支持向量机

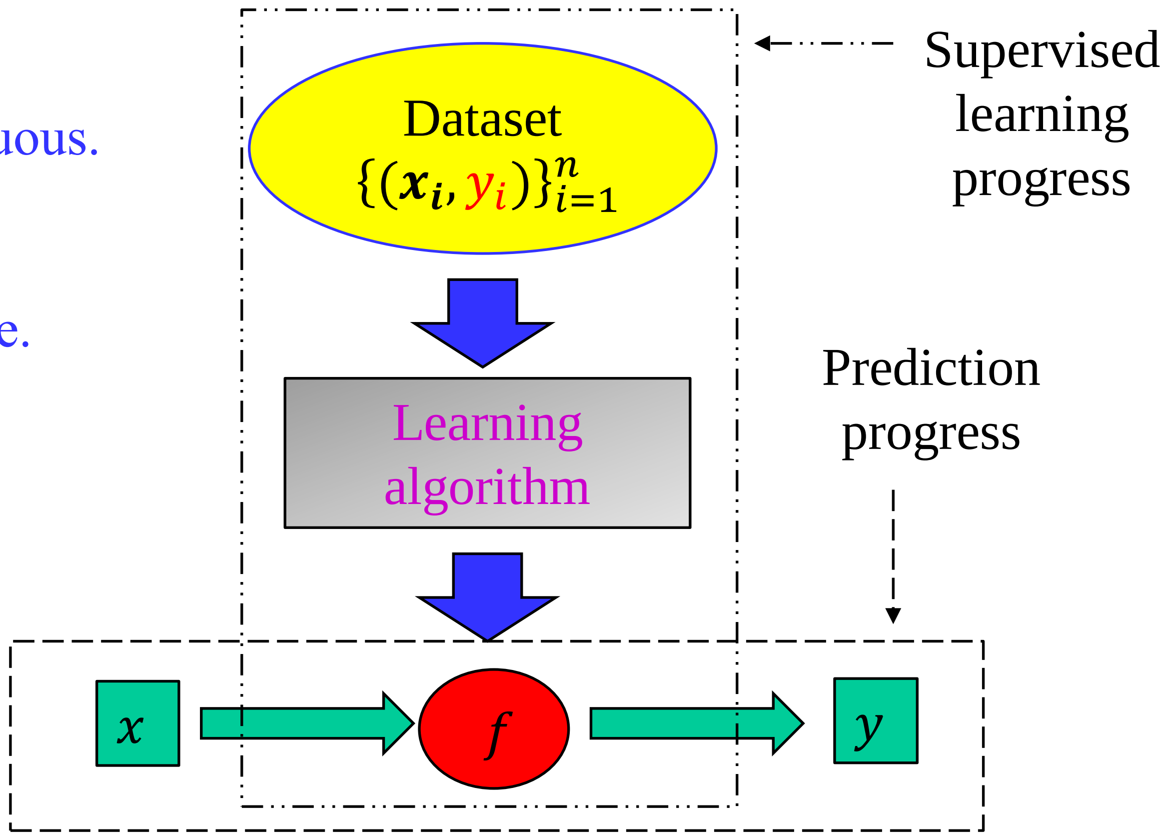
陈飞宇
fchen@cqu.edu.cn
办公室：软件学院529

监督学习（Supervised Learning）

Learning with a teacher

Regression:
y_i's are continuous.

Classification:
y_i's are discrete.

