支持向量机入门指北

Apple Zhang

Shenzhen University

April 8, 2023

目录

目录

SHENZHEN UNIVERSITY 深圳大学

符号定义

- 第i个训练样本: $\mathbf{x}_i = [x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(d)}]^T \in \mathbb{R}^d$.
- 第i个训练样本的标签: $y_i \in \{-1, 1\}$.
- 训练样本矩阵: $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{d\times n}$.
- 训练样本标签向量: $\mathbf{y} = [y_1, y_2, \cdots, y_n] \in \{-1, 1\}^n$.
- 超平面的法向量: $\mathbf{w} \in \mathbb{R}^d$.
- 超平面的偏移: $b \in \mathbb{R}$
- L_2范数:

$$\|x\|_2 = \sqrt{\sum_{i=1}^{d} x_i^2} = \sqrt{\mathbf{x}^T\mathbf{x}}.$$

- 损失函数: $\mathcal{L}(\cdot)$.
- 拉格朗日函数: $L(\cdot)$.
- 求梯度/求导: ∇f.

什么是超平面？

- 在$\mathbb{R}^2$空间中，一条直线可表示为点集

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

- 在$\mathbb{R}^3$空间中，一个平面可表示为点集

$$\left\{(x^{(1)}, x^{(2)}, x^{(3)}) \mid w_1 x^{(1)} + w_2 x^{(2)} + w_3 x^{(3)} + b = 0\right\}.$$

- 在$\mathbb{R}^d$空间中，一个超平面可表示为点集

$$\{\mathbf{x} \mid \mathbf{w}^T \mathbf{x} + b = 0\}$$

即：

$$\left\{(x^{(1)}, x^{(2)}, \cdots, x^{(d)}) \mid w_1 x^{(1)} + w_2 x^{(2)} + \cdots + w_d x^{(d)} + b = 0\right\}$$

数据点到超平面的距离

- 在 $\mathbb{R}^2$ 空间中, 数据点 $\mathbf{x}_i = [x_i^{(1)}, x_i^{(2)}]^T$ 到一条直线的距离:

$$\frac{|w_1 x_i^{(1)} + w_2 x_i^{(2)} + b|}{\sqrt{w_1^2 + w_2^2}}.$$

- 在 $\mathbb{R}^3$ 空间中, 数据点 $\mathbf{x}_i = [x_i^{(1)}, x_i^{(2)}, x_i^{(3)}]^T$ 到一个平面的距离:

$$\frac{|w_1 x_i^{(1)} + w_2 x_i^{(3)} + w_2 x_i^{(3)} + b|}{\sqrt{w_1^2 + w_2^2 + w_3^2}}.$$

- 在 $\mathbb{R}^d$ 空间中, 数据点 $\mathbf{x}_i = [x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(d)}]^T$ 到超平面的距离:

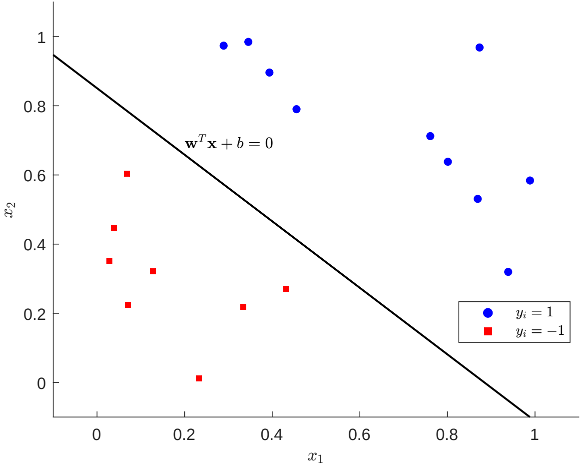
$$\frac{|w_1 x_i^{(1)} + w_2 x_i^{(2)} + \cdots + w_d x_i^{(d)} + b|}{\sqrt{w_1^2 + w_2^2 + \cdots + w_d^2}}$$

$$= \frac{|\mathbf{w}^T \mathbf{x}_i + b|}{\|\mathbf{w}\|_2}.$$

分类问题——线性可分

问题:

