第6章 支持向量机

王博：自动化（人工智能）学院
wangbo@hdu. edu. cn

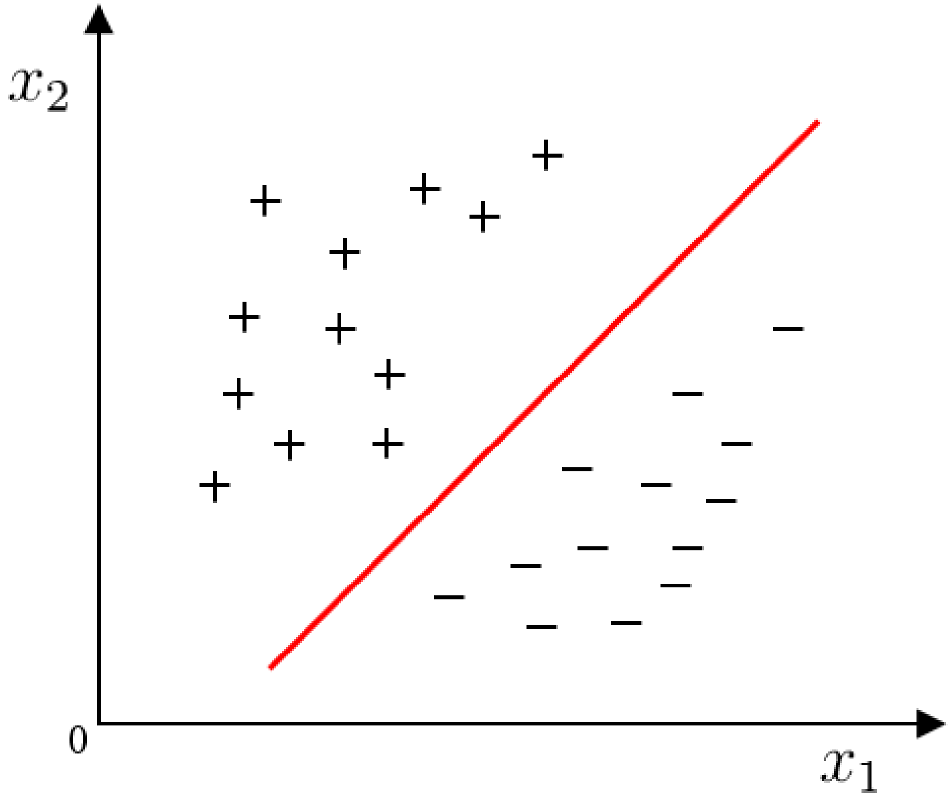
章节目录

- 间隔与支持向量
- 对偶问题
- 核函数
- 软间隔与正则化
- 支持向量回归
- 核方法

章节目录

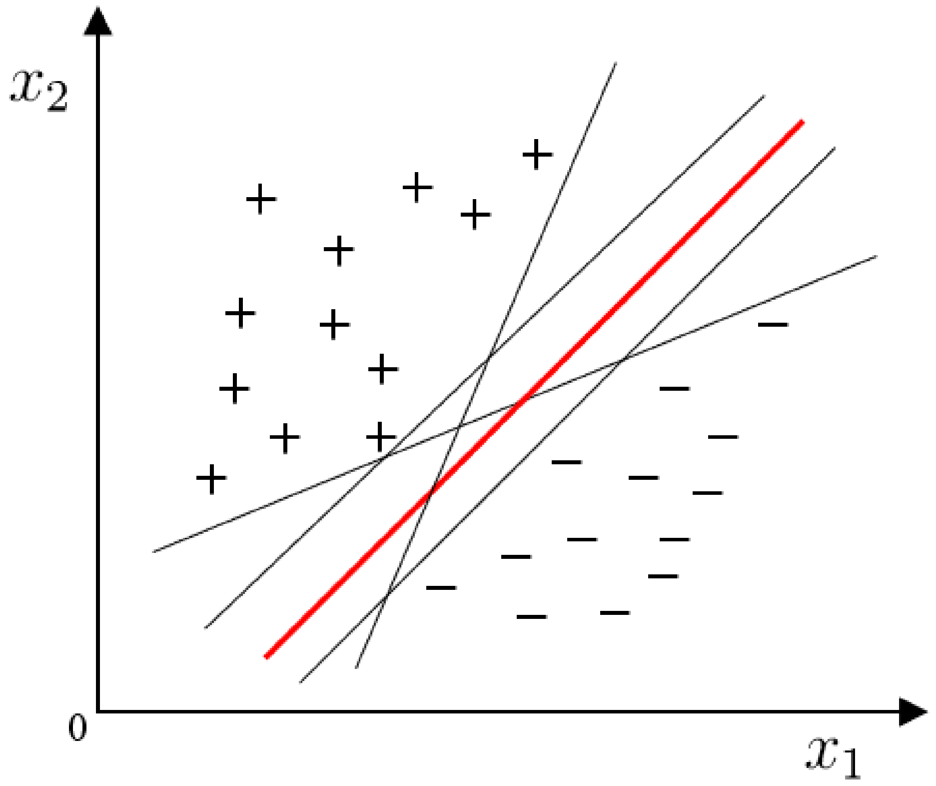
间隔与支持向量

线性模型：在样本空间中寻找一个超平面, 将不同类别的样本分开.



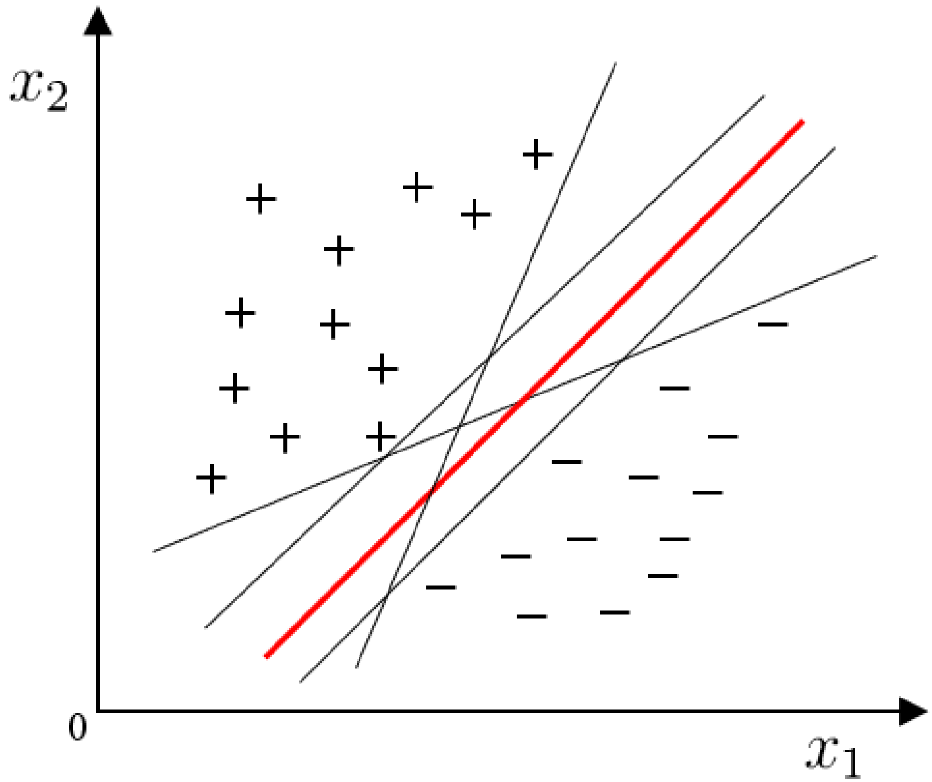
间隔与支持向量

Q:将训练样本分开的超平面可能有很多, 哪一个好呢?



间隔与支持向量

Q:将训练样本分开的超平面可能有很多, 哪一个好呢?



A:应选择"正中间", 容忍性好, 鲁棒性高, 泛化能力最强.

