第六章：支持向量机

张自力

大纲

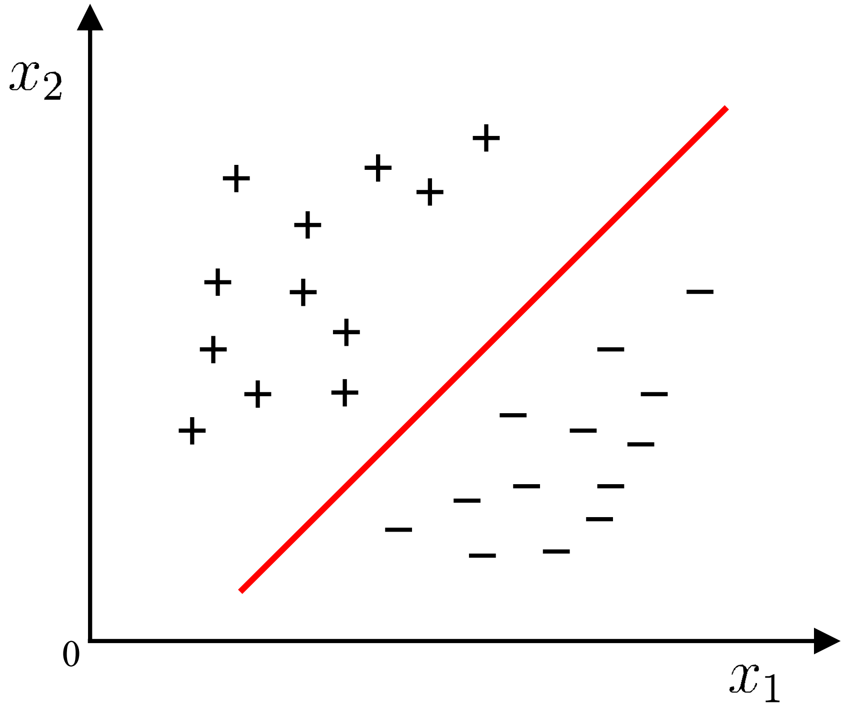
- 6.1 间隔与支持向量
- 6.2 对偶问题
- 6.3 核函数
- 6.4 软间隔与正则化
- 6.5 支持向量回归
- 6.6 核方法

大纲

- 6.1 间隔与支持向量
- 6.2 对偶问题
- 6.3 核函数
- 6.4 软间隔与正则化
- 6.5 支持向量回归
- 6.6 核方法

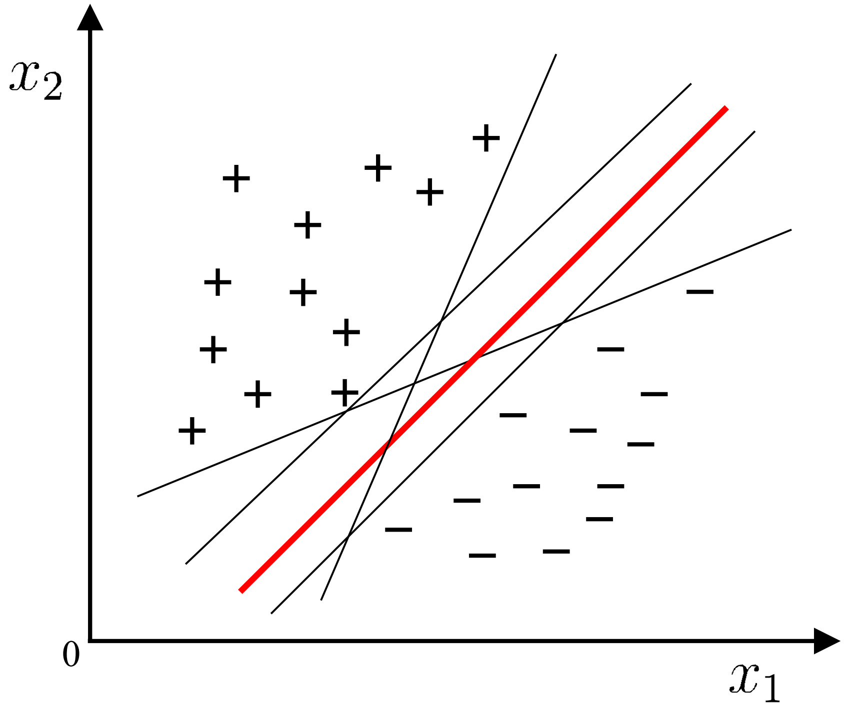
引子

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.



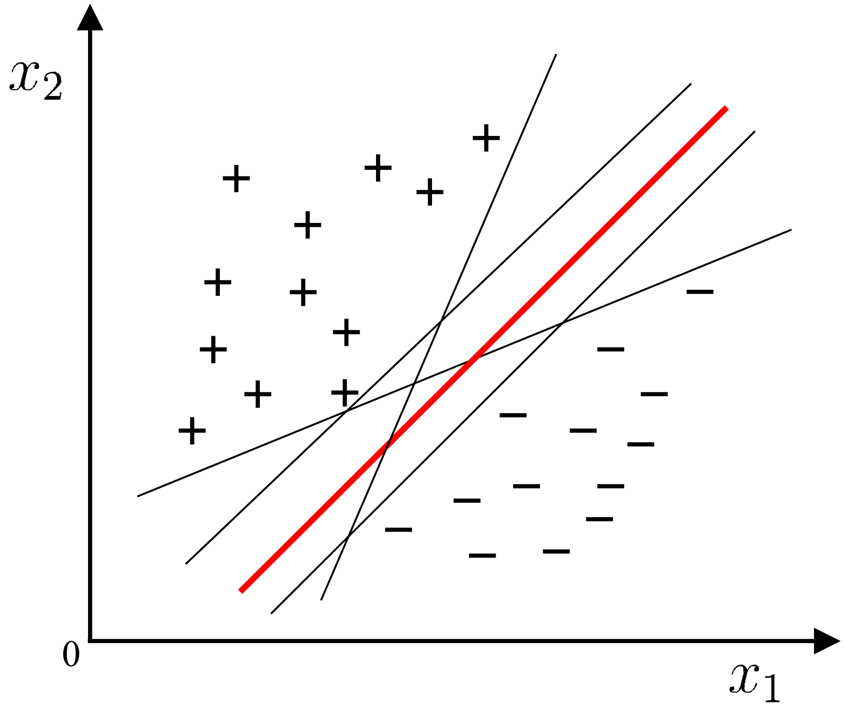
引子

-Q:将训练样本分开的超平面可能有很多, 哪一个好呢?



