机器学习

李成龙

安徽大学人工智能学院

“多模态认知计算”安徽省重点实验室

内容安排

- 什么是机器学习
- 机器如何学习
- 如何让机器学习的更好
- 为什么机器能学习

内容安排

- **机器如何学习**
  - 有监督学习
    - 感知机
    - 支持向量机
    - 朴素贝叶斯分类
    - 决策树
    - 集成学习（Bagging算法与随机森林、Boosting算法）
    - 线性回归
    - 逻辑回归
    - Softmax回归
    - 神经网络与深度学习
  - 无监督学习
    - 聚类
    - 主成分分析

本节目录

- 概述
- 线性可分SVM与硬间隔最大化
- 线性SVM与软间隔最大化
- 非线性SVM与核函数

本节目录

- **概述**
- 线性可分SVM与硬间隔最大化
- 线性SVM与软间隔最大化
- 非线性SVM与核函数

概述

1、感知机的分类超平面不唯一问题

– 增加约束，如SVM中的最大化间隔

2、感知机无法解决非线性问题

– 使用核方法，映射到高维空间

概述

- 支持向量机(support vector machines. SVM)
  - 二分类模型
  - 基本模型是定义在特征空间上的间隔最大的线性分类器，间隔最大使它有别于感知机
  - 支持向量机还包括核技巧，这使它成为实质上的非线性分类器
  - 支持向量机的学习策略就是间隔最大化，可形式化为一个求解凸二次规划(convex quadratic programming)的问题，也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

概述

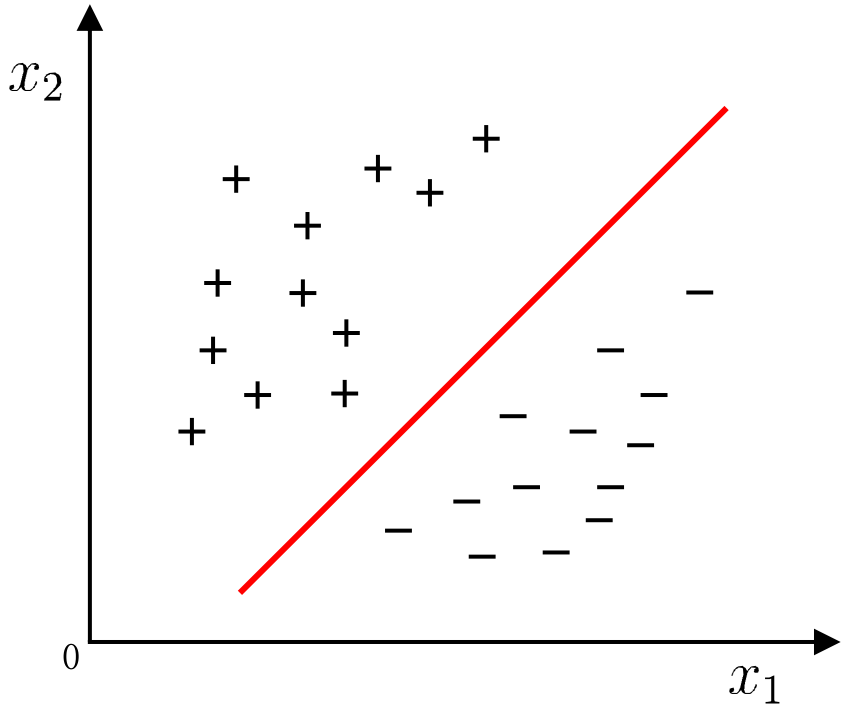
- **支持向量机**
  - 线性可分支持向量机(linear support vector machine in linearly separable case)
    - 硬间隔最大化(hard margin maximization)
  - 线性支持向量机(linear support vector machine)
    - 训练数据近似线性可分时，通过软间隔最大化(soft margin maximization)
  - 非线性支持向量机(non-linear support vector machine)
    - 当训练数据线性不可分时，通过使用核技巧(kernel trick)及软间隔最大化