间隔与支持向量

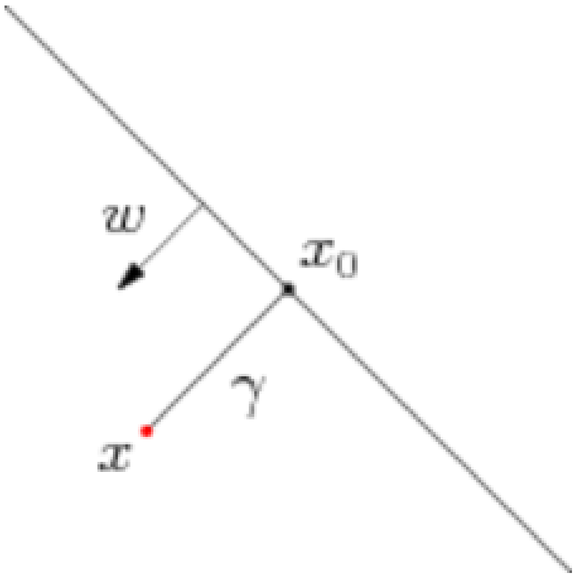
超平面方程: $\boldsymbol{w}^\top \boldsymbol{x} + b = 0$

点到平面的距离:

$$x = x_0 + \gamma \frac{w}{\|w\|}$$

$$\Downarrow$$

$$\gamma = \frac{w^T x + b}{\|w\|} = \frac{f(x)}{\|w\|}$$

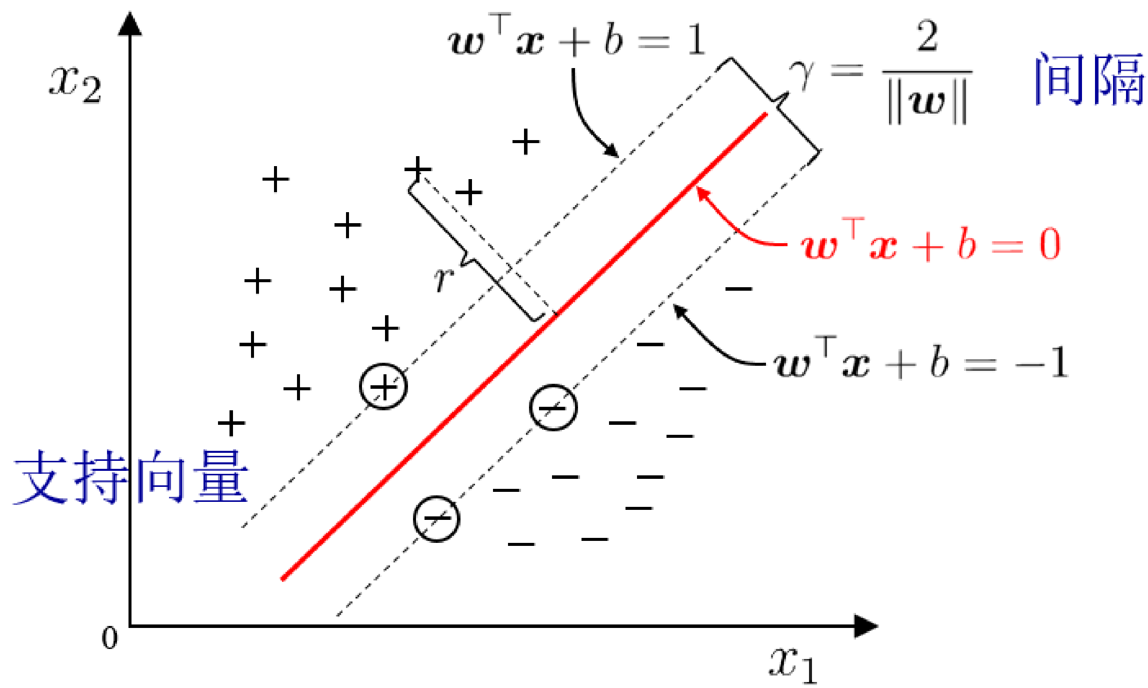


间隔与支持向量

超平面方程: $\boldsymbol{w}^{\top}\boldsymbol{x}+b=0$

假设线性可分下引入间隔y为标签:

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



目标函数为:

$$\max \tilde{\gamma}$$

$$s.t.,\ y_i(w^T x_i+b)=\hat{\gamma}_i \geq \hat{\gamma},\quad i=1,\ldots,n$$

支持向量机基本型

最大间隔: 寻找参数$\boldsymbol{w}$ 和b, 使得γ 最大.

$$\begin{aligned} \arg\max_{\boldsymbol{w},b} \quad & \frac{2}{\|\boldsymbol{w}\|} \\ \text{s.t.} \quad & y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq 1, \ i = 1, 2, \ldots, m. \end{aligned}$$

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

凸优化: 凸集上的凸函数的最小化问题

章节目录

- 间隔与支持向量
- **对偶问题**
- 核函数
- 软间隔与正则化
- 支持向量回归
- 核方法

对偶问题

- 标准优化问题，假定目标函数及约束函数都是可微的：

$$
\begin{aligned}
&\min_{\mathbf{x}} f_0(\mathbf{x}) \\
&s.t.\ \ f_i(\mathbf{x}) \le 0,\ i = 1, \ldots, m \\
&\qquad\ h_i(\mathbf{x}) = 0,\ i = 1, \ldots, p
\end{aligned}
$$

- 拉格朗日函数：

$$
L(\mathbf{x}, \lambda, \nu) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})
$$

其中，$\lambda_i \ge 0$，ν_i 分别为不等式约束和等式约束的拉格朗日乘子

对偶问题

- 拉格朗日对偶函数：

$$g(\lambda, \nu) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda, \nu)$$

其中 $\lambda = [\lambda_1, \ldots, \lambda_m]^T$ 和 $\nu = [\nu_1, \ldots, \nu_p]^T$

- 对偶问题：

$$\max_{\lambda,\, \nu} g(\lambda, \nu)$$
$$s.t.\ \lambda \succeq \mathbf{0}$$

KKT条件

记 $\mathbf{x}^\star$ 和 $(\lambda^\star, \nu^\star)$ 分别是原问题和对偶问题的一对最优解

假设强对偶性成立(原对偶问题有相同的最优值)，则有如下一组重要的结论

$$
\begin{aligned}
f_i(\mathbf{x}^\star) &\le 0,\ i=1,\ldots,m \\
h_i(\mathbf{x}^\star) &= 0,\ i=1,\ldots,p \\
\lambda_i^\star &\ge 0,\ i=1,\ldots,m \\
\lambda_i^\star f_i(\mathbf{x}^\star) &= 0,\ i=1,\ldots,m
\end{aligned}
$$

$$
\nabla f_0(\mathbf{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\mathbf{x}^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(\mathbf{x}^\star) = \mathbf{0}
$$

KKT条件是最优解的必要条件。

若原问题为凸优化问题，且目标函数和约束函数可微，KKT条件也是充分条件。

对偶问题

- 目标函数：

$$\min \frac{1}{2}\|w\|^2 \quad s.t.,\, y_i(w^T x_i + b) \geq 1, i = 1, \ldots, n$$

- 拉格朗日乘子法：

$$\mathcal{L}(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i(w^T x_i + b) - 1 \right)$$

- 引入函数：

$$\theta(w) = \max_{\alpha_i \geq 0} \mathcal{L}(w,b,\alpha)$$

- 原目标函数变成：

$$\min_{w,b} \theta(w) = \min_{w,b} \max_{\alpha_i \geq 0} \mathcal{L}(w,b,\alpha) = p^*$$

- 对偶问题：

$$\max_{\alpha_i \geq 0} \min_{w,b} \mathcal{L}(w,b,\alpha) = d^*$$

对偶问题

- 对偶函数：$$\max_{\alpha_i \geq 0} \min_{w,b} \mathcal{L}(w,b,\alpha) = d^*$$
- 首先 $\mathcal{L}(w,b,\alpha)$ 关于$\boldsymbol{w}$ 和 b最小化

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0$$

- 代入 $$\mathcal{L}(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i(w^T x_i + b) - 1 \right)$$

$$\mathcal{L}(w,b,\alpha) = \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j - b\sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n} \alpha_i$$