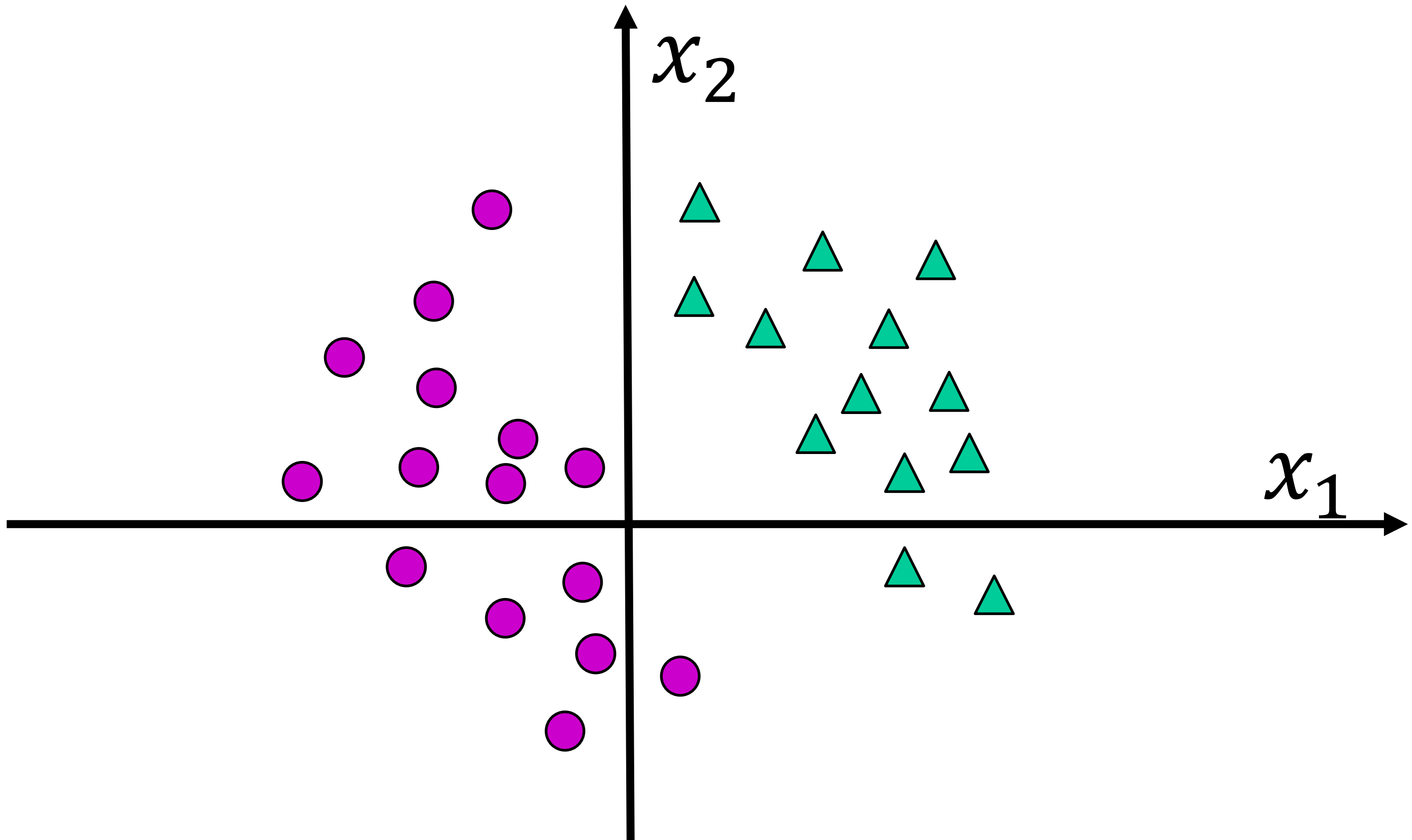


5.1 SVM模型

陈飞宇
fchen@cqu.edu.cn
办公室：软件学院529