本节目录

- 概述
- 线性可分SVM与硬间隔最大化
- 线性SVM与软间隔最大化
- 非线性SVM与核函数

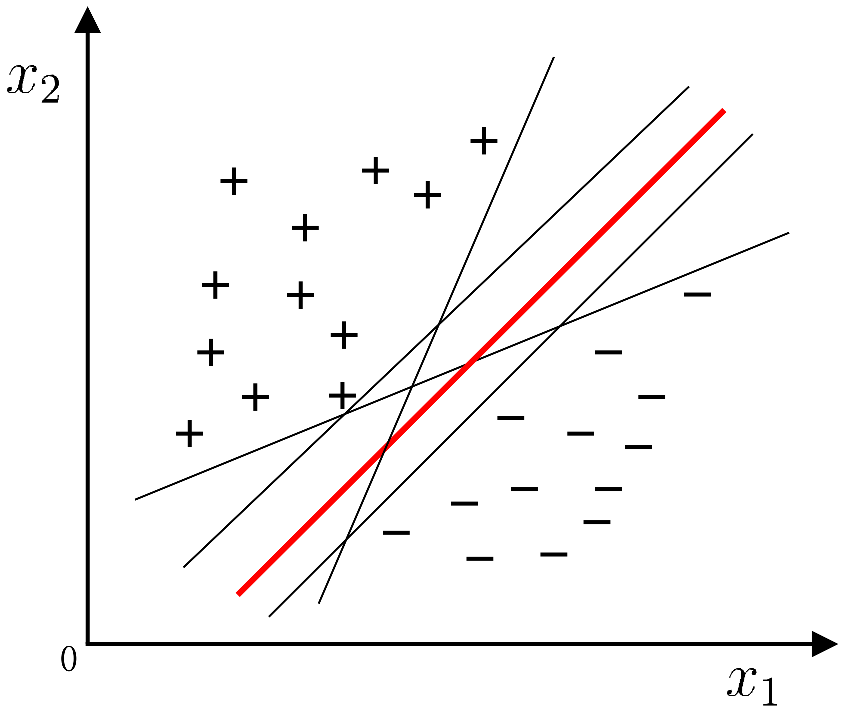
线性可分SVM与硬间隔最大化

- 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



线性可分SVM与硬间隔最大化

- 将训练样本分开的超平面可能有很多， 哪一个好呢?



线性可分SVM与硬间隔最大化

- **函数间隔**
  - 超平面方程: $\boldsymbol{w}^\top \boldsymbol{x} + b = 0$
  - 样本数据点 $\boldsymbol{x}$ 到该超平面的距离表示为: $|\mathbf{w}^{\mathrm{T}}\mathbf{x} + b|$
  - 判断 $\mathbf{w}^{\mathrm{T}}\mathbf{x} + b$ 与标记 y 符号是否一致
    - 若一致，则分类正确；否则，分类错误
  - 可用 $y(\mathbf{w}^{\mathrm{T}}\mathbf{x} + b)$ 替代 $|\mathbf{w}^{\mathrm{T}}\mathbf{x} + b|$ 表示样本数据点到超平面的距离
  - 函数间隔: $d_i = y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b)$
  - 样本数据集 D 的函数间隔 d 定义为 D 中所有样本数据点到超平面函数间隔的最小值，即有: $d = \min d_i$
  - d 作为优化指标构造SVM模型:

$$\max_{\mathbf{w},b} d, s.t.\ d_i \geq d$$

线性可分SVM与硬间隔最大化

- 几何间隔
  - 函数间隔的缺点：若同时同比例缩放其参数向量$\boldsymbol{w}$和偏置量b，会改变函数间隔d_i的取值
  - 几何间隔
    - 解决函数间隔d_i取值混乱的问题，对参数向量$\boldsymbol{w}$作归一化后计算得到的函数间隔
    - $\hat{d}_i$为函数间隔d_i所对应的几何间隔，则有

$$\hat{d}_i = \frac{y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b)}{\|\boldsymbol{w}\|} = \frac{d_i}{\|\boldsymbol{w}\|}$$

    - 优化问题

$$\max_{\boldsymbol{w},b} \hat{d} = \frac{d}{\|\boldsymbol{w}\|}, s.t. \hat{d}_i \geq \hat{d}$$

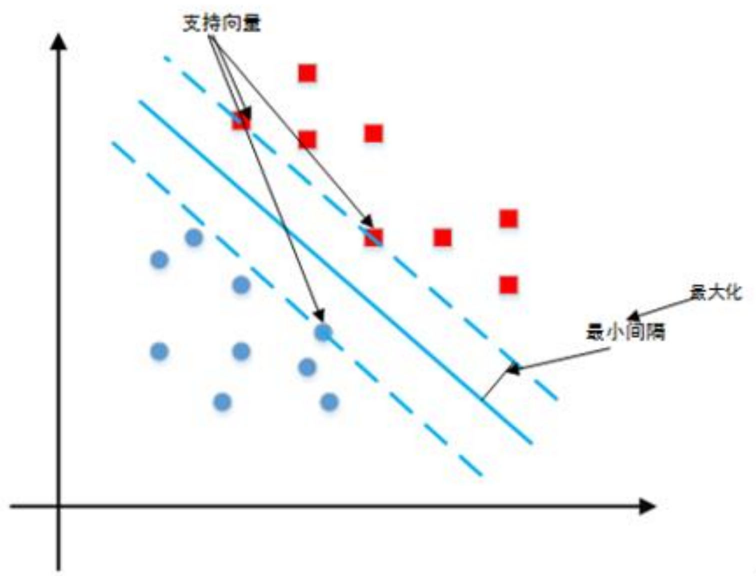
线性可分SVM与硬间隔最大化

- **函数间隔**
  - 求得该优化问题的解$\boldsymbol{w}^*$和b^*，得到最优分离超平面:

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

  - SVM模型：

$$f(\boldsymbol{x})=\mathrm{sgn}(\boldsymbol{w}^{*\mathrm{T}}\boldsymbol{x}+b^*)$$



线性可分SVM与硬间隔最大化

- 几何间隔
  - 函数间隔的缺点：若同时同比例缩放其参数向量w和偏置量b，会改变函数间隔d_i的取值
  - 几何间隔
    - 可在对参数向量w作归一化缩放基础上对其做进一步适当缩放，使得

