
支持向量机

陈飞宇

fchen@cqu.edu.cn

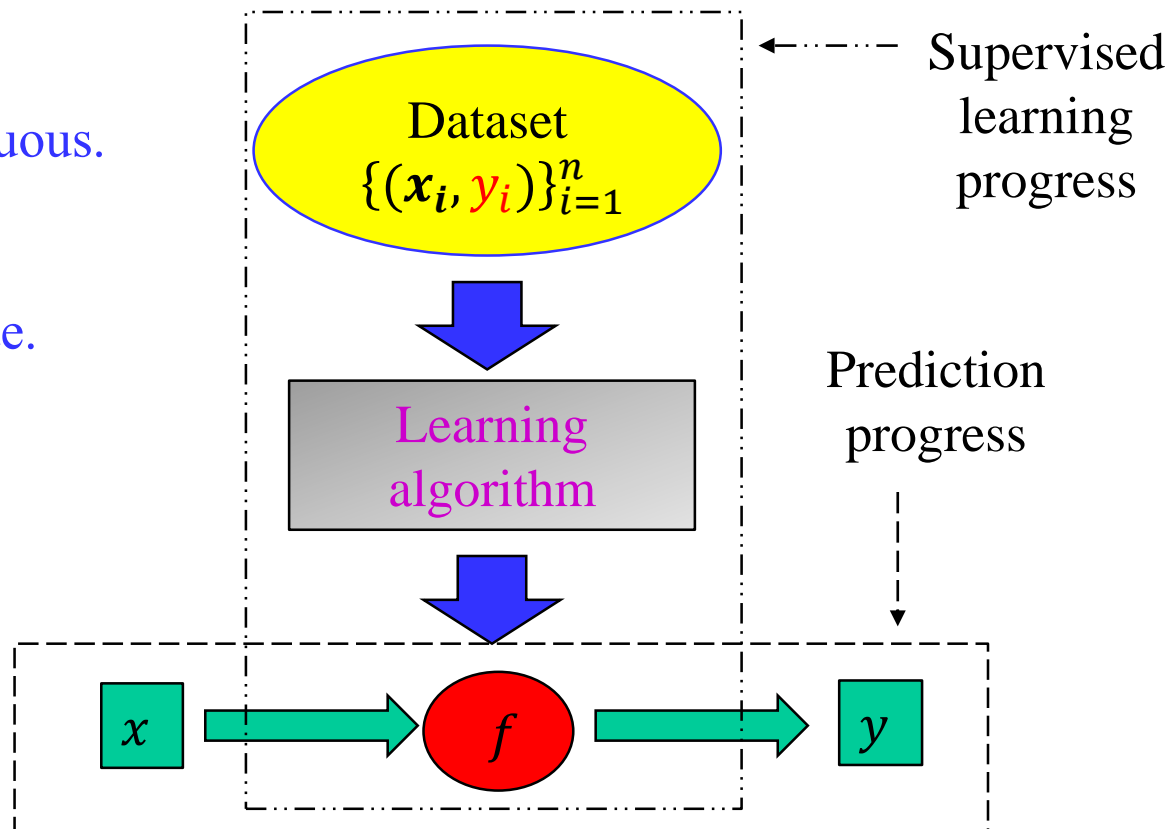
办公室：软件学院529

监督学习 (Supervised Learning)

Learning with a **teacher**

Regression:
 y_i 's are continuous.

Classification:
 y_i 's are discrete.



5.1 SVM模型

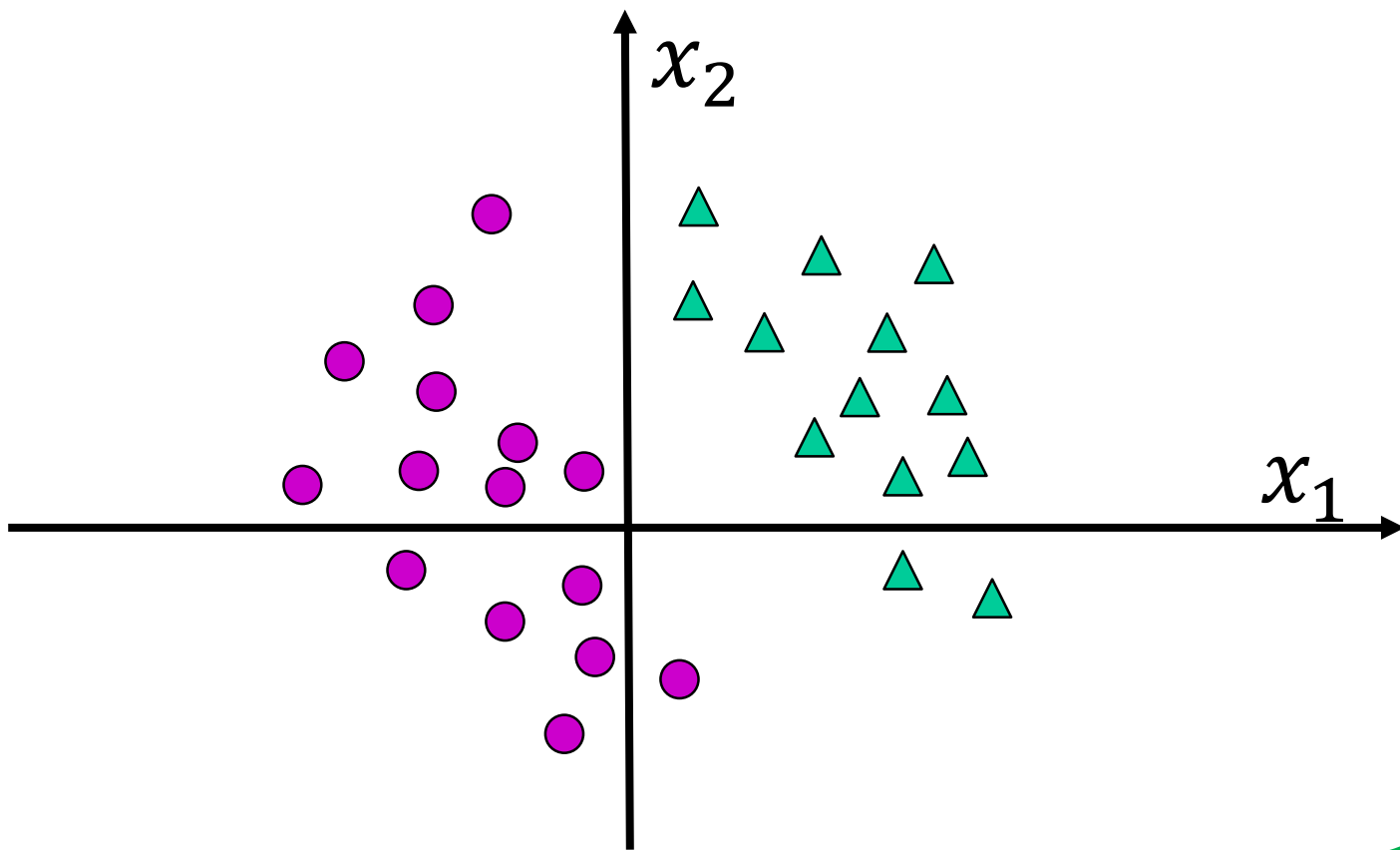
陈飞宇

fchen@cqu.edu.cn

办公室：软件学院529

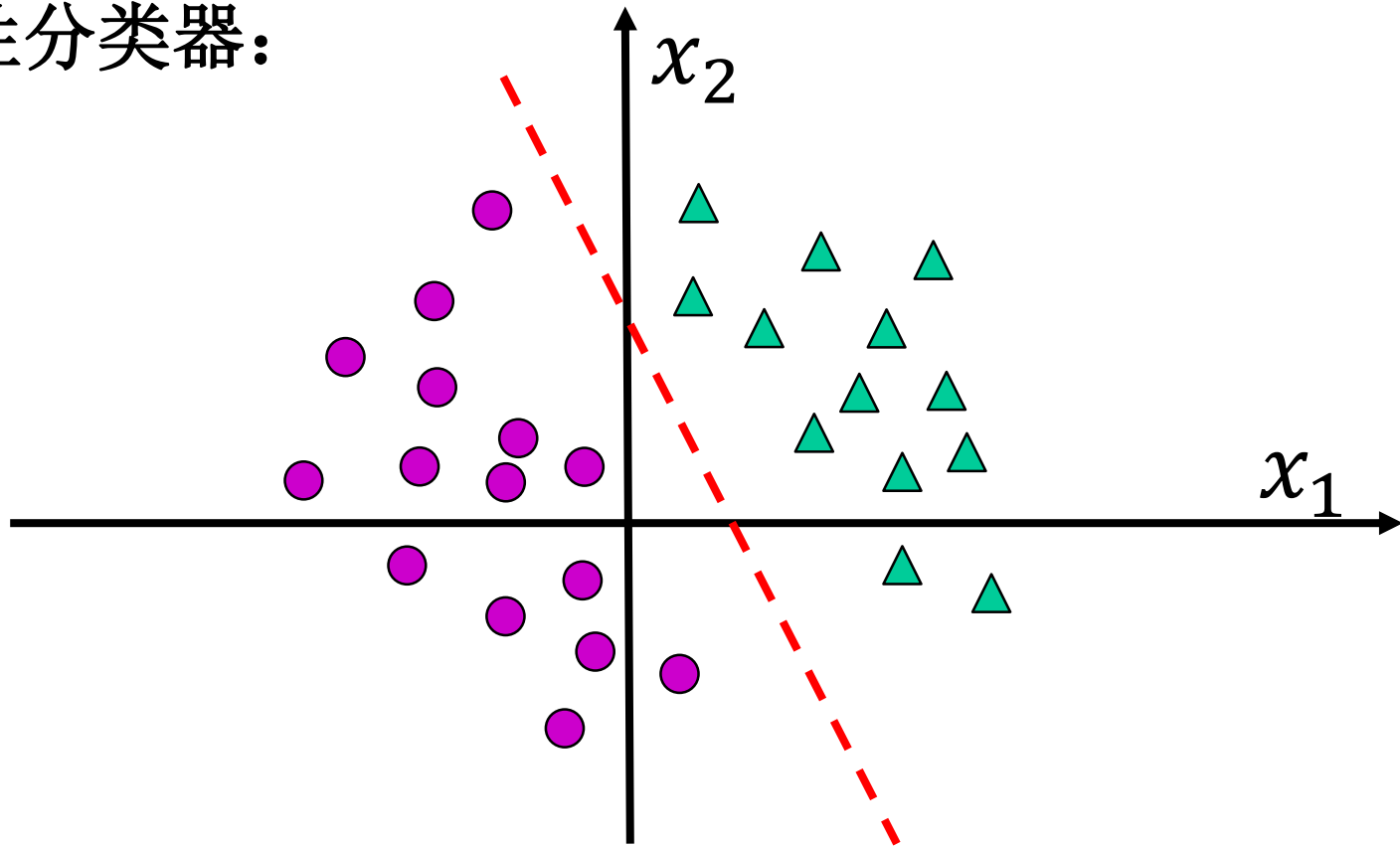
二分类 (Binary Classification)

给定训练集 $\{(x_i, y_i)\}_{i=1}^n$, 寻求合适的分类器?