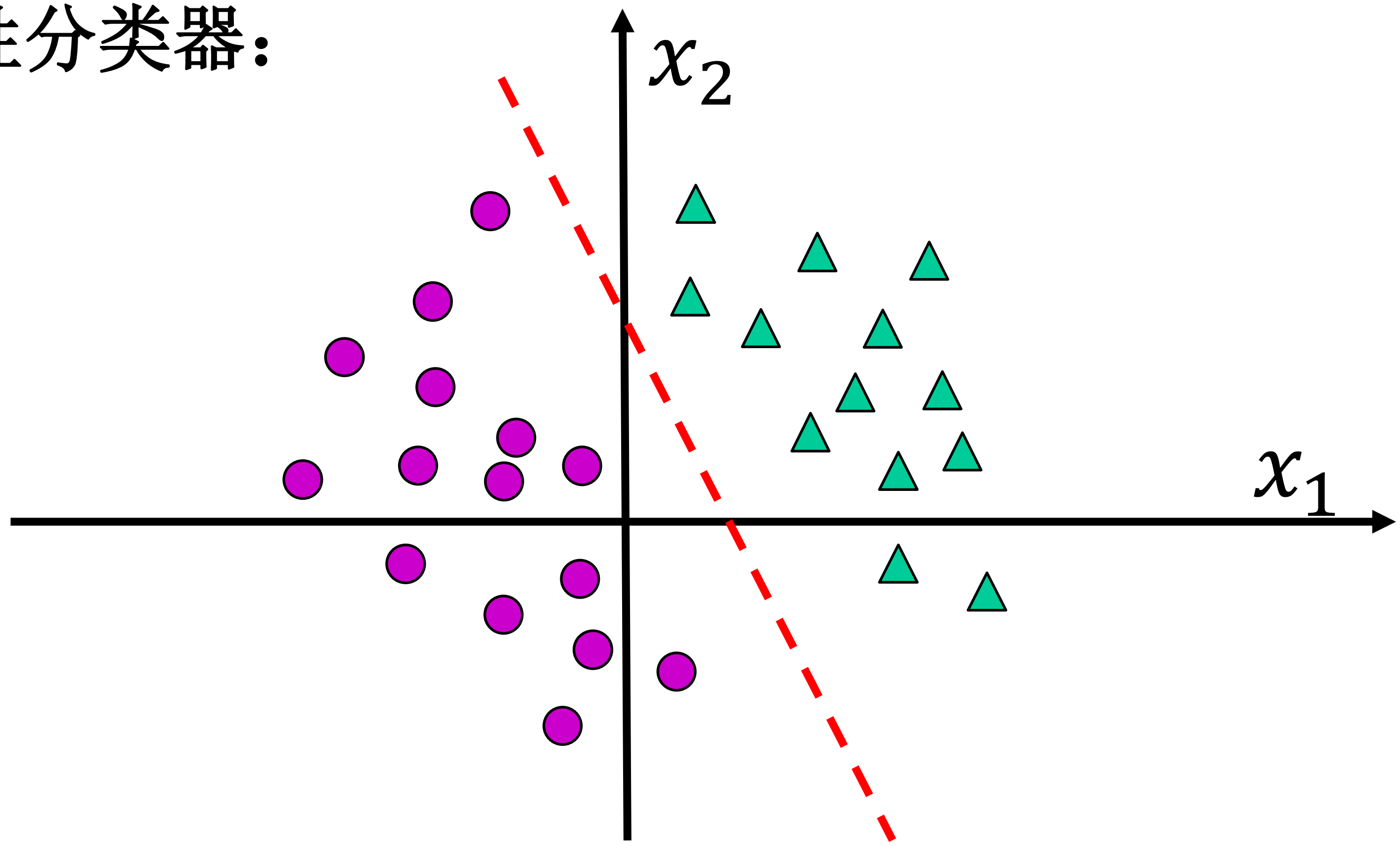
二分类（Binary Classification）

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