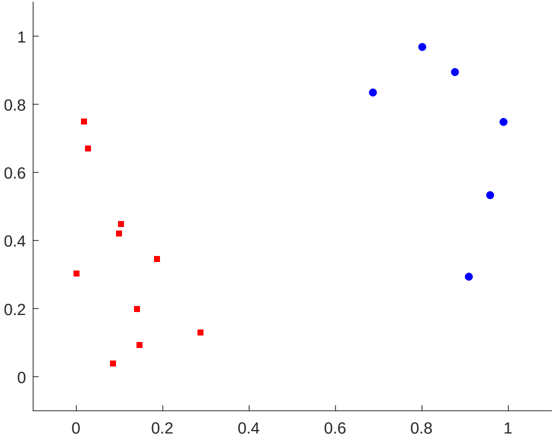
对于一个二分类问题，假设数据是线性可分的 (linear separatable)，如何找到一个超平面把两类数据分开?



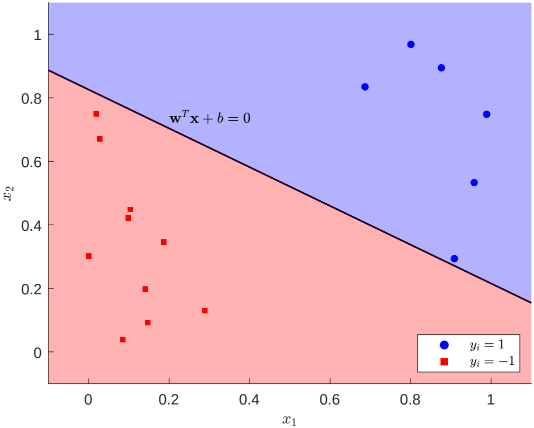
关键词:

- 超平面: $\mathbf{w}^T\mathbf{x}+b=0$.
- 预测函数: $h(\mathbf{x})=\text{sgn}(\mathbf{w}^T\mathbf{x}+b)$.
- 线性可分: 存在至少一个超平面，使得 $\forall i,\ y_i=h(\mathbf{x}_i)$.

如何对这个数据分类



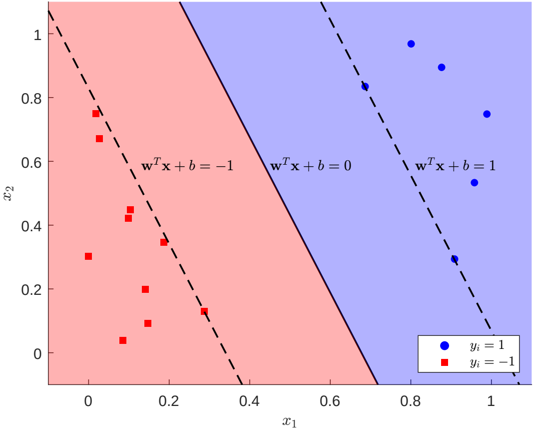
如何对这个数据分类



目录

最大间隔分类器: 支持向量机

SVM 寻找最大化两个类间的间隔的超平面[1].

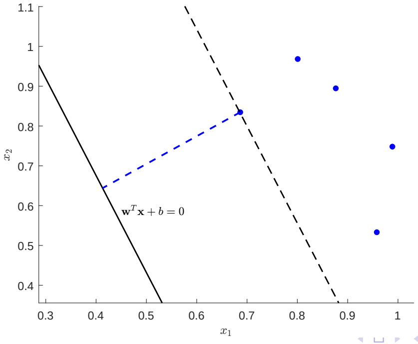


SVM 的优化目标

定义 (间隔)

样本集 $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ 到超平面 $\mathbf{w}^T\mathbf{x} + b = 0$ 的间隔，等于离超平面最近的那个点到超平面的距离:

