}^{m}\alpha_i y_i b}_{0}\right)$$

$$\max_{\alpha}\left(\min_{b,\boldsymbol{\omega}}\frac{1}{2}\|\boldsymbol{\omega}\|^2+\sum_{i=1}^{m}\alpha_i\left(1-y_i\boldsymbol{\omega}^T\boldsymbol{x}_i\right)\right)$$

$$\max_{\alpha}\left(\min_{b,\boldsymbol{\omega}}\frac{1}{2}\|\boldsymbol{\omega}\|^2+\sum_{i=1}^{m}\alpha_i-\underbrace{\sum_{i=1}^{m}\alpha_i y_i\boldsymbol{\omega}^T\boldsymbol{x}_i}_{\boldsymbol{\omega}^T\boldsymbol{\omega}}\right)$$

$$\max_{\alpha}\left(-\frac{1}{2}\|\boldsymbol{\omega}\|^2+\sum_{i=1}^{m}\alpha_i\right) \quad \max_{\alpha}\left(-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j {x_i}^T x_j+\sum_{i=1}^{m}\alpha_i\right)$$

对偶问题

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

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

$$\alpha_i \geq 0, i = 1, \cdots, m$$

KKT条件:

$$\begin{cases} \nabla_{\boldsymbol{\omega}} \mathcal{L}(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = 0 \\ \nabla_b \mathcal{L}(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = 0 \\ \alpha_i \left(1 - \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b)\right) = 0 \\ \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) \geq 1 \\ \alpha_i \geq 0, i = 1, \ldots m \end{cases}$$

对偶问题

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

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

$$\alpha_i \geq 0, i = 1, \cdots, m$$

$$\boldsymbol{\omega}^* = \sum_{i=1}^{m} \alpha_i^* y_i \boldsymbol{x}_i$$

KKT条件

$$\begin{cases} \alpha_i \geq 0 \\ \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) \geq 1 \\ \alpha_i \left(1 - \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b)\right) = 0 \end{cases}$$

对任意训练样本$(\boldsymbol{x}_i, y_i)$，

有$\alpha_i = 0$ 或 $\boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) = 1$.

对偶问题

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

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

$$\alpha_i \geq 0, i = 1, \cdots, m$$

$$\boldsymbol{\omega}^* = \sum_{i=1}^{m} \alpha_i^* y_i \boldsymbol{x}_i$$

若$\alpha_i = 0$，则不会对$\boldsymbol{\omega}^*$有任何影响；

KKT条件

$$\begin{cases} \alpha_i \geq 0 \\ \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) \geq 1 \\ \alpha_i\,(1 - \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b)) = 0 \end{cases}$$

对任意训练样本$(\boldsymbol{x}_i, y_i)$，

有$\alpha_i = 0$ 或 $\boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) = 1.$

对偶问题

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

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

$$\alpha_i \geq 0, i = 1, \cdots, m$$

$$\boldsymbol{\omega}^* = \sum_{i=1}^{m} \alpha_i^* y_i \boldsymbol{x}_i$$

若$\alpha_i = 0$，则不会对$\boldsymbol{\omega}^*$有任何影响；

KKT条件

$$\begin{cases} \alpha_i \geq 0 \\ \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) \geq 1 \\ \alpha_i \left(1 - \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b)\right) = 0 \end{cases}$$

若$\alpha_i > 0$，则$\boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) = 1$，所对应的样本点位于最大间隔边界上，是一个支持向量.

对任意训练样本$(\boldsymbol{x}_i, y_i)$，

有$\alpha_i = 0$ 或 $\boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) = 1$.

对偶问题

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

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

$$\alpha_i \geq 0, i = 1, \cdots, m$$

KKT条件:

$$\begin{cases} \underline{\nabla_{\omega} \mathcal{L}(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = 0} \\ \nabla_{b} \mathcal{L}(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = 0 \\ \underline{\alpha_i \left(1 - \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b)\right) = 0} \\ \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) \geq 1 \\ \alpha_i \geq 0, i = 1, \dots m \end{cases} \Rightarrow \quad \boldsymbol{\omega}^* = \sum_{i=1}^{m} \alpha_i^* y_i \boldsymbol{x}_i,$$

对偶问题

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

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

$$\alpha_i \geq 0, i = 1, \cdots, m$$

KKT条件:

$$\begin{cases} \nabla_{\boldsymbol{\omega}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = 0 \\ \nabla_b \mathcal{L}(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = 0 \\ \alpha_i\left(1 - \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b)\right) = 0 \\ \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) \geq 1 \\ \alpha_i \geq 0, i = 1, \dots m \end{cases}$$

$$\boldsymbol{\omega}^* = \sum_{i=1}^{m} \alpha_i^* y_i \boldsymbol{x}_i,$$

$$b^* = y_j - \sum_{i=1}^{m} \alpha_i^* y_i \boldsymbol{x}_i^T \boldsymbol{x}_j$$

对偶问题

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

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

$$\alpha_i \geq 0, i = 1, \cdots, m$$

5：求解

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

j为$\boldsymbol{\alpha}^*$中的一个正分量，存在下标j，使得$\alpha_j^* > 0$，

对于线性可分问题根据 KKT条件 求出$(\boldsymbol{\omega}^*, b^*)$

$$\boldsymbol{\omega}^* = \sum_{i=1}^{m} \alpha_i^* y_i \boldsymbol{x}_i, \qquad b^* = y_j - \sum_{i=1}^{m} \alpha_i^* y_i \boldsymbol{x}_i^T \boldsymbol{x}_j$$