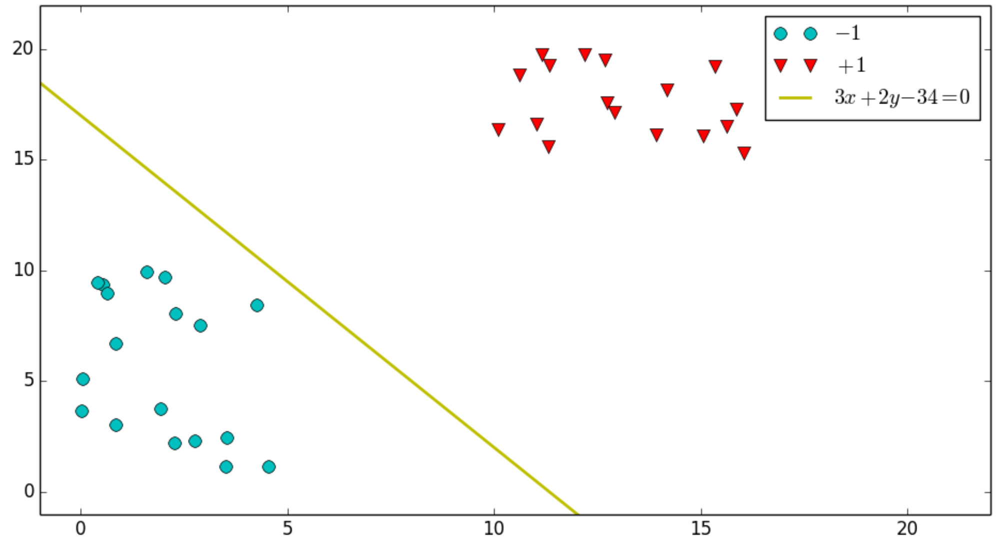
引子

-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.

6.1 间隔与支持向量



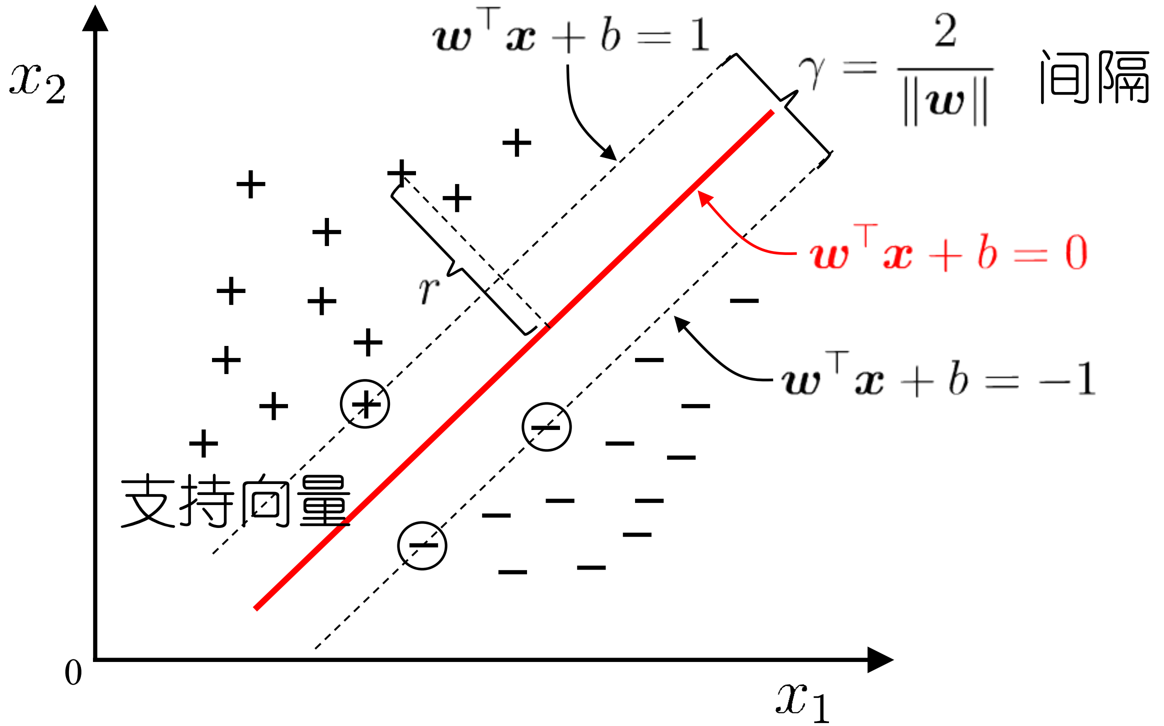
已知超平面$w_1x_1 + w_2x_2 + \dots + w_Nx_N + b = 0$ 和点x，其中$x = (x_1, x_2, \dots, x_N)$，则这个点到这个超平面的距离为：

$$\frac{|\omega_1 x_1 + \omega_2 x_2 + \dots + \omega_N x_N + b|}{\sqrt{\omega_1^2 + \omega_2^2 + \dots + \omega_N^2}}$$