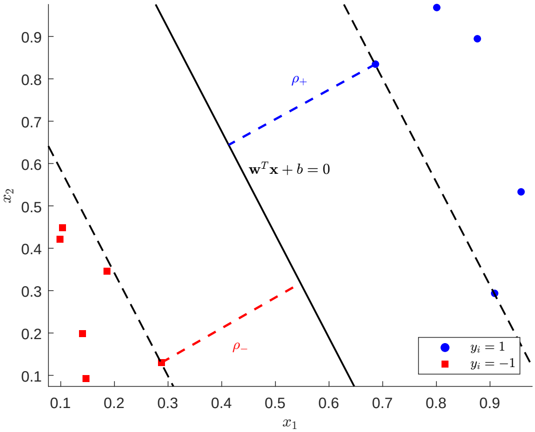
$$\rho = \min_{i=1,2,\cdots,n} \frac{|\mathbf{w}^T\mathbf{x}_i + b|}{\|\mathbf{w}\|_2} = \min_{i=1,2,\cdots,n} \frac{y_i(\mathbf{w}^T\mathbf{x}_i + b)}{\|\mathbf{w}\|_2} \tag{1}$$



SVM 的优化目标

总间隔等于正类样本和负类样本到超平面 $\mathbf{w}^T\mathbf{x} + b = 0$ 的间隔之和:

$$\rho = \rho_+ + \rho_-. \tag{2}$$



SVM 的优化目标

- 假设 $\rho_+ = \rho_-$，即分类超平面处于中间；（为什么？）
- 则可以设超平面（虚线）为 $\mathbf{w}^T\mathbf{x} + b = 1$ 和 $\mathbf{w}^T\mathbf{x} + b = -1$；（为什么？）
- 所以超平面有以下约束；（为什么？）

$$\begin{cases} \mathbf{w}^T\mathbf{x}_i + b \geq 1, & \text{if } y_i = 1, \\ \mathbf{w}^T\mathbf{x}_i + b \leq -1, & \text{if } y_i = -1. \end{cases} \Leftrightarrow y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1. \tag{3}$$

- 因此，两类样本到超平面 $\mathbf{w}^T\mathbf{x} + b = 0$ 的间隔均为：（为什么？）

$$\rho_+ = \rho_- = \frac{1}{\|\mathbf{w}\|_2}. \tag{4}$$

SVM 的优化问题

要最大化两个类之间的间隔, 则有

$$\max_{\mathbf{w},b} \frac{2}{\|\mathbf{w}\|_2}, \quad \text{s.t. } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1. \tag{5}$$

同时，由于

$$\max_{\mathbf{w},b} \frac{2}{\|\mathbf{w}\|_2} \Leftrightarrow \min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|_2^2.$$

所以可以将最大化间隔的问题转化为如下的优化问题

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|_2^2, \quad \text{s.t. } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1.$$

带不等式约束的优化问题

定义以下带不等式约束的优化问题:

$$\min_{\mathbf{u}} f_0(\mathbf{u}), \quad \text{s.t. } f_i(\mathbf{u}) \leq 0,\ i = 1, 2, \cdots, n. \tag{6}$$

定义 (6) 对应的拉格朗日函数为

$$L(\mathbf{u}, \boldsymbol{\alpha}) = f_0(\mathbf{u}) + \sum_{i=1}^{n} \alpha_i f_i(\mathbf{u}). \tag{7}$$

在 SVM 的问题中， $\mathbf{u} = (\mathbf{w}, b)$，以及

$$f_0(\mathbf{u}) = \frac{1}{2}\|\mathbf{w}\|^2, \quad f_i(\mathbf{u}) = 1 - y_i(\mathbf{w}^T\mathbf{x}_i + b). \tag{8}$$

用对偶问题求解 SVM

SVM 的拉格朗日函数为

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|_2^2 + \sum_{i=1}^{n} \alpha_i [1 - y_i(\mathbf{w}^T \mathbf{x}_i + b)]. \tag{9}$$

定理

求解 *SVM* 原问题等价于求解如下的**对偶问题**:

$$\max_{\boldsymbol{\alpha} \geq \mathbf{0}} \min_{\mathbf{w}, b} \; L(\mathbf{w}, b, \boldsymbol{\alpha}). \tag{10}$$

$$= \max_{\boldsymbol{\alpha} \geq \mathbf{0}} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j, \quad \text{s.t.} \; \sum_{i=1}^{n} \alpha_i y_i = 0. \tag{11}$$

