第7章 支持向量机

SVM and Vapnik

- Vapnik 百度
- Vapnik Wiki

支持向量机(Support Vector Machines, SVM)

- SVM
  - **二类分类模型**
  - 定义在特征空间上的间隔最大的线性分类器
    - 间隔最大，有别于感知机
  - 核技巧，使之成为实质上的非线性分类器

- 支持向量机的学习策略为间隔最大化，可形式化为一个求解凸二次规划(convex quadratic programming)的问题
  - 等价于正则化的合页损失函数的最小化问题
  - 支持向量机的学习算法是求解凸二次规划的最优化算法

SVM分类

- 线性可分支持向量机(linear support vector machine in linearly separable case)
  - 硬间隔最大化(hard margin maximization)
- 线性支持向量机(linear support vector machine)
  - 训练数据近似线性可分时，通过软间隔最大化(soft margin maximization)
- 非线性支持向量机(non-linear support vector machine)
  - 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

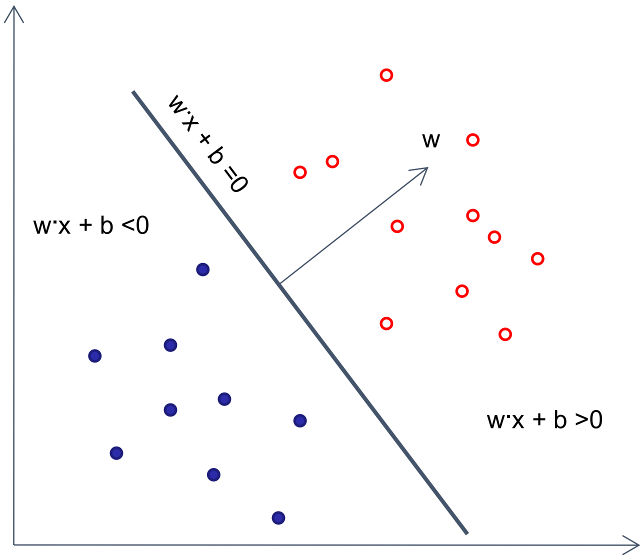
1 线性可分支持向量机与硬间隔最大化

线性可分支持向量机

- 二分类问题
  - 输入空间：欧式空间或离散集合
  - 特征空间：欧式空间或希尔伯特空间
- 线性可分支持向量机、线性支持向量机
  - 假设这两个空间的元素一一对应，并将输入空间中的输入映射为特征空间中的特征向量
- 非线性支持向量机
  - 利用一个从输入空间到特征空间的非线性映射，将输入映射为特征向量

线性可分支持向量机与硬间隔最大化

➢红圈正例，蓝圈负例



线性可分支持向量机

特征空间上的训练数据集： $T=\{(x_1,y_1),(x_2,y_2),\dots,(x_N,y_N)\}$, $x_i\in\mathcal{X}\subseteq \mathrm{R}^n, y_i\in\mathcal{Y}=\{+1,-1\}, i=1,2,\dots,N$

学习的目标：找到分类超平面(法向量指向的一侧为正类+1，另一侧为负类−1)

定义 7.1 (线性可分支持向量机) 给定线性可分训练数据集, 通过间隔最大化或等价地求解相应的凸二次规划问题学习得到的分离超平面为

$$w^*\cdot x+b^*=0$$

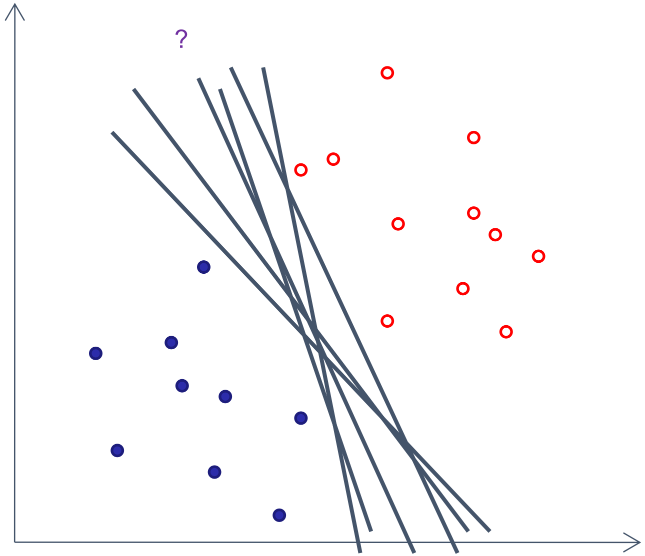
以及相应的分类决策函数

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

称为线性可分支持向量机。

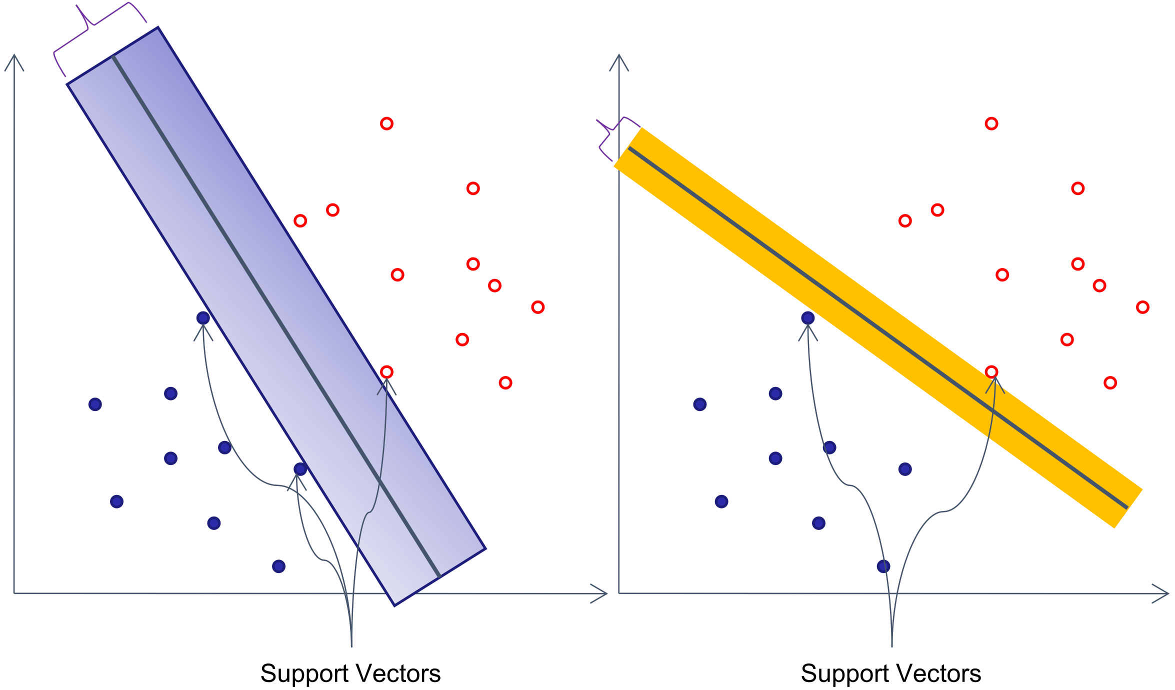
超平面选择

- 存在很多符合条件的超平面，有没有最好的？



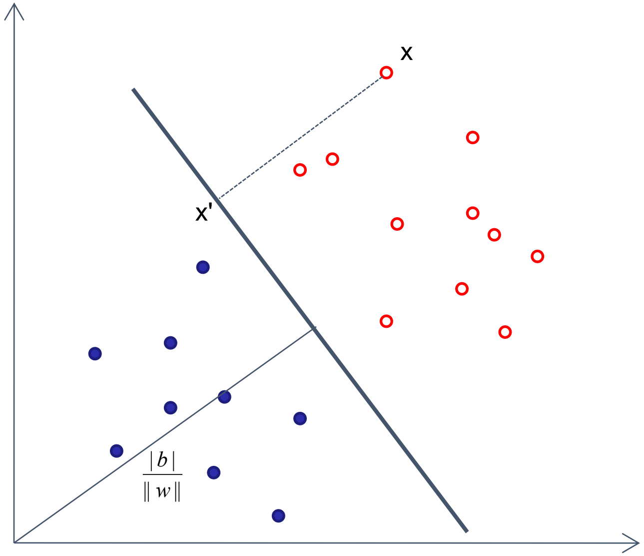
Margins

➢函数间隔，最大的函数间隔



点到超平面的距离

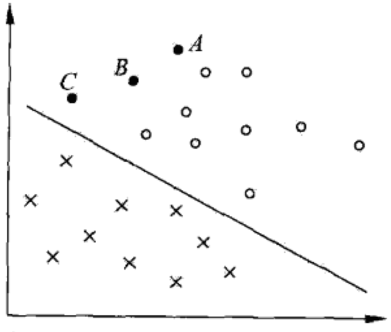
➤点到超平面的距离



$$g(x) = w \cdot x + b$$

函数间隔和几何间隔

- 点距离分离超平面的远近可以表示分类预测的确信程度
  - $|\omega \cdot x + b|$可以用来相对地表示点到分离超平面的远近
  - $\omega \cdot x + b$的符号与类标记y的符号是否一致能够表示分类是否正确
    - 一致，正类
    - 不一致，负类
  - 分类正确的情况下，$|\omega \cdot x + b| = y(w \cdot x + b)$
    - 【注】绝对值运算不可导，应消除