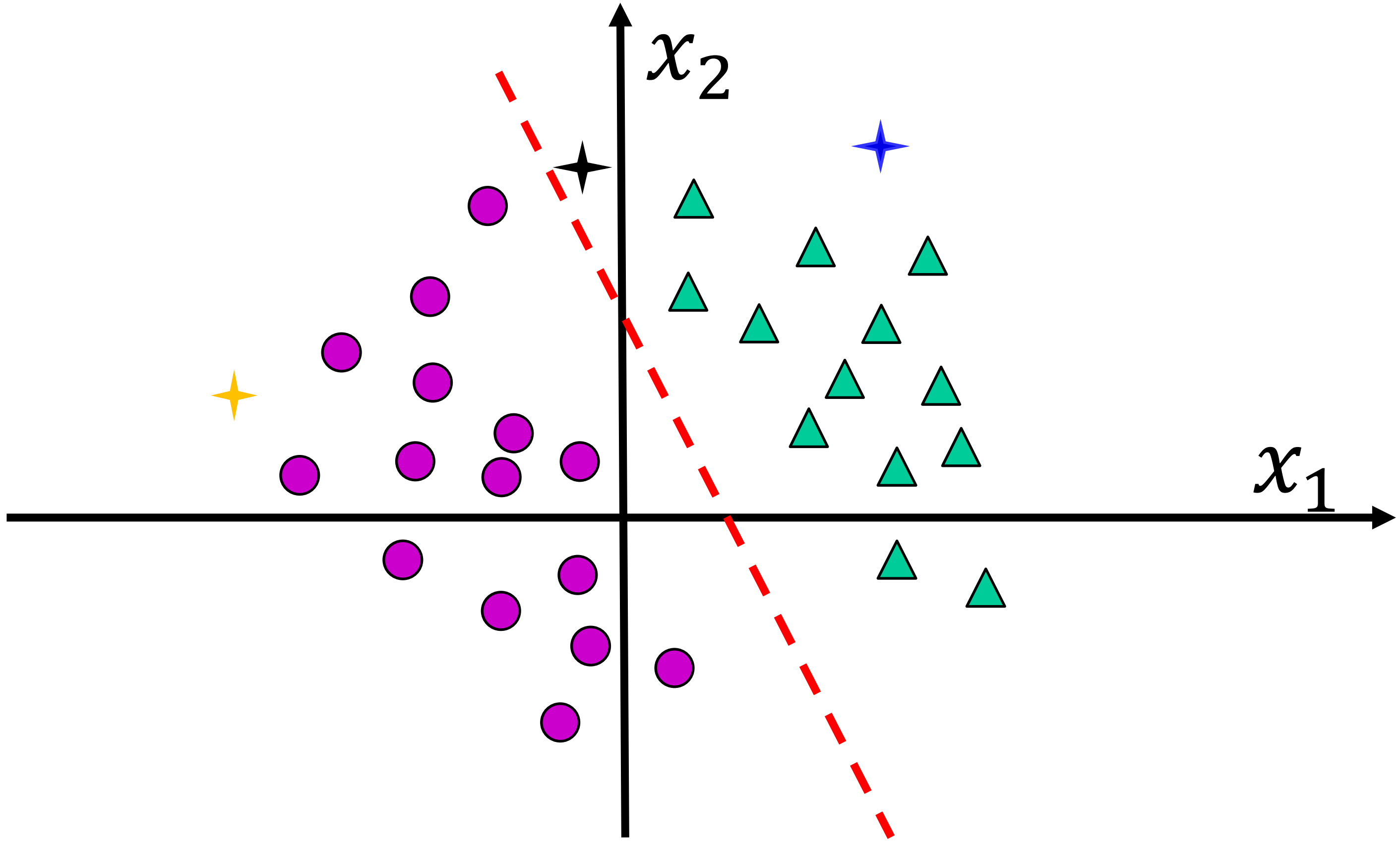


归纳（Induction）

线性分类器：



演绎（Deduction）



问题?