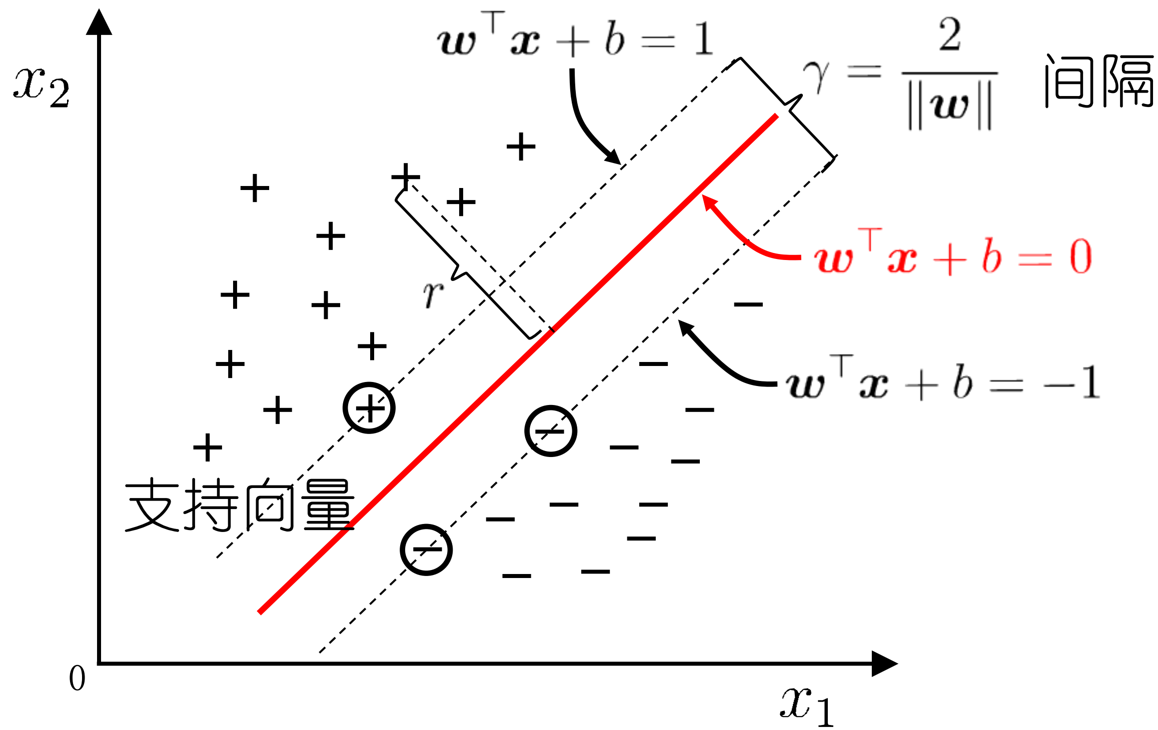
$$d = 1$$

    - 优化问题转化为

$$\max_{w,b} \frac{1}{\|\boldsymbol{w}\|}, s.t.\ y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1$$

线性可分SVM与硬间隔最大化

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



线性可分SVM与硬间隔最大化

安徽大学 ANHUI UNIVERSITY

- **支持向量机基本型：线性可分SVM**
  - 最大间隔：寻找参数w和b, 使得 γ 最大

$$\max_{w,b} \frac{1}{\|\boldsymbol{w}\|}$$

$$s.t.\ y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1, i = 1,2,\ldots,n$$

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

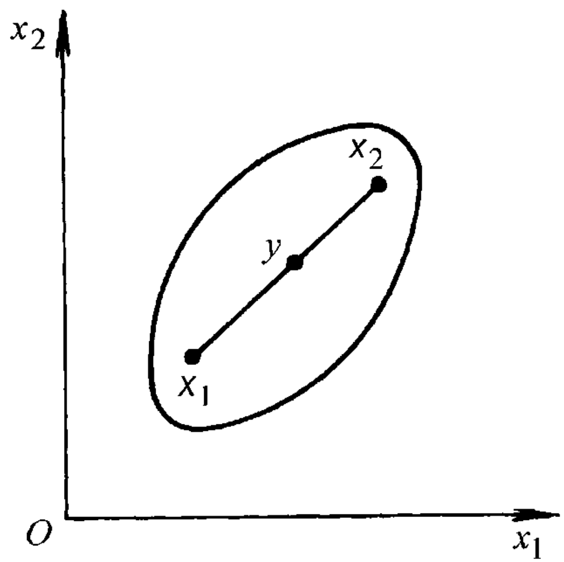
$$s.t.\ y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b)\text{-}1 \geq 0, i = 1,2,\ldots,n$$

  - 凸二次规划问题(convex quadratic programming)

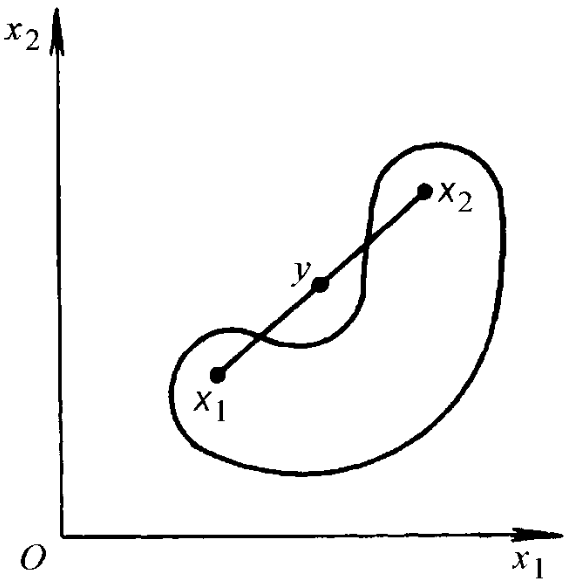
线性可分SVM与硬间隔最大化

- **补充知识**
  - 凸集

一个点集（或区域），如果连接其中任意两点x_1, x_2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



