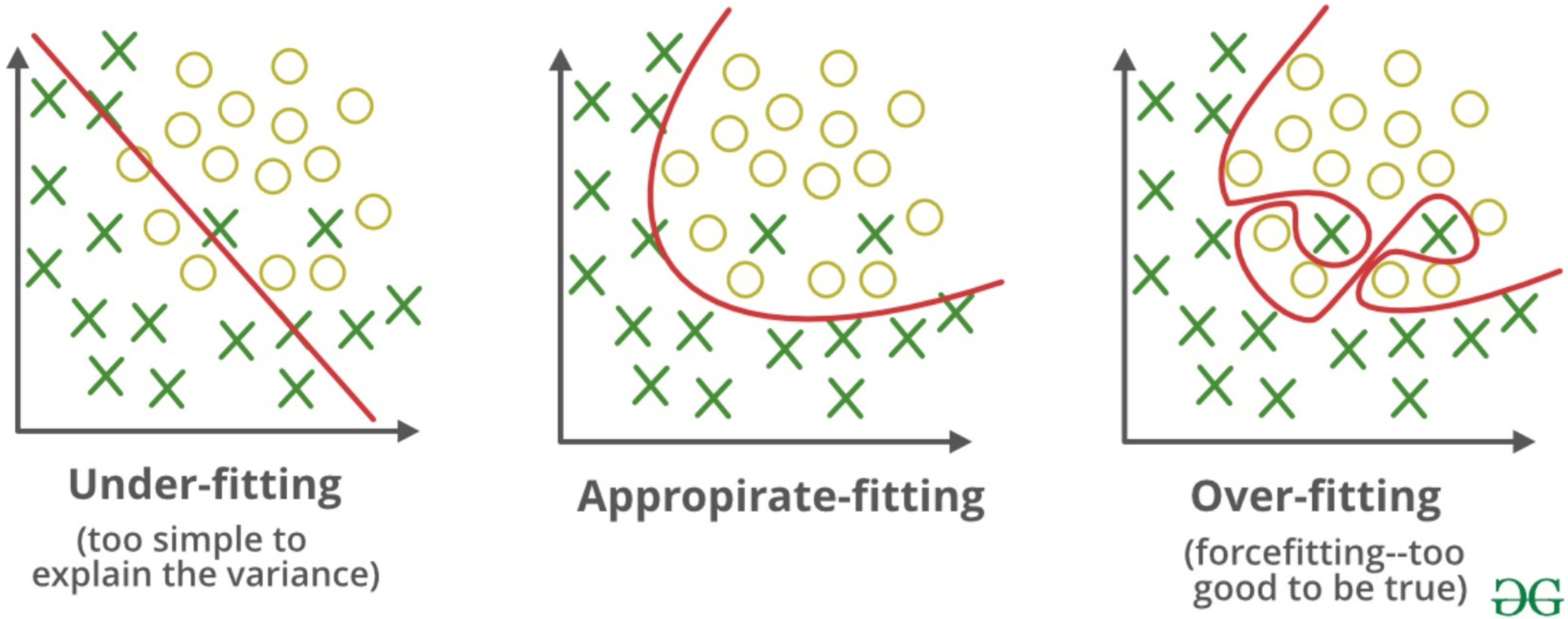
中山大學 SUN YAT-SEN UNIVERSITY

第5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《机器学习方法》第7章

软间隔



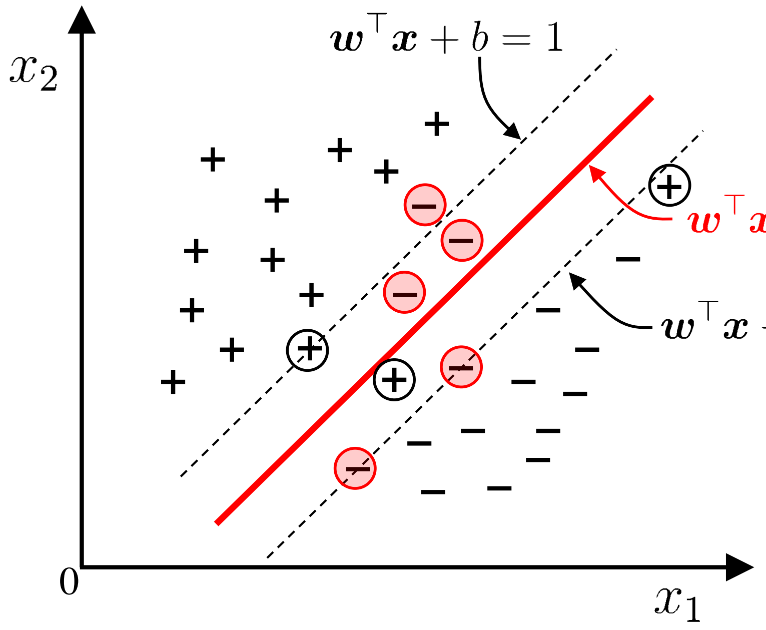
https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/