用对偶问题求解 SVM

$$\max_{\alpha \geq \mathbf{0}} \min_{\mathbf{w}, b} \frac{1}{2}\|\mathbf{w}\|_2^2 + \sum_{i=1}^{n} \alpha_i \left[1 - y_i \left(\mathbf{w}^T \mathbf{x}_i + b\right)\right].$$

首先对于 $\mathbf{w}, b$ 最小化 $L(\mathbf{w}, b, \boldsymbol{\alpha})$:

$$\nabla_{\mathbf{w}} L = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i,$$

$$\nabla_b L = 0 \Rightarrow 0 = \sum_{i=1}^{n} \alpha_i y_i.$$

将以上条件代入，得到如下的二次规划问题:

$$\max_{\boldsymbol{\alpha} \geq \mathbf{0}} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j, \quad \text{s.t.} \sum_{i=1}^{n} \alpha_i y_i = 0.$$

用对偶问题求解 SVM

化为标准二次规划的矩阵形式:

$$\min_{\boldsymbol{\alpha}} \ \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha} - \mathbf{1}^T\boldsymbol{\alpha}, \quad \text{s.t. } \alpha_i \geq 0, \ \mathbf{y}^T\boldsymbol{\alpha} = 0.$$

这里 $\mathbf{1}$ 是全1的列向量, $\mathbf{Q} \in \mathbb{R}^{n\times n}$的每一个元素是

$$Q_{ij} = y_i y_j \mathbf{x}_i^T \mathbf{x}_j.$$

实际实现中可以调用 Matlab 的函数: `quadprog`[1], 对于Python, 可以使用`cvxopt` 中的 `solvers.qp`[2] 函数.
最后, 我们可以把预测函数写为

$$h(\mathbf{x}) = \operatorname{sgn}(\mathbf{w}^T\mathbf{x} + b) = \operatorname{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b\right).$$

[1] 需要 Matlab Optimization toolbox
[2] 需要 numpy-mkl

小贴士

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{Q}\boldsymbol{\alpha} - \mathbf{1}^T\boldsymbol{\alpha}, \quad \text{s.t. } \alpha_i \geq 0,\ \mathbf{y}^T\boldsymbol{\alpha} = 0.$$

- 上式解出的α是稀疏的. 如果 $\alpha_i > 0$, 则其对应的训练样本 $\mathbf{x}_i$ 称为支持向量(**Support Vector**, SV).
- b可以用下面的公式算出:

$$b = \frac{1}{\|\boldsymbol{\alpha}\|_0} \sum_{\alpha_i > 0} \left(\frac{1}{y_i} - \sum_{j=1}^{n} \alpha_j y_j \mathbf{x}_j^T \mathbf{x}_i \right).$$

其中$\|\cdot\|_0$表示L_0范数算子, 返回向量非零元素个数.

小贴士

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha} - \mathbf{1}^T\boldsymbol{\alpha}, \quad \text{s.t. } \alpha_i \geq 0,\ \mathbf{y}^T\boldsymbol{\alpha} = 0.$$

Matlab函数: `x = quadprog(H, f, A, b, Aeq, Beq, lb, ub, x0);`

$$\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{f}^T\mathbf{x},\ \text{s.t.} \begin{cases} \mathbf{A}\cdot\mathbf{x} \leq \mathbf{b} \\ \mathbf{Aeq}\cdot\mathbf{x} = \mathbf{beq} \\ lb \leq \mathbf{x} \leq ub \end{cases}$$

因此对偶问题可以按以下方式求解:

$$\texttt{alpha = quadprog}(\mathbf{Q},\ -\mathbf{1},\ \texttt{[]},\ \texttt{[]},\ \mathbf{y}^T,\ 0,\ \mathbf{0},\ \texttt{[]},\ \mathbf{x}_0)$$

注意: 需要安装Matlab Optimization toolbox.