$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

对偶问题

- 原问题：

$$\min \frac{1}{2}\|w\|^2 \quad s.t.\,, y_i(w^T x_i + b) \geq 1, i = 1, \ldots, n$$

- 对偶问题：

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

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

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

解的稀疏性

最终模型： $f(\boldsymbol{x}) = \boldsymbol{w}^\top \boldsymbol{x} + b = \boxed{\sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i^\top \boldsymbol{x}} + b$

KKT条件：

$$\begin{cases} \alpha_i \geq 0, \\ y_i f(\boldsymbol{x}_i) \geq 1, \\ \alpha_i (y_i f(\boldsymbol{x}_i) - 1) = 0. \end{cases}$$

+

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^n \alpha_i y_i x_i$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \Rightarrow \sum_{i=1}^n \alpha_i y_i = 0$$

$$\boxed{y_i f(\boldsymbol{x}_i) > 1 \quad \Rightarrow \quad \alpha_i = 0}$$

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

偏移项的确定

偏移项 b：通过支持向量来确定

任意的支持向量 (x_s, y_s)：

$$y_s f(x_s) = 1 \quad \Rightarrow \quad y_s(\sum_{i \in S} \alpha_i y_i x_i^T x_s + b) = 1$$

更鲁棒的方法：

$$b = \frac{1}{|S|} \sum_{s \in S} (y_s - \sum_{i \in S} \alpha_i y_i x_i^T x_s)$$

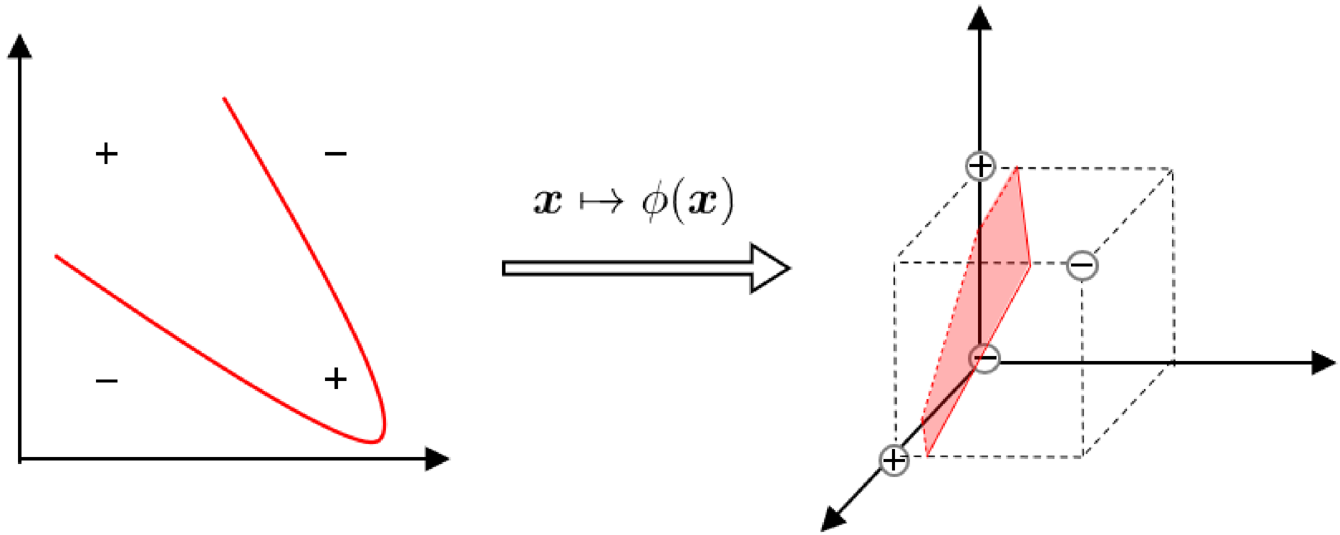
章节目录

- 间隔与支持向量
- 对偶问题
- **核函数**
- 软间隔与正则化
- 支持向量回归
- 核方法

线性不可分

Q:若不存在一个能正确划分两类样本的超平面，怎么办?

A:将样本从原始空间映射到一个更高维的特征空间，使得样本在这个特征空间内线性可分.



异或问题与非线性映射

核支持向量机

设样本$\boldsymbol{x}$映射后的向量为$\phi(\boldsymbol{x})$，划分超平面为 $f(\boldsymbol{x}) = \boldsymbol{w}^\top \phi(\boldsymbol{x}) + b$

原始问题：
$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2$$
$$\text{s.t. } y_i(\boldsymbol{w}^\top \phi(\boldsymbol{x}_i) + b) \geq 1,\ i = 1, 2, \ldots, m.$$

对偶问题：
$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^m \alpha_i - \frac{1}{2}\sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \boxed{\phi(\boldsymbol{x}_i)^\top \phi(\boldsymbol{x}_j)}$$
$$\text{s.t. } \sum_{i=1}^m \alpha_i y_i = 0,\ \alpha_i \geq 0,\ i = 1, 2, \ldots, m.$$

求解后模型：
$$f(\boldsymbol{x}) = \boldsymbol{w}^\top \phi(\boldsymbol{x}) + b = \sum_{i=1}^m \alpha_i y_i \boxed{\phi(\boldsymbol{x}_i)^\top \phi(\boldsymbol{x})} + b$$

核函数

基本想法：不显式地设计核映射, 而是设计核函数.

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^\top \phi(\boldsymbol{x}_j)$$

定义 核函数[2] 设 $\mathcal{X}$ 是输入空间（即 $x_i \in \mathcal{X}$， $\mathcal{X}$ 是 $\mathbb{R}^n$ 的子集或离散集合），又设 $\mathcal{H}$ 为特征空间（$\mathcal{H}$ 是希尔伯特空间[3]），如果存在一个从 $\mathcal{X}$ 到 $\mathcal{H}$ 的映射

$$\phi(x) : \mathcal{X} \rightarrow \mathcal{H}$$

使得对所有 $x, z \in \mathcal{X}$，函数 $K(x, z)$ 满足条件

$$K(x, z) = \langle \phi(x), \phi(z) \rangle$$

则称 K 为核函数。其中 $\phi(x)$ 为映射函数，$\langle \cdot, \cdot \rangle$ 为内积。

核函数

定理 6.1 (核函数) 令 $\mathcal{X}$ 为输入空间, $\kappa(\cdot,\cdot)$ 是定义在 $\mathcal{X}\times\mathcal{X}$ 上的对称函数, 则 κ 是核函数当且仅当对于任意数据 $D=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_m\}$, "核矩阵" (kernel matrix) $\mathbf{K}$ 总是半正定的:

$$
\mathbf{K}=\begin{bmatrix}
\kappa(\boldsymbol{x}_1,\boldsymbol{x}_1) & \cdots & \kappa(\boldsymbol{x}_1,\boldsymbol{x}_j) & \cdots & \kappa(\boldsymbol{x}_1,\boldsymbol{x}_m)\\
\vdots & \ddots & \vdots & \ddots & \vdots\\
\kappa(\boldsymbol{x}_i,\boldsymbol{x}_1) & \cdots & \kappa(\boldsymbol{x}_i,\boldsymbol{x}_j) & \cdots & \kappa(\boldsymbol{x}_i,\boldsymbol{x}_m)\\
\vdots & \ddots & \vdots & \ddots & \vdots\\
\kappa(\boldsymbol{x}_m,\boldsymbol{x}_1) & \cdots & \kappa(\boldsymbol{x}_m,\boldsymbol{x}_j) & \cdots & \kappa(\boldsymbol{x}_m,\boldsymbol{x}_m)
\end{bmatrix}.
$$