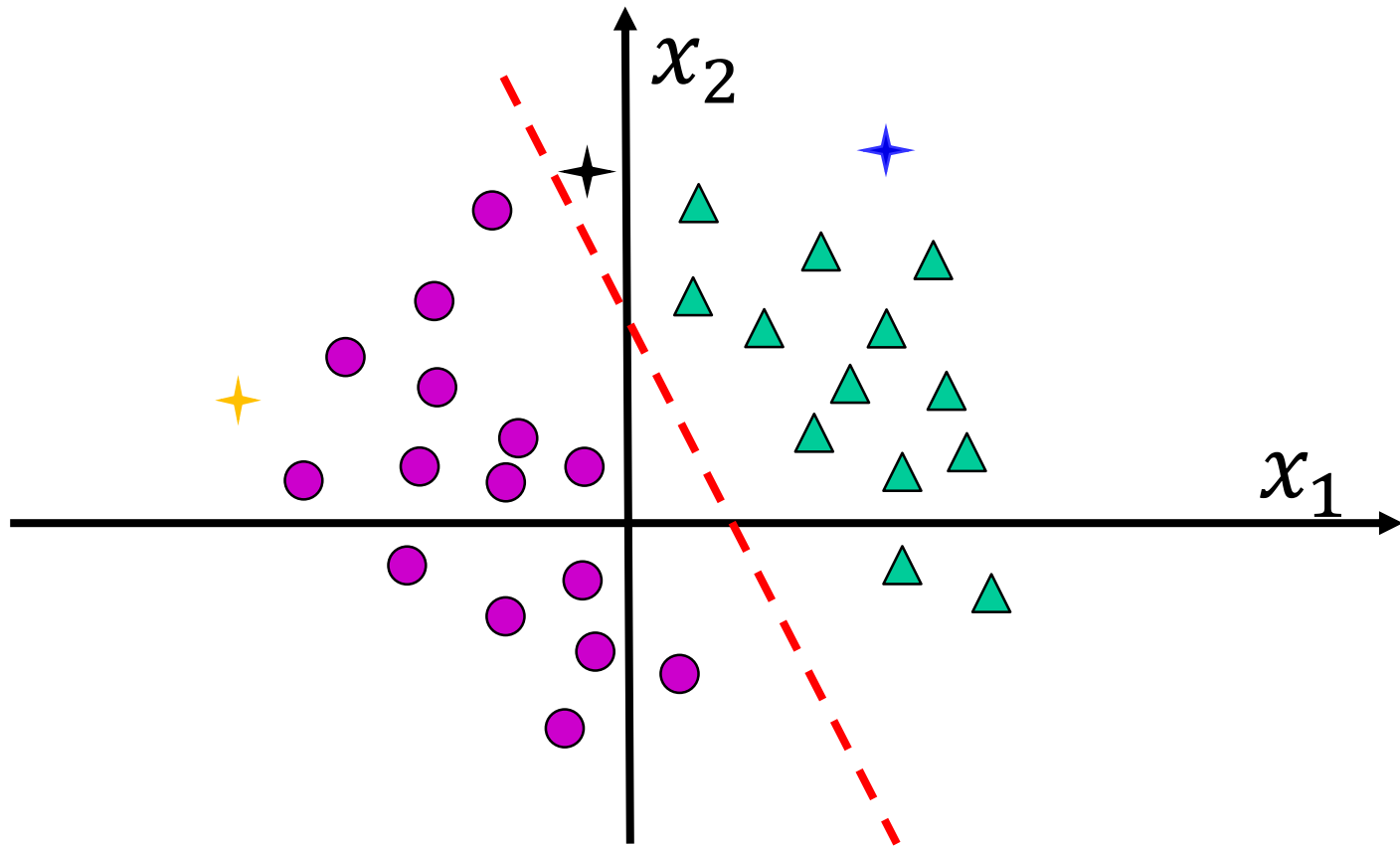


归纳 (Induction)

线性分类器:



演绎 (Deduction)



问题？

数据点 \star , \star 在直线上方，为正样本。

数据点 \star 在直线下方，为负样本。

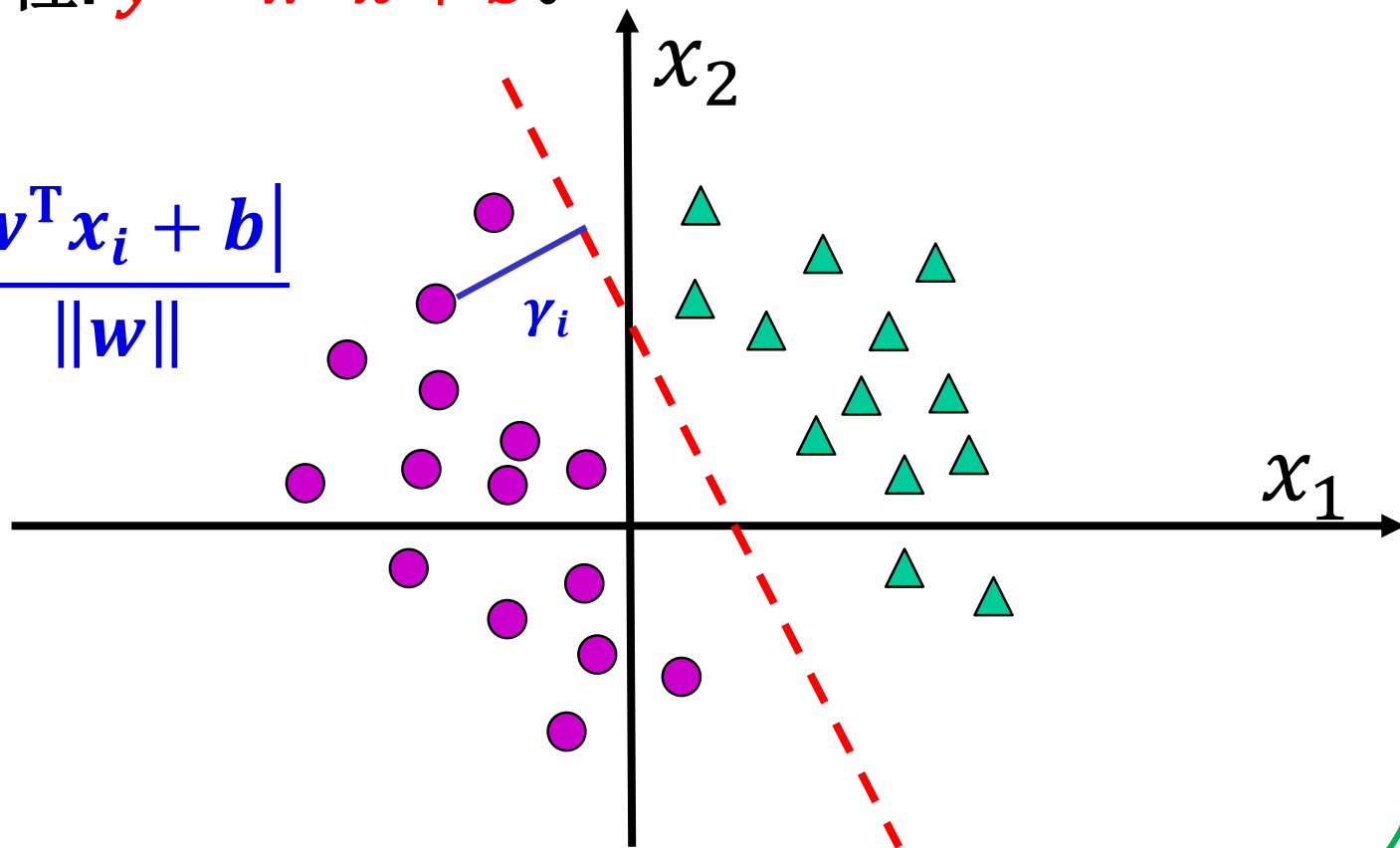
数据点 \star \star 都被判定为正样本，但是置信程度（**confident**）不同。

主观上说，离分类边界远的样本的置信度高，而离分类边界近的样本置信度低。

间隔 (Margin)

直线方程: $y = w^T x + b$ 。

$$\gamma_i = \frac{|w^T x_i + b|}{\|w\|}$$



间隔 (Margin)

考虑输出变量 $y_i \in \{-1, 1\}$ 。

$$\gamma_i = \frac{|w^T x_i + b|}{\|w\|} = \frac{y_i(w^T x_i + b)}{\|w\|}$$

定义数据集 $\{(x_i, y_i)\}_{i=1}^n$ 关于直线 $y = w^T x + b$ 的间隔为：

$$\gamma = \min_{i=1, \dots, n} \gamma_i$$

γ 表示所有数据点到直线的最短距离。

间隔最大化 (Maximum Margin)

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

$$\text{s. t. } \frac{y_i(w^T x_i + b)}{\|w\|} \geq \gamma$$

我们希望找到最大的 γ ，使得对于所有数据点 $\{(x_i, y_i)\}_{i=1}^n$ 到超平面的距离至少是 γ 。

问题：给定数据点和直线， γ 可以被唯一确定，换句话说 γ 仅依赖于变量 w, b 的选择。能否把 γ 消去？

支持向量机

可以对 w, b 进行适当缩放, 使得 $\gamma = \frac{1}{\|w\|}$, 则

$$\begin{aligned} \max_{\gamma, w, b} \quad & \gamma \\ \text{s. t.} \quad & \frac{y_i(w^T x_i + b)}{\|w\|} \geq \gamma \end{aligned}$$