函数间隔

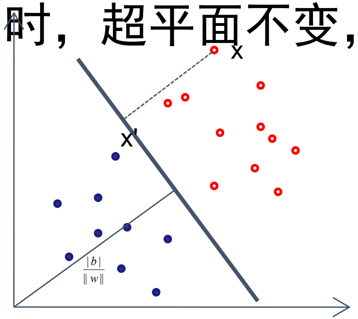
定义 7.2 (函数间隔) 对于给定的训练数据集 T 和超平面 (w,b), 定义超平面 (w,b) 关于样本点 (x_i,y_i) 的函数间隔为

$$\hat{\gamma}_i = y_i(w \cdot x_i + b)$$

定义超平面 (w,b) 关于训练数据集 T 的**函数间隔**为超平面 (w,b) 关于 T 中所有 样本点 (x_i,y_i) 的函数间隔之最小值, 即

$$\hat{\gamma} = \min_{i=1,\cdots,N} \hat{\gamma}_i$$

函数间隔缺点：成比例地改变(w,b)时，超平面不变，但函数间隔变化



几何间隔

定义 7.3 (几何间隔) 对于给定的训练数据集 T 和超平面 (w,b), 定义超平面 (w,b) 关于样本点 (x_i,y_i) 的**几何间隔**为

$$\gamma_i = y_i\left(\frac{w}{\| w \|}\cdot x_i + \frac{b}{\| w \|}\right)$$

超平面关于T的几何间隔，为超平面关于所有样本点 (x_i,y_i) 的几何间隔之最小值

$$\gamma = \min_{i=1,\cdots,N} \gamma_i$$

几何间隔：实例点到超平面的带符号的距离(signed distance); 当样本点被超平面正确分类时,为实例点到超平面的距离

函数间隔和几何间隔之间关系： $\gamma_i = \frac{\hat{\gamma}_i}{\|w\|}, \gamma = \frac{\hat{\gamma}}{\|w\|}$

如 $\| w \| = 1$, 函数间隔和几何间隔相等

如超平面参数 w 和 b 成比例地改变， 超平面不变, 函数间隔按比例变,几何间隔不变

间隔最大化

- 对线性可分的训练数据集，线性可分分离超平面有无穷多个(等价于感知机)
  - 几何间隔最大的分离超平面是唯一的
    - 此间隔最大化又称为硬间隔最大化

- 间隔最大化的直观解释
  - 对训练数据集找到几何间隔最大的超平面，意味着以充分大的确信度对训练数据进行分类
    - 不仅将正负实例点分开，而且对最难分的实例点(离超平面最近的点)也有足够大的确信度将它们分开
    - 这样的超平面应该对未知的新实例有很好的分类预测能力

间隔最大化

(最大间隔分离超平面)问题表示为约束最优化问题：

$$\max_{w,b} \quad \gamma$$

$$\text{s.t.} \quad y_i\left(\frac{w}{\| w \|} \cdot x_i + \frac{b}{\| w \|}\right) \geqslant \gamma,\ i = 1,2,\cdots,N$$

优化目标，最大化几何间隔 γ；约束条件，超平面关于每个训练样本点的几何间隔至少是γ;考虑到几何间隔和函数间隔的关系 $\gamma = \hat{\gamma}/(\| w \|)$，问题改写为

$$\max_{w,b} \quad \frac{\hat{\gamma}}{\| w \|}$$

$$\text{s.t.} \quad y_i(w \cdot x_i + b) \geqslant \hat{\gamma},\ i = 1,2,\cdots,N$$

间隔最大化

$$\max_{w,b} \quad \frac{\hat{\gamma}}{\| w \|}$$

$$\text{s.t.} \quad y_i(w \cdot x_i + b) \geqslant \hat{\gamma},\ i = 1,2,\cdots,N$$

【注】函数间隔 $\hat{\gamma}$ 的取值并不影响最优化问题的解【缩放系数可以放到$\| w \|$，$w^{'} = w/\hat{\gamma}$】，取 $\hat{\gamma} = 1$，注意到最大化 $\frac{1}{\|w\|}$ 和最小化 $\frac{1}{2} \| w \|^2$ 等价，得到优化问题【经典数学优化问题，最小化二次函数】

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

$$\text{s.t.} \quad y_i(w \cdot x_i + b) - 1 \geqslant 0,\ i = 1,2,\cdots,N$$

【注】数学里的方程习惯上$\| w \| = 1$，和此处不一致

线性可分支持向量机学习算法-最大间隔法

算法 7.1 (线性可分支持向量机学习算法-最大间隔法)

输入：训练集$T=\{(x_1,y_1),\cdots,(x_N,y_N)\}$,$x_i\in\mathcal{X}=\mathbf{R}^n$, $y_i\in\mathcal{Y}=\{-1,+1\}$, $i=1,\cdots,N$;

输出：最大间隔分离超平面和分类决策函数。

(1)构造并求解约束最优化问题：

$$\begin{aligned}\min_{w,b}\quad & \frac{1}{2}\|w\|^2\\ \text{s.t.}\quad & y_i(w\cdot x_i+b)-1\geqslant 0,\ i=1,2,\cdots,N\end{aligned}$$

求得最优解 w^*, b^*

(2)由此得到分离超平面：

$$w^*\cdot x+b^*=0$$

分类决策函数

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

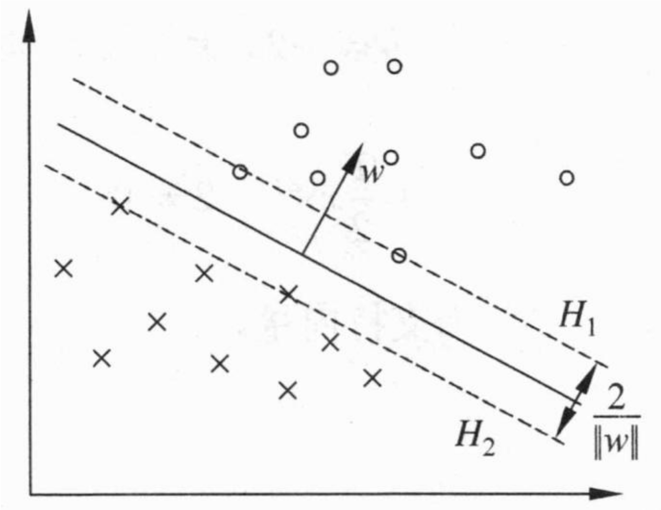
最大间隔分离超平面的存在唯一性

线性可分训练数据集的最大间隔分离超平面是存在且唯一的。

定理 7.1 (最大间隔分离超平面的存在唯一性)
若训练数据集 T 线性可分, 则可将训练数据集中的样本点完全正确分开的最大间隔分离超平面存在且唯一。

支持向量和间隔边界

- 支持向量(support vector)。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的实例称为支持向量
- 分离超平面与它们平行且位于它们中央。长带的宽度，称为间隔(margin)



线性可分支持向量机的对偶算法

原始最优化问题（ 最大间隔分离超平面 ）

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$
$$s.t. \quad 1 - y_i(w \cdot x_i + b) \le 0, \qquad i = 1, \ldots, N$$

线性可分支持向量机的对偶算法 (dual algorithm)： 应用拉格朗日对偶性（附录 C),

通过求解对偶问题 (dual problem) 得到原始问题 (primal problem) 的最优解

此为凸二次规划(convex quadratic programming)问题

$$\min_{x} \quad f(\mathrm{x})$$
$$s.t. \quad \mathrm{c}_i(x) \le 0, \qquad i = 1,2,\ldots,k$$
$$h_i(x) = 0, \qquad i = 1,2,\ldots,l$$

- 其中，条件
  - $f(x), c_i(x)$连续可微，凸函数；$h_i(x)$仿射函数；$f(x)$二次函数

凸二次规划问题

凸二次规划(convex quadratic programming)问题

$$
\begin{aligned}
\min_{x} \quad & f(\mathrm{x}) \\
s.t. \quad & \mathrm{c}_i(x) \le 0, \qquad i = 1,2,\ldots,k \\
& h_i(x) = 0, \qquad i = 1,2,\ldots,l
\end{aligned}
$$