目录

SHENZHEN UNIVERSITY 深圳大学

初探对偶原理

对于下面的不等式优化问题以及拉格朗日函数

$$\min_{\mathbf{u}} f_0(\mathbf{u}), \quad \text{s.t. } f_i(\mathbf{u}) \le 0,\ i = 1, 2, \cdots, n, \tag{12}$$

$$L(\mathbf{u}, \boldsymbol{\alpha}) = f_0(\mathbf{u}) + \sum_{i=1}^{n} \alpha_i f_i(\mathbf{u}). \tag{13}$$

我们有以下结论

定理

设原问题(12)的最优值为 p^*，则

$$p^* = \min_{\mathbf{u}} \max_{\boldsymbol{\alpha}} L(\mathbf{u}, \boldsymbol{\alpha}), \quad \text{s.t. } \alpha_i \ge 0. \tag{14}$$

初探对偶原理

证明: 定义 $\theta(\mathbf{u})$ 为

$$\begin{aligned}\theta(\mathbf{u}) &= \max_{\boldsymbol{\alpha}\geq\mathbf{0}} L(\mathbf{u},\boldsymbol{\alpha}) \\ &= \max_{\boldsymbol{\alpha}\geq\mathbf{0}} f_0(\mathbf{u}) + \sum_{i=1}^{n} \alpha_i f_i(\mathbf{u}).\end{aligned} \tag{15}$$

可以看到，拉格朗日函数是关于 α_i 的线性函数，于是分以下两种情况讨论：

- $\forall i \in \{1,2,\cdots,n\},\ f_i(\mathbf{u}) \leq 0 \Rightarrow \theta(\mathbf{u}) = f_0(\mathbf{u})$.
- $\exists i \in \{1,2,\cdots,n\},\ f_i(\mathbf{u}) > 0 \Rightarrow \theta(\mathbf{u}) = +\infty$.

所以

$$\theta(\mathbf{u}) = \begin{cases} f_0(\mathbf{u}), & \forall i \in \{1,2,\cdots,n\},\ f_i(\mathbf{u}) \leq 0, \\ +\infty, & \text{otherwise}. \end{cases}$$

深圳大学 SHENZHEN UNIVERSITY

初探对偶原理

$$\theta(\mathbf{u}) = \begin{cases} f_0(\mathbf{u}), & \forall i \in \{1, 2, \cdots, n\},\ f_i(\mathbf{u}) \le 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

考虑 $\theta(\mathbf{u})$ 的最小值，由于必定有 $f_0(\mathbf{u}) < +\infty$，所以

$$\begin{aligned} \min_{\mathbf{u}} \theta(\mathbf{u}) &= \min_{\mathbf{u}} f_0(\mathbf{u}), \quad \text{s.t. } f_i(\mathbf{u}) \le 0,\ i = 1, 2, \cdots, n \\ &= p^*. \end{aligned} \tag{16}$$

也就是说，

$$\min_{\mathbf{u}} \max_{\boldsymbol{\alpha} \ge 0} L(\mathbf{u}, \boldsymbol{\alpha}) = \min_{\mathbf{u}} \theta(\mathbf{u}) = p^*. \tag{17}$$

初探对偶原理

我们假设对偶问题的最优值为

$$d^* = \max_{\boldsymbol{\alpha} \geq \mathbf{0}} \min_{\mathbf{u}} L(\mathbf{u}, \boldsymbol{\alpha}) \tag{18}$$

问题： 下面的等式一定成立吗？

$$p^* = \underbrace{\min_{\mathbf{u}} \max_{\boldsymbol{\alpha} \geq \mathbf{0}} L(\mathbf{u}, \boldsymbol{\alpha})}_{\text{原问题}} = \underbrace{\max_{\boldsymbol{\alpha} \geq \mathbf{0}} \min_{\mathbf{u}} L(\mathbf{u}, \boldsymbol{\alpha})}_{\text{对偶问题}} = d^* \tag{19}$$

下面这个不等式呢？

$$p^* = \underbrace{\min_{\mathbf{u}} \max_{\boldsymbol{\alpha} \geq \mathbf{0}} L(\mathbf{u}, \boldsymbol{\alpha})}_{\text{原问题}} \geq \underbrace{\max_{\boldsymbol{\alpha} \geq \mathbf{0}} \min_{\mathbf{u}} L(\mathbf{u}, \boldsymbol{\alpha})}_{\text{对偶问题}} = d^* \tag{20}$$

初探对偶原理

- $p^* \geq d^*$ 称为**弱对偶性**，这个性质对任何优化问题**总是成立**.
- $p^* = d^*$ 称为**强对偶性**，只有一部分优化问题满足这个性质.