数据点✦,✦ 在直线上方，为正样本。

数据点✦ 在直线下方，为负样本。

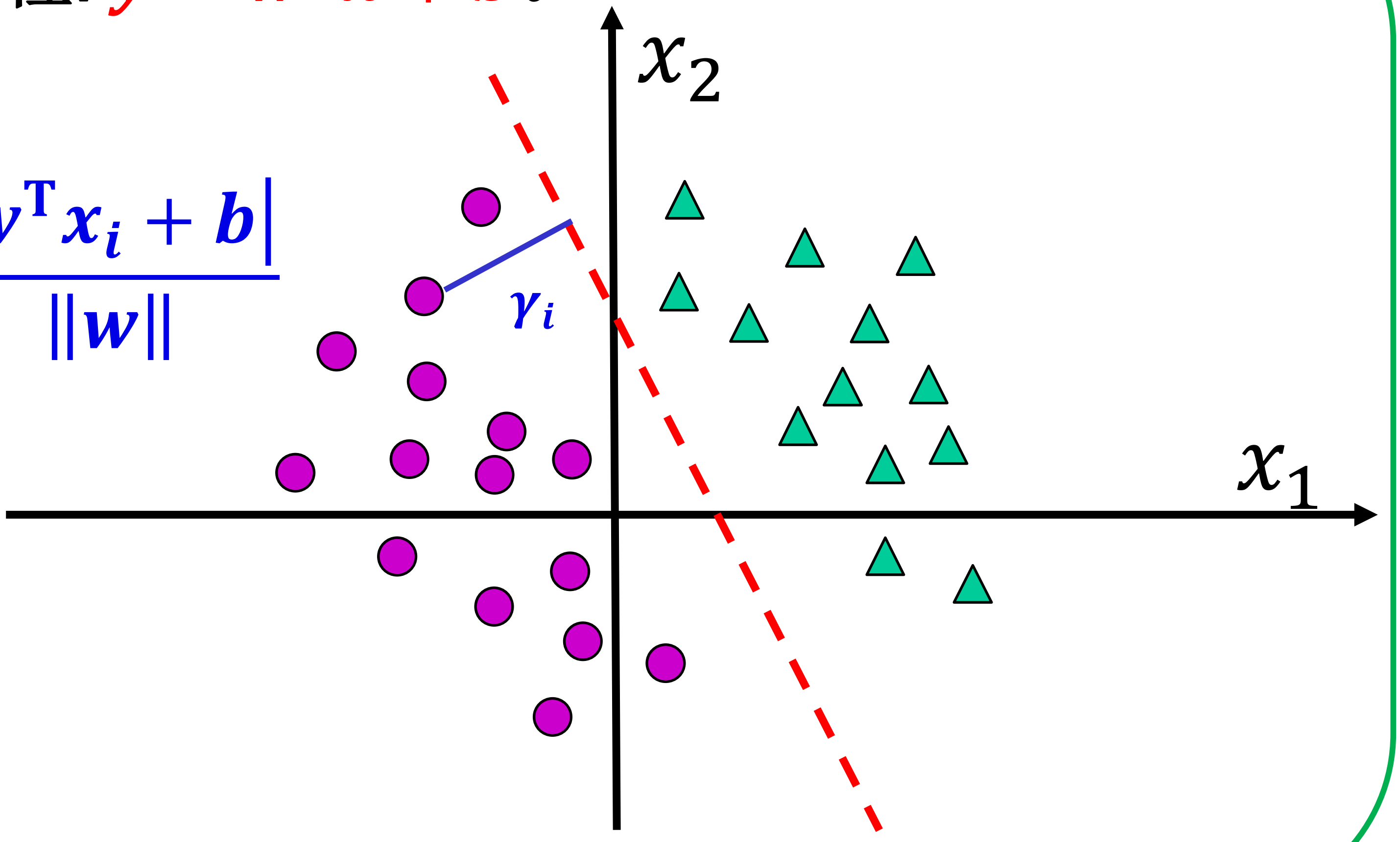
数据点✦ ✦都被判定为正样本，但是置信程度（confident）不同。

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: $y = w^{\mathrm{T}}x + b$。

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



间隔（Margin）

考虑输出变量 $\boldsymbol{y_i} \in \{-\mathbf{1}, \mathbf{1}\}$。

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

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

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

$\boldsymbol{\gamma}$ 表示所有数据点到直线的最短距离。

间隔最大化（Maximum Margin）

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

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

我们希望找到最大的γ，使得对于所有数据点$\{(x_i, y_i)\}_{i=1}^{m}$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$\boldsymbol{w}, \boldsymbol{b}$进行适当缩放，使得$\gamma = \frac{1}{\|\boldsymbol{w}\|}$，则