软间隔

-Q:现实中, 很难确定合适的函数使得训练样本在特征空间中线性可分; 同时一个线性可分的结果也很难断定是否是有过拟合造成的.

-A:引入”软间隔”的概念, 允许支持向量机在一些样本上不满足约束.

不满足约束的样本 越少越好



例子

- 数据非完全可分的情况

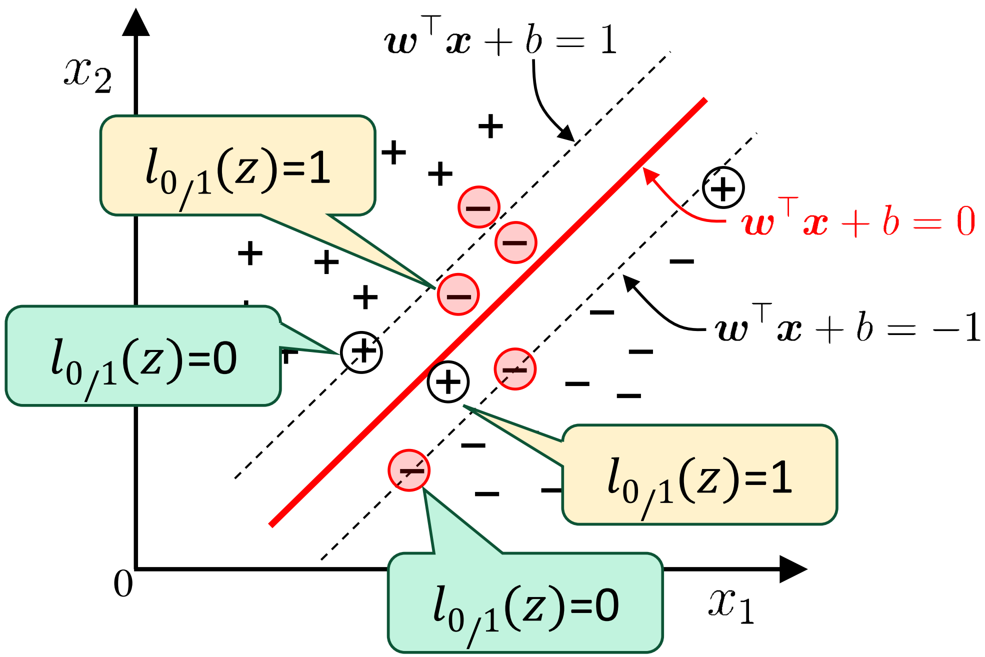
支持向量机 SVM

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

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

不满足约束的样本　越少越好

$$\min_{b,\boldsymbol{\omega}}\left(\sum_{i=1}^{m} l_{0/1}(y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) - 1) \right)$$



$$l_{0/1}(z) = \begin{cases} 1, & if\ z < 0 \\ 0, & othereise \end{cases}$$

软间隔

- 数据非完全可分的情况

支持向量机 SVM

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

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

不满足约束的样本 越少越好

$$\min_{b,w}\left(\sum_{i=1}^{m} l_{0/1}(y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) - 1) \right)$$

两者结合一下？

$$\min_{b,w} \quad \frac{1}{2}\|w\|^2 + \boxed{C * \sum_{i=1}^{m} l_{0/1}(y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) - 1)}$$