可以向量表示为 $\frac{|\omega^T x + b|}{\|\omega\|}$ 。

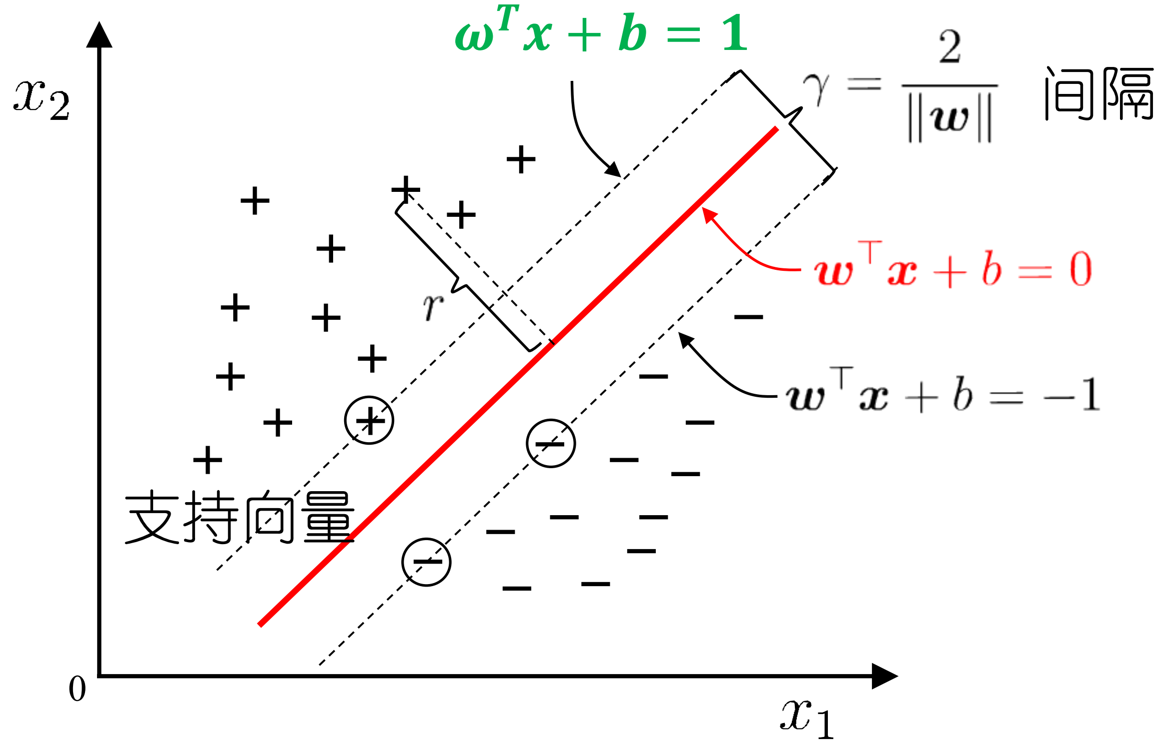
6.1 间隔与支持向量

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



线性可分情况下的线性分类器，这是最原始的SVM，最核心的思想就是最大的分类间隔。

6.1 间隔与支持向量



当 $w^Tx + b >= 1$时，$y = 1$，正样本;

当 $w^Tx + b <= -1$时，$y = -1$，负样本；

将两个式子统一起来，$y(w^Tx+b) >= 1$。

6.1 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.

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

$$\Downarrow$$

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

大纲

- 6.1 间隔与支持向量
- 6.2 对偶问题
- 6.3 核函数
- 6.4 软间隔与正则化
- 6.5 支持向量回归
- 6.6 核方法

6.2 对偶问题

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

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

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

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

  - 第三步：回代

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^\top \boldsymbol{x}_j - \sum_{i=1}^{m} \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^{m} \alpha_i y_i = 0, \ \alpha_i \geq 0, \ i = 1, 2, \ldots, m. \end{aligned}$$

6.2 对偶问题

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

那么问题就变成了：

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

所谓的对偶问题就是：

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

6.2 对偶问题

- 第二步 -> 第三步

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

$$L(w,b,a) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} a_i (y_i (w^T \bullet x_i + b) - 1)$$

$$= \frac{1}{2} w^T w - w^T \sum_{i=1}^{n} a_i y_i x_i - b \sum_{i=1}^{n} a_i y_i + \sum_{i=1}^{n} a_i$$

$$= \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} a_i a_j y_i y_j x_i x_j - \sum_{j=1}^{n} \sum_{i=1}^{n} a_i a_j y_i y_j x_i x_j + \sum_{i=1}^{n} a_i$$

$$= \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} a_i a_j y_i y_j x_i x_j$$

6.2 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：

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

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

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.

6.2 求解方法 - SMO

- 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 .