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

$$\text{s.t.} \quad \frac{y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + \boldsymbol{b})}{\|\boldsymbol{w}\|} \geq \gamma$$

等价于

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

$$\text{s.t.} \quad y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + \boldsymbol{b}) \geq 1$$

支持向量机

支持向量机（**Support Vector Machine**）

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

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

- 满足 $\boldsymbol{y_i}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x_i}+\boldsymbol{b}) = \mathbf{1}$ 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直线。

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

5.2 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：

$$\min_{\boldsymbol{x}\in\mathbb{R}^{\boldsymbol{n}}} \quad \boldsymbol{f}(\boldsymbol{x})$$

$$\mathbf{s.t.} \quad \boldsymbol{g}_{\boldsymbol{i}}(\boldsymbol{x}) \le \mathbf{0} \qquad \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{m}$$

拉格朗日函数(Lagrange function)：

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{f}(\boldsymbol{x}) + \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{m}} \boldsymbol{\lambda}_{\boldsymbol{i}}\, \boldsymbol{g}_{\boldsymbol{i}}(\boldsymbol{x})$$

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：

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \ \boldsymbol{f}(\boldsymbol{x}) \quad \mathbf{s.t.} \quad \boldsymbol{g_i}(\boldsymbol{x}) \leq \mathbf{0} \qquad \boldsymbol{i} = \mathbf{1}, \ldots, \boldsymbol{m}$$

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

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

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

$$\boldsymbol{f}(\boldsymbol{x}) = \min \ \max_{\boldsymbol{\lambda}\geq 0} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

拉格朗日乘子法

极大极小问题（Primal Problem）：

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

极小极大问题（Dual Problem）：

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

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）：

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

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

5.3 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）

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

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

- 变量: $\boldsymbol{w} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n)^{\mathrm{T}}$， $\boldsymbol{b}$
- 约束: m 个不等式约束。
- 数据: $x_i \in \mathbb{R}^n$， $y_i \in \{-1, 1\}$ $\quad i = 1, \dots, m$

对偶问题

极小极大问题（Dual Problem）：

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

其中，

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{b}, \boldsymbol{\lambda}) = \frac{\mathbf{1}}{\mathbf{2}} \|\boldsymbol{w}\|^2 + \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{m}} \boldsymbol{\lambda}_{\boldsymbol{i}} \left(\mathbf{1} - \boldsymbol{y}_{\boldsymbol{i}}(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_{\boldsymbol{i}} + \boldsymbol{b})\right)$$

极小问题

$$\min_{\boldsymbol{w},\boldsymbol{b}} \ \frac{1}{2}\|\boldsymbol{w}\|^2 + \sum_{i=1}^{m} \lambda_i \left(1 - y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + \boldsymbol{b})\right)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{i=1}^{m} \lambda_i y_i \boldsymbol{x}_i\,; \qquad \frac{\partial \mathcal{L}}{\partial \boldsymbol{b}} = -\sum_{i=1}^{m} \lambda_i y_i$$

$$\boldsymbol{w} = \sum_{i=1}^{m} \lambda_i y_i \boldsymbol{x}_i; \qquad \sum_{i=1}^{m} \lambda_i y_i = 0$$

极小问题

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

$$\min_{\boldsymbol{w},b} \mathcal{L}(\boldsymbol{w}, b, \boldsymbol{\lambda})$$

$$= \frac{1}{2}\|\boldsymbol{w}\|^2 + \sum_{i=1}^{m}\lambda_i \left(1 - y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b)\right)$$

$$= \frac{1}{2}\left\|\sum_{i=1}^{m}\lambda_i y_i \boldsymbol{x}_i\right\|^2 + \sum_{i=1}^{m}\lambda_i \left(1 - y_i\left(\left(\sum_{j=1}^{m}\lambda_j y_j \boldsymbol{x}_j^{\mathrm{T}}\right)\boldsymbol{x}_i + b\right)\right)$$

$$= \sum_{i=1}^{m}\lambda_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j - \sum_{i=1}^{m}\lambda_i y_i b$$

$$= \sum_{i=1}^{m}\lambda_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$$