$f(x), c_i(x)$连续可微，凸函数

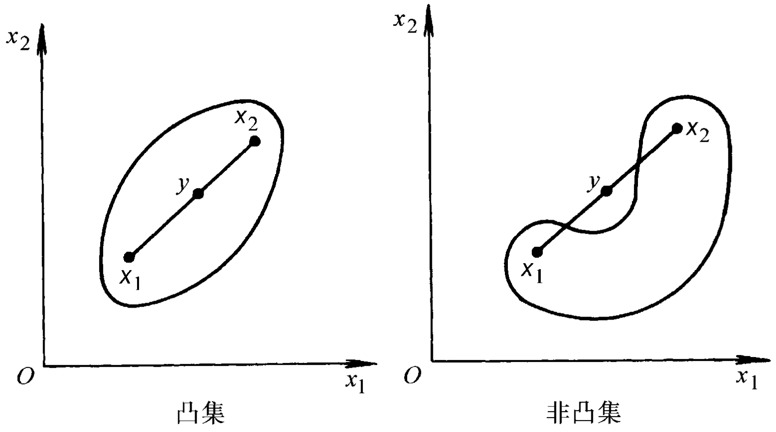
$h_i(x)$仿射函数

$f(x)$二次函数

凸二次规划问题求解

补充：凸集

一个点集(或区域)，如果连接其中任意两点的线段都全部包含在该集合内，就称该点集为凸集，否则为非凸集。



补充：凸性条件

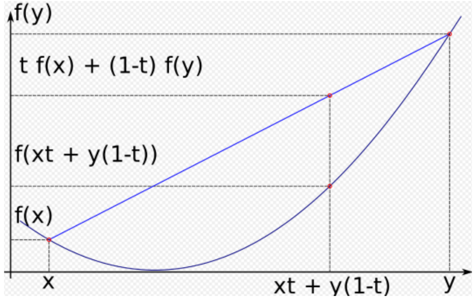
1.根据一阶导数(函数的梯度)来判断函数的凸性

设$f(x)$为定义在凸集R上，且具有连续的一阶导数的函数，则$f(x)$在R上为凸函数的充要条件是对凸集R内任意不同两点，以下不等式恒成立

$$f(x_2) \geq f(x_1) + (x_2 - x_1)^T \nabla f(x_1)$$

2.根据二阶导数(Hesse矩阵)来判断函数的凸性

设$f(x)$为定义在凸集R上且具有连续二阶导数的函数，则$f(x)$在R上为凸函数的充要条件：Hesse矩阵在R上处处半正定



凸二次规划问题 - 广义拉格朗日函数

原问题

$$\begin{aligned} \min_{x} \quad & f(\mathrm{x}) \\ s.t. \quad & \mathrm{c}_i(x) \le 0, \qquad i = 1,2,\dots,k \\ & h_i(x) = 0, \qquad i = 1,2,\dots,l \end{aligned}$$

广义拉格朗日函数

$$\mathrm{L}(\mathrm{x},\alpha,\beta) = f(x) + \sum_{\mathrm{i}=1}^{\mathrm{k}} \alpha_{\mathrm{i}} \mathrm{c}_{\mathrm{i}}(\mathrm{x}) + \sum_{\mathrm{j}=1}^{\mathrm{l}} \beta_{\mathrm{j}} \mathrm{h}_{\mathrm{j}}(\mathrm{x}), \alpha_{\mathrm{i}} > 0$$

记$\theta_P(\mathrm{x}) = \max\limits_{\alpha,\beta;\ \alpha_{\mathrm{i}}>0} \mathrm{L}(\mathrm{x},\alpha,\beta)$，则$\theta_P(x) = f(x)$

原问题⟺拉格朗日函数的极小极大问题：$\min\limits_{x} \max\limits_{\alpha,\beta:\ \alpha_{\mathrm{i}}\ge 0} \mathrm{L}(\mathrm{x},\alpha,\beta)$

进一步转换成对偶形式：$\max\limits_{\alpha,\beta:\ \alpha_i\ge 0} \min\limits_{x} L(x,\alpha,\beta)$

凸二次规划问题 - Karush-Kuhn-Tucker (KKT) 条件

函数$f(x)$和$c_i(x)$是凸函数， $h_i(x)$是仿射函数

则$\mathrm{x}^*, \alpha^*, \beta^*$是原问题和对偶问题的充要条件是， $\mathrm{x}^*, \alpha^*, \beta^*$满足

$$\boldsymbol{\nabla_x} \mathbf{L}(\boldsymbol{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = \mathbf{0}$$

$$\boldsymbol{\alpha_i^* c_i}(\boldsymbol{x}^*) = \mathbf{0}, \quad \mathbf{i} = \mathbf{1}, \mathbf{2}, \cdots, \mathbf{k}$$

$$\mathrm{c_i}(\mathrm{x}^*) \le 0, \quad \mathrm{i} = 1,2, \cdots, \mathrm{k}$$

$$\alpha_\mathrm{i}^* \ge 0, \quad \mathrm{i} = 1,2, \cdots, \mathrm{k}$$

$$\mathrm{h_j}(\mathrm{x}^*) = 0 \quad \mathrm{j} = 1,2, \cdots, \mathrm{l}$$

其中$\alpha_\mathrm{i}^* \mathrm{c_i}(\mathrm{x}^*) = 0$对偶互补条件： $\alpha_\mathrm{i}^* > 0$则$\mathrm{c_i}(\mathrm{x}^*) = 0$

【注】KKT条件给出了对偶问题的求解方式

学习的对偶算法

线性可分支持向量机的对偶算法

求解线性可分支持向量机的最优化问题作为原始最优化问题，应用拉格朗日对偶性, 通过求解对偶问题得到原始问题的最优解。

- 对偶问题往往更容易求解;
- 自然引入核函数, 推广到非线性分类问题。

首先，对每一个不等式约束，引进拉格朗日乘子 $\alpha_i \geqslant 0, i = 1,2,\cdots,N$ ，定义拉格朗日函数：

$$L(w,b,\alpha) = \frac{1}{2}\parallel w \parallel^2 - \sum_{i=1}^{N} \alpha_i y_i (w \cdot x_i + b) + \sum_{i=1}^{N} \alpha_i$$

其中, $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_N)^{\mathrm{T}}$ 为拉格朗日乘子向量。

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：

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

所以, 为了得到对偶问题的解, 需要先求 $L(w,b,\alpha)$ 对 w,b 的极小, 再求对 α 的极大。

学习的对偶算法 - 对偶问题的极小问题

(1) 求$\min\limits_{w,b} L(w,b,\alpha)$

将拉格朗日函数 $L(w,b,\alpha)$ 分别对 w,b 求偏导数并令其等于0，

$$\nabla_w L(w,b,\alpha) = w - \sum_{i=1}^{N} \alpha_i y_i x_i = 0$$

$$\nabla_b L(w,b,\alpha) = -\sum_{i=1}^{N} \alpha y_i = 0$$

得$w = \sum_{i=1}^{N} \alpha_i y_i x_i$, $\sum_{i=1}^{N} \alpha_i y_i = 0$，代入拉格朗日函数, 即得

$$L(w,b,\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N} \alpha_i y_i \left(\left(\sum_{j=1}^{N} \alpha_j y_j x_j\right)\cdot x_i + b\right) + \sum_{i=1}^{N} \alpha_i$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N} \alpha_i$$

即

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

学习的对偶算法 - 对偶问题的极大问题

(2) 求 $\min_{w,b} L(w,b,\alpha)$ 对 α 的极大, 即是对偶问题

$$\max_{\alpha} \quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N}\alpha_i$$

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

$$\alpha_i \geqslant 0, \ i = 1,2,\cdots,N$$

将上式目标函数由求极大转换成求极小, 得:

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\alpha_i$$

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

$$\alpha_i \geqslant 0, \ i = 1,2,\cdots,N$$

学习的对偶算法 - 回溯

根据max问题，构造并求解约束最优化问题，求得最优解$\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_N^*)^{\mathrm{T}}$

代入min问题，求w^*，b^*

$$\mathrm{w}^* = \sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_{\mathrm{i}}^* \mathrm{y}_{\mathrm{i}} \mathrm{x}_{\mathrm{i}}$$

取某一$\alpha_{\mathrm{j}}^* > 0$

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j)$$

得到$\mathrm{w}^* \cdot \mathrm{x} + \mathrm{b}^* = 0$

对偶优化算法的理论

定理 7.2 设 $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_l^*)^{\mathrm{T}}$ 是对偶最优化问题的解, 则存在下标 j（$\alpha_j^* > 0$） 并求得原始问题的解：

$$w^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$$

$$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j)$$

证明 根据定理 C.3, KKT 条件成立, 即得

$$\nabla_w L(w^*, b^*, \alpha^*) = w^* - \sum_{i=1}^{N} \alpha_i^* y_i x_i = 0$$

$$\nabla_b L(w^*, b^*, \alpha^*) = -\sum_{i=1}^{N} \alpha_i^* y_i = 0$$

$$\alpha_i^* (y_i (w^* \cdot x_i + b^*) - 1) = 0,\ i = 1,2,\cdots,N$$

$$y_i (w^* \cdot x_i + b^*) - 1 \geqslant 0,\ i = 1,2,\cdots,N$$

$$\alpha_i^* \geqslant 0,\ i = 1,2,\cdots,N$$

由此得：$w^* = \sum_i \alpha_i^* y_i x_i$，其中至少有一个 $\alpha_j^* > 0$ (用反证法, 假设 $\alpha^* = 0$, 可知 $w^* = 0$, 而 $w^* = 0$ 不是原始最优化问题的解, 产生矛盾), 对此 j 有$y_j (w^* \cdot x_j + b^*) - 1 = 0$