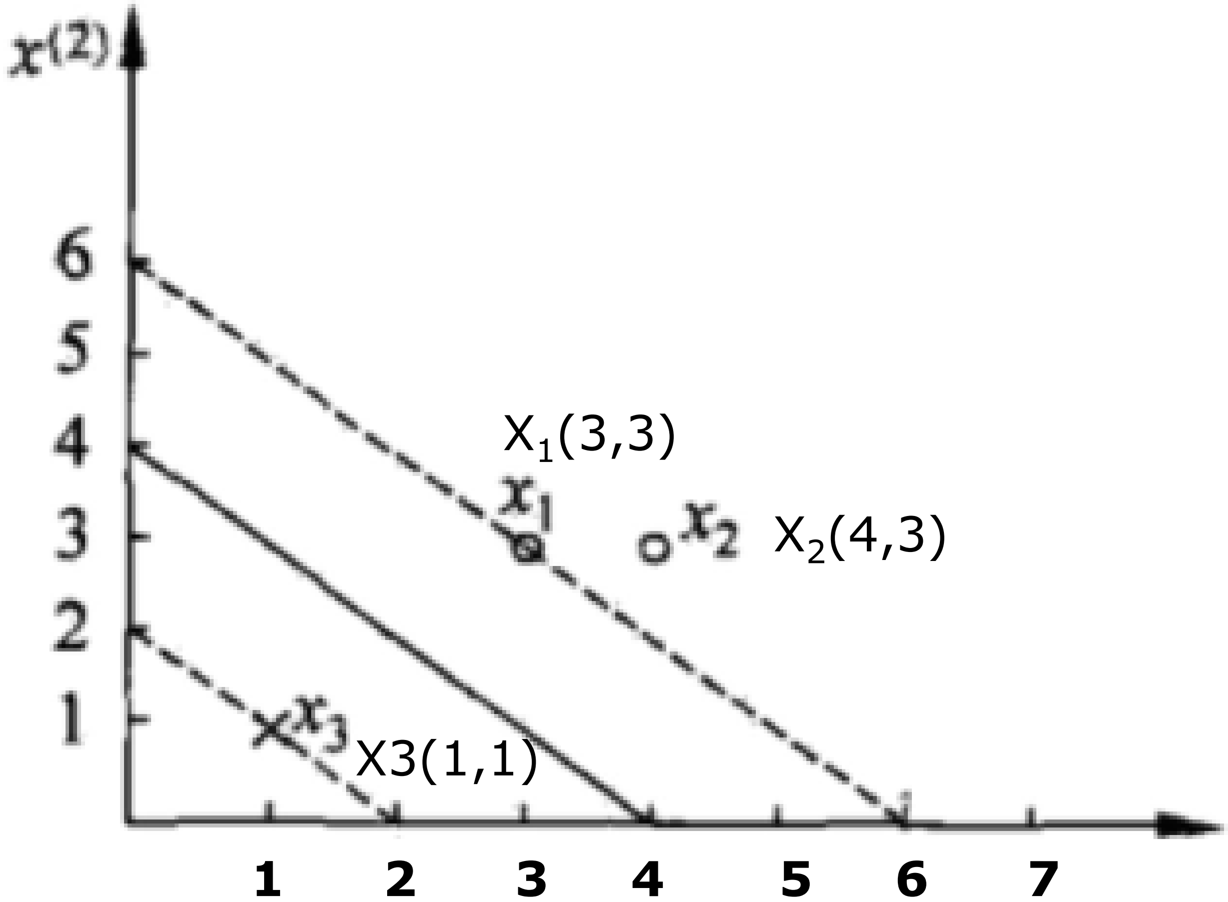
- 仅考虑 α_i 和 α_j 时，对偶问题的约束变为

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

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.

- 偏移项 b：通过支持向量来确定.

6.2 实例



$y_1=y_2=1$，$y_3=-1$

6.2 实例

$X_1(3,3)$, $x_2(4,3)$, $x_3(1,1)$, $y_1 = y_2 = 1$, $y_3 = -1$.

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

i=1, j = 1: $\alpha_1\alpha_1 y_1 y_1 x_1^T x_1 = \alpha_1\alpha_1(1)(1)(3;3)^T(3;3)$

$= \alpha_1\alpha_1(9+9) = 18\alpha_1^2$

j = 2: $\alpha_1\alpha_2 y_1 y_2 x_1^T x_2 = \alpha_1\alpha_2(1)(1)(3;3)^T(4;3)$

$= \alpha_1\alpha_2(12+9) = 21\alpha_1\alpha_2$

j = 3: $\alpha_1\alpha_3 y_1 y_3 x_1^T x_3 = \alpha_1\alpha_3(1)(-1)(3;3)^T(1;1)$

$= \alpha_1\alpha_3(-1)(3+3) = -6\alpha_1\alpha_3$

6.2 实例

$X_1(3,3)$, $x_2(4,3)$, $x_3(1,1)$, $y_1 = y_2 = 1$, $y_3 = -1$.

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

i=2, j = 1: $$\alpha_2\alpha_1 y_2 y_1 x_2^T x_1 = \alpha_2\alpha_1(1)(1)(4;3)^T(3;3) = \alpha_1\alpha_2(12+9) = 21\alpha_1\alpha_2$$

j = 2: $$\alpha_2\alpha_2 y_2 y_2 x_2^T x_2 = \alpha_2\alpha_2(4;3)^T(4;3) = \alpha_2\alpha_2(16+9) = 25\alpha_2^2$$

j = 3: $$\alpha_2\alpha_3 y_2 y_3 x_2^T x_3 = \alpha_2\alpha_3(1)(-1)(4;3)^T(1;1) = \alpha_2\alpha_3(-1)(4+3) = -7\alpha_2\alpha_3$$

6.2 实例

$X_1(3,3)$, $x_2(4,3)$, $x_3(1,1)$, $y_1 = y_2 = 1$, $y_3 = -1$.

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

i=3, j = 1: $\alpha_3\alpha_1 y_3 y_1 x_3^T x_1 = \alpha_3\alpha_1(-1)(1)(1;1)^T(3;3)$
$= \alpha_3\alpha_1(-1)(3+3) = -6\alpha_1\alpha_3$

j = 2: $\alpha_3\alpha_2 y_3 y_2 x_3^T x_2 = \alpha_3\alpha_2(-1)(1;1)^T(4;3)$
$= \alpha_2\alpha_3(-1)(4+3) = -7\alpha_2\alpha_3$

j = 3: $\alpha_3\alpha_3 y_3 y_3 x_3^T x_3 = \alpha_3\alpha_3(-1)(-1)(1;1)^T(1;1)$
$= \alpha_3\alpha_3(1+1) = 2\alpha_3^2$

6.2 实例

$X_1(3,3)$, $x_2(4,3)$, $x_3(1,1)$, $y_1 = y_2 = 1$, $y_3 = -1$.