0/1损失函数

存在的问题：0/1损失函数非凸、非连续, 不易优化！

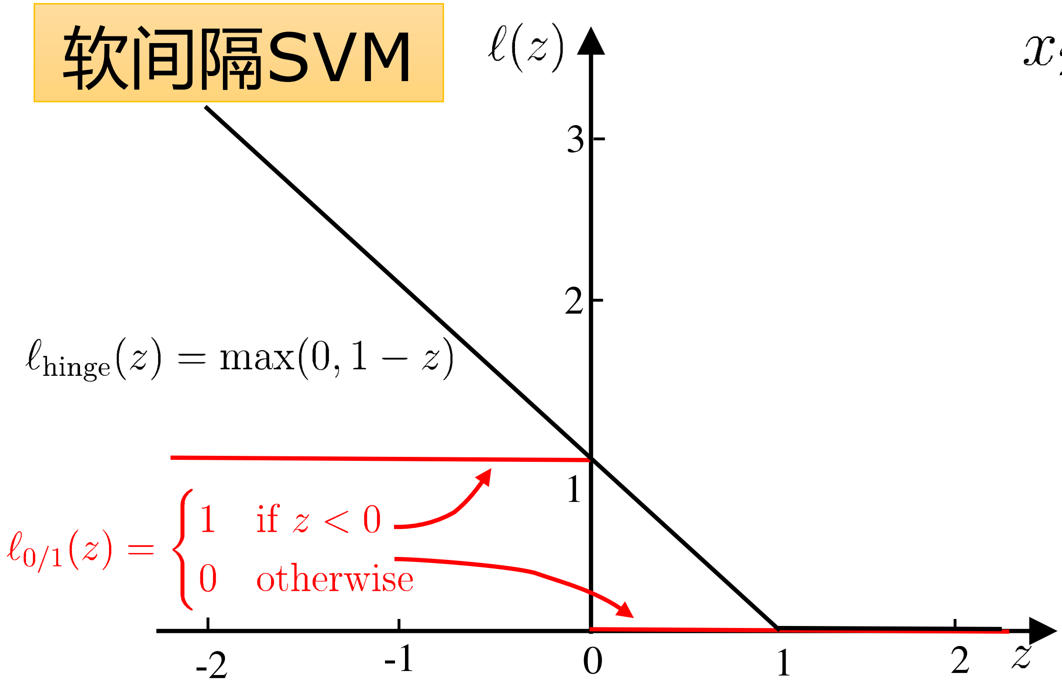
替代损失

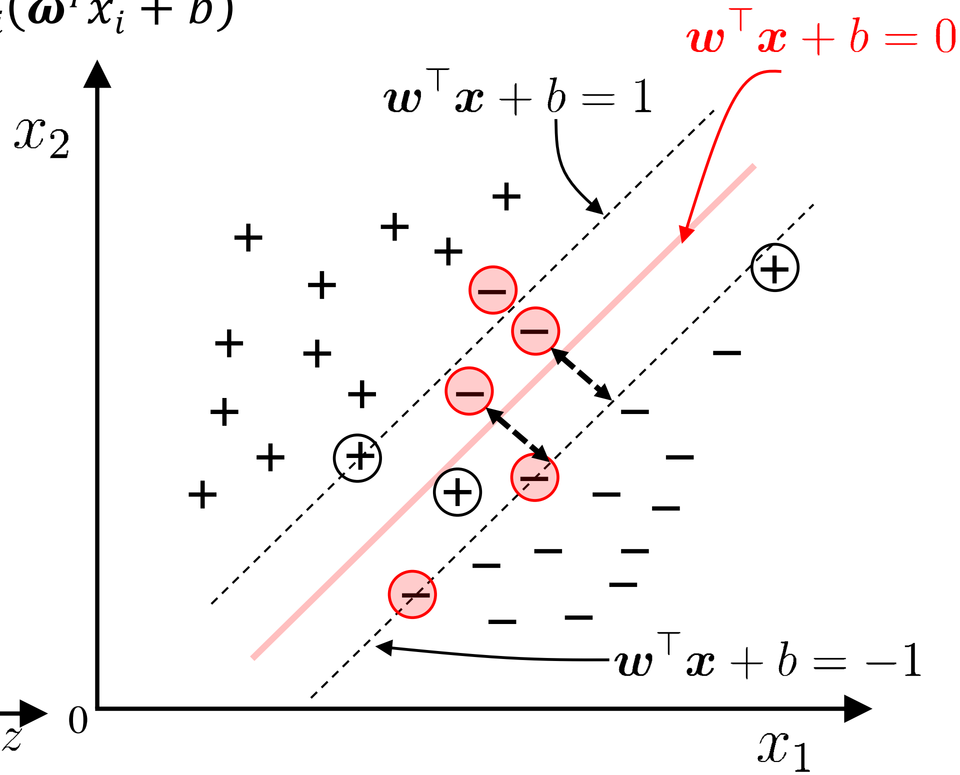
$$\min_{b,w} \quad \frac{1}{2}\|w\|^2 + C * \sum_{i=1}^{m} l_{0/1}(y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) - 1)$$