学习的对偶求解

- 线性可分支持向量机

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

通过求解对偶问题得到原始问题的最优解

第一步：引入拉格朗日乘子$\alpha_i \geq 0$，得到拉格朗日函数

$$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 + \sum_{i=1}^{m} \alpha_i \left(1 - y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right)\right)$$

$$\boldsymbol{\alpha} = (\alpha_1; \cdots; \alpha_m)$$

根据拉格朗日对偶性，原始问题的对偶问题是极大极小问题：

$$\max_{\boldsymbol{\alpha}} \min_{\boldsymbol{\omega},b} L(\boldsymbol{\omega}, b, \boldsymbol{\alpha})$$

为了得到对偶问题的解，需要先求$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha})$对$\boldsymbol{\omega}, b$的极小，再求对$\boldsymbol{\alpha}$的极大

学习的对偶求解

线性可分支持向量机

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

通过求解对偶问题得到原始问题的最优解

第一步：引入拉格朗日乘子$\alpha_i \geq 0$，得到拉格朗日函数

$$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 + \sum_{i=1}^{m}\alpha_i\left(1 - y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right)\right)$$

$$\boldsymbol{\alpha} = (\alpha_1;\cdots;\alpha_m)$$

第二步：令$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha})$对$\boldsymbol{\omega}$和$b$的偏导为零可得

$$\boldsymbol{\omega} = \sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i, \sum_{i=1}^{m}\alpha_i y_i = 0$$

学习的对偶算法

□ 线性可分支持向量机

第三步：回代，得到对偶问题

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

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

$$\alpha_i \geq 0, i = 1, \cdots, m$$

第四步：解出$\boldsymbol{\alpha}^*$，对于线性可分问题根据KKT条件求出$(\boldsymbol{\omega}^*, b^*)$

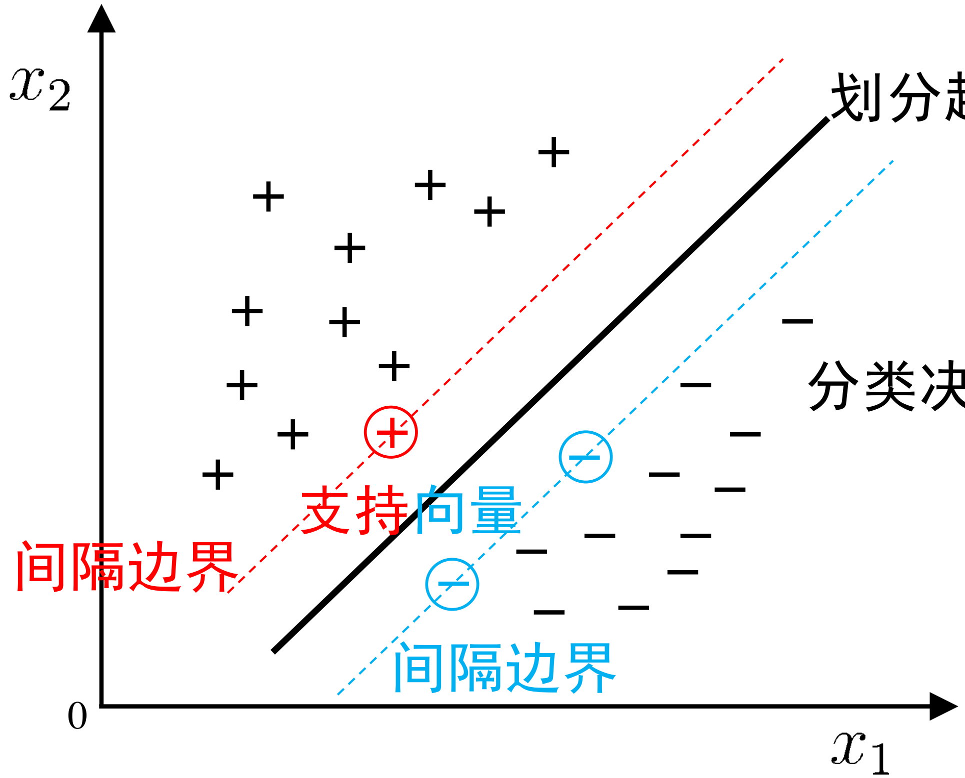
$$\boldsymbol{\omega}^* = \sum_{i=1}^{m} \alpha_i^* y_i \boldsymbol{x}_i \,, b^* = y_j - \sum_{i=1}^{m} \alpha_i^* y_i \boldsymbol{x}_i^T \boldsymbol{x}_j$$

j为$\boldsymbol{\alpha}^*$中的一个正分量 $\alpha_j^* > 0$所对应的下标

学习的对偶算法

☐ 线性可分支持向量机

第五步： 最终模型$f(\boldsymbol{x}) = \mathrm{sign}\left(\boldsymbol{\omega}^{*\mathrm{T}}\boldsymbol{x} + b^*\right)$



划分超平面：$\sum_{i=1}^{m} \alpha_i^* y_i \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x} + b^* = 0$

分类决策函数：$f(\boldsymbol{x}) = \mathrm{sign}\left(\sum_{i=1}^{m} \alpha_i^* y_i \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x} + b^*\right)$

现象：分类决策函数只依赖于输入$\boldsymbol{x}$和训练样本输入的内积。

仅依赖于支持向量

谢谢！