Slater 条件是原问题满足**强对偶性**的充分条件:

定理 (*Slater 条件)

设定义在 $\mathcal{D}$ 上的函数 $f_i(\cdot), i = 0, 1, \cdots, n$ 均为凸函数，对于优化问题

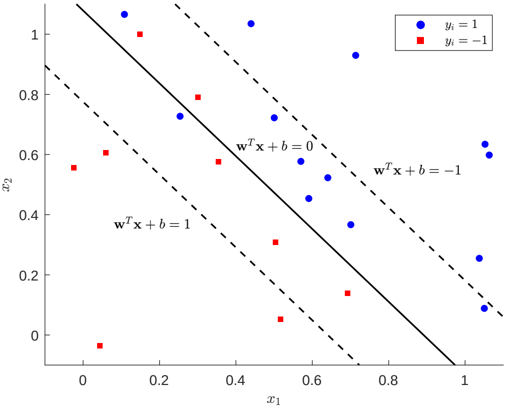
$$\min_{\mathbf{u}} f_0(\mathbf{u}), \quad \text{s.t. } f_i(\mathbf{u}) \leq 0.$$

如果存在点 $\tilde{\mathbf{u}} \in \mathbf{relint}\mathcal{D}$ （即定义域 $\mathcal{D}$ 的相对内点），使得 $f_i(\tilde{\mathbf{u}}) < 0$, 则强对偶性成立.

目录

如果数据分布不理想...

硬间隔SVM 的要求是数据严格线性可分，但一般这个条件不成立.
我们更多使用的是**软间隔SVM**.



软间隔SVM

软间隔 SVM 的优化问题:

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \frac{1}{2}\|\mathbf{w}\|_2^2 + c\sum_{i=1}^{n}\xi_i,$$
$$\text{s.t. } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0.$$

其中 c 是惩罚参数. 上述问题的拉格朗日函数:

$$L(\mathbf{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\|\mathbf{w}\|_2^2 + c\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}\left[\alpha_i\left(1 - \xi_i - y_i(\mathbf{w}^T\mathbf{x}_i + b)\right)\right] - \sum_{i=1}^{n}\beta_i\xi_i. \tag{21}$$

对偶问题:

$$\max_{\boldsymbol{\alpha}} \mathbf{1}^T\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha}, \quad \text{s.t. } 0 \leq \alpha_i \leq c,\ \mathbf{y}^T\boldsymbol{\alpha} = 0.$$

软间隔SVM

软间隔 SVM 等价于以下的结构化损失最小化模型

$$\min_{\mathbf{w},b} \underbrace{\frac{1}{2}\|\mathbf{w}\|_2^2}_{L_2\text{正则}} + c\sum_{i=1}^{n} \underbrace{\max\left(0, 1 - y_i(\mathbf{w}^T\mathbf{x}_i + b)\right)}_{L_1\text{铰链损失}}. \tag{22}$$

从这个优化模型出发，可以得出很多 SVM 的变种模型：

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|_2^2 + c\sum_{i=1}^{n} \max\left(0, 1 - y_i(\mathbf{w}^T\mathbf{x}_i + b)\right)^2, \tag{23}$$

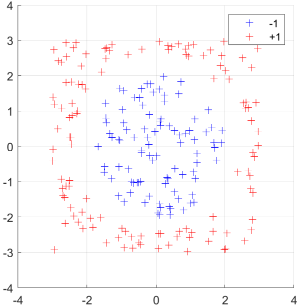
$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|_2^2 + c\sum_{i=1}^{n} \max\left(0, 1 - y_i\mathbf{w}^T\mathbf{x}_i\right). \tag{24}$$