核函数

常用核函数：

名称	表达式	参数
线性核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^\top \boldsymbol{x}_j$	
多项式核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^\top \boldsymbol{x}_j)^d$	$d \geq 1$为多项式的次数
高斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\\|\boldsymbol{x}_i - \boldsymbol{x}_j\\|^2}{2\delta^2}\right)$	$\delta > 0$为高斯核的带宽(width)
拉普拉斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\\|\boldsymbol{x}_i - \boldsymbol{x}_j\\|}{\delta}\right)$	$\delta > 0$
Sigmoid核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh(\beta \boldsymbol{x}_i^\top \boldsymbol{x}_j + \theta)$	tanh为双曲正切函数, $\beta > 0$, $\theta < 0$

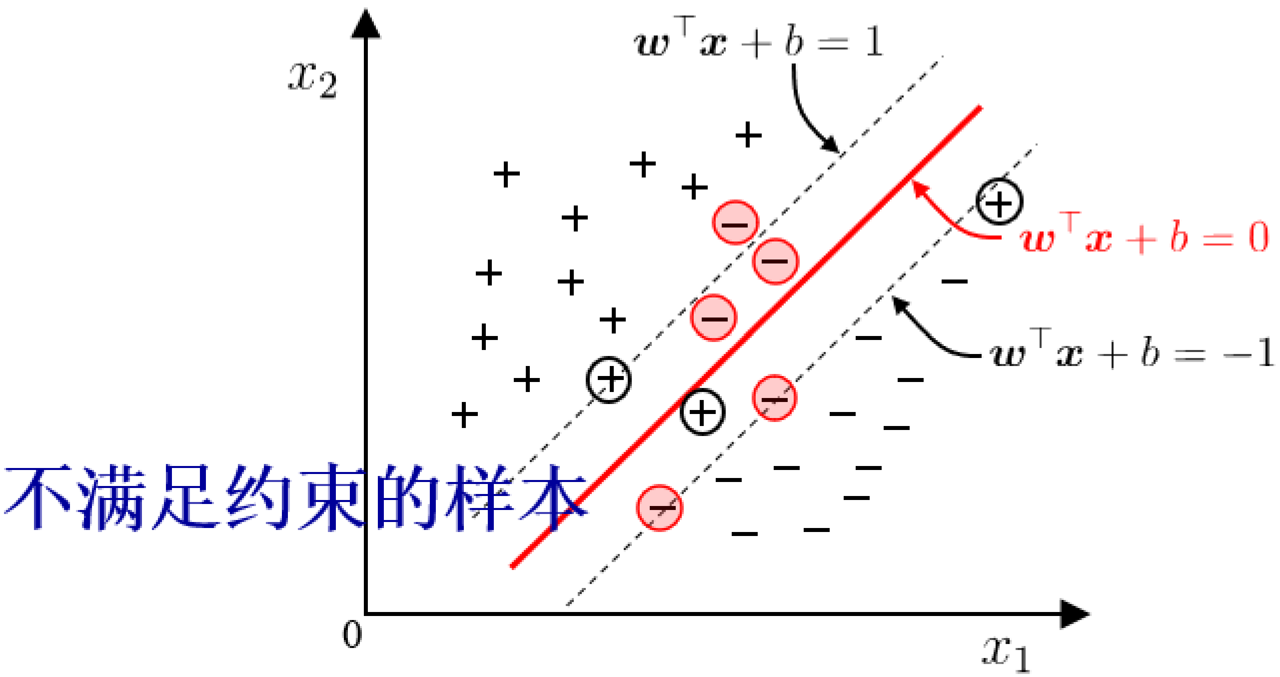
章节目录

- 间隔与支持向量
- 对偶问题
- 核函数
- **软间隔与正则化**
- 支持向量回归
- 核方法

软间隔

Q:现实中, 很难确定合适的核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.



0/1损失函数

基本想法：最大化间隔的同时, 让不满足约束的样本应尽可能少.

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

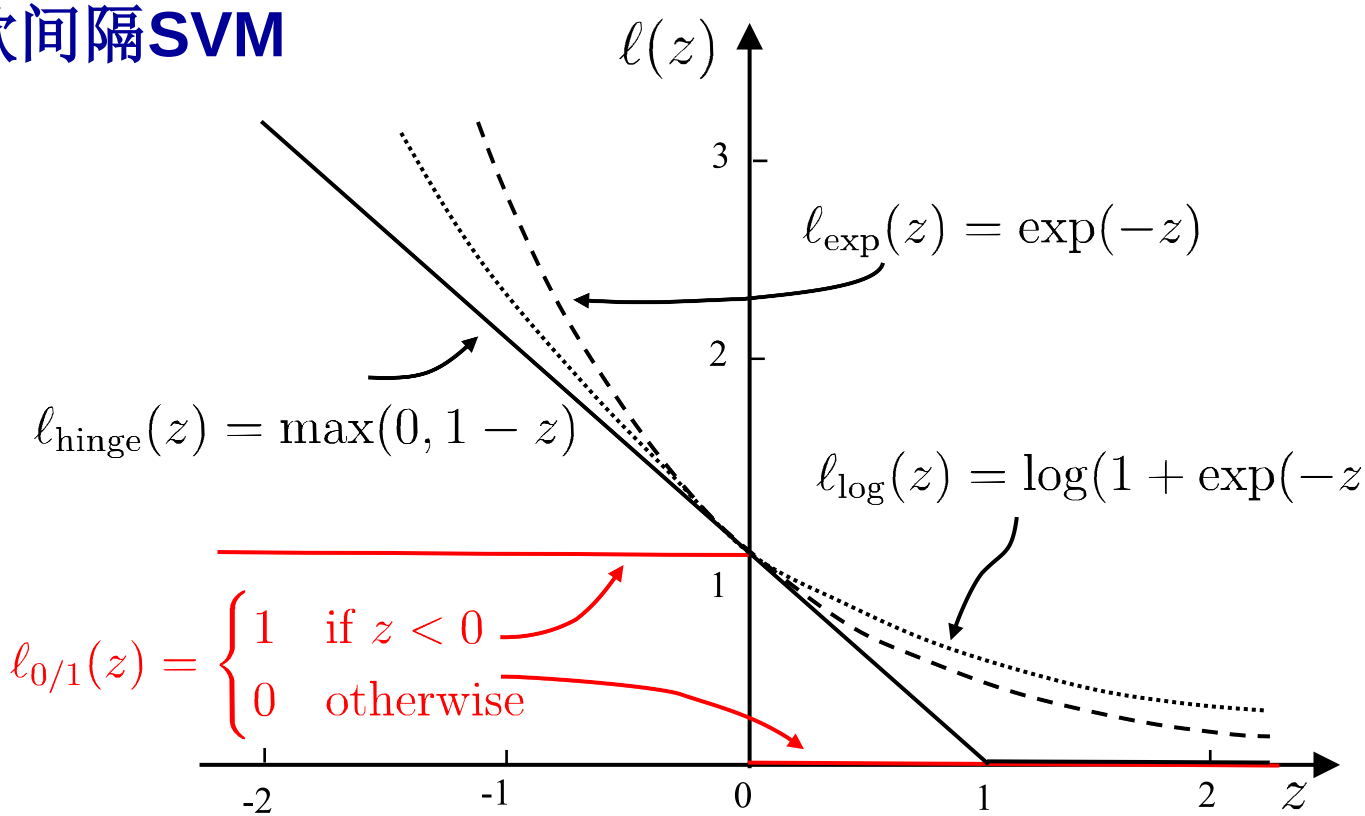
其中 $l_{0/1}$是"**0/1**损失函数"

$$l_{0/1} = \begin{cases} 1 & z < 0 \\ 0 & otherwise \end{cases}$$

存在的问题：**0/1**损失函数非凸、非连续, 不易优化 !