等价于

$$\begin{aligned} \min_{w, b} \quad & \frac{1}{2} \|w\|^2 \\ \text{s. t.} \quad & y_i(w^T x_i + b) \geq 1 \end{aligned}$$

支持向量机

支持向量机 (Support Vector Machine)

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

$$\text{s. t.} \quad y_i(w^T x_i + b) \geq 1$$

- 满足 $y_i(w^T x_i + b) = 1$ 的样本被称为支持向量。
- SVM是一个凸的带不等式约束的凸二次规划问题。

支持向量

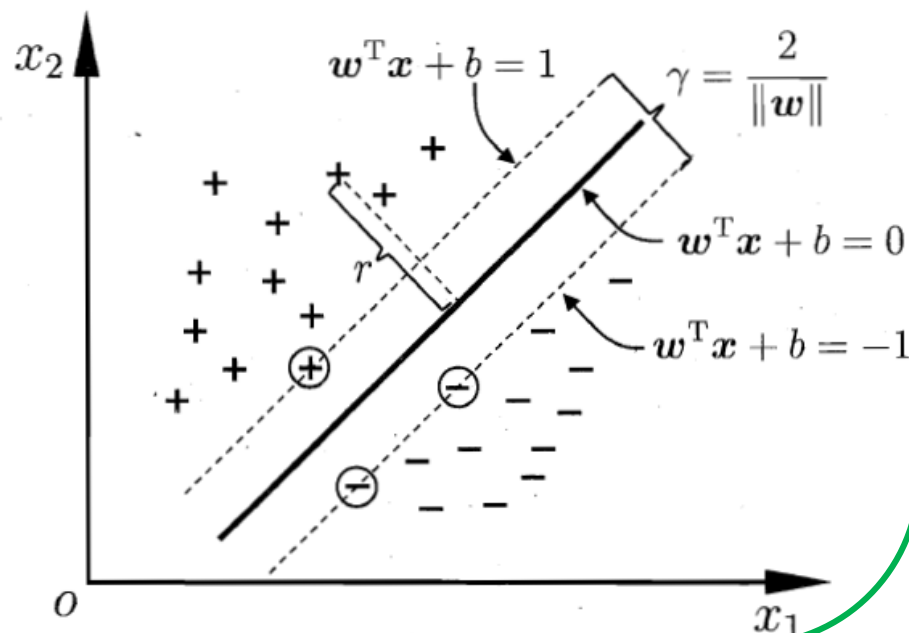
分离超平面: $w^T x + b = 0$

间隔边界: $w^T x + b = \pm 1$

支持向量:

$$y_i(w^T x_i + b) - 1 = 0$$

- 在间隔边界上。



支持向量机小结

支持向量机的基本思想是寻找两类样本之间**最中间的超平面**。

支持向量机的目的是使划分平面对于样本的**扰动容忍性好**。

逻辑回归算法是基于全部样本的二分类器：考虑全部样本的平均似然性。--任课教师

支持向量机算法是基于部分样本的二分类器：考虑部分靠近边界的支持向量。--辅导员

6.2 拉格朗日乘子法

陈飞宇

fchen@cqu.edu.cn

办公室：软件学院529

拉格朗日乘子法

考虑约束优化问题:

$$\begin{aligned} \min_{x \in \mathbb{R}^d} \quad & f(x) \\ \text{s. t.} \quad & g_i(x) \leq 0 \quad i = 1, \dots, n \end{aligned}$$

拉格朗日函数(Lagrange function):

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^n \lambda_i g_i(x)$$

其中 $\lambda_i \geq 0$ 是拉格朗日乘子(Lagrange multiplier).

拉格朗日乘子法

约束优化问题:

$$\min_{x \in \mathbb{R}^d} f(x) \quad \text{s.t.} \quad g_i(x) \leq 0 \quad i = 1, \dots, n$$

极大问题: $\max_{\lambda \geq 0} \mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^n \lambda_i g_i(x)$

若对于某个 i , $g_i(x) > 0$, 则 $\max_{\lambda \in \mathbb{R}^m} \mathcal{L}(x, \lambda) = +\infty$

若对于所有 i , $g_i(x) \leq 0$, 则 $\max_{\lambda \in \mathbb{R}^m} \mathcal{L}(x, \lambda) = f(x)$

$$\min_{\lambda \geq 0} \max \mathcal{L}(x, \lambda) = \min(+\infty, f(x)) = f(x)$$

拉格朗日乘子法

约束优化问题:

$$\begin{aligned} \min_{x \in \mathbb{R}^d} \quad & f(x) \\ \text{s. t.} \quad & g_i(x) \leq 0 \quad i = 1, \dots, n \end{aligned}$$

等价于极大极小问题:

$$\min_{x \in \mathbb{R}^d} \max_{\lambda \geq 0} \mathcal{L}(x, \lambda)$$

其中: $\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^n \lambda_i g_i(x)$

拉格朗日乘子法

极大极小问题（Primal Problem）：

$$p = \min_{x \in \mathbb{R}^d} \max_{\lambda \geq 0} \mathcal{L}(x, \lambda)$$

极小极大问题（Dual Problem）：

$$d = \max_{\lambda \geq 0} \min_{x \in \mathbb{R}^d} \mathcal{L}(x, \lambda)$$

其中P问题的最优解和D问题的最优解成对出现

拉格朗日乘子法

原问题和对偶问题等价的充分必要条件被称作KKT条件（Karush-Kuhn-Tucher）：

$$\begin{cases} \nabla_x \mathcal{L}(x, \lambda) = 0 \\ \lambda^T g(x) = 0 \\ \lambda \geq 0 \\ g(x) \leq 0 \end{cases}$$

拉格朗日乘子法小结

- 拉格朗日乘子法是求解约束优化问题常用的方法之一，其基本思想是求解与之等价的无约束对偶问题。
- 较之原问题来说，对偶问题可能更方便求解
- 较之原问题来说，对偶问题也可能更有意义

6.3 SVM对偶模型

陈飞宇

fchen@cqu.edu.cn

办公室：软件学院529

支持向量机

支持向量机 (SVM)

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

$$\text{s. t. } 1 - y_i(w^T x_i + b) \leq 0 \quad i = 1, \dots, n$$

- 变量: $w = (w_1, w_2, \dots, w_d)^T$, b
- 约束: n 个不等式约束。
- 数据: $x_i \in \mathbb{R}^d$, $y_i \in \{-1, 1\} \quad i = 1, \dots, n$

对偶问题

极小极大问题（Dual Problem）：

$$d = \max_{\lambda \geq 0} \min_{\mathbf{w}, b} \mathcal{L}(\mathbf{w}, b, \lambda)$$

其中，

$$\mathcal{L}(\mathbf{w}, b, \lambda) = \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^n \lambda_i (1 - \mathbf{y}_i(\mathbf{w}^T \mathbf{x}_i + b))$$

极小问题: $\min_{w,b} \mathcal{L}(w, b, \lambda)$

$$\min_{w,b} \frac{1}{2} \|w\|^2 + \sum_{i=1}^n \lambda_i (1 - y_i(w^T x_i + b))$$

$$\frac{\partial \mathcal{L}}{\partial w} = w - \sum_{i=1}^n \lambda_i y_i x_i; \quad \frac{\partial \mathcal{L}}{\partial b} = - \sum_{i=1}^n \lambda_i y_i$$

$$w = \sum_{i=1}^n \lambda_i y_i x_i; \quad \sum_{i=1}^n \lambda_i y_i = 0$$

极小问题: $\min_{\mathbf{w}, b} \mathcal{L}(\mathbf{w}, b, \lambda)$

把 $\mathbf{w} = \sum_{i=1}^n \lambda_i y_i \mathbf{x}_i$ 代入拉格朗日函数有:

$$\begin{aligned} & \min_{\mathbf{w}, b} \mathcal{L}(\mathbf{w}, b, \lambda) \\ &= \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^n \lambda_i (1 - y_i(\mathbf{w}^T \mathbf{x}_i + b)) \\ &= \frac{1}{2} \|\sum_{i=1}^n \lambda_i y_i \mathbf{x}_i\|^2 + \sum_{i=1}^n \lambda_i [1 - y_i((\sum_{j=1}^n \lambda_j y_j \mathbf{x}_j^T) \mathbf{x}_i + b)] \\ &= \sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{i=1}^n \lambda_i y_i b \\ &= \sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \end{aligned}$$

极小极大问题

极小极大问题（Dual Problem）：

$$d = \max_{\lambda \geq 0} \min_{w, b} \mathcal{L}(w, b, \lambda)$$

把 $\min_{w, b} \mathcal{L}(w, b, \lambda)$ 代入上述问题有

$$\max_{\lambda \geq 0} \sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j$$

$$\text{s. t. } \sum_{i=1}^n \lambda_i y_i = 0$$

对偶问题

SVM 对偶问题:

$$\min_{\lambda} \quad \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j - \sum_{i=1}^n \lambda_i$$

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

$$\lambda_i \geq 0 \quad i = 1, \dots, n$$

变量: $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)^T$

约束: 1 个等式约束, n 个不等式约束。

数据: $x_i^T x_j \in \mathbb{R}, \quad y_i y_j \in \{-1, 1\} \quad i = 1, \dots, n$

KKT 条件

KKT 条件:

$$\left\{ \begin{array}{l} \mathbf{w} - \sum_{i=1}^n \lambda_i y_i \mathbf{x}_i = \mathbf{0} \\ \sum_{i=1}^n \lambda_i y_i = 0 \\ \lambda_i [y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1] = 0 \\ \lambda_i \geq 0 \\ y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 \geq 0 \end{array} \right.$$

支持向量 $\lambda_i \neq 0$

对于任意的数据点 x_i , 我们有 $\lambda_i = 0$ 或者 $y_i(w^T x_i + b) - 1 = 0$.