将式 (7.25) 代入式 (7.28) 并注意到 $y_j^2 = 1$, 即得$b^* = y_j - \sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j)$

线性可分支持向量机学习算法

算法 7.2 (线性可分支持向量机学习算法)

输入： 线性可分训练集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$, 其中 $x_i\in\mathcal{X}=\mathbf{R}^n$, $y_i\in\mathcal{Y}=\{-1,+1\}, i=1,2,\cdots,N$;

输出： 分离超平面和分类决策函数。

(1) 构造并求解约束最优化问题

$$\min_{\alpha}\quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\left(x_i\cdot x_j\right)-\sum_{i=1}^{N}\alpha_i$$

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

$$\alpha_i\geqslant 0,\ i=1,2,\cdots,N$$

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

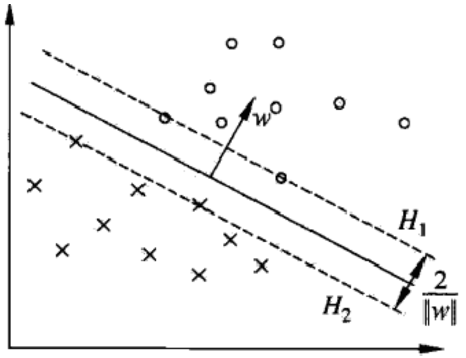
(2) 计算： $w^*=\sum_{i=1}^{N}\alpha_i^* y_i x_i$

并选择 α^* 的一个正分量 $\alpha_j^*>0$, 计算： $b^*=y_j-\sum_{i=1}^{N}\alpha_i^* y_i\left(x_i\cdot x_j\right)$

(3) 求得分离超平面和分类决策函数 $w^*\cdot x+b^*=0$， $f(x)=\operatorname{sign}(w^*\cdot x+b^*)$

支持向量

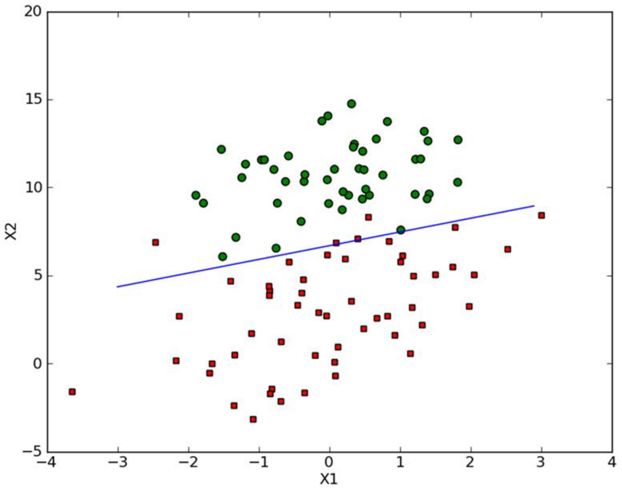
- 定义 7.4 (支持向量）
  - 将训练数据集中对应于$\alpha_i^* > 0$的样本点$(\mathrm{x_i}, \mathrm{y_i})$的实例$\mathrm{x_i} \in \mathrm{R^n}$称为支持向量
  - $\alpha_i^* = 0$ 对$\mathrm{w^*}$ $\mathrm{b^*}$没有支持(贡献)
- 由KKT互补条件
  - $\alpha_\mathrm{i}^*(\mathrm{y_i}(\mathrm{w^*} \cdot \mathrm{x_i} + \mathrm{b^*}) - 1) = 0,\ \mathrm{i} = 1,2,\cdots,\mathrm{N}$
  - 有$y_i(w^* \cdot x_i + b^*) - 1 = 0$ 或者$w^* \cdot x_i + b^* = \pm 1$



2 线性支持向量机与软间隔最大化

线性支持向量机-软间隔最大化

- 问题：训练数据中有一些特异点(outlier)，不能满足函数间隔大于等于1的约束条件
- 解决方法：对每个样本点引进一个松弛变量$\xi_i \geq 0$，使得函数间隔加上松弛变量大于等于1
  - 约束条件：$y_i(w \cdot x_i + b) \geq 1 - \xi_i$
  - 目标函数：$\frac{1}{2}|w|^2 + C\sum_{i=1}^{N}\xi_i$，$C$惩罚参数
    - 间隔尽量大，$\frac{1}{2}|w|^2$尽量小
    - 误分类点的个数尽量小
    - 其中，C是调和二者的系数
- 【注】目标函数最小化，松弛项当损失看
  - 可以证明w的解是唯一的
  - b的解可能不唯一，存在于一个区间



线性不可分的线性支持向量机的学习问题

原始问题

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

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1,2,\cdots,N$$

$$\xi_i \geq 0(i = 1,2,\cdots,N)$$

拉格朗日函数

$$\mathrm{L}(\mathrm{w},\mathrm{b},\xi,\alpha,\mu) = \frac{1}{2}|\mathrm{w}|^2 + \mathrm{C}\sum_{\mathrm{i}=1}^{\mathrm{N}}\xi_{\mathrm{i}} - \sum_{\mathrm{i}=1}^{\mathrm{N}}\alpha_{\mathrm{i}}(\mathrm{y}_{\mathrm{i}}(\mathrm{w}\cdot\mathrm{x}_{\mathrm{i}} + \mathrm{b}) - 1 + \xi_{\mathrm{i}}) - \sum_{\mathrm{i}=1}^{\mathrm{N}}\mu_{\mathrm{i}}\xi_{\mathrm{i}}$$

$$\alpha_{\mathrm{i}} \geq 0, \mu_{\mathrm{i}} \geq 0$$

对偶问题极小化

对偶问题step1，拉格朗日函数$L(w, b, \xi, \alpha, \mu)$极小化，对$w, b, \xi$偏导

$$\nabla_w L(w, b, \xi, \alpha, \mu) = w - \sum_{i=1}^{N} \alpha_i y_i x_i = 0$$

$$\nabla_b L(w, b, \xi, \alpha, \mu) = -\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\nabla_{\xi_i} L(w, b, \xi, \alpha, \mu) = C - \alpha_i - \mu_i = 0$$

得

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

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

$$C - \alpha_i - \mu_i = 0$$

这样，得到

$$\min_{w,b,\xi} L(w, b, \xi, \alpha, \mu) = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{N} \alpha_i$$

对偶问题极大化

对偶问题step2， $\min\limits_{w,b,\xi} L(w,b,\xi,\alpha,\mu)$极大

$$\max_{\alpha} \ -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\left(x_i\cdot x_j\right) \ +\sum_{i=1}^{N}\alpha_i$$

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

$$C-\alpha_i-\mu_i=0$$

$$\alpha_i \geqslant 0$$

$$\mu_i \geqslant 0, i=1,2,\cdots,N$$

整理，得到对偶问题

$$\min_{\alpha}\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\left(x_i\cdot x_j\right)-\sum_{i=1}^{N}\alpha_i$$

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

$$0\leq\alpha_i\leq C,\quad i=1,2,\cdots,N$$

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

计算$w^*=\sum_{i=1}^{N}\alpha_i^* y_i x_i$ ；选择α_j^*适合条件$0<\alpha_j^*<C$，计算$b^*=y_j-\sum_{i=1}^{N}y_i\alpha_i^*\left(x_i\cdot x_j\right)$

求得分离超平面：$w^*\cdot x+b^*=0$；分类决策函数：$f(x)=\mathrm{sign}(w^*\cdot x+b^*)$

对偶算法定理

定理 7.3 设 $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_N^*)^{\mathrm{T}}$ 是对偶问题的一个解, 若存在α^*的一个分量$\alpha_j^*, 0 < \alpha_j^* < C$, 则原始问题的解$w^*, b^*$为:

$$w^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$$

$$b^* = y_j - \sum_{i=1}^{N} y_i \alpha_i^* (x_i \cdot x_j)$$

【证明】 原始问题是凸二次规划问题, 解满足 KKT 条件。即得

$$\nabla_w L(w^*, b^*, \xi^*, \alpha^*, \mu^*) = w^* - \sum_{i=1}^{N} \alpha_i^* y_i x_i = 0$$

$$\nabla_b L(w^*, b^*, \xi^*, \alpha^*, \mu^*) = -\sum_{i=1}^{N} \alpha_i^* y_i = 0$$

$$\nabla_\xi L(w^*, b^*, \xi^*, \alpha^*, \mu^*) = C - \alpha^* - \mu^* = 0$$

$$\alpha_i^* (y_i (w^* \cdot x_i + b^*) - 1 + \xi_i^*) = 0$$

$$\mu_i^* \xi_i^* = 0$$

$$y_i (w^* \cdot x_i + b^*) - 1 + \xi_i^* \geqslant 0$$

$$\xi_i^* \geqslant 0, \alpha_i^* \geqslant 0, \mu_i^* \geqslant 0, \ i = 1,2, \cdots, N$$