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

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

$$s.t.\ \ \alpha_1 + \alpha_2 - \alpha_3 = 0$$

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

6.2 实例

将 $\alpha_1+\alpha_2=\alpha_3$ 代入下式

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

化简之后，得到如下等式：

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

6.2 实例

对S求偏导得到如下等式

$$\frac{\partial S(\alpha_1,\alpha_2)}{\partial \alpha_1} = 8\alpha_1 + 10\alpha_2 - 2 = 0$$

$$\frac{\partial S(\alpha_1,\alpha_2)}{\partial \alpha_2} = 13\alpha_2 + 10\alpha_1 - 2 = 0$$

联立上述2个方程组，解得

$$\alpha_1 = 1.5$$
$$\alpha_2 = -1$$

6.2 实例

- 而a_2=-1的点不满于a_2>0的条件，所以最小值在边界上取得。
- 边界情况要么是a_1=0，要么是a_2=0，
- 当a_1=0时，把a_1的值往s对a_2的偏导里面带入计算得到a_2=2/13 (满足条件)。

S(0, 2/13) = -2/13 = -0.1538。

- 当a_2=0时，把a_2的值往s对a_1的偏导里面带入计算得到a_1=1/4 （满足条件）。

S(1/4, 0) = -1/4 = -0.25

显然后面的结果更小，所以:a_1 = 1/4; a_2 = 0; a_3 = 1/4

6.2 实例

- 到这里就能验证上面的结论了，a_1和a_3是x_1和x_3的系数，x_1和x_3是支持向量，而x_2不是，所以前面的系数是0。

- 因为根据w求解公式，$a_2=0$，所以x_2对w权的最后取值没有影响，所以x_2不是支持向量。

- 最后一步，带到上面的式子中求w，b：

$$\omega=\sum_{i=1}^{n}\alpha_i x_i y_i=\frac{1}{4}\times(3,3)\times 1+\frac{1}{4}\times(1,1)\times(-1)$$

$$=\frac{1}{4}\times(2,2)=(\frac{1}{2},\frac{1}{2})$$

6.2 实例

$$\omega=\sum_{i=1}^{n}\alpha_i x_i y_i=\frac{1}{4}\times(3,3)\times1+\frac{1}{4}\times(1,1)\times(-1)$$
$$=\frac{1}{4}\times(2,2)=(\frac{1}{2},\frac{1}{2})$$

得到$w_1=w_2=0.5$。
对于支持向量x_1,计算b的值：

$$b=y_i-\sum_{i=1}^{n}\alpha_i y_i(\Phi(x_i)\Phi(x_j))=1-(\alpha_1 y_1 {x_1}^T x_1+\alpha_3 y_3 {x_3}^T x_1)$$

$$=1-(\frac{1}{4}\times1\times{x_1}^T x_1+\frac{1}{4}\times(-1)\times{x_3}^T x_1)$$

$$=1-\frac{1}{4}(18-6)=-2$$

6.2 实例

$$\omega = \sum_{i=1}^{n} \alpha_i x_i y_i = \frac{1}{4} \times (3,3) \times 1 + \frac{1}{4} \times (1,1) \times (-1)$$

$$= \frac{1}{4} \times (2,2) = (\frac{1}{2}, \frac{1}{2})$$

对于非支持向量x2,计算b的值：

$$b = y_i - \sum_{i=1}^{n} \alpha_i y_i (\Phi(x_i)\Phi(x_j)) = 1 - (\alpha_1 y_1 {x_1}^T x_2 + \alpha_3 y_3 {x_3}^T x_2)$$

$$= 1 - (\frac{1}{4} \times 1 \times {x_1}^T x_2 + \frac{1}{4} \times (-1) \times {x_3}^T x_2)$$