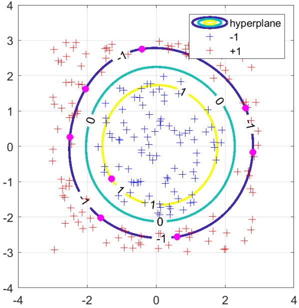
深圳大学
SHENZHEN UNIVERSITY

目录

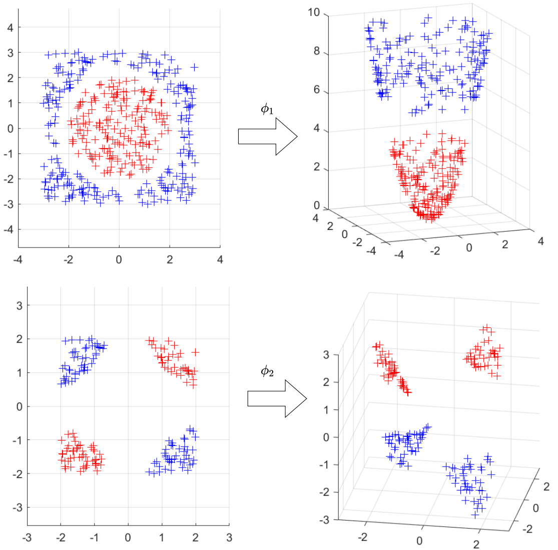
非线性分类: 核 SVM

前面使用的都是线性模型. 如果数据是非线性分布的, 则效果会较差.
核 SVM: 通过核方法完成非线性分类.





非线性分类: 核 SVM



深圳大学
SHENZHEN UNIVERSITY

核技巧

定义 (正定对称核)

定义函数 $\mathcal{K}:\mathbb{R}^d\times\mathbb{R}^d\mapsto\mathbb{R}$. 若对任意在 $\mathbb{R}^d$ 上的数据集: $\{\mathbf{x}_1,\cdots,\mathbf{x}_n\}$, 矩阵 $\mathbf{K}\in\mathbb{R}^{n\times n}$ 满足 $K_{ij}=\mathcal{K}(\mathbf{x}_i,\mathbf{x}_j)$, 且 $\mathcal{K}$ 是对称正定的, 则称 $\mathcal{K}$ 是正定对称核.

对任意正定对称核, 存在唯一的特征空间 $\mathcal{H}$ 以及映射 $\phi:\mathbb{R}^d\mapsto\mathcal{H}$, 使得

$$\mathcal{K}(\mathbf{x}_i,\mathbf{x}_j)=\langle\phi(\mathbf{x}_i),\phi(\mathbf{x}_j)\rangle=\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j). \tag{25}$$

高斯核函数（或 RBF 核函数）是最常见的核函数之一:

$$\mathcal{K}(\mathbf{x}_i,\mathbf{x}_j)=\exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|_2^2}{t}\right),\quad t>0. \tag{26}$$

核函数可以隐式确定一个非线性映射 $\phi(\cdot)$, 于是不需要知道 $\phi(\cdot)$ 的具体表达式就可以计算非线性映射后的内积.

核 SVM

对于经过非线性映射后的数据 $\phi(\mathbf{x})$, 运用软间隔 SVM, 求以下超平面:

$$\mathbf{w}^T\phi(\mathbf{x}) + b = 0. \tag{27}$$

对应的 SVM 优化问题是:

$$\begin{aligned}&\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|_{\mathcal{H}}^2 + c\sum_{i=1}^{n}\xi_i,\\ &\text{s.t. } y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i,\ \xi_i \geq 0.\end{aligned} \tag{28}$$

实际上, 只是将软间隔 SVM 中的 $\mathbf{x}_i$ 都替换成了 $\phi(\mathbf{x}_i)$.