替代损失

软间隔SVM



“替代损失”函数数学性质较好，一般是**凸的**且是**0/1**损失函数的上界

软间隔支持向量机

采用hinge损失原始问题：

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

引入松弛变量 $\xi_i \geq 0$

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

$$s.t. \quad y_i(w^T x_i + b) \geq 1 - \xi_i$$

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

软间隔支持向量机

拉格朗日函数：

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

$L(\boldsymbol{w}, b, \boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\mu})$对 w,b,ξ_i 求偏导

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

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

$$C = \alpha_i + \mu_i \ .$$

软间隔支持向量机

对偶函数：

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \quad \sum_{i=1}^{m}\alpha_i y_i = 0, \ \boxed{0 \leq \alpha_i \leq C}, \ i = 1, 2, \ldots, m.$$

KKT条件：

$$\begin{aligned}
y_i\left(\mathbf{w}^{\mathbf{T}}\mathbf{x_i} + b\right) - 1 + \xi_i &\geq 0 \\
\xi_i &\geq 0 \\
\alpha_i &\geq 0 \\
\mu_i &\geq 0 \\
\alpha_i\left[y_i\left(\mathbf{w}^{\mathbf{T}}\mathbf{x_i} + b\right) - 1 + \xi_i\right] &= 0 \\
\mu_i\xi_i &= 0
\end{aligned}$$

$$+$$

$$\begin{aligned}
\frac{\partial L}{\partial \mathbf{w}} &= \mathbf{w} - \sum_{i=1}^{N}\alpha_i y_i \mathbf{x_i} = \mathbf{0} \\
\frac{\partial L}{\partial \xi} &= C\cdot\mathbf{1} - \alpha - \mu = \mathbf{0} \\
\frac{\partial L}{\partial b} &= \sum_{i=1}^{N}\alpha_i y_i = \alpha^{\mathbf{T}}\mathbf{y} = 0
\end{aligned}$$

软间隔支持向量机

KKT条件

$$
\begin{aligned}
y_i\left(\mathbf{w}^{\mathbf{T}}\mathbf{x_i}+b\right)-1+\xi_i &\geq 0 \\
\xi_i &\geq 0 \\
\alpha_i &\geq 0 \\
\mu_i &\geq 0 \\
\alpha_i\left[y_i\left(\mathbf{w}^{\mathbf{T}}\mathbf{x_i}+b\right)-1+\xi_i\right] &= 0 \\
\mu_i\xi_i &= 0
\end{aligned}
$$

\+

$$
\begin{aligned}
\frac{\partial L}{\partial \mathbf{w}} &= \mathbf{w}-\sum_{i=1}^{N}\alpha_i y_i \mathbf{x_i}=\mathbf{0} \\
\frac{\partial L}{\partial \xi} &= C\cdot\mathbf{1}-\alpha-\mu=\mathbf{0} \\
\frac{\partial L}{\partial b} &= \sum_{i=1}^{N}\alpha_i y_i=\alpha^{\mathbf{T}}\mathbf{y}=0
\end{aligned}
$$

分析：如果 $\alpha_i \in (0, C)$
$$
\begin{cases}
y_i f(x_i)-1+\xi_i=0 \\
C-\alpha_i-\mu_i=0 \\
\mu_i\xi_i=0
\end{cases}
\Rightarrow y_i f(x_i)=1 \text{(支持向量点）}
$$

如果 $\alpha_i=0$
$$
\begin{cases}
y_i f(x_i)-1+\xi_i\geq 0 \\
C-\alpha_i-\mu_i=0 \\
\mu_i\xi_i=0
\end{cases}
\Rightarrow y_i f(x_i)\geq 1
$$

此外 $\alpha_i=C \Leftrightarrow y_i f(x_i)\geq 1$

SMO算法

序列最小优化算法(SMO: Sequential minimal optimization)

求解如下的最优问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{m} \alpha_i$$

$$\text{s.t.} \quad \sum_{i=1}^{m} \alpha_i y_i = 0, \boxed{0 \le \alpha_i \le C}, \; i = 1, 2, \ldots, m.$$

SMO算法

基本思路：不断执行如下两个步骤直至收敛.

- 第一步：选取一对需更新的变量 α_i 和 α_j.
- 第二步：固定 α_i 和 α_j 以外的参数, 求解对偶问题更新 α_i 和 α_j.

仅考虑 α_i 和 α_j 时, 对偶问题的约束变为

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

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

SMO算法

仅考虑 α_i 和 α_j 时，原问题可以变成只含2个变量的优化问题

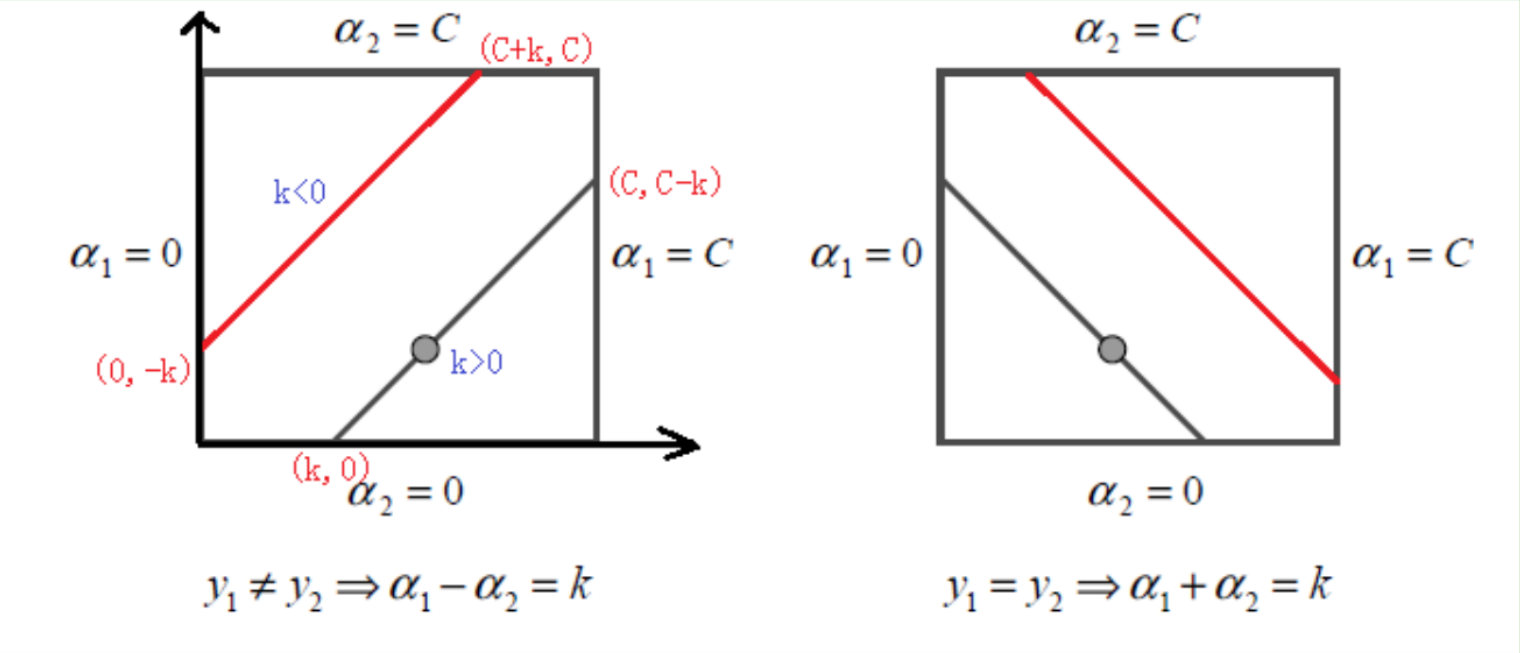
$$\min_{\alpha_1,\alpha_1} \frac{1}{2}K_{11}\alpha_1^2 + \frac{1}{2}K_{22}\alpha_2^2 + y_1y_2K_{12}\alpha_1\alpha_2 - (\alpha_1 + \alpha_2)$$
$$+y_1\alpha_1\sum_{i=3}^{m} y_i\alpha_iK_{i1} + y_2\alpha_2\sum_{i=3}^{m} y_i\alpha_iK_{i2}$$

$$s.t.\ \ \alpha_1y_1 + \alpha_2y_2 = -\sum_{i=3}^{m} y_i\alpha_i = \varsigma,\ \ 0 \le \alpha_i \le C$$