凸集



非凸集

线性可分SVM与硬间隔最大化

- **补充知识**
  - 凸性条件
    - 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 x_1, x_2，不等式 $f(x_2) \geq f(x_1) + (x_2 - x_1)^T \nabla f(x_1)$ 恒成立

    - 根据二阶导数（ Hessen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：Hessen矩阵在R上处处半正定

线性可分SVM与硬间隔最大化

- **补充知识**
  - 凸规划

对于约束优化问题

$$\min f(x)$$
$$s.t.\quad g_j(x) \le 0 \qquad j = 1, 2, \ldots, m$$

若 $f(x)$ 和 $g_j(x)$ 都为凸函数，则此问题为凸规划

线性可分SVM与硬间隔最大化

安徽大学 ANHUI UNIVERSITY

- 补充知识
  - 凸规划的性质
    - 若给定一点x_0, 则集合 $R = \{x \mid f(x) \le f(x_0)\}$ 为凸集
    - 可行域 $R = \{x \mid g_j(x) \le 0, j = 1, 2, \ldots, m\}$ 为凸集
    - 凸规划的任何局部最优解就是全局最优解

线性可分SVM与硬间隔最大化

- **补充知识**
  - 凸优化问题
    - 凸优化问题: 指约束最优化问题

$$\min_{w} f(\boldsymbol{w}), s.t.\ g_i(\boldsymbol{w}) \le 0, i=1,2,\ldots,k,\ \ h_i(\boldsymbol{w})=0, i=1,2,\ldots,l$$

其中，目标函数$f(\boldsymbol{w})$和约束函数$g_i(\boldsymbol{w})$都是R^n上连续可微的凸函数，约束函数$h_j(\boldsymbol{w})$是R^n上的仿射函数

- 当目标函数为二次函数，g函数为仿射函数时，为凸二次规划问题

线性可分SVM与硬间隔最大化

- **学习算法流程**
  - 输入：线性可分训练数据集$T=\{(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)\}$

  其中 $x \in \mathcal{X} \in R^n \quad y \in \mathcal{Y} \in \{-1,1\}$
  - 输出：最大间隔分离超平面和分类决策函数
    - 构造并求解约束最优化问题，求得$\boldsymbol{w}^*$和b^*

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

$$s.t.\ y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\text{-}1 \ge 0, i=1,2,\dots,n$$

    - 得到分离超平面

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

    - 分类决策函数

$$f(\boldsymbol{x}_i)=\mathrm{sgn}(\boldsymbol{w}^{*\mathrm{T}}\boldsymbol{x}_i+b^*)$$

线性可分SVM与硬间隔最大化

安徽大学 ANHUI UNIVERSITY

- **对偶问题**
  - 目标函数是二次的，约束条件是线性的，所以它是一个凸二次规划问题
  - QP（Quadratic Programming）的优化包都可以进行求解：在一定的约束条件下，目标最优，损失最小
  - 结构特殊：通过 Lagrange Duality 变换到对偶变量（dual variable）的优化问题之后，可以找到一种更加有效的方法来进行求解
  - 通常情况下这种方法比直接使用通用的 QP 优化包进行优化要高效得多
  - 可以自然的引入核函数，进而推广到非线性分类问题

线性可分SVM与硬间隔最大化

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

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

  - 上式可以转换为

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

  - 满足以下KKT条件时，上式等价于 $\max_{\alpha_i \geq 0} \min_{w,b} \mathcal{L}(w, b, \alpha)$

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

线性可分SVM与硬间隔最大化

安徽大學 ANHUI UNIVERSITY

- **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

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

线性可分SVM与硬间隔最大化

- 使用序列最小化算法SMO求解，得到最优参数向量α^*，则

$$\boldsymbol{w}^*=\sum_{i=1}^{m}\alpha_i^* y_i \boldsymbol{x}_i; b_j^* = y_j - \sum_{i=1}^{m}\alpha_i^* y_i \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j, where\ \alpha_j \neq 0$$

其中，偏移项b：通过某个支持向量来确定

- **最终模型**

$$f(\boldsymbol{x}_j) = \mathrm{sgn}(\boldsymbol{w}^{*\mathrm{T}} \boldsymbol{x}_j + b^*) = \mathrm{sgn}(\sum_{i=1}^{m}\alpha_i^* y_i \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j + b^*)$$

线性可分SVM与硬间隔最大化

- **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.

线性可分SVM与硬间隔最大化

- **求解算法-SMO**（Sequential Minimal Optimization）
  - 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**.