若存在 $\alpha_j^*, 0 < \alpha_j^* < C$, 则 $y_i (w^* \cdot x_i + b^*) - 1 = 0$ 。由此即得

线性支持向量机学习算法

算法 7.3 (线性支持向量机学习算法)

输入： 训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$, 其中, $x_i\in\mathcal{X}=\mathbf{R}^n$, $y_i\in\mathcal{Y}=\{-1,+1\}, i=1,2,\cdots,N$;

输出： 分离超平面和分类决策函数。

(1) 选择惩罚参数 $C>0$, 构造并求解凸二次规划问题

$$\begin{aligned}\min_{\alpha}\quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\left(x_i\cdot x_j\right)-\sum_{i=1}^{N}\alpha_i\\ \text{s.t.}\quad & \sum_{i=1}^{N}\alpha_i y_i=0\\ & 0\leqslant\alpha_i\leqslant C,\ i=1,2,\cdots,N\end{aligned}$$

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

(2) 计算 $w^*=\sum_{i=1}^{N}\alpha_i^* y_i x_i$

选择 α^* 的一个分量 α_j^* 适合条件 $0<\alpha_j^*<C$, 计算$b^*=y_j-\sum_{i=1}^{N}y_i\alpha_i^*\left(x_i\cdot x_j\right)$

(3) 求得分离超平面：$w^*\cdot x+b^*=0$；分类决策函数 ： $f(x)=\operatorname{sign}(w^*\cdot x+b^*)$

支持向量

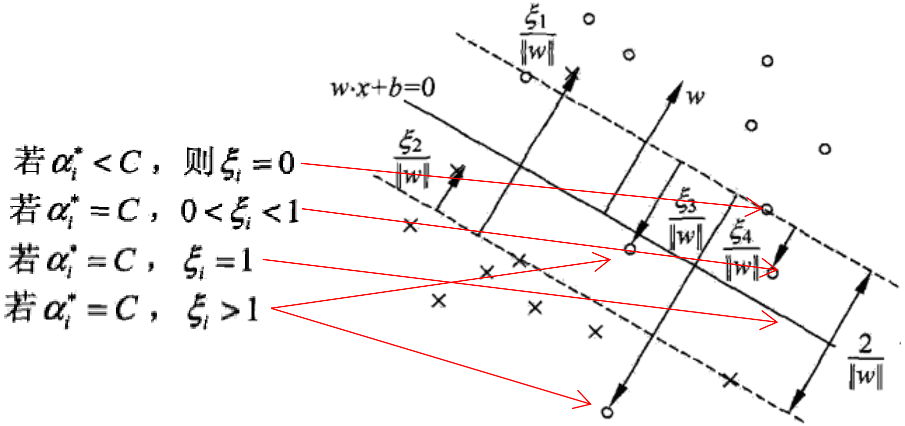
定义 7.5 (线性支持向量机) 对于给定的线性不可分的训练数据集, 通过求解凸 二次规划问题, 即软间隔最大化问题，得到的分离超平面为

$$w^* \cdot x + b^* = 0$$

以及相应的分类决策函数

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

称为线性支持向量机。



另一个角度：合页损失函数hinge loss function

选错分个数，不连续，求导有问题。引入合页损失函数$[z]_+ = \begin{cases} z, & z > 0 \\ 0, & z \le 0 \end{cases}$

$$[1 - y_i(w \cdot x_i + b)]_+ = \begin{cases} 1 - y_i(w \cdot x_i + b), & y_i(w \cdot x_i + b) \le 1 \\ 0, & y_i(w \cdot x_i + b) > 1 \text{ (正确分类)} \end{cases}$$

目标函数：$\frac{1}{2}\|w\|^2 + \mathrm{C}\sum_{\mathrm{i}=1}^{N}[1 - y_i(w \cdot x_i + b)]_+$

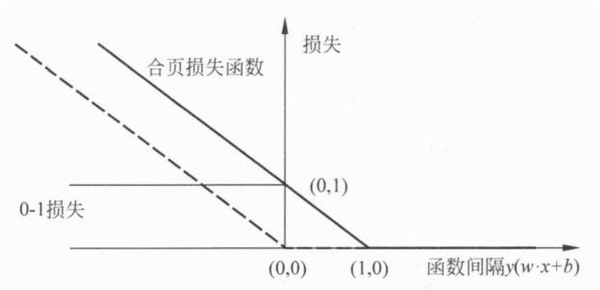
令$\xi_i = 1 - y_i(w \cdot x_i + b)$，则$\begin{cases} \xi_i \le 0\text{，正确分类} \\ \xi_i \ge 0, \quad \text{错分} \end{cases}$

则$[\xi_i]_+$为合页损失函数

0，$\xi_i \le 0$【正确】

$\xi_i = 1 - y_i(w \cdot x_i + b)$，$\xi_i \ge 0$【错分】

$\xi_i \ge 0$表示离±1(支持平面)的距离



合页损失函数

定理 7.4 线性支持向量机原始最优化问题：

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

$$\text{s.t.} \quad y_i(w \cdot x_i + b) \geqslant 1 - \xi_i,\ i = 1,2,\cdots,N \ (2)$$

$$\xi_i \geqslant 0,\ i = 1,2,\cdots,N \ (3)$$

等价于最优化问题

$$\min_{w,b} \sum_{i=1}^{N}[1 - y_i(w \cdot x_i + b)]_+ + \lambda \| w \|^2 \quad (4)$$

证明 可将最优化问题 (4) 写成问题 (1)- (3)。令

$$[1 - y_i(w \cdot x_i + b)]_+ = \xi_i \quad (5)$$

则 $\xi_i \geqslant 0$, 式 (3) 成立。由式 (5), 当 $1 - y_i(w \cdot x_i + b) > 0$ 时, 有 $y_i(w \cdot x_i + b) = 1 - \xi_i$; 当 $1 - y_i(w \cdot x_i + b) \leqslant 0$ 时, $\xi_i = 0$, 有 $y_i(w \cdot x_i + b) \geqslant 1 - \xi_i$ 。故式 (2) 成 立。于是 w, b, ξ_i 满足约束条件 (2) – (3) 。所以最优化问题 (4) 可写成

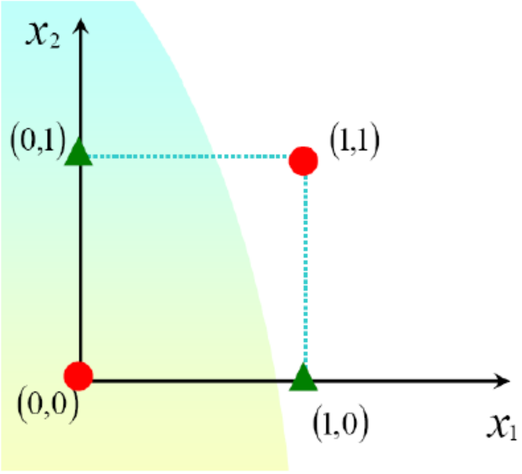
$$\min_{w,b} \sum_{i=1}^{N}\xi_i + \lambda \| w \|^2$$

若取 $\lambda = \frac{1}{2C}$, 则$\min_{w,b} \frac{1}{C}\left(\frac{1}{2}\| w \|^2 + C\sum_{i=1}^{N}\xi_i\right)$与式 (1) 等价。