$$s.t. \quad y_i(\boldsymbol{\omega}^T x_i + b) \geq 1 \text{ for all } i \text{ <= m}$$

- Loss
  - IF $\quad y_i(\boldsymbol{\omega}^T x_i + b) \geq 1 \quad loss = 0$
  - IF $\quad y_i(\boldsymbol{\omega}^T x_i + b) < 1 \quad loss = 1 - y_i(\boldsymbol{\omega}^T x_i + b)$

软间隔SVM





软间隔

- 引入松弛变量

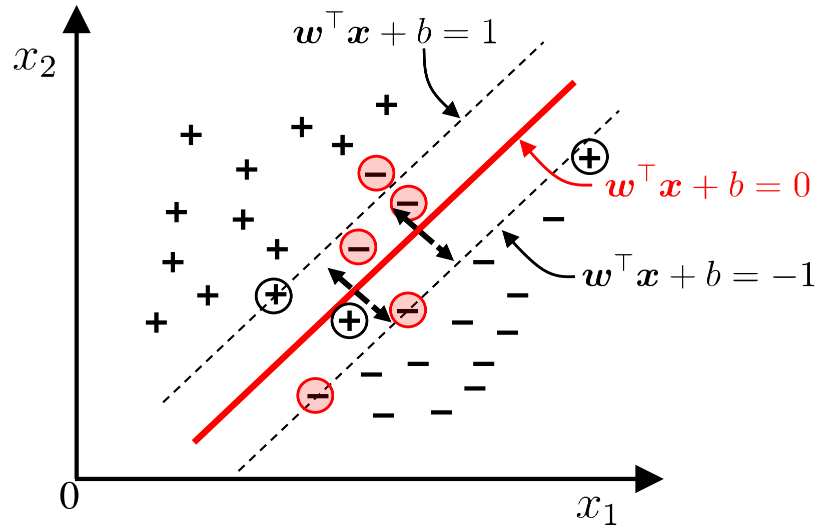
$$\min_{b,w} \quad \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C * \sum_{i=1}^{m} max(0,1 - y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b))$$

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

$\xi_i \geq 0$

$$\min_{b,w} \quad \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C * \sum_{i=1}^{m} \xi_i$$

$$s.t. \quad y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1 - \xi_i, \; \xi_i \geq 0$$



软间隔

$$\min_{b,w} \quad \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C * \sum_{i=1}^{m} l_{0/1}(yi(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) - 1)$$

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

$$\min_{b,w} \quad \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C * \sum_{i=1}^{m} \xi_i$$

$$s.t. \quad y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

$$L(\boldsymbol{\omega}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C * \sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \alpha_i (1 - \xi_i - y_i (\boldsymbol{\omega}^T \boldsymbol{x}_i + b)) + \sum_{i=1}^{m} \beta_i (-\xi_i)$$

$$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\xi}) = \frac{1}{2}\|w\|^2 + C * \sum_{i=1}^{m} \xi_i$$
$$+ \sum_{i=1}^{m} \alpha_i \left(1 - \xi_i - y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + b)\right) + \sum_{i=1}^{m} \beta_i (-\xi_i)$$

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

对 $\boldsymbol{\omega}, b, \xi$ 求偏导

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

$s.t.\ \ \alpha_i \geq 0, \beta_i \geq 0, i = 1, \ldots m$

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

s.t. $\sum_{i=1}^{m} \alpha_i y_i = 0$, $C \geq \alpha_i \geq 0$, i= 1, 2, ..., m