其中 $K_{ij} = K(x_i, x_j)$

SMO算法

根据上面的约束条件$\alpha_1 y_1 + \alpha_2 y_2 = \varsigma \quad 0 \le \alpha_i \le C \quad i = 1, 2$，又由于$y_1, y_2$均只能取值1或者-1，这样$\alpha_1, \alpha_2$在[0,C]和[0,C]形成的盒子里面，并且两者的关系直线的斜率只能为1或者-1，也就是说α_1, α_2的关系直线平行于[0,C]和[0,C]形成的盒子的对角线，如下图所示：



由于α_1, α_2的关系被限制在盒子里的一条线段上，所以两变量的优化问题实际上仅仅是一个变量的优化问题。不妨我们假设最终是α_2的优化问题。由于我们采用的是启发式的迭代法，假设我们上一轮迭代得到的解是$\alpha_1^{old}, \alpha_2^{old}$，假设沿着约束方向$\alpha_2$未经剪辑的解是$\alpha_2^{new,unc}$.本轮迭代完成后的解为$\alpha_1^{new}, \alpha_2^{new}$

SMO算法

由于α_2^{new}必须满足上图中的线段约束。假设L和H分别是上图中α_2^{new}所在的线段的边界。那么很显然我们有：

$$L \leq \alpha_2^{new} \leq H$$

而对于L和H，我们也有限制条件如果是上面左图中的情况，则

$$L = max(0, \alpha_2^{old} - \alpha_1^{old}) \quad H = min(C, C + \alpha_2^{old} - \alpha_1^{old})$$

如果是上面右图中的情况，我们有：

$$L = max(0, \alpha_2^{old} + \alpha_1^{old} - C) \quad H = min(C, \alpha_2^{old} + \alpha_1^{old})$$

也就是说，假如我们通过求导得到的$\alpha_2^{new,unc}$，则最终的α_2^{new}应该为：

$$\alpha_2^{new} = \begin{cases} H & \alpha_2^{new,unc} > H \\ \alpha_2^{new,unc} & L \leq \alpha_2^{new,unc} \leq H \\ L & \alpha_2^{new,unc} < L \end{cases}$$

SMO算法

仅考虑 α_2 的二次规划问题，导数为0

$$E_i = g(x_i) - y_i = \sum_{j=1}^{m} \alpha_j^* y_j K(x_i, x_j) + b - y_i$$

定理 7.6 最优化问题沿着约束方向未经剪辑时的解是

$$\alpha_2^{\text{new,unc}} = \alpha_2^{\text{old}} + \frac{y_2(E_1 - E_2)}{\eta}$$

其中，

$$\eta = K_{11} + K_{22} - 2K_{12} = \|\Phi(x_1) - \Phi(x_2)\|^2$$

$\Phi(x)$ 是输入空间到特征空间的映射，

经剪辑后 α_2 的解是

$$\alpha_2^{\text{new}} = \begin{cases} H, & \alpha_2^{\text{new,unc}} > H \\ \alpha_2^{\text{new,unc}}, & L \leqslant \alpha_2^{\text{new,unc}} \leqslant H \\ L, & \alpha_2^{\text{new,unc}} < L \end{cases}$$

由 α_2^{new} 求得 α_1^{new} 是

$$\alpha_1^{\text{new}} = \alpha_1^{\text{old}} + y_1 y_2 (\alpha_2^{\text{old}} - \alpha_2^{\text{new}})$$

SMO-变量的选择方法

第1个变量的选择（外循环实现）：选违反**KKT**最严重的样本点，并将其对应的变量作为第一个变量

$$y_i g(x_i) = \begin{cases} \geq 1, \Leftrightarrow \alpha_i = 0; \\ = 1, \Leftrightarrow \alpha_i \in (0, C); \\ \leq 1, \Leftrightarrow \alpha_i = C \end{cases}$$

$$g(x_i) = \sum_{j=1}^{m} \alpha_j y_j K(x_i, x_j) + b$$

首先找违反 $\alpha_i \in (0, C) \Rightarrow y_i g(x_i) \leq 1$ 的点，如果都满足，再找其它情况 $\alpha_i = 0, \alpha_i = C$

SMO-变量的选择方法

第2个变量α_2的选择（内循环实现）：希望能使 $|E_1 - E_2|$ 有足够多的变化

由于α_1定了的时候, E_1也确定了，所以要想 $|E_1 - E_2|$最大

- 只需要在 E_1为正时，选择最小的 E_i 作为 E_2，
- 在 E_1为负时，选择最大的E_i 作为E_2，可以将所有的 E_i 保存下来加快迭代

如果内存循环找到的点不能让目标函数有足够的下降， 可以采用遍历支持向量点来做α_2, 直到目标函数有足够的下降， 如果所有的支持向量做 α_2 都不能让目标函数有足够的下降，可以跳出循环，重新选择α_1

SMO-变量的选择方法

第2个变量α_2的选择（内循环实现）：希望能使 $|E_1 - E_2|$ 有足够多的变化

由于α_1定了的时候, E_1也确定了，所以要想 $|E_1 - E_2|$最大

- 只需要在 E_1为正时，选择最小的 E_i 作为 E_2，
- 在 E_1为负时，选择最大的E_i 作为E_2，可以将所有的 E_i 保存下来加快迭代

如果内存循环找到的点不能让目标函数有足够的下降， 可以采用遍历支持向量点来做α_2, 直到目标函数有足够的下降， 如果所有的支持向量做 α_2 都不能让目标函数有足够的下降，可以跳出循环，重新选择α_1

SMO-计算偏移项b和偏差E_i

在每次完成两个变量的优化之后，需要重新计算阈值b。当$0 < \alpha_1^{new} < C$时，我们有

$$y_1 - \sum_{i=1}^{m} \alpha_i y_i K_{i1} - b_1 = 0$$

于是新的b_1^{new}为：

$$b_1^{new} = y_1 - \sum_{i=3}^{m} \alpha_i y_i K_{i1} - \alpha_1^{new} y_1 K_{11} - \alpha_2^{new} y_2 K_{21}$$

计算出E_1为：

$$E_1 = g(x_1) - y_1 = \sum_{i=3}^{m} \alpha_i y_i K_{i1} + \alpha_1^{old} y_1 K_{11} + \alpha_2^{old} y_2 K_{21} + b^{old} - y_1$$

可以看到上两式都有$y_1 - \sum_{i=3}^{m} \alpha_i y_i K_{i1}$，因此可以将$b_1^{new}$用$E_1$表示为：

$$b_1^{new} = -E_1 - y_1 K_{11}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{21}(\alpha_2^{new} - \alpha_2^{old}) + b^{old}$$

SMO-计算偏移项b和偏差E_i

同样的，如果$0 < \alpha_2^{new} < C$, 那么有：

$$b_2^{new} = -E_2 - y_1 K_{12}(\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{22}(\alpha_2^{new} - \alpha_2^{old}) + b^{old}$$

最终的b^{new}为：

$$b^{new} = \frac{b_1^{new} + b_2^{new}}{2}$$

得到了b^{new}我们需要更新E_i：

$$E_i = \sum_S y_j \alpha_j K(x_i, x_j) + b^{new} - y_i$$

其中，S是所有支持向量x_j的集合。

SMO-伪代码

输入是m个样本$(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$
y为二元输出，值为1，或者-1.精度e。

输出是近似解α

1)取初值$\alpha^0 = 0, k = 0$

2)选择α_1^k，α_2^k，求出新的$\alpha_2^{new,unc}$。

$$\alpha_2^{new,unc} = \alpha_2^k + \frac{y_2(E_1 - E_2)}{K_{11} + K_{22} - 2K_{12})}$$

3)按照下式求出α_2^{k+1}

$$\alpha_2^{k+1} = \begin{cases} H & \alpha_2^{new,unc} > H \\ \alpha_2^{new,unc} & L \leq \alpha_2^{new,unc} \leq H \\ L & \alpha_2^{new,unc} < L \end{cases}$$