3 非线性支持向量机与核函数

非线性支持向量机与核函数

➢例子（线性二分类问题）



$\omega_1 : \{(0,0),(1,1)\}$

$\omega_2 : \{(1,0),(0,1)\}$

线性识别函数不存在

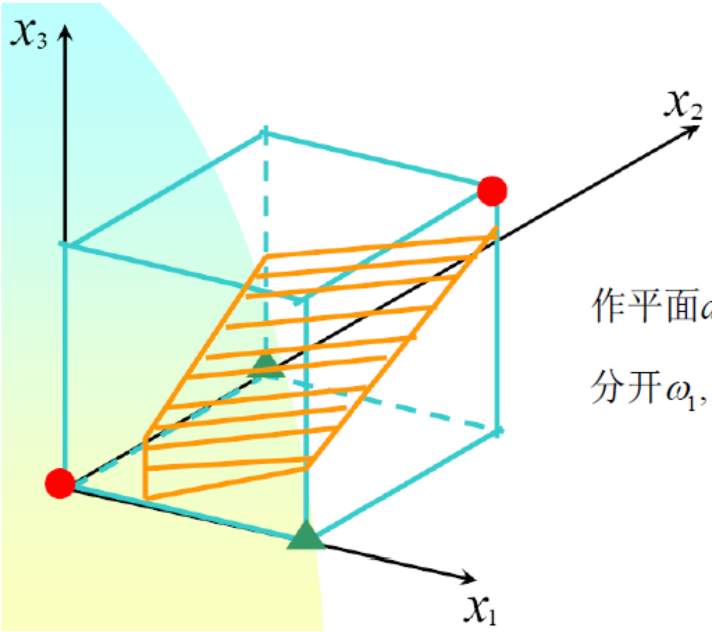
把2维空间(x_1,x_2)

变为3维空间(x_1,x_2,x_3)

$$x_3 = x_1 x_2$$

(x_1,x_2)	(x_1,x_2,x_3)
$(0,0)$	$(0,0,0)$
$(1,1)$	$(1,1,1)$
$(1,0)$	$(1,0,0)$
$(0,1)$	$(0,1,0)$

升维

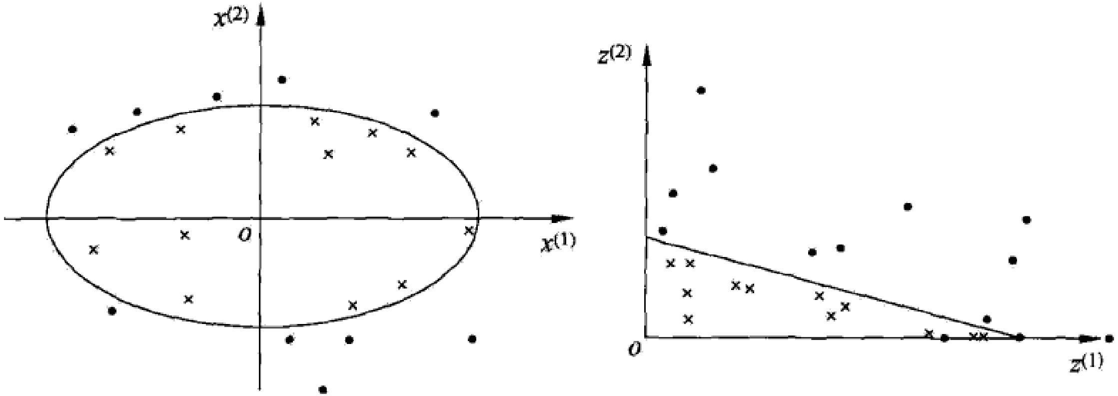


作平面$d(X) = x_1 + x_2 - 2x_3 - \frac{1}{3} = 0$

分开ω_1, ω_2

例7.7：

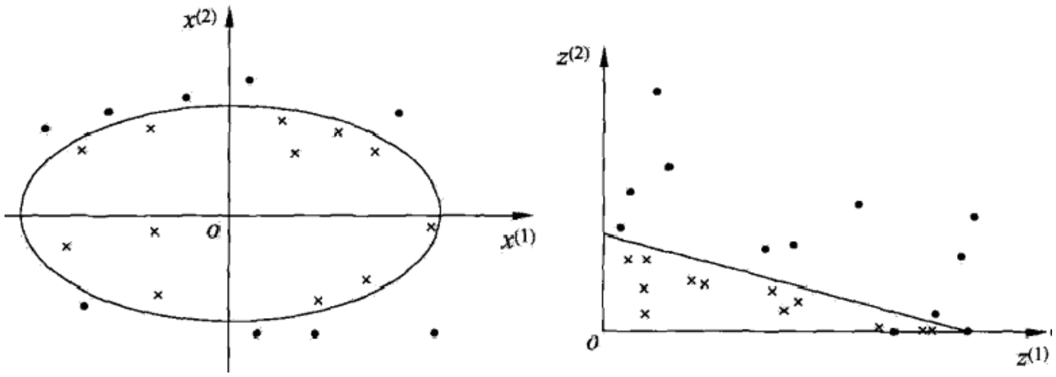
- 变换以后成线性



非线性支持向量机与核函数

➢非线性分类问题：

➢如果能用R^n中的一个超曲面将正负例正确分开，则称这个问题为非线性可分问题.



非线性支持向量机与核函数

- 非线性问题往往不好求解，希望能用解线性分类问题的方法解决这个问题。
- 原空间：

$$\mathcal{X} \subset R^2, x = \left(x^{(1)}, x^{(2)}\right)^{\mathrm{T}} \in \mathcal{X}$$

- 进行一个非线性变换，新空间：

$$\mathcal{Z} \subset R^2, z = \left(z^{(1)}, z^{(2)}\right)^{\mathrm{T}} = \mathcal{Z}\ z = \phi(x) = \left(\left(x^{(1)}\right)^2, \left(x^{(2)}\right)^2\right)^{\mathrm{T}}$$

- 原空间中的椭圆

$$w_1\left(x^{(1)}\right)^2 + w_2\left(x^{(2)}\right)^2 + b = 0$$

- 变换成为新空间中的直线

$$w_1 z^{(1)} + w_2 z^{(2)} + b = 0$$

非线性支持向量机与核函数

- 用线性分类方法求解非线性分类问题
  - 首先，使用一个变换将原空间的数据映射到新空间
  - 然后，在新空间里用线性分类学习方法从训练数据中学习分类模型
- 核技巧应用到支持向量机，其基本想法
  - 通过一个非线性变换将输入空间(欧氏空间R”或离散集合)对应于一个特征空间(希尔伯特空间)，使得在输入空间中的超曲面模型对应于特征空间中的超平面模型(支持向量机)
  - 分类问题的学习任务通过在特征空间中求解线性支持向量机就可以完成.

核函数定义

定义7.6 (核函数) 设 $\mathcal{X}$ 是输入空间 (欧氏空间 $\mathbf{R}^n$ 的子集或离散集合), 又设 $\mathcal{H}$ 为特征空间 (希尔伯特空间), 如果存在一个从 $\mathcal{X}$ 到 $\mathcal{H}$ 的映射

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

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

$$K(x,z) = \phi(x) \cdot \phi(z)$$

则称 $K(x,z)$ 为核函数, $\phi(x)$ 为映射函数, 式中 $\phi(x) \cdot \phi(z)$ 为 $\phi(x)$ 和 $\phi(z)$ 的内积。

核技巧的思想：

在学习与预测中只定义核函数$K(x,z)$【原空间中的】，而不显式地定义映射函数$\phi(x)$。通常，直接计算$K(x,z)$比较容易，而通过$\phi(x)$和$\phi(z)$计算$K(x,z)$并不容易。

注意： $\phi(x)$是输入空间R^n到特征空间$\mathcal{H}$的映射，特征空间$\mathcal{H}$一般是高维，映射$\phi(x)$并不唯一。

【注】核函数定义在原空间中

例7.3

例 7.3 假设输入空间是 $\mathbf{R}^2$, 核函数是 $K(x,z)=(x\cdot z)^2$, 试找出其相关的特征 空间 $\mathcal{H}$ 和映射 $\phi(x)$：$\mathbf{R}^2\to\mathcal{H}$ 。

解 取特征空间 $\mathcal{H}=\mathbf{R}^3$, 记 $x=\left(x^{(1)},x^{(2)}\right)^{\mathrm{T}}, z=\left(z^{(1)},z^{(2)}\right)^{\mathrm{T}}$, 由于

$$(x\cdot z)^2=\left(x^{(1)}z^{(1)}+x^{(2)}z^{(2)}\right)^2=\left(x^{(1)}z^{(1)}\right)^2+2x^{(1)}z^{(1)}x^{(2)}z^{(2)}+\left(x^{(2)}z^{(2)}\right)^2$$

所以可以取映射

$$\phi(x)=\left(\left(x^{(1)}\right)^2,\sqrt{2}x^{(1)}x^{(2)},\left(x^{(2)}\right)^2\right)^{\mathrm{T}}$$

容易验证$\phi(x)\cdot\phi(z)=(x\cdot z)^2=K(x,z)$

仍取 $\mathcal{H}=\mathbf{R}^3$ 以及

$$\phi(x)=\frac{1}{\sqrt{2}}\left(\left(x^{(1)}\right)^2-\left(x^{(2)}\right)^2,2x^{(1)}x^{(2)},\left(x^{(1)}\right)^2+\left(x^{(2)}\right)^2\right)^{\mathrm{T}}$$