线性支持向量机

线性支持向量机

线性不可分

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

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

引进一个松弛变量$\xi_i\geq 0$

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

对每一个松弛变量支付一个代价引入目标函数

$$\begin{aligned}\min_{\boldsymbol{\omega},b,\xi}\quad & \frac{1}{2}\|\boldsymbol{\omega}\|^2+C\sum_{i=1}^{m}\xi_i\\ \text{s.t.}\quad & y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+b\right)\geq 1-\xi_i,\\ & \xi_i\geq 0\\ & i=1,\cdots,m\end{aligned}$$

$C>0$为惩罚参数：

当C取值增大时对误分类的惩罚增大；

当C取值小时对误分类的惩罚减小

线性支持向量机

□ 线性支持向量机

线性不可分

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$\boldsymbol{\omega}\in\mathbb{R}^d$

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（引入松弛变量和惩罚参数）

$$(\boldsymbol{\omega}^*,b^*,\boldsymbol{\xi}^*)=\min_{\boldsymbol{\omega},b,\boldsymbol{\xi}}\frac{1}{2}\|\boldsymbol{\omega}\|^2+C\sum_{i=1}^{m}\xi_i$$

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

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

软间隔最大化

soft margin maximization

算法：对偶算法

学得模型：分离超平面$\boldsymbol{\omega}^*\cdot\boldsymbol{x}+b^*=0$和分类决策函数为$f(\boldsymbol{x})=\mathrm{sign}(\boldsymbol{\omega}^*\cdot\boldsymbol{x}+b^*)$，即线性支持向量机。

学习的对偶算法

- 线性支持向量机

$$\min_{\boldsymbol{\omega},b,\boldsymbol{\xi}} \quad \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{m}\xi_i$$

$$\text{s.t.} \quad y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right) \geq 1 - \xi_i, i = 1,\cdots,m;\ \xi_i \geq 0, i = 1,\cdots,m$$

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

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

$$\boldsymbol{\alpha} = (\alpha_1;\cdots;\alpha_m), \boldsymbol{\beta} = (\beta_1;\cdots;\beta_m)$$

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：

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

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

学习的对偶算法

- 线性支持向量机

$$\min_{\boldsymbol{\omega},b,\xi} \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{m}\xi_i$$

$$\text{s.t.} \quad y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right) \geq 1 - \xi_i, i = 1,\cdots,m;\ \xi_i \geq 0, i = 1,\cdots,m$$

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

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

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

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

学习的对偶算法

线性支持向量机

第三步：回代，得到对偶问题

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

$$\boxed{\begin{aligned} \text{s.t.} \sum_{i=1}^{m}\alpha_i y_i &= 0, C-\alpha_i-\beta_i = 0 \\ \alpha_i \geq 0, \beta_i &\geq 0, i=1,\cdots,m \end{aligned}} \Rightarrow \boxed{\begin{aligned} \text{s.t.} \sum_{i=1}^{m}\alpha_i y_i &= 0, \\ 0 \leq \alpha_i &\leq C, i=1,\cdots,m \end{aligned}}$$