线性可分SVM与硬间隔最大化

安徽大学 ANHUI UNIVERSITY

- 练习：给定正例x_1=(3,3)，x_2=(4,3)，负例x_3=(1,1)，计算支持向量、分类决策函数和分类超平面方程
  - 解：对偶形式

$$
\begin{aligned}
\min_{\alpha}\quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j)-\sum_{i=1}^{N}\alpha_i \\
& =\frac{1}{2}(18\alpha_1^2+25\alpha_2^2+2\alpha_3^2+42\alpha_1\alpha_2-12\alpha_1\alpha_3-14\alpha_2\alpha_3)-\alpha_1-\alpha_2-\alpha_3 \\
\text{s.t.}\quad & \alpha_1+\alpha_2-\alpha_3=0 \\
& \alpha_i\geqslant 0,\quad i=1,2,3
\end{aligned}
$$

把 $\alpha_3=\alpha_1+\alpha_2$ 代入目标函数，得到：

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

对 α_1,α_2 求偏导，得到(α_1,α_2) 在 $\left(\frac{3}{2},-1\right)^{\mathrm{T}}$ 取极值，但该点不满足约束条件 $\alpha_2\geqslant 0$，所以最小值应在边界上达到

线性可分SVM与硬间隔最大化

安徽大学 ANHUI UNIVERSITY

- 练习：给定正例x_1=(3,3)，x_2=(4,3)，负例x_3=(1,1)，计算支持向量、分类决策函数和分类超平面方程
  - 当$\alpha_1=0$时，最小值 $s\left(0,\frac{2}{13}\right)=-\frac{2}{13}$；当$\alpha_2=0$ 时，最小值 $s\left(\frac{1}{4},0\right)=-\frac{1}{4}$

于是 $s(\alpha_1,\alpha_2)$ 在 $\alpha_1=\frac{1}{4},\alpha_2=0$ 获得极小，则 $\alpha_3=\alpha_1+\alpha_2=\frac{1}{4}$

$\alpha_1^*=\alpha_3^*=\frac{1}{4}$ 对应的实例向量为支持向量

计算得到：
$$w_1^*=w_2^*=\frac{1}{2}$$
$$b^*=-2$$

分类决策函数为：$f(x)=\mathrm{sign}\left(\frac{1}{2}x^{(1)}+\frac{1}{2}x^{(2)}-2\right)$

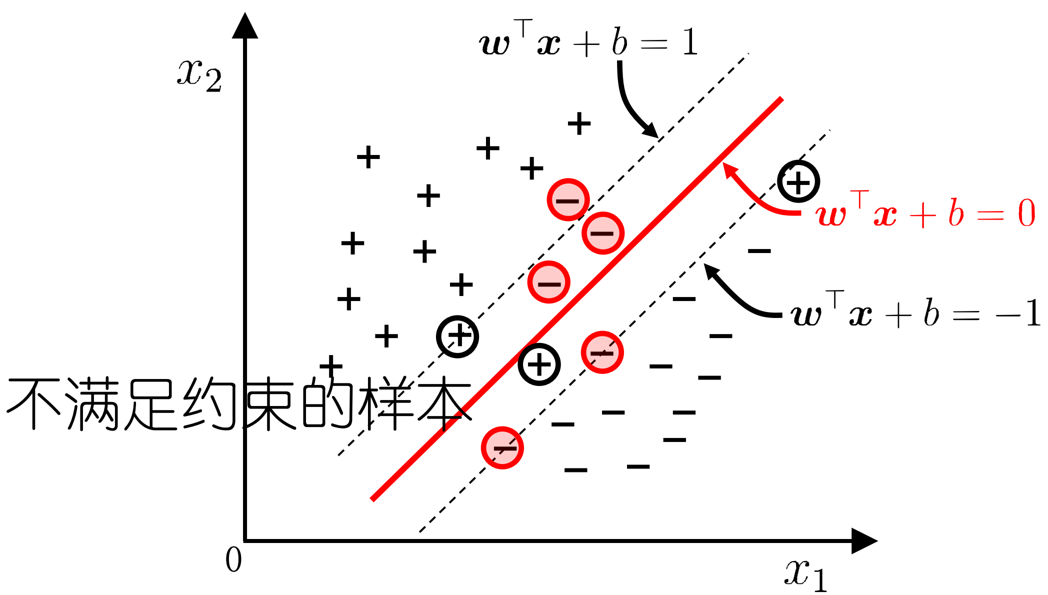
分类超平面方程为：$\frac{1}{2}x^{(1)}+\frac{1}{2}x^{(2)}-2=0$

本节目录

- 概述
- 线性可分SVM与硬间隔最大化
- 线性SVM与软间隔最大化
- 非线性SVM与核函数

线性SVM与软间隔最大化

- 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
- 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SVM与软间隔最大化

安徽大学 ANHUI UNIVERSITY

- **线性SVM模型**
  - 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
  - 引入一个松弛变量 ξ_i，将约束条件转化为

    $$y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1\text{-}\xi_i$$

    其中，ξ_i的取值越大，模型对错误分类的容忍程度越高
  - 线性SVM模型的目标函数

    $$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}\xi_i, s.t.\ y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0$$

    其中$C > 0$是为惩罚因子

线性SVM与软间隔最大化

- **软间隔最大化**
  - 上述优化问题对应的拉格朗日函数为

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

  - 即有

$$\min_{\alpha}\frac{1}{2}\sum_{i=1}^{m}\sum_{i=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j-\sum_{i=1}^{m}\alpha_i, s.t.\sum_{i=1}^{m}\alpha_i y_i=0,\ 0\le\alpha_i\le C,\mu_i=C-\alpha_i$$

  - 计算最优参数向量α^*后，可求得对应的软间隔支持向量机模型

$$f(\boldsymbol{x}_j)=\mathrm{sgn}(\boldsymbol{w}^{*\mathrm{T}}\boldsymbol{x}_j+b^*)=\mathrm{sgn}(\sum_{i=1}^{m}\alpha_i^* y_i\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j+b^*)$$