- $\lambda_i = 0$:

对偶问题中变量 λ_i 不在求和中出现。

- $y_i(w^T x_i + b) - 1 = 0$:

x_i 是一个支持向量, $b = y_i - w^T x_i$

求解分类平面

SVM 对偶问题:

$$\min_{\lambda} \quad \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j - \sum_{i=1}^n \lambda_i$$

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

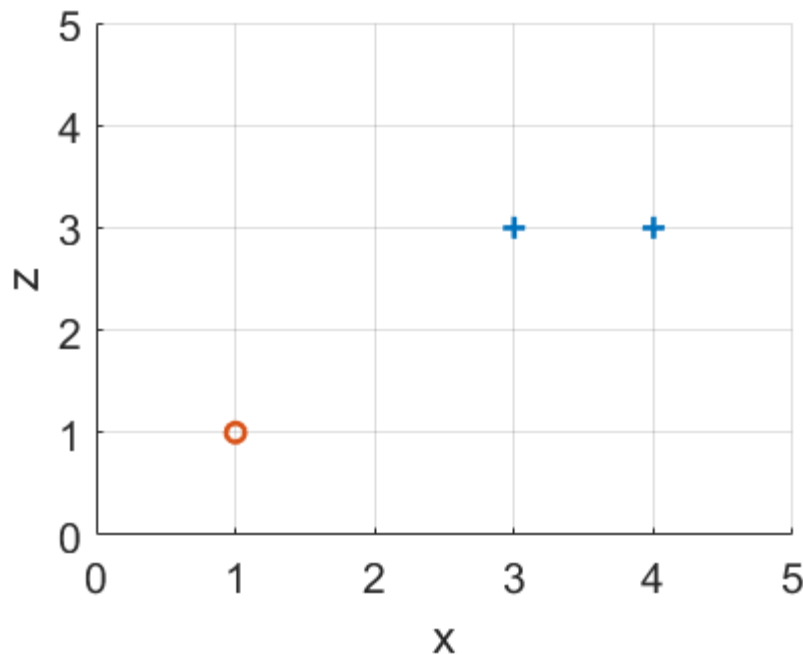
$$\lambda_i \geq 0 \quad i = 1, \dots, n$$

$$w = \sum_{i=1}^n \lambda_i y_i x_i, \quad b = y_j - w^T x_j$$

$$f(x) = w^T x + b = \left(\sum_{i=1}^n \lambda_i y_i x_i^T \right) (x - x_j) + y_j$$

例子

假设正例点是 $x_1 = (3, 3)^T$, $x_2 = (4, 3)^T$, 负例点是 $x_3 = (1, 1)^T$ 。试利用支持向量机算法求出分类超平面。



例子

解: $x_1 = (3, 3)^T, x_2 = (4, 3)^T, x_3 = (1, 1)^T$ 。

$$y_1 = 1, \quad y_2 = 1, \quad y_3 = -1。$$

计算向量内积,

$$K = \begin{pmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{pmatrix} = \begin{pmatrix} 18 & 21 & 6 \\ 21 & 25 & 7 \\ 6 & 7 & 2 \end{pmatrix}$$

计算目标函数,

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j \\ &= 18\lambda_1^2 + 25\lambda_2^2 + 2\lambda_3^2 + 42\lambda_1\lambda_2 - 12\lambda_1\lambda_3 - 14\lambda_2\lambda_3 \end{aligned}$$

例子

$$\begin{aligned} \min_{\lambda} \quad & 9\lambda_1^2 + \frac{25}{2}\lambda_2^2 + \lambda_3^2 + 21\lambda_1\lambda_2 - 6\lambda_1\lambda_3 - 7\lambda_2\lambda_3 - \sum_{i=1}^3 \lambda_i \\ \text{s. t.} \quad & \lambda_1 + \lambda_2 - \lambda_3 = 0, \quad \lambda_i \geq 0 \quad i = 1, \dots, 3 \end{aligned}$$

令 $\lambda_3 = \lambda_1 + \lambda_2$ ，代入上式有

$$\begin{aligned} \min_{\lambda} \quad & 4\lambda_1^2 + \frac{13}{2}\lambda_2^2 + 10\lambda_1\lambda_2 - 2\lambda_1 - 2\lambda_2 \\ \text{s. t.} \quad & \lambda_i \geq 0 \quad i = 1, 2 \end{aligned}$$

例子

二次规划问题:

$$\min_{\lambda} \quad 4\lambda_1^2 + \frac{13}{2}\lambda_2^2 + 10\lambda_1\lambda_2 - 2\lambda_1 - 2\lambda_2$$

$$\text{s. t.} \quad \lambda_i \geq 0 \quad i = 1, 2$$

目标函数分别对 λ_1 和 λ_2 求导有

$$8\lambda_1 + 10\lambda_2 - 2 = 0 \quad 10\lambda_1 + 13\lambda_2 - 2 = 0$$

$$\text{求解得到: } \lambda_1 = \frac{3}{2}, \quad \lambda_2 = -1$$

此时 λ_2 不满足约束, 最优解在边界上取得。

例子

当 $\lambda_1 = 0$, $\lambda_2 = \frac{2}{13}$, 目标函数值是 $-\frac{2}{13}$ 。

当 $\lambda_2 = 0$, $\lambda_1 = \frac{1}{4}$, 目标函数值是 $-\frac{1}{4}$ 。

当 $\lambda_2 = 0$, $\lambda_1 = \frac{1}{4}$ 时取得最小值, $\lambda_3 = \lambda_1 + \lambda_2 = \frac{1}{4}$

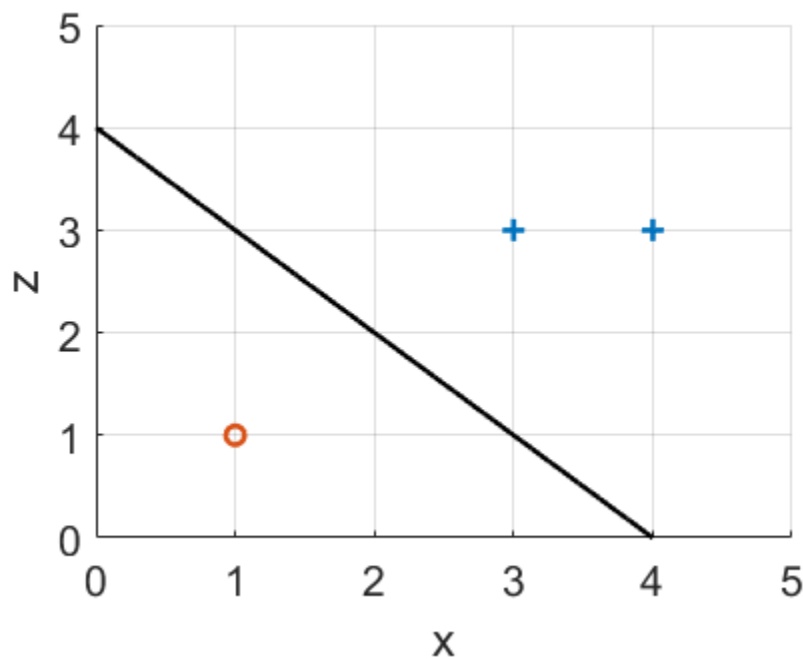
$$w = \sum_{i=1}^n \lambda_i y_i x_i = \frac{1}{4} x_1 + 0 - \frac{1}{4} x_3 = \left(\frac{1}{2}, \frac{1}{2}\right)^T$$

$$b = y_j - \sum_{i=1}^n \lambda_i y_i x_i^T x_j = 1 - \frac{1}{4} \times 18 + \frac{1}{4} \times 6 = -2$$

分离超平面为: $\frac{1}{2}x + \frac{1}{2}z - 2 = 0$

例子

假设正例点是 $x_1 = (3, 3)^T$, $x_2 = (4, 3)^T$, 负例点是 $x_3 = (1, 1)^T$ 。试利用支持向量机算法求出分类超平面。



原始问题？

假设正例点是 $x_1 = (3, 3)^T$, $x_2 = (4, 3)^T$, 负例点是 $x_3 = (1, 1)^T$ 。试利用支持向量机算法求出分类超平面。

$$\min_{w, b} \quad \frac{1}{2} (w_1^2 + w_2^2)$$

$$\text{s. t.} \quad 1 - (3w_1 + 3w_2 + b) \leq 0$$

$$1 - (4w_1 + 3w_2 + b) \leq 0$$

$$1 + (w_1 + w_2 + b) \leq 0$$

SVM对偶模型小结

- 我们可以利用拉格朗日乘子法得到SVM的对偶模型。
- 对偶模型更能反映该问题的特点，只有支持向量才对优化问题起作用。
- 求解对偶模型的比求解原问题简单，计算复杂度更低。