同样有 $\phi(x)\cdot\phi(z)=(x\cdot z)^2=K(x,z)$ 。

还可以取 $\mathcal{H}=\mathbf{R}^4$ 和

$$\phi(x)=\left(\left(x^{(1)}\right)^2,x^{(1)}x^{(2)},x^{(1)}x^{(2)},\left(x^{(2)}\right)^2\right)^{\mathrm{T}}$$

核函数在支持向量机的应用

线性支持向量机对偶问题中，目标函数、决策函数都只涉及输入实例和实例之间的内积

内积 $x_i \cdot x_j$用核函数$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$代替

目标函数：

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

同样, 分类决策函数中的内积也可以用核函数代替, 而分类决策函数式成为

$$\begin{aligned} f(x) &= \operatorname{sign}\left(\sum_{i=1}^{N_s} a_i^* y_i \phi(x_i) \cdot \phi(x) + b^*\right) \\ &= \operatorname{sign}\left(\sum_{i=1}^{N_s} a_i^* y_i K(x_i, x) + b^*\right) \end{aligned}$$

等价于经过映射函数 ϕ 将原来的输入空间变换到一个新的特征空间, 将输入空间中的内积 $x_i \cdot x_j$ 变换为特征空间中的内积 $\phi(x_i) \cdot \phi(x_j)$, 在新的特征空间里从训练样本中学习线性支持向量机。

定义核函数

- 问题：
  - 已知映射函数ϕ, 可以通过 $\phi(x)$ 和 $\phi(z)$ 的内积求得核函数 $K(x,z)$ 。
    - 不用构造映射 $\phi(x)$ 能否直接判断一个给定的函数 $K(x,z)$ 是不是核函数?
    - 或者说, 函数 $K(x,z)$ 满足什么条件才能成为核函数?

- 假设 $K(x,z)$ 是定义在 $\mathcal{X}\times\mathcal{X}$上的对称函数, 并且对任意的 $x_1,x_2,\cdots,x_m\in\mathcal{X}$, $K(x,z)$ 关于 $x_1,x_2,\cdots,x_m$的Gram矩阵是半正定的。可以依据函数 $K(x,z)$, 构成一个希尔伯特空间(Hilbert space),其步骤是
  - 首先定义映射ϕ并构成向量空间$\mathcal{S}$
  - 然后在$\mathcal{S}$上定义内积构成内积空间
  - 最后将$\mathcal{S}$完备化构成希尔伯特空间

1. 定义映射，构成向量空间$\mathcal{S}$

定义映射, 松成向量空间 $\mathcal{S}$

先定义映射

$$\phi:\ x \rightarrow K(\cdot, x)$$

根据这一映射, 对任意 $x_i \in \mathcal{X}, \alpha_i \in \mathbf{R}, i = 1,2,\cdots,m$, 定义线性组合

$$f(\cdot) = \sum_{i=1}^{m} \alpha_i K(\cdot, x_i)$$

考虑由线性组合为元素的集合 $\mathcal{S}$ 。由于集合 $\mathcal{S}$ 对加法和数乘运算是封闭的, 所以 $\mathcal{S}$ 构成一个向量空间。

2. 在$\mathcal{S}$上定义内积，构成内积空间

在 $\mathcal{S}$ 上定义一个运算 $*$： 对任意 $f, g \in \mathcal{S}$,

$$f(\cdot) = \sum_{i=1}^{m} \alpha_i K(\cdot, x_i)$$

$$g(\cdot) = \sum_{j=1}^{l} \beta_j K(\cdot, z_j)$$

定义运算 *

$$f * g = \sum_{i=1}^{m} \sum_{j=1}^{l} \alpha_i \beta_j K(x_i, z_j)$$

3. 将内积空间$\mathcal{S}$完备化为希尔伯特空间

由内积得到范数： $|f| = \sqrt{f \cdot f}$

因此，$\mathcal{S}$是一个赋范向量空间。根据泛函分析理论，对于不完备的赋范向量空间$\mathcal{S}$，一定可以使之完备化，得到完备的赋范向量空间$\mathcal{H}$；一个内积空间，当作为一个赋范向量空间是完备的时候，就是希尔伯特空间，这样就得到了希尔伯特空间$\mathcal{H}$

$\mathcal{H}$称为再生核希尔伯特空间 (reproducing kernel Hilbert space, RKHS)。这是由于核 K 具有再生性, 即满足

$$K(\cdot, x) \cdot f = f(x)$$

及

$$K(\cdot, x) \cdot K(\cdot, z) = K(x, z)$$

称为再生核。

正定核的充要条件

定理**7.5** (正定核的充要条件) 设 K: $\mathcal{X} \times \mathcal{X} \to \mathbf{R}$ 是对称函数, 则 $K(x, z)$ 为正 定核函数的充要条件是对任意 $x_i \in \mathcal{X}, i = 1,2, \cdots, m, K(x, z)$ 对应的 Gram 矩阵：

$$K = \left[K\left(x_i, x_j\right)\right]_{m \times m}$$

是半正定矩阵

定义7.7 (正定核的等价定义) 设 $\mathcal{X} \subset \mathbf{R}^n, K(x, z)$ 是定义在 $\mathcal{X} \times \mathcal{X}$ 上的对称 函数, 如果对任意 $x_i \in \mathcal{X}, i = 1,2, \cdots, m, K(x, z)$ 对应的 Gram 矩阵

$$K = \left[K\left(x_i, x_j\right)\right]_{m \times m}$$

是半正定矩阵, 则称 $K(x, z)$ 是正定核。

常用核函数

1、多项式核函数(Polynomial kernel function)

$$K(x,z)=(x\cdot z+1)^p$$

对应的支持向量机为P次多项式分类器，分类决策函数：

$$f(x)=\mathrm{sign}\left(\sum_{i=1}^{N_s}a_i^* y_i(x_i\cdot x+1)^p\ +b^*\right)$$

2、高斯核函数 (Gaussian Kernel Function)

$$K(x,z)=\exp\left(-\frac{\| x-z \|^2}{2\sigma^2}\right)$$

决策函数：

$$f(x)=\mathrm{sign}\left(\sum_{i=1}^{N_s}a_i^* y_i \exp\left(\frac{-\| x-x_i \|^2}{2\sigma^2}\right)+b^*\right)$$

3、字符串核函数 (string kernel function)

非线性支持向量机

将线性支持向量机扩展到非线性支持向量机，只需将对偶形式中的内积换成核函数。

定义 7.8 (非线性支持向量机) 从非线性分类训练集，通过核函数与软间隔最大 化，或凸二次规划，学习得到的分类决策函数

$$f(x) = \operatorname{sign}\left(\sum_{i=1}^{N} \alpha_i^* y_i K(x, x_i) + b^*\right)$$

称为非线性支持向量机，$K(x, z)$ 是正定核函数。

非线性支持向量机学习算法

算法7.4 (非线性支持向量机学习算法)

输入：训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$, 其中 $x_i\in\mathcal{X}=\mathbf{R}^n, y_i\in\mathcal{Y}=\{-1,+1\}, i=1,2,\cdots,N$;

输出：分类决策函数。

(1) 选取适当的核函数 $K(x,z)$ 和适当的参数 C, 构造并求解最优化问题

$$\begin{aligned}\min_{\alpha}\quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i,x_j)-\sum_{i=1}^{N}\alpha_i\\ \text{s.t.}\quad & \sum_{i=1}^{N}\alpha_i y_i=0\\ & 0\leqslant\alpha_i\leqslant C,\ i=1,2,\cdots,N\end{aligned}$$

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

(2) 选择 α^* 的一个正分量 $0<\alpha_j^*<C$, 计算：$b^*=y_j-\sum_{i=1}^{N}\alpha_i^* y_i K(x_i,x_j)$

(3) 构造决策函数：$f(x)=\mathrm{sign}\left(\sum_{i=1}^{N}\alpha_i^* y_i K(x,x_i)+b^*\right)$

当$K(x,z)$ 是正定核函数时, 问题 (7.95) ~ (7.97) 是凸二次规划问题, 解是存在的。

附录：泛函分析基础

泛函基础

- 度量空间
- 赋范空间
- 向量空间
- 希尔伯特空间
- 内积空间

泛函分析

- 形成于20世纪30年代的数学分支
  - 从变分问题，积分方程和理论物理的研究中发展而来
  - 综合运用了函数论，几何学，代数学的观点
  - 可看成是无限维向量空间的解析几何及数学分析
- 函数概念被赋予了更为一般的意义，推广
  - 古典分析，函数概念是指两个数集之间所建立的一种对应关系
  - 现代数学，建立两个任意集合之间的某种对应关系
  - 在数学上，把无限维空间到无限维空间的变换叫做算子
  - 研究无限维线性空间上的泛函和算子理论，产生一门新的分析数学，泛函分析
- 泛函分析自身
  - 算子谱理论、巴拿赫代数、拓扑线性空间理论、广义函数论
- 与其他数学学科的关联
  - 微分方程、概率论、函数论、连续介质力学、计算数学、控制论、最优化理论等学科中都有重要的应用，建立群上调和分析理论的基本工具
- 与其他科学学科的关联
  - 连续介质力学、量子物理学，是研究无限个自由度物理系统的重要而自然的工具之一