其中，$\boldsymbol{w}^*=\sum_{i=1}^{m}\alpha_i^* y_i\boldsymbol{x}_i; b_j^*=y_j-\sum_{i=1}^{m}\alpha_i^* y_i\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j, where\ \alpha_j\neq 0,\xi_j=0$

线性SVM与软间隔最大化

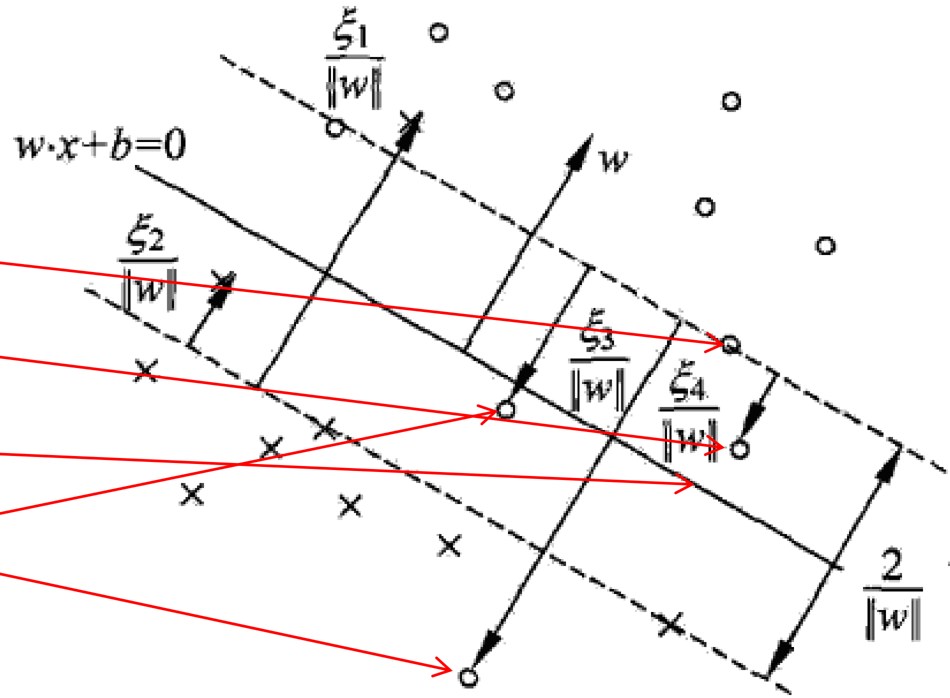
- 线性SVM的支持向量

若 $\alpha_i^* < C$，则 $\xi_i = 0$

若 $\alpha_i^* = C$，$0 < \xi_i < 1$

若 $\alpha_i^* = C$，$\xi_i = 1$

若 $\alpha_i^* = C$，$\xi_i > 1$



线性SVM与软间隔最大化

- **线性SVM的另一种解释：合页损失函数**
  - 对于优化问题：

$$\min_{w,b}\sum_{i=1}^{m}L(y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b))+\lambda\|\boldsymbol{w}\|^2$$

其中L称为合页损失函数，定义如下：

$$L(y(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b))=[1\text{-}y(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b)]_+,\quad [z]_+=\begin{cases} z, z>0 \\ 0, z\le 0\end{cases}$$