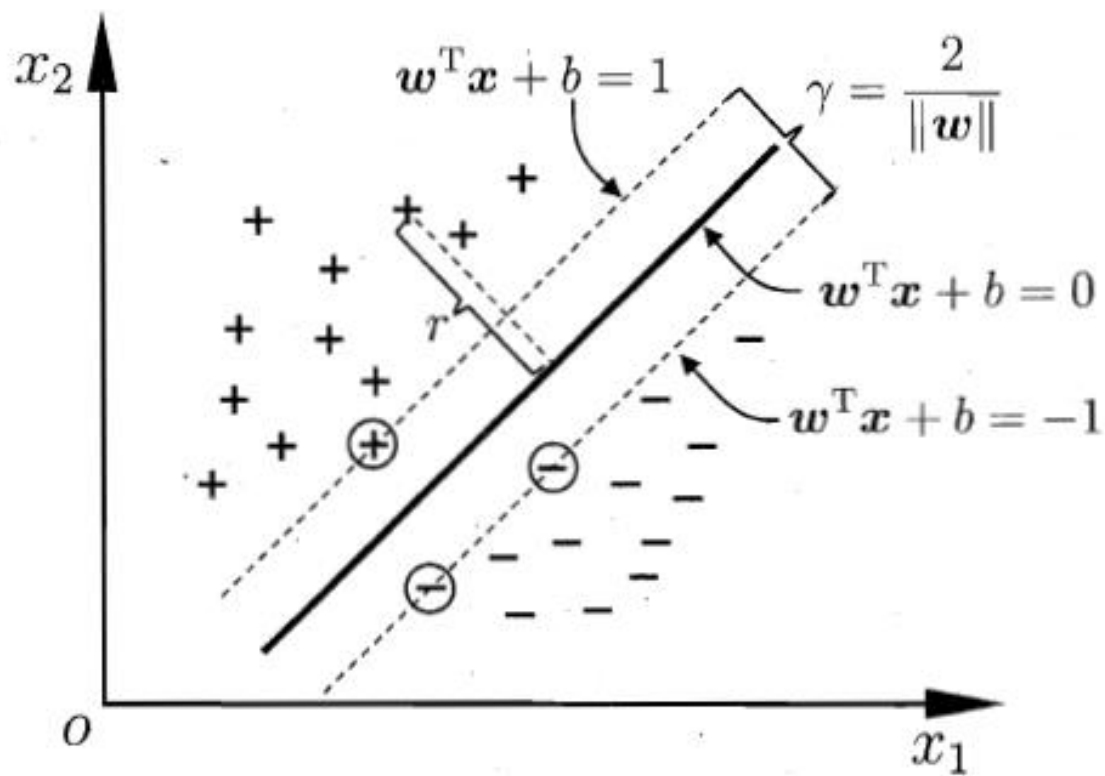
6.4 软间隔SVM

陈飞宇

fchen@cqu.edu.cn

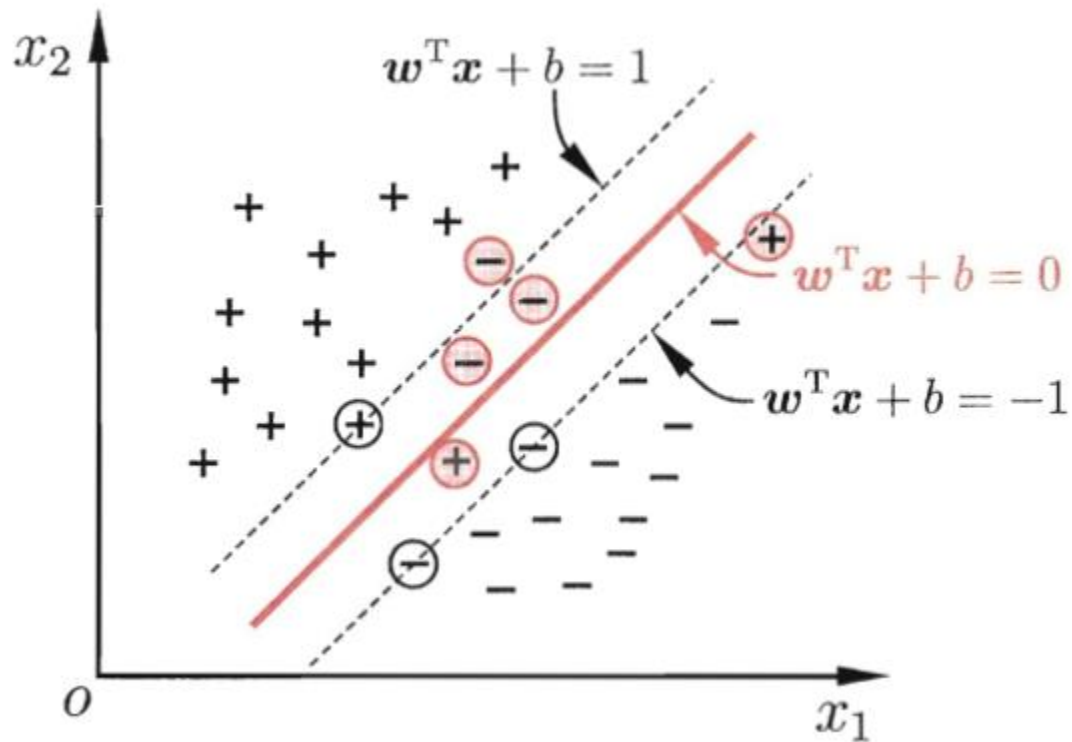
办公室：软件学院529

SVM图例



软间隔SVM

大部分数据线性可分，但存在一些outliers。



软间隔SVM

软间隔支持向量机 (SVM)

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

$$\text{s. t. } y_i(w^T x_i + b) \geq 1 - \xi_i \quad i = 1, \dots, n$$

$$\xi_i \geq 0 \quad i = 1, \dots, n$$

- 变量: w, b, ξ_i (松弛变量)
- 约束: $2n$ 个不等式约束。
- 数据: $x_i \in \mathbb{R}^d, y_i \in \{-1, 1\} \quad i = 1, \dots, n$

对偶问题

极小极大问题（Dual Problem）：

$$d = \max_{\lambda, \mu \geq 0} \min_{w, b, \xi} \mathcal{L}(w, b, \xi, \lambda, \mu)$$

其中，

$$\begin{aligned} & \mathcal{L}(w, b, \xi, \lambda, \mu) \\ &= \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i - \sum_{i=1}^n \mu_i \xi_i \\ &+ \sum_{i=1}^n \lambda_i (1 - \xi_i - y_i(w^T x_i + b)) \end{aligned}$$

极小问题 $\min_{w, b, \xi} \mathcal{L}(w, b, \xi, \lambda, \mu)$

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \mu_i \xi_i + \sum_{i=1}^n \lambda_i (1 - \xi_i - y_i(w^T x_i + b))$$

$$\frac{\partial \mathcal{L}}{\partial w} = w - \sum_{i=1}^n \lambda_i y_i x_i \quad \Rightarrow \quad w = \sum_{i=1}^n \lambda_i y_i x_i$$

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

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = C - \lambda_i - \mu_i \quad \Rightarrow \quad C = \lambda_i + \mu_i$$

极小问题 $\min_{w, b, \xi} \mathcal{L}(w, b, \xi, \lambda, \mu)$

把 $w = \sum_{i=1}^n \lambda_i y_i x_i$ 代入拉格朗日函数有：

$$\begin{aligned} & \min_{w, b, \xi} \mathcal{L}(w, b, \xi, \lambda, \mu) \\ &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i + \sum_{i=1}^n \lambda_i (1 - \xi_i - y_i(w^T x_i + b)) - \sum_{i=1}^n \mu_i \xi_i \\ &= \sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j + \sum_{i=1}^n (C - \lambda_i - \mu_i) \xi_i \\ &= \sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j \end{aligned}$$

因此，SVM和软间隔SVM的对偶问题有相同的目标函数！

极小极大问题

极小极大问题（Dual Problem）：

$$d = \max_{\lambda \geq 0} \min_{w, b} \mathcal{L}(w, b, \lambda)$$

把 $\min_{w, b, \xi} \mathcal{L}(w, b, \xi, \lambda, \mu)$ 代入上述问题有

$$\begin{aligned} \max_{\lambda, \mu \geq 0} \quad & \sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j \\ \text{s. t.} \quad & \sum_{i=1}^n \lambda_i y_i = 0 \\ & \lambda_i + \mu_i = C \quad i = 1, \dots, n \end{aligned}$$

对偶问题

软间隔SVM 对偶问题:

$$\min_{\lambda} \quad \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j - \sum_{i=1}^n \lambda_i$$

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

$$\lambda_i + \mu_i = C \quad i = 1, \dots, n$$

$$\lambda_i \geq 0 \quad i = 1, \dots, n$$

$$\mu_i \geq 0 \quad i = 1, \dots, n$$

注意, 后面三个条件等价于 $0 \leq \lambda_i \leq C$

对偶问题

软间隔SVM 对偶问题:

$$\min_{\lambda} \quad \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j - \sum_{i=1}^n \lambda_i$$

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

$$0 \leq \lambda_i \leq C \quad i = 1, \dots, n$$

变量: $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)^T$

约束: 1 个等式约束, n 个不等式约束。

数据: $x_i^T x_j \in \mathbb{R}, \quad y_i y_j \in \{-1, 1\} \quad i = 1, \dots, n$

KKT 条件

KKT 条件:

$$\left\{ \begin{array}{l} w - \sum \lambda_i y_i x_i = 0 \\ \sum \lambda_i y_i = 0 \\ \lambda_i + \mu_i = C \\ \lambda_i (y_i (w^T x_i + b) - 1 + \xi_i) = 0 \\ \mu_i \xi_i = 0 \\ \lambda_i \geq 0 \\ \mu_i \geq 0 \\ y_i (w^T x_i + b) - 1 \geq 0 \\ \xi_i \geq 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} w - \sum \lambda_i y_i x_i = 0 \\ \sum \lambda_i y_i = 0 \\ \lambda_i (y_i (w^T x_i + b) - 1 + \xi_i) = 0 \\ (C - \lambda_i) \xi_i = 0 \\ 0 \leq \lambda_i \leq C \\ y_i (w^T x_i + b) - 1 \geq 0 \\ \xi_i \geq 0 \end{array} \right.$$

支持向量

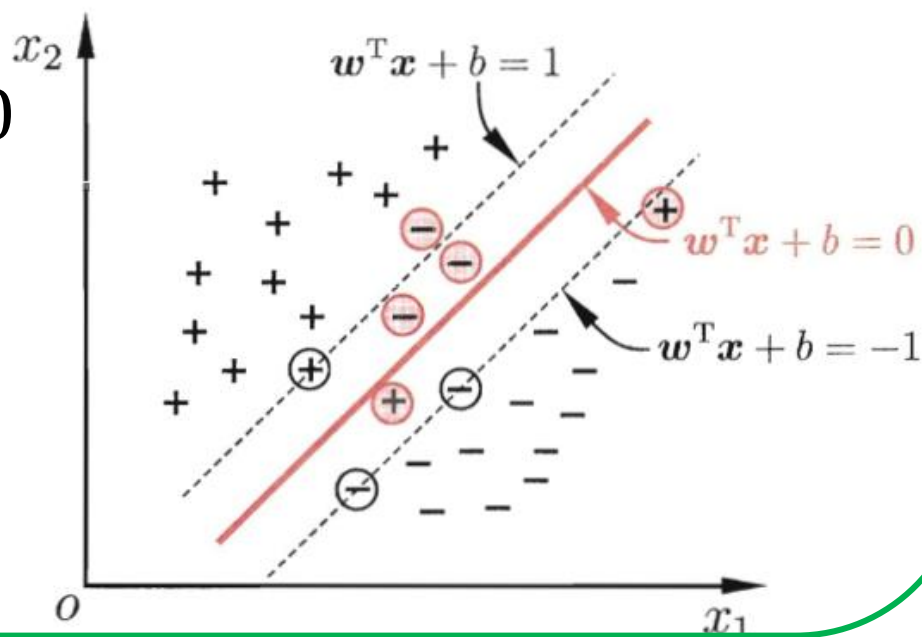
分离超平面: $w^T x + b = 0$

间隔边界: $w^T x + b = \pm 1$

支持向量:

$$y_i(w^T x_i + b) - 1 \leq 0$$

- 间隔边界上。
- 间隔边界间。
- 分类错误点。



支持向量 $\lambda_i \neq 0$

- 若 $\lambda_i < C$, 则 $\xi_i = 0$, x_i 落在间隔边界上。
- 若 $\lambda_i = C$, $0 < \xi_i < 1$ 时, x_i 落在间隔边界和分离超平面之间。
- 若 $\lambda_i = C$, $\xi_i = 1$ 时, x_i 落在分离超平面上。
- 若 $\lambda_i = C$, $\xi_i > 1$ 时, x_i 落在分离超平面另一侧。