学习的对偶算法

□ 线性支持向量机

第三步：回代，得到对偶问题

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

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

$$0 \le \alpha_i \le C, i = 1, \cdots, m$$

第四步：解出$\boldsymbol{\alpha}^*$，根据KKT条件求出$(\boldsymbol{\omega}^*, b^*)$

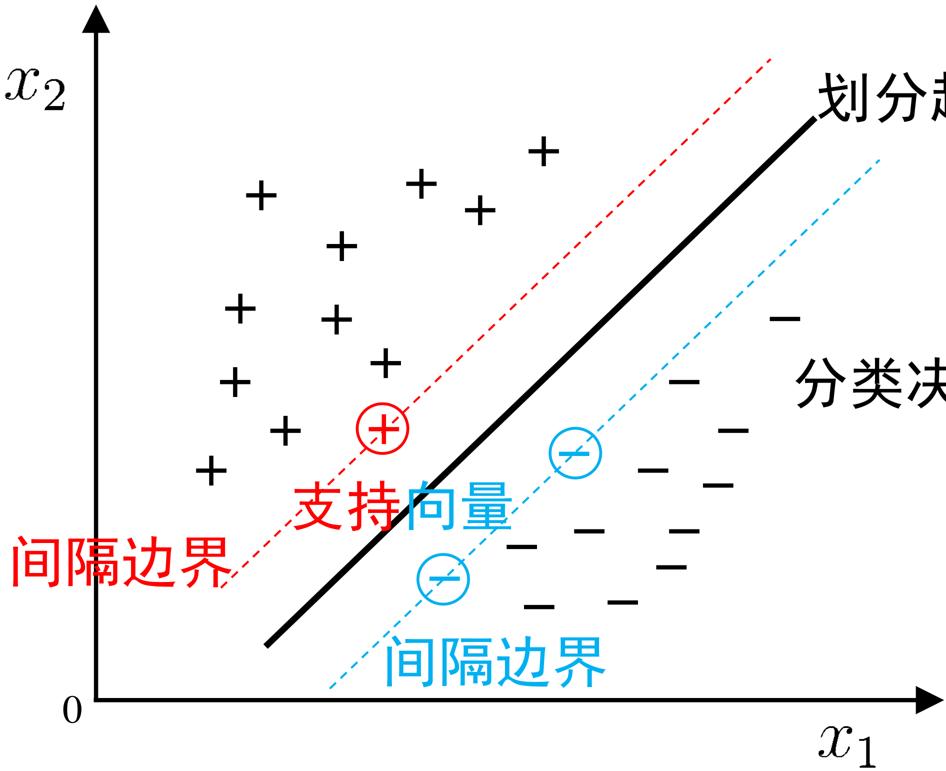
$$\boldsymbol{\omega}^* = \sum_{i=1}^{m}\alpha_i^* y_i \boldsymbol{x}_i, b^* = y_j - \sum_{i=1}^{m}\alpha_i^* y_i \boldsymbol{x}_i^T\boldsymbol{x}_j$$

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

学习的对偶算法

□ 线性支持向量机

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。

线性支持向量机学习算法

◻ 线性支持向量机

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

输出： 划分超平面和分类决策函数

（1） 选择惩罚参数$C>0$，构造并求解约束最优化问题

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

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

$$0\le\alpha_i\le C, i=1,\cdots,m$$

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

线性支持向量机学习算法

□ 线性支持向量机

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

输出： 划分超平面和分类决策函数

（2） 计算

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

并选择$\boldsymbol{\alpha}^*$中的一个正分量$0<\alpha_j^*<C$，计算

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

线性支持向量机学习算法

□ 线性可分支持向量机

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

输出： 划分超平面和分类决策函数

（3） 求得划分超平面

$$\boldsymbol{\omega}^{*\mathrm{T}}\boldsymbol{x}+b^*=0 \text{ 即} \sum_{i=1}^{m}\alpha_i^* y_i \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}+b^*=0$$

和分类决策函数

$$f(\boldsymbol{x})=\operatorname{sign}\left(\sum_{i=1}^{m}\alpha_i^* y_i \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}+b^*\right)$$

硬间隔与软间隔KKT条件

硬间隔：

$$\min_{a}\left(\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j y_i y_j x_i^T x_i - \sum_{i=1}^{m} a_i\right)$$

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

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

KKT条件：

$$1 - \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) \leq 0$$

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

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

软间隔：

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

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

$$0 \leq \alpha_i \leq C, i = 1, \cdots, m$$

KKT条件：

$$y_i\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b\right) \geq 1 - \varepsilon_i, i = 1, \cdots, m$$

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

$$0 < \alpha_i < C \Rightarrow \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) = 1$$

$$\alpha_i = 0 \Rightarrow \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) \geq 1$$

$$\alpha_i = C \Rightarrow \boldsymbol{y_i}(\boldsymbol{\omega}^T x_i + b) \leq 1$$

线性支持向量机

□ 线性支持向量机

线性不可分

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

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（引入松弛变量和惩罚参数）

$$(\boldsymbol{\omega}^*,b^*,\boldsymbol{\xi}^*)=\min_{\boldsymbol{\omega},b,\boldsymbol{\xi}}\frac{1}{2}\|\boldsymbol{\omega}\|^2+C\sum_{i=1}^{m}\xi_i$$

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

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

软间隔最大化

soft margin maximization

算法：对偶算法

学得模型：分离超平面$\boldsymbol{\omega}^*\cdot\boldsymbol{x}+b^*=0$和分类决策函数为$f(\boldsymbol{x})=\operatorname{sign}(\boldsymbol{\omega}^*\cdot\boldsymbol{x}+b^*)$，即线性支持向量机。

谢谢！