$$= 1 - \frac{1}{4}(21 - 7) = -5/2$$

于是得到最后的决策边界为：$0.5*x_1+0.5*x_2-2=0$

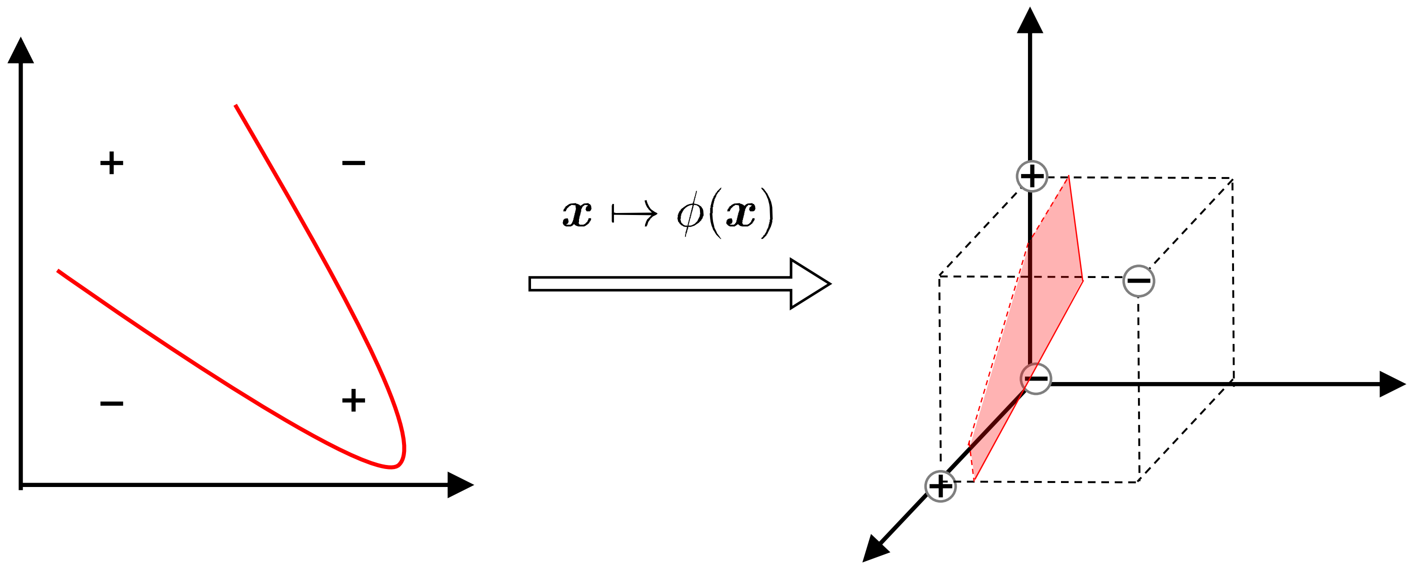
大纲

- 6.1 间隔与支持向量
- 6.2 对偶问题
- 6.3 核函数
- 6.4 软间隔与正则化
- 6.5 支持向量回归
- 6.6 核方法

6.3 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.



6.3 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

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

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

对偶问题

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boxed{\phi(\boldsymbol{x}_i)^\top \phi(\boldsymbol{x}_j)} - \sum_{i=1}^{m} \alpha_i$$

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

只以内积的形式出现

预测

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

6.3 核函数

- 基本想法：不显式地设计核映射，而是设计核函数.

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

- Mercer定理(充分非必要)：只要一个对称函数所对应的核矩阵半正定，则它就能作为核函数来使用.

- 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	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\Vert\boldsymbol{x}_i - \boldsymbol{x}_j\Vert^2}{2\delta^2}\right)$	$\delta > 0$为高斯核的带宽(width)
拉普拉斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\Vert\boldsymbol{x}_i - \boldsymbol{x}_j\Vert}{\delta}\right)$	$\delta > 0$
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，$\beta > 0$, $\theta < 0$

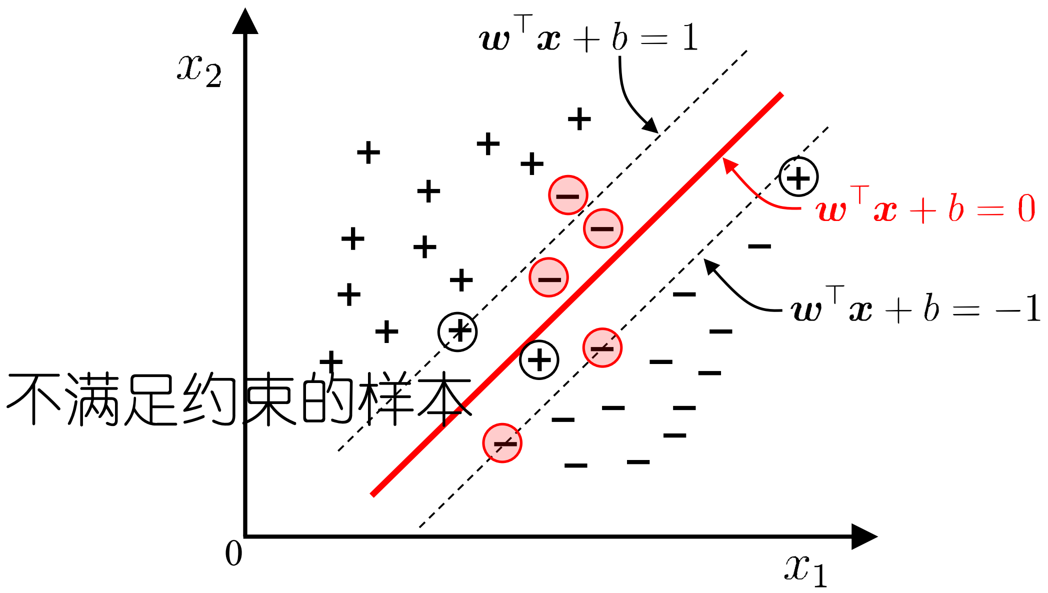
大纲

- 6.1 间隔与支持向量
- 6.2 对偶问题
- 6.3 核函数
- 6.4 软间隔与正则化
- 6.5 支持向量回归
- 6.6 核方法

6.4 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.



线性不可分情况下的线性分类器。

6.4 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.