当 $C \rightarrow +\infty$ 时

支持向量机 (SVM)

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

$$\text{s. t. } y_i(w^T x_i + b) \geq 1 \quad i = 1, \dots, n$$

软间隔支持向量机 (SVM)

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

$$\text{s. t. } y_i(w^T x_i + b) \geq 1 - \xi_i \quad i = 1, \dots, n$$

$$\xi_i \geq 0 \quad i = 1, \dots, n$$

当 $C \rightarrow +\infty$ 时

SVM 对偶问题:

$$\min_{\lambda} \quad \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j - \sum_{i=1}^n \lambda_i$$

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

$$0 \leq \lambda_i \leq +\infty \quad i = 1, \dots, n$$

软间隔SVM 对偶问题:

$$\min_{\lambda} \quad \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j - \sum_{i=1}^n \lambda_i$$

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

$$0 \leq \lambda_i \leq C \quad i = 1, \dots, n$$

软间隔SVM

软间隔支持向量机 (SVM)

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

$$\text{s. t. } y_i(w^T x_i + b) \geq 1 - \xi_i \quad i = 1, \dots, n$$

$$\xi_i \geq 0 \quad i = 1, \dots, n$$

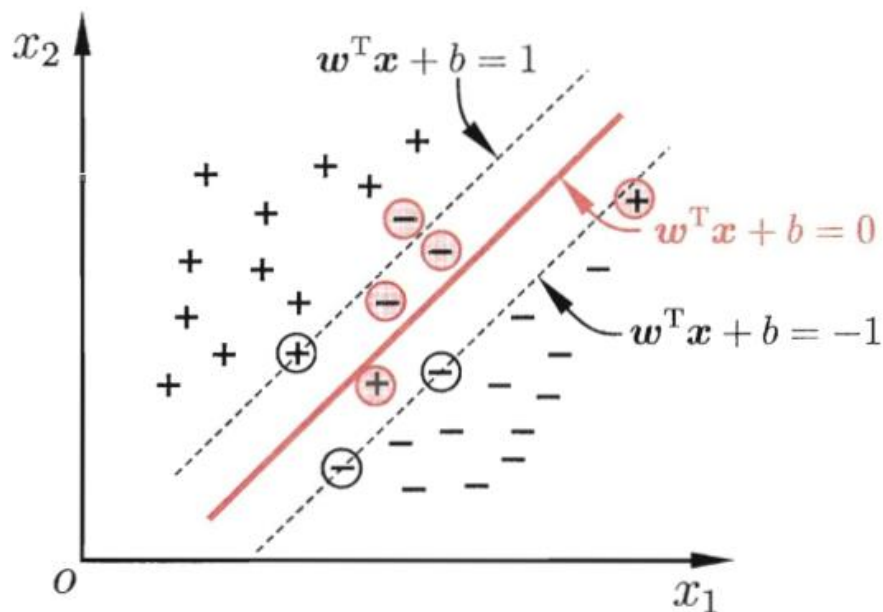
显然, $\xi_i \geq \max(0, 1 - y_i(w^T x_i + b))$

$$\min_{w, b} \quad \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n [1 - y_i(w^T x_i + b)]_+$$

软间隔SVM

$$\min_{w,b} \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n [1 - y_i(w^T x_i + b)]_+$$

$[1 - y_i(w^T x_i + b)]_+$: 对支持向量 x_i 到其对应的间隔平面距离的惩罚。



软间隔SVM

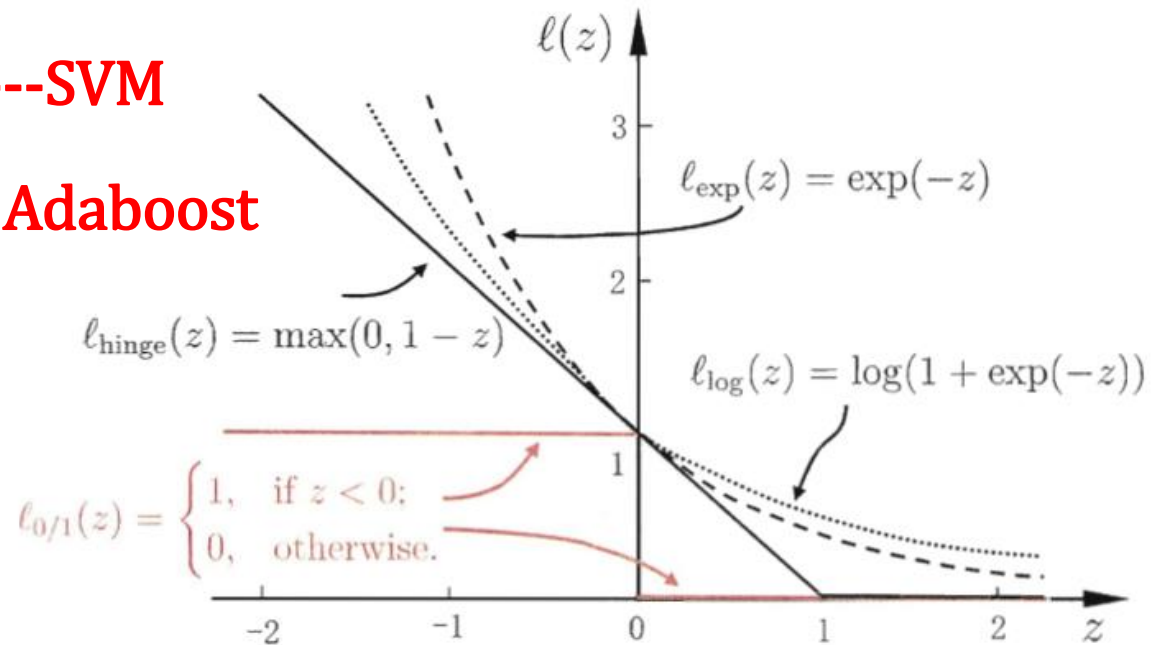
监督学习任务模型：

$$\min_f \sum_{i=1}^n \ell(f(x_i), y_i) + \alpha \Omega(f)$$

- $\ell(f(x_i), y_i)$ 为误差项，衡量预测值与真实值的误差。
- $\Omega(f)$ 为正则项，提供了关于 f 的先验信息，防止过拟合。
- α 是正则参数，为了平衡两项的大小。

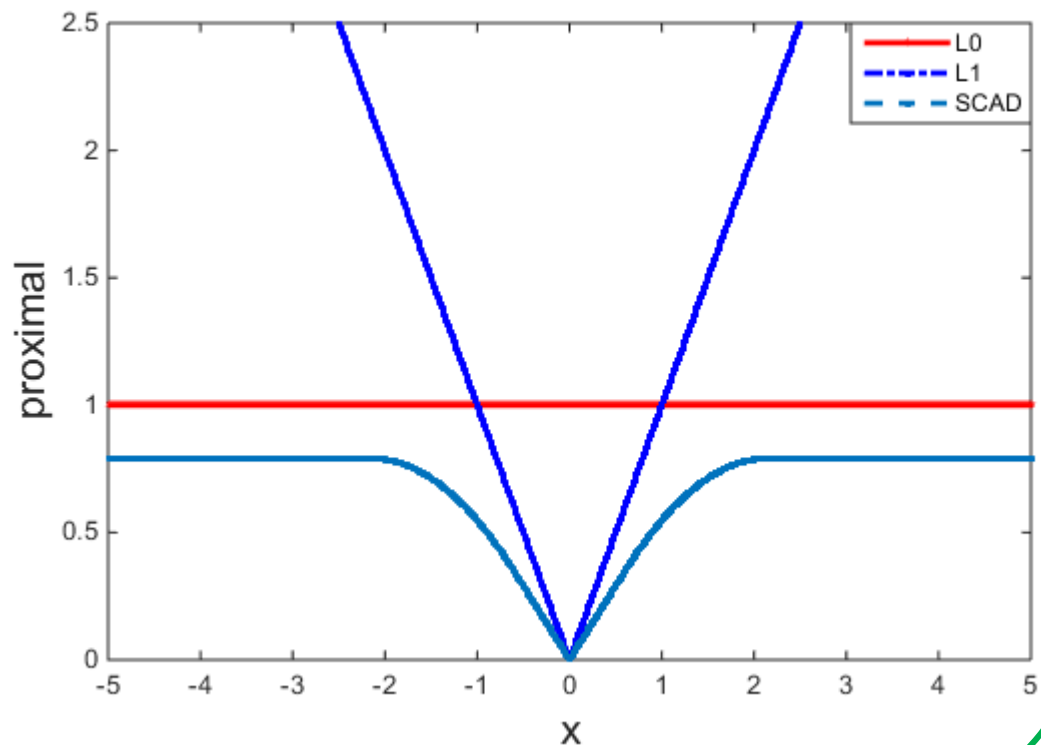
损失函数: $\ell(f(x_i), y_i)$

- 0/1 loss --- 非凸非连续
- Log loss --- LR
- Hinge loss --- SVM
- Exp loss --- Adaboost