深圳大学 SHENZHEN UNIVERSITY

核技巧

拉格朗日函数:

$$L(\mathbf{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\|\mathbf{w}\|_{\mathcal{H}}^2 + c\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}\left[\alpha_i\left(1 - \xi_i - y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b)\right)\right] - \sum_{i=1}^{n}\beta_i\xi_i. \tag{29}$$

$$\nabla_{\mathbf{w}} L = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^{n}\alpha_i y_i \phi(\mathbf{x}_i),$$

$$\nabla_b L = 0 \Rightarrow \sum_{i=1}^{n}\alpha_i y_i = 0,$$

$$\nabla_{\xi_i} L = 0 \Rightarrow c - \alpha_i - \beta_i = 0.$$

核方法

对偶问题:

$$\max_{\boldsymbol{\alpha}} \mathbf{1}^T\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Q}^{\varphi}\boldsymbol{\alpha},$$
$$\text{s.t. } 0 \le \alpha_i \le c,\ \mathbf{y}^T\boldsymbol{\alpha} = 0.$$

其中 $\mathbf{Q}^{\varphi} \in \mathbb{R}^{n\times n}$ 定义为:

$$Q_{ij}^{\varphi} = y_i y_j \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) = y_i y_j \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j).$$

注意: 在核 SVM 中, 测试样本 $\mathbf{x}$ 的标签需要通过以下方式计算:

$$h(\mathbf{x}) = \operatorname{sgn}(\mathbf{w}^T\phi(\mathbf{x}) + b) = \operatorname{sgn}\left(\sum_{i=1}^{n}\alpha_i y_i \mathcal{K}(\mathbf{x}_i, \mathbf{x}) + b\right).$$

核 SVM 的预测必须依赖训练集.

目录

更多有关SVM...

高效的SVM解决方案(C++实现): **LIBSVM, LIBLINEAR**[2, 3].

- 已有C++, Java, MATLAB, Python接口.
- 贼快.

其他SVM的变种:

- (1999 NPL) Least Square Support Vector Machine (LSSVM) [4].
- (2001 JMLR) Crammar-Singer Support Vector Machine (CSSVM) [5].
- (2006 JMLR) Laplacian Support Vector Machine (Lap-SVM) [6].
- (2007 PAMI) Twin Support Vector Machine (TWSVM) [7].
- (2019 NN) Capped-norm Twin Support Vector Machine (CTWSVM)[8].
- (2022 TCYB) Robust Manifold Twin Bounded SVM [9].
- (2023 TCYB) Maximal Margin SVM [10]

结束

谢谢!

https://github.com/Apple-Zhang/SVM-Intro

深圳大学 SHENZHEN UNIVERSITY

参考文献

[1] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[2] C.-C. Chang and C.-J. Lin, “Libsvm: a library for support vector machines,” *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, pp. 1–27, 2011.

[3] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, “Liblinear: A library for large linear classification,” *the Journal of machine Learning research*, vol. 9, pp. 1871–1874, 2008.

[4] J. A. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[5] K. Crammer and Y. Singer, “On the algorithmic implementation of multiclass kernel-based vector machines,” *Journal of machine learning research*, vol. 2, no. Dec, pp. 265–292, 2001.

[6] M. Belkin, P. Niyogi, and V. Sindhwani, “Manifold regularization: A geometric framework for learning from labeled and unlabeled examples.,” *Journal of machine learning research*, vol. 7, no. 11, 2006.

[7] R. Khemchandani, S. Chandra, *et al.*, “Twin support vector machines for pattern classification,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 29, no. 5, pp. 905–910, 2007.

[8] C. Wang, Q. Ye, P. Luo, N. Ye, and L. Fu, “Robust capped l1-norm twin support vector machine,” *Neural Networks*, vol. 114, pp. 47–59, 2019.

[9] J. Zhang, Z. Lai, H. Kong, and L. Shen, “Robust twin bounded support vector classifier with manifold regularization,” *IEEE Transactions on Cybernetics*, pp. 1–16, 2022.

[10] Z. Lai, X. Chen, J. Zhang, H. Kong, and J. Wen, “Maximal margin support vector machine for feature representation and classification,” *IEEE Transactions on Cybernetics*, 2023.