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

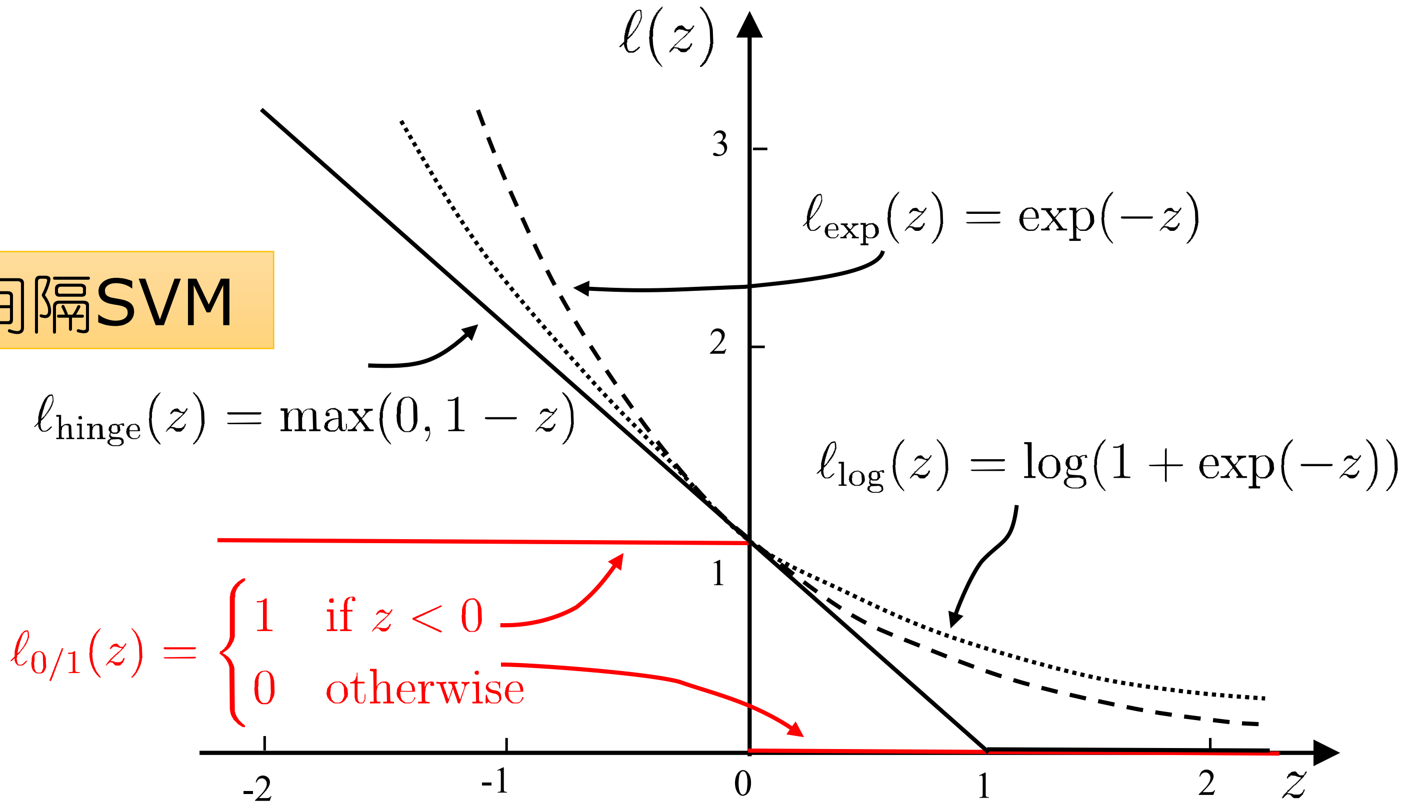
其中 $l_{0/1}$ 是"0/1损失函数"

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

- 存在的问题：0/1损失函数非凸、非连续，不易优化！

6.4 替代损失

软间隔SVM



替代损失函数数学性质较好，一般是0/1损失函数的上界

6.4 软间隔支持向量机

原始问题

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

对偶问题

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

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

根据KKT条件可推得最终模型仅与支持向量有关，也即hinge损失函数依然保持了支持向量机解的稀疏性.

6.4 正则化

- 支持向量机学习模型的更一般形式

$$\min_f \; \Omega(f) + C\sum_{i=1}^{m} l(f(\boldsymbol{x}_i), y_i)$$

结构风险，描述模型的某些性质

经验风险，描述模型与训练数据的契合程度

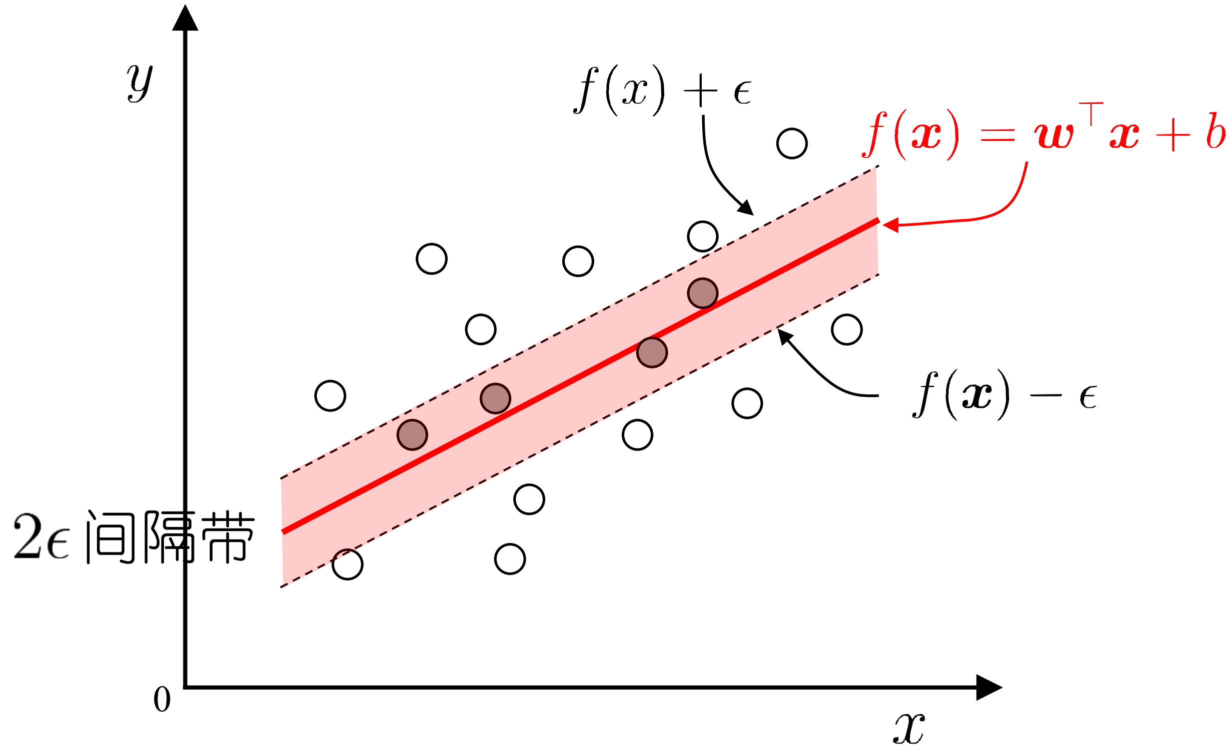
- 通过替换上面两个部分，可以得到许多其他学习模型
  - 对数几率回归(Logistic Regression)
  - 最小绝对收缩选择算子(LASSO)
  - ……