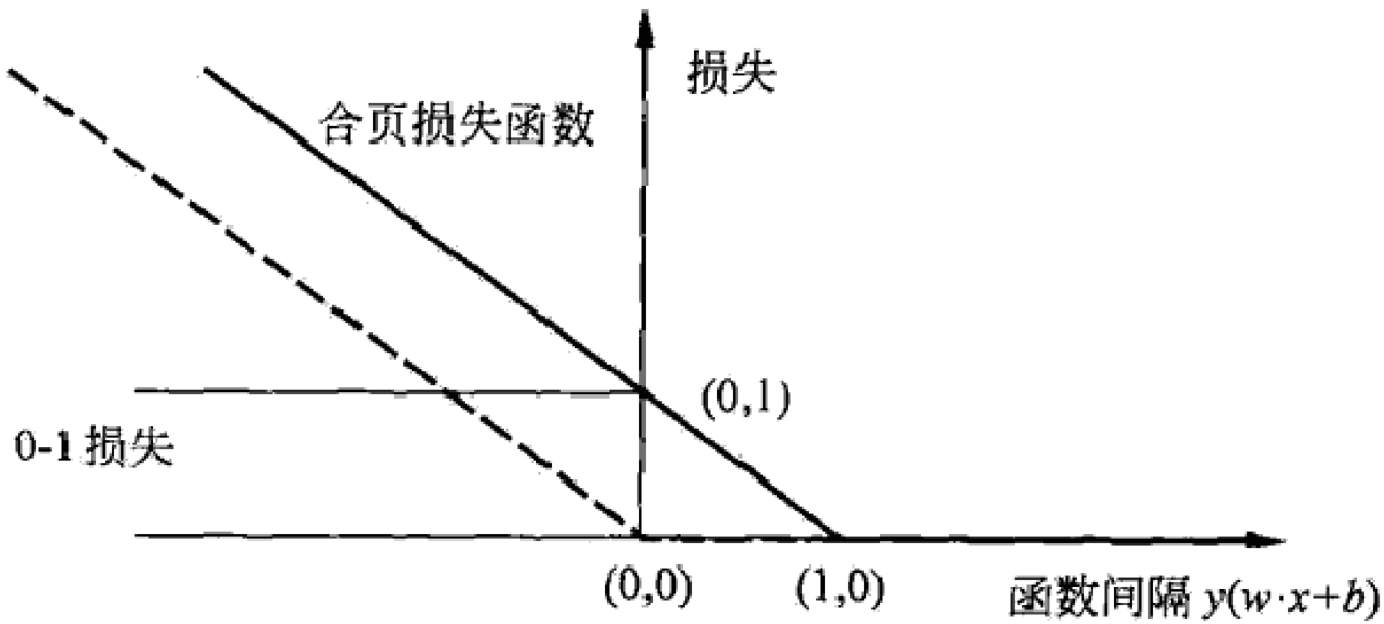
上述优化问题和线性SVM优化问题是等价的

$$\min_{w,b}\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{m}\xi_i, s.t.\ y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)\ge 1-\xi_i,\ \xi_i\ge 0$$

思考： 如何证明等价性？

线性SVM与软间隔最大化

- 线性SVM的另一种解释：**合页损失函数**



本节目录

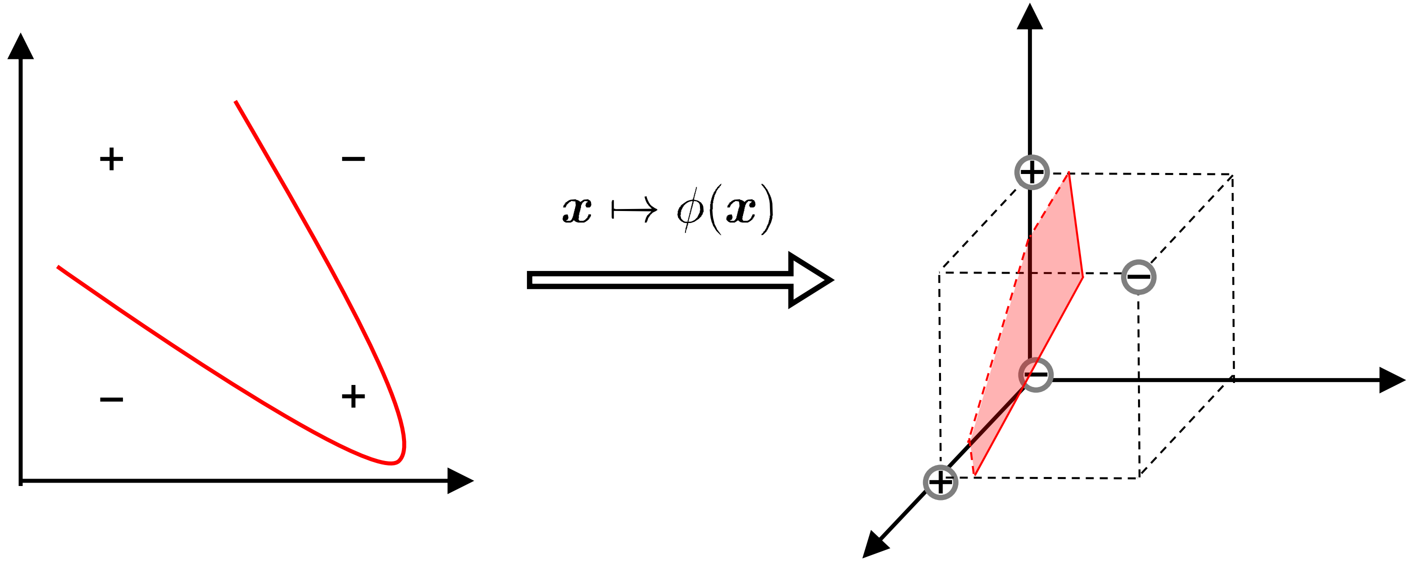
- 概述
- 线性可分SVM与硬间隔最大化
- 线性SVM与软间隔最大化
- 非线性SVM与核函数

非线性SVM与核函数

- **线性不可分情况**
  - 若不存在一个能正确划分两类样本的超平面，怎么办？
  - 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