正则项 $\Omega(f)$

- L0 范数--- $\|w\|_0$ 是 w 中非零元的个数
- L1 范数--- $\|w\|_1 = \sum_j |w_j|$
- L2范数 ---凸连续
- SCAD函数



软间隔SVM小结

- 软间隔SVM可以对有outlier的数据分类。
- 软间隔SVM对偶模型与SVM对偶模型非常相似，可以用相同算法求解。
- 软间隔SVM模型可以看作是最小化hinge损失函数的正则化模型。
- 当参数C趋向无穷大时，软间隔SVM退化成普通的SVM。

6.5 SMO算法

陈飞宇

fchen@cqu.edu.cn

办公室：软件学院529

SMO算法

SMO(Sequential Minimal Optimization):

$$\min_{\lambda} \quad \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j - \sum_{i=1}^n \lambda_i$$

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

$$0 \leq \lambda_i \leq C \quad i = 1, \dots, n$$

SMO的两个部分:

- 解析地求解两个变量二次规划问题。
- 启发式地选择变量。

SMO算法

固定其他变量，仅考虑 λ_i 和 λ_j 时，目标函数：

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j x_i^T x_j \lambda_i \lambda_j - \sum_{i=1}^n \lambda_i \\ &= \frac{1}{2} (x_i^T x_i) \lambda_i^2 + \frac{1}{2} (x_j^T x_j) \lambda_j^2 + (y_i y_j x_i^T x_j) \lambda_i \lambda_j - \lambda_i \\ & \quad - \lambda_j + (\sum_{k \neq i, j}^n y_i y_k x_i^T x_k \lambda_k) \lambda_i + (\sum_{k \neq i, j}^n y_j y_k x_j^T x_k \lambda_k) \lambda_j \end{aligned}$$

约束：

$$y_i \lambda_i + y_j \lambda_j = - \sum_{k \neq i, j}^n y_k \lambda_k$$

SMO算法

$$\min_{\lambda_i, \lambda_j} \frac{1}{2} K_{ii} \lambda_i^2 + \frac{1}{2} K_{jj} \lambda_j^2 + c_{ij} \lambda_i \lambda_j + c_i \lambda_i + c_j \lambda_j$$

$$\text{s. t. } y_i \lambda_i + y_j \lambda_j = c, \quad 0 \leq \lambda_i, \lambda_j \leq C$$

其中:

$$K_{ii} = x_i^T x_i \quad K_{jj} = x_j^T x_j$$

$$c_{ij} = y_i y_j K_{ij} \quad c = -\sum_{k \neq i, j}^n y_k \lambda_k$$

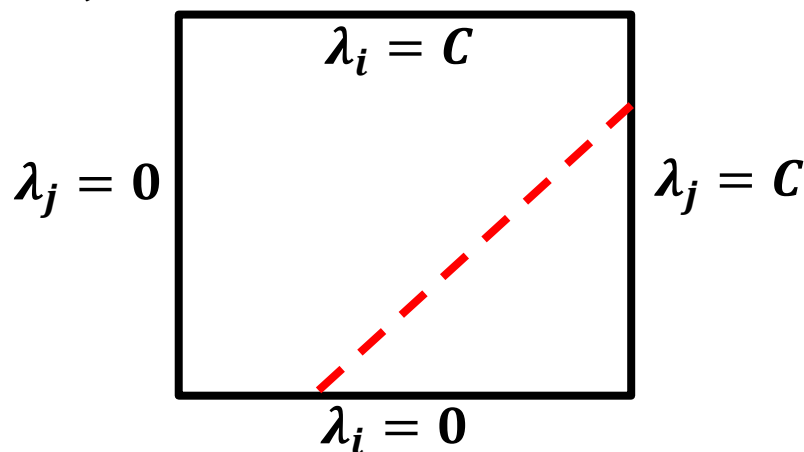
$$c_i = \sum_{k \neq i, j}^n y_i y_k x_i^T x_k \lambda_k - 1 \quad c_j = \sum_{k \neq i, j}^n y_j y_k x_j^T x_k \lambda_k - 1$$

SMO算法

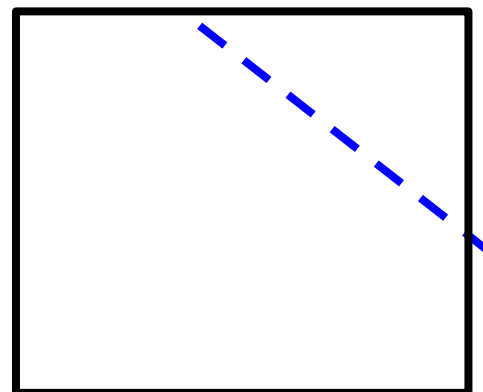
$$\min_{\lambda_i, \lambda_j} \frac{1}{2} K_{ii} \lambda_i^2 + \frac{1}{2} K_{jj} \lambda_j^2 + c_{ij} \lambda_i \lambda_j + c_i \lambda_i + c_j \lambda_j$$

$$\text{s. t. } y_i \lambda_i + y_j \lambda_j = c, \quad 0 \leq \lambda_i, \lambda_j \leq C$$

约束:



$$y_i \neq y_j \Rightarrow \lambda_i - \lambda_j = c$$



$$y_i = y_j \Rightarrow \lambda_i + \lambda_j = c$$

SMO算法

$$\min_{\lambda_i, \lambda_j} \frac{1}{2} K_{ii} \lambda_i^2 + \frac{1}{2} K_{jj} \lambda_j^2 + c_{ij} \lambda_i \lambda_j + c_i \lambda_i + c_j \lambda_j$$

$$\text{s. t. } y_i \lambda_i + y_j \lambda_j = c, \quad 0 \leq \lambda_i, \lambda_j \leq C$$

- 令 $\lambda_j = y_j(c - y_i \lambda_i)$ 代入上述方程, 求解关于 λ_i 的单变量二次优化问题。

$$\begin{aligned} \min_{0 \leq \lambda_i \leq C} \quad & \frac{1}{2} K_{ii} \lambda_i^2 + \frac{1}{2} K_{jj} (c - y_i \lambda_i)^2 + c_{ij} \lambda_i y_j (c - y_i \lambda_i) \\ & + c_i \lambda_i + c_j y_j (c - y_i \lambda_i) \end{aligned}$$

- 并验证 $\lambda_j = y_j(c - y_i \lambda_i)$ 是否在区间 $[0, C]$, 否则最小值在边界处求得。

SMO算法小结

- 序列极小化优化算法（SMO）是求解SVM模型最高效的算法。
- SMO每次迭代解两个变量的二次优化问题，其最优解可以显式表达。
- SMO利用了启发式算法根据数据点违反KKT条件的大小选取需要迭代的变量。

6.6 核化SVM

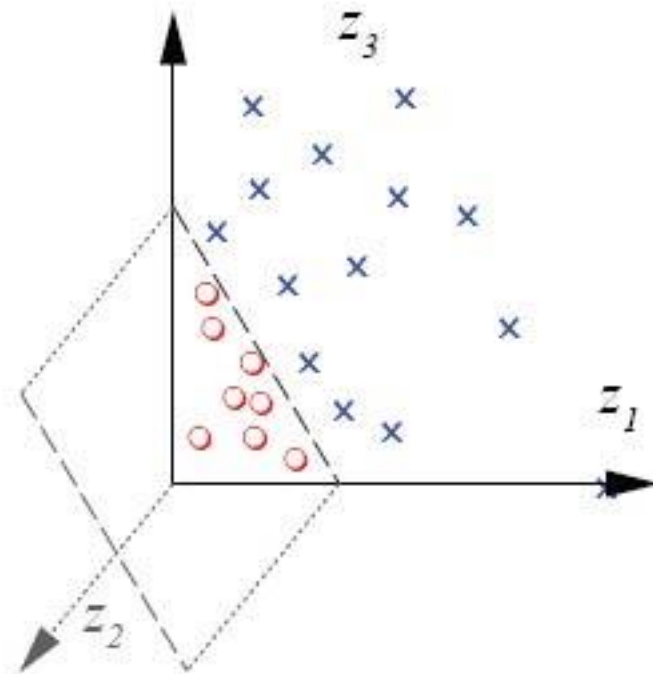
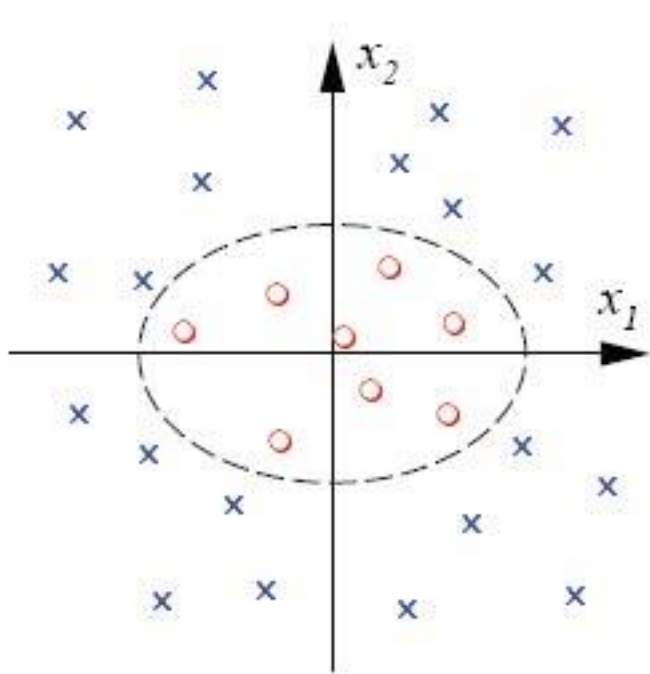
陈飞宇

fchen@cqu.edu.cn

办公室：软件学院529

非线性SVM

非线性可分数据:



解决方案

对于非线性可分的数据 $\{x_i\}_{i=1}^n$ ，尝试找到一个非线性映射 ϕ ，使得数据 $\{\phi(x_i)\}_{i=1}^n$ 在新的空间（通常为高维空间）是线性可分的，然后再使用线性SVM进行分类。