大纲

- 6.1 间隔与支持向量
- 6.2 对偶问题
- 6.3 核函数
- 6.4 软间隔与正则化
- 6.5 支持向量回归
- 6.6 核方法

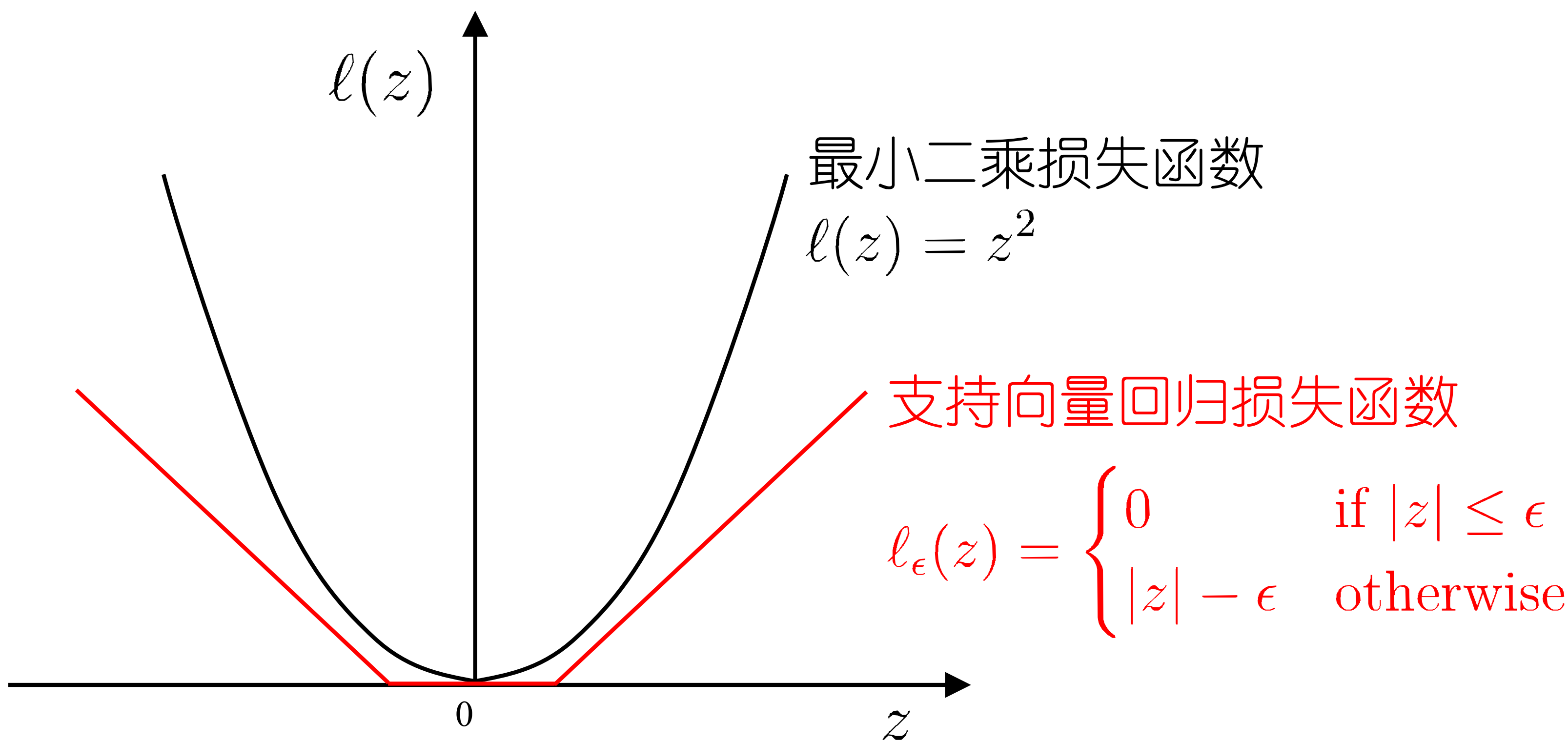
6.5 支持向量回归

特点：允许模型输出和实际输出间存在 2ϵ 的偏差.



6.5 损失函数

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.



6.5 形式化

原始问题

$$\min_{\boldsymbol{w},b,\xi_i,\hat{\xi}_i} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}(\xi_i + \hat{\xi}_i)$$

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

对偶问题

$$\min_{\boldsymbol{\alpha},\hat{\boldsymbol{\alpha}}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j)\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j) + \sum_{i=1}^{m}(\alpha_i(\epsilon - y_i) + \hat{\alpha}_i(\epsilon + y_i))$$

$$\begin{aligned}\text{s.t.}\quad & \sum_{i=1}^{m}(\alpha_i - \hat{\alpha}_i) = 0,\\ & 0 \le \alpha_i \le C,\ 0 \le \hat{\alpha}_i \le C.\end{aligned}$$

预测

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

大纲

- 6.1 间隔与支持向量
- 6.2 对偶问题
- 6.3 核函数
- 6.4 软间隔与正则化
- 6.5 支持向量回归
- 6.6 核方法

6.6 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 $h^* = \sum_{i=1}^{m} \alpha_i \kappa(\cdot, \boldsymbol{x}_i)$.

6.6 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

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

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

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

线性不可分情况下的非线性分类器。

Take Home Message

- 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