4)利用α_2^{k+1}和α_1^{k+1}的关系求出α_1^{k+1}

5)计算b^{k+1}和E_i

6）在精度e范围内检查是否满足如下的终止条件：

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

$$0 \leq \alpha_i \leq C, i = 1, 2...m$$

$$\alpha_i^{k+1} = 0 \Rightarrow y_i g(x_i) \geq 1$$

$$0 < \alpha_i^{k+1} < C \Rightarrow y_i g(x_i) = 1$$

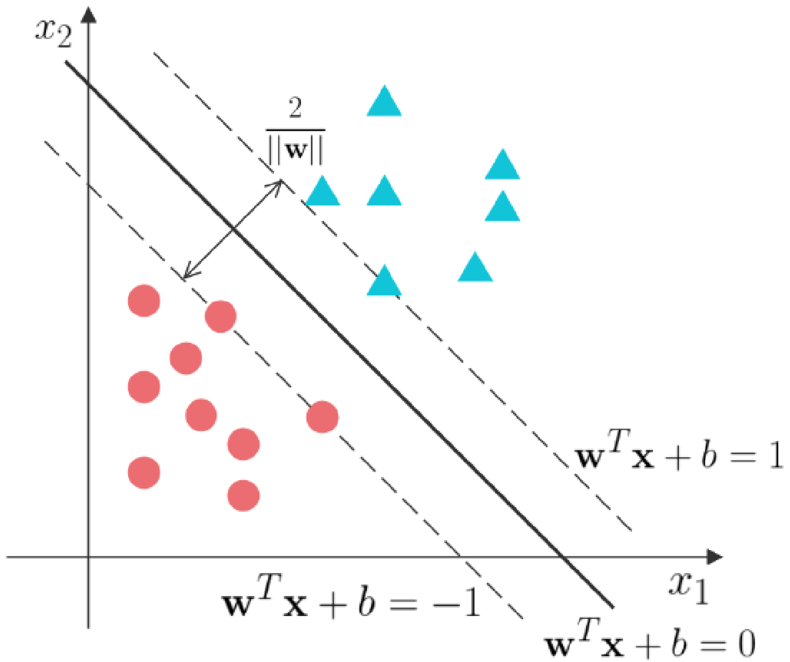
$$\alpha_i^{k+1} = C \Rightarrow y_i g(x_i) \leq 1$$

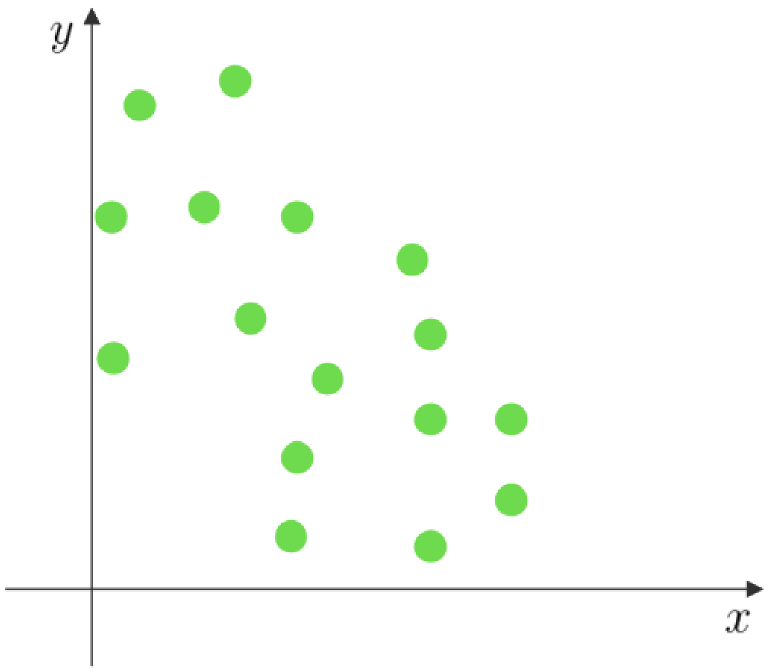
7)如果满足则结束，返回α^{k+1},否则转到步骤2）。

章节目录

- 间隔与支持向量
- 对偶问题
- 核函数
- 软间隔与正则化
- **支持向量回归**
- 核方法

支持向量回归



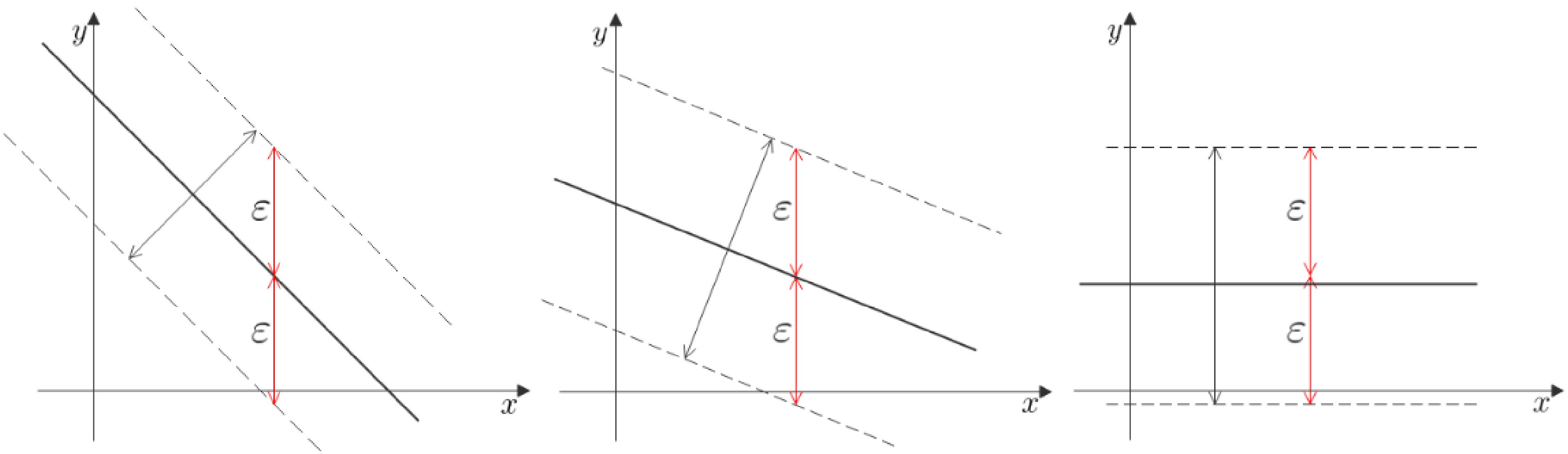


SVM使靠超平面最近的样本点之间的间隔最大

间隔最大？满足误差约束

支持向量回归

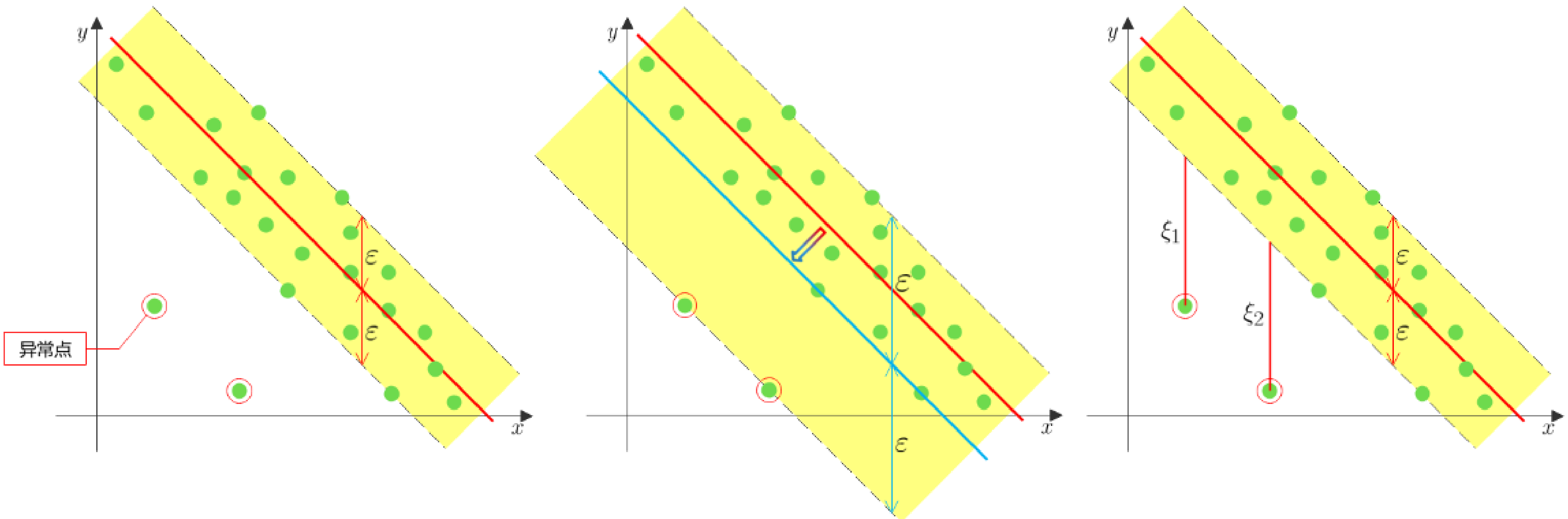
满足误差约束下间隔最大： $|y_i-(\mathbf{w}^T\mathbf{x}_i+b)|\le\varepsilon,\quad i=1,2,\cdots,N$