椭圆方程: $w_1(x^{(1)})^2 + w_2(x^{(2)})^2 + b = 0$

直线方程: $w_1x^{(1)} + w_2x^{(2)} + b = 0$

特征空间SVM

支持向量机 (SVM)

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

$$\text{s. t. } 1 - y_i(w^T \phi(x_i) + b) \leq 0 \quad i = 1, \dots, n$$

SVM 对偶问题:

$$\min_{\lambda} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \phi(x_i)^T \phi(x_j) \lambda_i \lambda_j - \sum_{i=1}^n \lambda_i$$

$$\text{s. t. } \sum_{i=1}^n \lambda_i y_i = 0$$

$$\lambda_i \geq 0 \quad i = 1, \dots, n$$

核技巧 (Kernel Trick)

然而，对于任意给定数据寻找合适的非线性映射 ϕ 是不现实的，而且内积 $\phi(x_i)^T \phi(x_j)$ 的计算代价太大。

定义核函数 (Kernel Function) :

$$\kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

这样， ϕ 可以隐式地由核函数表示出来，而且内积的计算也变得更加容易。

特征空间SVM

SVM 对偶问题:

$$\min_{\lambda} \quad \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \kappa(x_i, x_j) \lambda_i \lambda_j - \sum_{i=1}^n \lambda_i$$

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

$$\lambda_i \geq 0 \quad i = 1, \dots, n$$

分割超平面:

$$\begin{aligned} f(x) = w^T \phi(x) + b &= \sum_{i=1}^n \lambda_i y_i \phi(x_i)^T \phi(x) + b \\ &= \sum_{i=1}^n \lambda_i y_i \kappa(x, x_i) + b \end{aligned}$$

如何选择核函数

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 $D = \{x_1, x_2, \dots, x_n\}$ ，核矩阵(Kernel matrix) K 总是半正定的：

$$K = \begin{bmatrix} \kappa(x_1, x_1) & \cdots & \kappa(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \kappa(x_n, x_1) & \cdots & \kappa(x_n, x_n) \end{bmatrix}$$

- 给定核函数 κ ，其对应的核矩阵 K 是对称半正定的
- 给定核矩阵 K ，总能找到一个与之对应的核函数 κ

核函数

常见的核函数（核矩阵对称半正定）：

名称	表达式
线性核	$\kappa(x_i, x_j) = x_i^T x_j$
多项式核	$\kappa(x_i, x_j) = (x_i^T x_j)^d$
高斯核	$\kappa(x_i, x_j) = \exp(-\frac{\ x_i - x_j\ ^2}{2\sigma^2})$
拉普拉斯核	$\kappa(x_i, x_j) = \exp(-\frac{\ x_i - x_j\ }{\sigma})$
Sigmoid 核	$\kappa(x_i, x_j) = \tanh(\beta x_i^T x_j + \theta)$

核函数

- 若 κ_1 和 κ_2 为核函数，则对于任意正数 γ_1 、 γ_2 ，其线性组合

$$\gamma_1 \kappa_1 + \gamma_2 \kappa_2$$

也是核函数。

- 若 κ_1 和 κ_2 为核函数，则核函数的直积

$$\kappa_1 \otimes \kappa_2(x, z) = \kappa_1(x, z) \kappa_2(x, z)$$

也是核函数。

- 若 κ_1 为核函数，则对于任意函数 g

$$\kappa(x, z) = g(x) \kappa_1(x, z) g(z)$$

也是核函数。

非线性SVM

对偶问题:

$$\begin{aligned} \min_{\lambda} \quad & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \mathbf{x}_i^T \mathbf{x}_j \lambda_i \lambda_j - \sum_{i=1}^n \lambda_i \\ \text{s. t.} \quad & \sum_{i=1}^n \lambda_i y_i = 0 \\ & 0 \leq \lambda_i \leq C \quad i = 1, \dots, n \end{aligned}$$

非线性SVM

对偶问题:

$$\begin{aligned} \min_{\lambda} \quad & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j K_{ij} \lambda_i \lambda_j - \sum_{i=1}^n \lambda_i \\ \text{s. t.} \quad & \sum_{i=1}^n \lambda_i y_i = 0 \\ & 0 \leq \lambda_i \leq C \quad i = 1, \dots, n \end{aligned}$$

- Step 1: 选取合适的核函数，求出矩阵 K 。
- Step 2: 应用SMO算法求解对偶问题。

核化SVM小结

- 核技巧是处理非线性分布数据处理问题最常见的方法之一。
- 核技巧最早出线在SVM模型中，随后在许多机器学习领域都有应用。。
- 核化SVM的效果主要取决于核函数的选取。

6.7 支持向量回归

陈飞宇

fchen@cqu.edu.cn

办公室：软件学院529

支持向量回归

- 线性回归问题：寻找最佳直线（超平面），使得预测值 $f(x_i)$ 和真实标签的均方误差尽量小。
- 支持向量机思想：寻找最佳直线（超平面），使得数据点到直线的间隔尽可能的大。
- 如何结合两者呢？

支持向量回归

支持向量机 (SVM)

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

$$\text{s. t. } 1 - y_i(w^T x_i + b) \leq 0 \quad i = 1, \dots, n$$

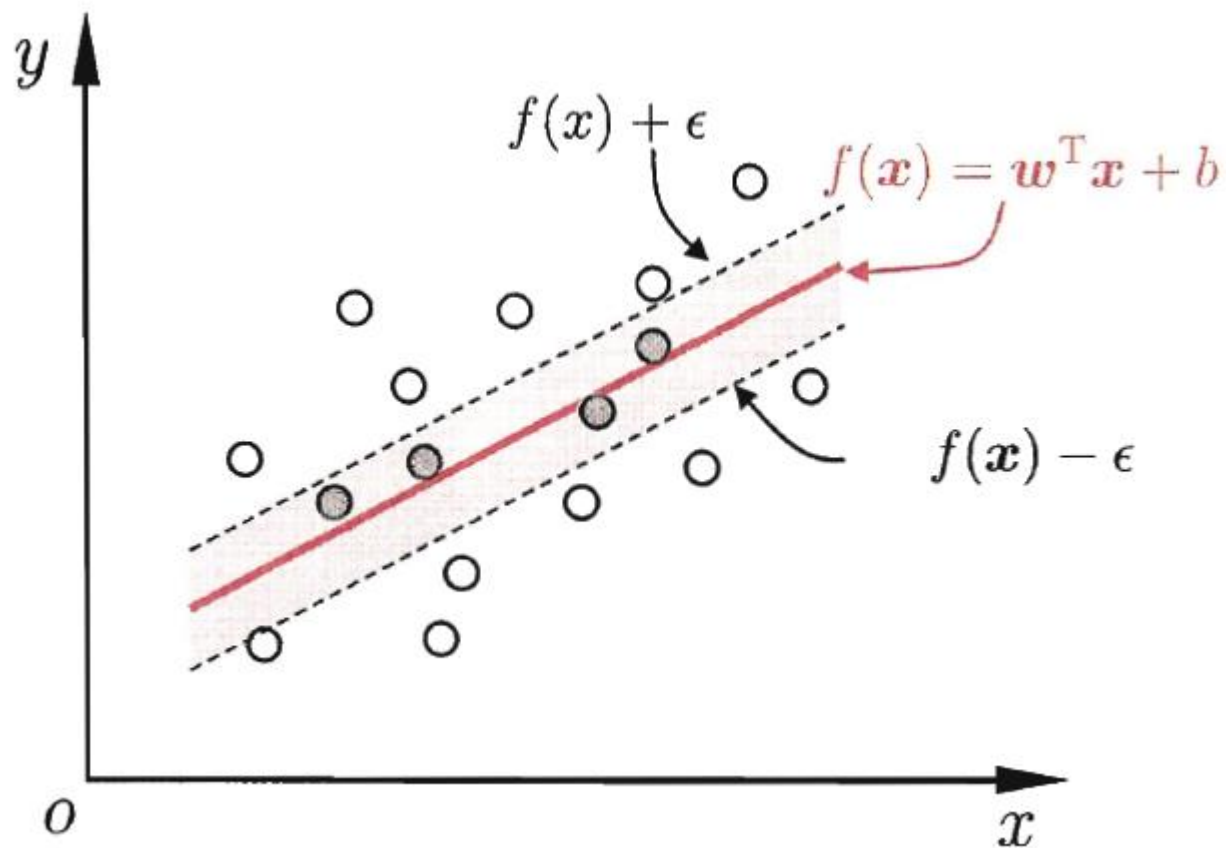
支持向量回归 (SVR)

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

$$\text{s. t. } |y_i - (w^T x_i + b)| \leq \epsilon \quad i = 1, \dots, n$$

即所有点的真实值 y_i 都落在 $f(x_i)$ 的 ϵ 邻域内。

支持向量回归



支持向量回归

支持向量回归 (SVR)

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

$$\text{s. t. } |y_i - (w^T x_i + b)| \leq \epsilon \quad i = 1, \dots, n$$

罚模型形式:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \ell_{\epsilon}(y_i - (w^T x_i + b))$$

其中

$$\ell_{\epsilon}(z) = \begin{cases} 0 & \text{if } |z| \leq \epsilon \\ |z| - \epsilon & \text{otherwise} \end{cases}$$

支持向量回归

支持向量回归 (SVR)

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

$$\text{s.t. } |y_i - (w^T x_i + b)| \leq \epsilon \quad i = 1, \dots, n$$

软间隔支持向量回归 (SVR)

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

$$\text{s.t. } y_i - (w^T x_i + b) \leq \epsilon + \xi_i \quad i = 1, \dots, n$$

$$-y_i + (w^T x_i + b) \leq \epsilon + \eta_i \quad i = 1, \dots, n$$

$$\xi_i, \eta_i \geq 0 \quad i = 1, \dots, n$$

练习题

- 求解P38页中的原始SVM优化问题。

提示： matlab function: quadprog.

- 利用SMO算法求解例子中的问题。

提示： SVM工具包---LIBSVM

- 在同一人工（或下载）数据集上比较 逻辑回归与支持向量机分类精度。
- 在SMO中，假设选定需要更新的变量为 λ_i 和 λ_j ，试求 λ_i 和 λ_j 的更新公式。