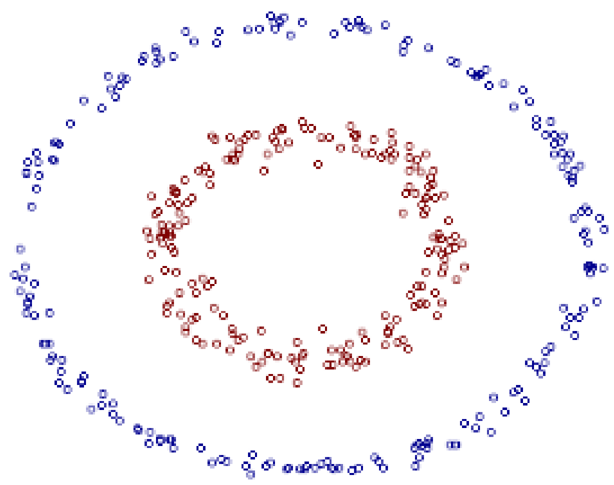
非线性SVM与核函数

- **线性不可分情况**
  – 举个简单例子
    - 线性分类面不存在
    - 从二维转换到三维：$x_3 = x_1 * x_2$



非线性SVM与核函数

- **处理非线性数据**
  - 举个复杂例子
    - 用两个半径不同的圆圈加上了少量的噪音生成得到如下数据



    - 一个理想的分界应该是一个“圆圈”而不是一条线（超平面）

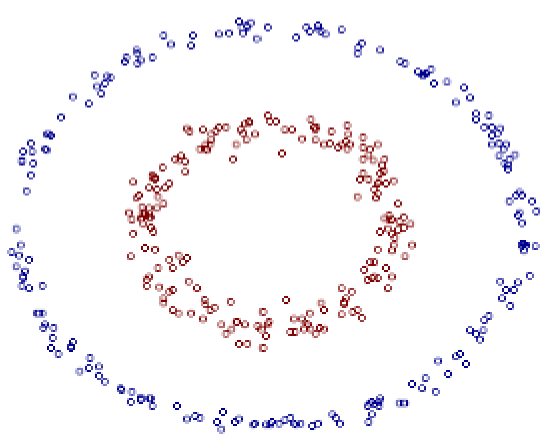
非线性SVM与核函数

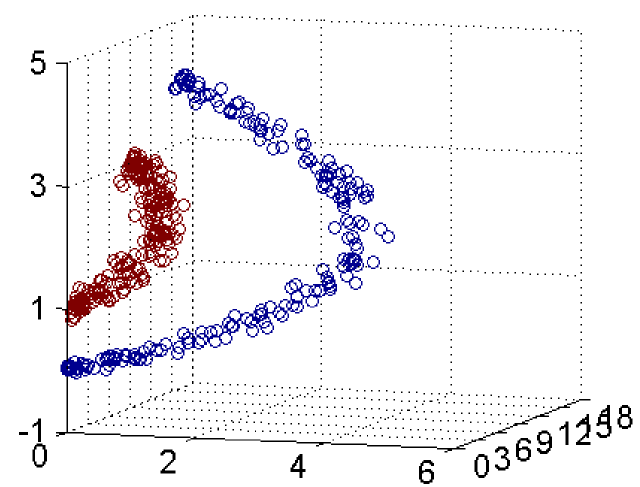
- **处理非线性数据**
  - 举个复杂例子
    - 一条二次曲线（圆圈是二次曲线的一种特殊情况）的方程可以写作这样的形式

$$a_1X_1 + a_2X_1^2 + a_3X_2 + a_4X_2^2 + a_5X_1X_2 + a_6 = 0$$

    - 构造另外一个五维的空间，其中五个坐标的值分别为X_1， X_1^2， X_2， X_2^2， X_1X_2, 那上述方程可以写成

$$\sum_{i=1}^{5} a_i Z_i + a_6 = 0$$





非线性SVM与核函数

- **处理非线性数据**
  - 维度的增大为计算带来了非常大的困难，而且如果遇到无穷维的情况，就根本无从计算了
  - 例子：设两个向量 $x_1=(\eta_1,\eta_2)^T$，$x_2=(\xi_1,\xi_2)^T$，$\phi(\cdot)$ 表示之前2维到5维的映射，映射过后的内积是

$$\langle\phi(x_1),\phi(x_2)\rangle=\eta_1\xi_1+\eta_1^2\xi_1^2+\eta_2\xi_2+\eta_2^2\xi_2^2+\eta_1\eta_2\xi_1\xi_2$$

    又注意到：$(\langle x_1,x_2\rangle+1)^2=2\eta_1\xi_1+\eta_1^2\xi_1^2+2\eta_2\xi_2+\eta_2^2\xi_2^2+2\eta_1\eta_2\xi_1\xi_2+1$
  - 只要把某几个维度线性缩放一下，然后再加上一个常数维度，两者相等，区别是
    - 一个是映射到高维空间中，然后再根据内积的公式进行计算
    - 另一个则直接在原来的低维空间中进行计算，而不需要显式地写出映射后的结果
  - 核函数：$\kappa(x_1,x_2)=(\langle x_1,x_2\rangle+1)^2$

支持向量机—核函数

- **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

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

只以**内积**的形式出现

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

  - 预测

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

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

支持向量机—核函数

- **核函数的本质**
  - 实际中，我们会经常遇到线性不可分的样例，此时，我们的常用做法是把样例特征映射到高维空间中去
  - 如果凡是遇到线性不可分的样例，一律映射到高维空间，那么这个维度大小是会高到可怕的
  - 核函数的价值在于它虽然也是将特征进行从低维到高维的转换，但核函数的本质是在它事先在低维上进行计算，而将实质上的分类效果表现在了高维上，也就避免了直接在高维空间中的复杂计算

总结

- **概述**
  - 解决了感知机遇到的两个关键问题
- **线性可分SVM与硬间隔最大化**
  - 函数间隔与几何间隔
  - SVM基本型
  - 对偶问题及其求解
- **线性SVM与软间隔最大化**
  - 特异点与软间隔
  - 线性SVM模型
- **非线性SVM与核函数**
  - 非线性问题
  - 核函数
  - 非线性SVM模型

思考题

- 线性可分SVM对噪声点敏感吗？为什么？
- 对于线性SVM通过引入软间隔的方式解决不满足约束条件的样本点，而非线性SVM能够处理这种情形吗？为什么？

练习题

- 推导线性SVM如下变形的对偶形式

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}\xi_i^2, s.t.\ y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1 - \xi_i,\ \xi_i > 0$$