SVR的目的是：保证所有样本点在ε管道内的前提下，回归超平面$f(\mathbf{x})$尽可能地平。

支持向量回归

ε设置太小无法保证所有样本点都在ε管道内，ε太大回归超平面会被一些异常点带偏。



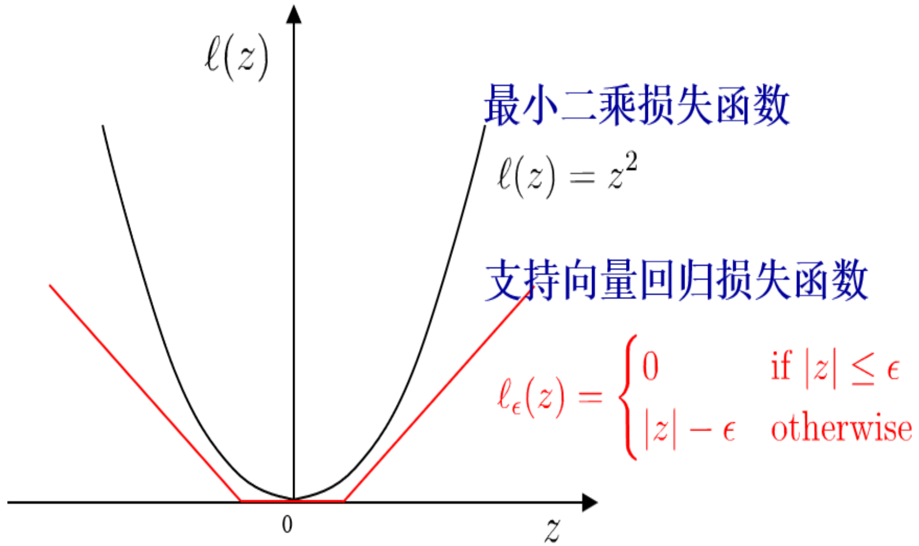
ε设得太小　　　ε设大一点呢？　　　添加松弛变量

损失函数

不仅需要最大化间隔带，
还有最小化损失

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} l_{\varepsilon}(f(x_i) - y_i)$$

允许间隔带外存在点，但
损失应尽可能的小



落入中间 2ϵ 间隔带的样本
不计算损失, 从而使得模型
获得稀疏性.

引入松弛变量SVR问题

原始问题

$$\begin{aligned}\min_{\boldsymbol{w},b,\xi_i,\hat{\xi}_i} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}(\xi_i + \hat{\xi}_i) \\ \text{s.t.} \quad & y_i - \boldsymbol{w}^\top \phi(\boldsymbol{x}_i) - b \leq \epsilon + \xi_i, \\ & y_i - \boldsymbol{w}^\top \phi(\boldsymbol{x}_i) - b \geq -\epsilon - \hat{\xi}_i, \\ & \xi_i \geq 0,\ \hat{\xi}_i \geq 0,\ i = 1, 2, \ldots, m.\end{aligned}$$

对偶问题

$$\begin{aligned}\min_{\boldsymbol{\alpha},\hat{\boldsymbol{\alpha}}} \quad & \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j)\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_{i=1}^{m}(\alpha_i(\epsilon - y_i) + \hat{\alpha}_i(\epsilon + y_i)) \\ \text{s.t.} \quad & \sum_{i=1}^{m}(\alpha_i - \hat{\alpha}_i) = 0, \\ & 0 \leq \alpha_i \leq C,\ 0 \leq \hat{\alpha}_i \leq C.\end{aligned}$$

预测

$$f(\boldsymbol{x}) = \boldsymbol{w}^\top \phi(\boldsymbol{x}) + b = \sum_{i=1}^{m}(\hat{\alpha}_i - \alpha_i) y_i \kappa(\boldsymbol{x}_i, \boldsymbol{x}) + b$$

章节目录

- 间隔与支持向量
- 对偶问题
- 核函数
- 软间隔与正则化
- 支持向量回归
- **核方法**

表示定理

支持向量机 $$f(\boldsymbol{x}) = \boldsymbol{w}^\top \phi(\boldsymbol{x}) + b = \sum_{i=1}^{m} \alpha_i y_i \kappa(\boldsymbol{x}_i, \boldsymbol{x}) + b$$

支持向量回归 $$f(\boldsymbol{x}) = \boldsymbol{w}^\top \phi(\boldsymbol{x}) + b = \sum_{i=1}^{m} (\hat{\alpha}_i - \alpha_i) y_i \kappa(\boldsymbol{x}_i, \boldsymbol{x}) + b$$

结论: 无论是支持向量机还是支持向量回归, 学得的模型总可以表示成**核函数的线性组合**.

更一般的结论(表示定理): 对于任意单调增函数Ω和任意非负损失函数 l, 优化问题

$$\min_{h \in \mathbb{H}} \quad F(h) = \Omega(\|h\|_{\mathbb{H}}) + l(h(\boldsymbol{x}_1), \ldots, h(\boldsymbol{x}_m))$$

的解总可以写为 $$h^* = \sum_{i=1}^{m} \alpha_i \kappa(\cdot, \boldsymbol{x}_i)$$

核线性判别分析

通过表示定理可以得到很多线性模型的"核化"版本

- 核SVM
- 核LDA
- 核PCA
- ……

核LDA：先将样本映射到高维特征空间，然后在此特征空间中做线性判别分析

$$\max_{\boldsymbol{w}} J(\boldsymbol{w}) = \frac{\boldsymbol{w}^\top \mathbf{S}_b^\phi \boldsymbol{w}}{\boldsymbol{w}^\top \mathbf{S}_w^\phi \boldsymbol{w}}$$

$$h(\boldsymbol{x}) = \boldsymbol{w}^\top \phi(\boldsymbol{x}) = \sum_{i=1}^{m} \alpha_i \kappa(\boldsymbol{x}_i, \boldsymbol{x})$$

$$\max_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}) = \frac{\boldsymbol{\alpha}^\top \mathbf{M} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^\top \mathbf{N} \boldsymbol{\alpha}}$$

总结

支持向量机的”最大间隔”思想

对偶问题及其解的稀疏性

通过向高维空间映射解决线性不可分的问题

引入”软间隔”缓解特征空间中线性不可分的问题

将支持向量的思想应用到回归问题上得到支持向量回归

将核方法推广到其他学习模型

成熟的SVM软件包

LIBSVM
http://www.csie.ntu.edu.tw/~cjlin/libsvm/

LIBLINEAR
http://www.csie.ntu.edu.tw/~cjlin/liblinear/

SVM^{light}、SVM^{perf}、SVM^{struct}
http://svmlight.joachims.org/svm_struct.html

Pegasos
http://www.cs.huji.ac.il/~shais/code/index.html