度量空间

- 距离空间
- Banach空间(完备的赋范线性空间)
- Hilbert空间(完备的内积空间)

度量(距离)空间

- 设X是非空集合，对于X中的任意两元素x与y，按某一法则都对应唯一的实数$\rho(x, y)$，并满足以下三条公理(距离公理)
  - 1.非负性：$\rho(x, y) \geq 0$，$\rho(x, y) = 0$当且仅当$x=y$；
  - 2.对称性：$\rho(x, y) = \rho(y, x)$；
  - 3.三角不等式： 对任意的x, y, z，$\rho(x, y) \leq \rho(x, z) + \rho(z, y)$

- 则称$\rho(x, y)$为x与y间的距离(或度量)，并称X是以ρ为距离的距离空间(或度量空间)，记成(X, ρ)，或简记为X；X中的元素称为X中的点

度量(距离)空间

- 是泛函分析最基本的研究框架，泛函分析中的空间均为度量空间
- 度量空间的概念起源于经典分析，人们在研究线性常微分方程、偏微分方程、变分法以及逼近论
- 泛函分析中，经常将符合一定要求的元素放在一起所构成的集合称之为一个“空间”，元素称之为“点”
  - “点”：包含真正意义下的点、数列和函数
- 泛函分析中，很少研究一个“点”的具体性质，而是研究一个空间中点与点之间的关系，以及空间中符合一定条件的点组成的该空间子集的一些性质

度量(距离)空间

- $L^p[a,b]$表示区间$[a,b]$绝对值的p次幂L可积函数的全体，并把几乎处处相等的函数看成是同一个函数，对于$x,y \in L^P[a,b]$，规定

$$\rho(x,y)=\left[\int_a^b |x(t)-y(t)|^p \mathrm{d}\,t\right]^{1/p}, p \geq 1$$

- 则$L^p[a,b]$构成一个距离空间，称之为p次幂可积函数空间

完备性概念

保证空间中的运算自洽

设 (X,ρ) 是度量空间，设 $\{x_n\}_{n=1}^{\infty}$ 是 X 中的点列, 如果对于任一正数ε,存在正数N, 使得当自然数 $n,m \geqslant N$ 时

$$\rho(x_n, x_m) < \varepsilon$$

就称 $\{x_n\}_{n=1}^{\infty}$ 是 X 中的基本点列, 或者称为 Cauchy,点列

如果度量空间 X 中每个基本,点列都收敛, 称 X 是完备度量空间

$C[a,b]$ 和 $L^P[a,b]$都是完备距离空间

线性空间和赋范空间

- 线性空间(线性空间)
  - 空间中的任意两点可以做加法或数乘，运算的结果仍未该空间的点
  - 并且该空间中的每个点可以定义长度，这个长度称为该点的范数，范数可以视为欧式空间中向量长度概念的推广。
- 由于赋范空间既有代数结构，又有拓扑结果，因此其空间结构较度量空间要丰富得多，其在实际中的应用也更加重要

线性空间

- 设V是一个非空集合，K是实(或复)数域，并可在其上定义“加法”，“数乘”运算，而且满足以下公理
  - 加法交换律：$x+y=y+x$
  - 加法结合律：$(x+y)+z= x+(y+z)$
  - 存在零元：$x+0=x$
  - 存在逆元：$x+(-x)=0$
  - 数乘：$1x=x$
  - $a(bx)= (ab)x$
  - $(a+b)x=ax+bx$
  - $a(x+y)=ax+ay$
- 则称V是数域K上的线性空间

赋范空间

- 设X是实(或复)线性空间，如果对于X中每个元素x，按照一定的法则对应于实数$||x||$，且满足
  - $||x||\geq 0$，$||x||=0$当且仅当x等于零元(x=0)
  - $||ax|| = |a|||x||$，a是实(或复)数
  - $||x+y|| \leq ||x|| + ||y||$
- 则称X是实(或复)赋范线性空间，$||x||$称为x的范数
- 赋范线性空间必然是距离空间
  - 定义$\rho(x, y) = ||x-y||$
- 与度量空间不同
  - 平移不变性：$d(x+a, y+a) = d(x, y)$，x,y,a 属于X
  - 齐次性：$d(ax, ay)=|a|d(x, y)$,x, y属于X，a属于K

巴拿赫(Banach)空间

- 如果赋范线性空间(X, ||.||)是完备的，则称(X, ||.||)是Banach空间
- 例子：n维Euclid空间 R^n 是Banach空间
  - $L^{\mathrm{p}}[a,b](p \geqslant 1)$ 是Banach空间, 定义范数
  - $\| x \| = \left[\int_a^b |x(t)|^p \,\mathrm{d}t\right]^{1/p}, p \geq 1$
- 例子： $C^k[a,b]$ 是Banach空间
  - $C^k[a,b]$ 表示定义在区间 $[a,b]$ 上 k 阶连续可导的函数全 体. 在 $C^k[a,b]$ 定义范数
  - $\| x \| = \sum_{j=0}^{k} \mathrm{m} \left|x^{(j)}(t)\right|, x^{(0)}(t) = x(t) \in C[a,b]$

有界线性算子

- T是由赋范线性空间X中的某个子集D到赋范线性空间中的一个映射，则称T 是算子
  - D是T 的定义域，记为D(T)，像集$\{y \mid y=Tx,\ x\in D\}$是T 的值域，记为T(D).
- 若T进一步满足
  - 可加性：$T(x+y)=Tx+Ty$
  - 齐次性：$T(ax)=aT(x)$
- 则T是线性算子
- 若存在正数M使得对于一切$x\in D(T)$，有$\|Tx\| \le M\|x\|$，则T是有界算子

有界线性算子

- 定理：设X 和X1都是赋范线性空间，在B(X, X1)中定义加法和数乘运算：
  - (T1+T2)x=T1x+T2x(T1,T2 ∈B(X, X1)，x ∈X)
  - (aT)x=a(Tx) (T∈B(X, X1)，a是数)
- 则B(X, X1)按照以上的线性运算是一个线性空间，并以前述算子范数的定义构成赋范线性空间
- 若X1是Banach空间，则B(X, X1)也是Banach空间

内积空间和Hilbert空间

- 几何化
  - 引入正交投影的概念
- 定义：设X 是定义在实(或复)数域K上的线性空间，若对于X 任意一对有序元素x,y, 恒对应数域K的值(x, y)，且满足
  - $(ax, y) = a(x, y)$；
  - $(x+y, z) = (x, z) + (x, z)$
  - $(x, y)=(y, x)$
  - $(x, x) \geq 0$，且$(x, x)=0$的充要条件是$x=0$
- 则称X为内积空间，(x, y)称为x, y的内积。

Hilbert空间

- 可由内积导出范数$|x| = \sqrt{(x, x)}$
- 完备的内积空间称为希尔伯特(Hilbert)空间
- Hilbert空间必为Banach空间

Hilbert空间

- $L^2[a,b]$ 中定义 x,y 的内积 (x,y) 为

$$(x,y)=\int_a^b x(t)\bar{y}(t)\mathrm{d}t, x(t),y(t)\in L^2[a,b]$$

- 因此, $L^2[a,b]$ 是变个可分的Hilbert空间
- $L^p[a,b](p\neq 2)$ 不可能诱导由范数诱导出内积空间

Hilbert空间

- 赋范线性空间X成为内积空间的充要条件是它的范数满足中线公式

$$\| x+y \|^2+\| x-y \|^2=2\| x \|^2+2\| y \|^2$$

- 而且内积可表示为

$$(x,y)=\frac{1}{4}(\| x+y \|^2-\| x-y \|^2+i\| x+iy \|^2-i\| x-y \|^2)$$

Hilbert空间

- 正交

$$(x, y) = 0$$

记为 $x \perp y$

- 正交补

$$A^{\perp} = \{x \mid x \perp A, x \in X\} \subset X$$

- 公股定理

$$x \perp y \Rightarrow \| x + y \|^2 = \| x \|^2 + \| y \|^2$$

- 正交分解：设 M 是内积空间 X 的完备子空间, 则对任意 $x \in X$, 均有以下唯一的正交分解

$$x = x_0 + z, x_0 \in M, z \in M^{\perp}$$

内积空间中的标准正交系

定义： 内积空间X中的元素列 $\{e_k\}$, 如果满足

$$(e_i, e_j) = 0, i \neq j$$

则称 $\{e_k\}$ 是一正交系. 进一步, 如果满足

$$(e_i, e_j) = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

则称 $\{e_k\}$ 是一标准正交系

有限维空间, 将给定向量展开成正交单位向量的线性组合 无限维空间, 将给定函数展开成Fourier级数