对偶问题

SVM 对偶问题：

$$\min_{\boldsymbol{\lambda}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j - \sum_{i=1}^{m}\lambda_i$$

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

$$\lambda_i \geq 0 \qquad i = 1,\dots,m$$

变量：$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_m)^{\mathrm{T}}$

约束：m 个等式约束，m 个不等式约束。

数据：$\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j \in \mathbb{R}, \qquad y_i y_j \in \{-1, 1\} \quad i = 1,\dots,m$

KKT 条件

KKT 条件：

$$\begin{cases} \boldsymbol{w} - \sum_{i=1}^{m} \lambda_i y_i \boldsymbol{x}_i = 0 \\ \sum_{i=1}^{m} \lambda_i y_i = 0 \\ \lambda_i (y_i (\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i + b) - 1) = 0 \\ \lambda_i \geq 0 \\ y_i (\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i + b) - 1 \geq 0 \end{cases}$$

对偶问题

对于任意的数据点 $\boldsymbol{x}_i$，我们有 $\boldsymbol{\lambda}_i = 0$ 或者 $\boldsymbol{y}_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + \boldsymbol{b}) - \mathbf{1} = 0$.

- $\boldsymbol{\lambda}_i = 0$：

  对偶问题中变量 $\boldsymbol{\lambda}_i$ 不在求和中出现。

- $\boldsymbol{y}_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + \boldsymbol{b}) - \mathbf{1} = 0$：

  $\boldsymbol{x}_i$ 是一个支持向量, $\boldsymbol{b} = \boldsymbol{y}_i - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i$

求解分类平面

SVM 对偶问题：

$$\min_{\lambda} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j - \sum_{i=1}^{m}\lambda_i$$

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

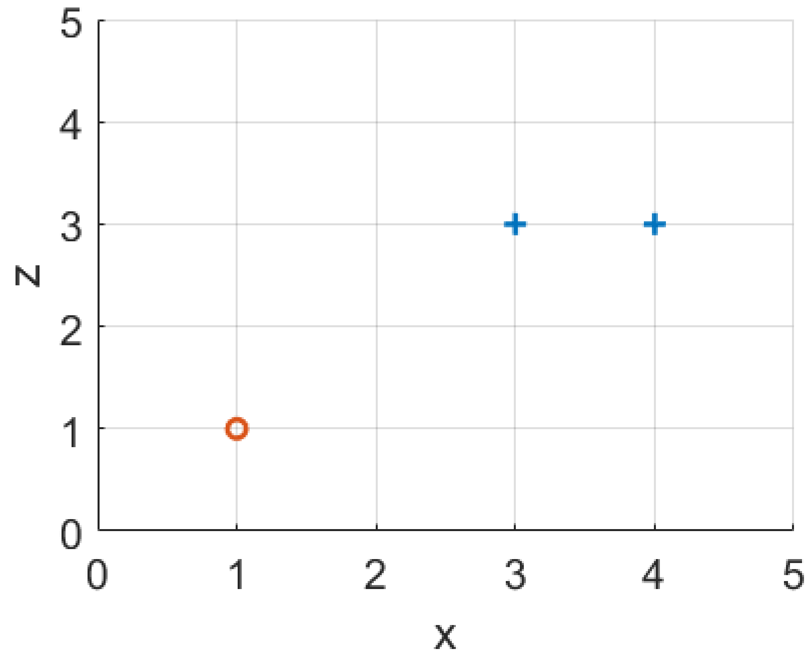
$$\lambda_i \geq 0 \qquad i = 1, \dots, m$$

$$\boldsymbol{w} = \sum_{i=1}^{m}\lambda_i y_i \boldsymbol{x}_i, \qquad b = y_j - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_j$$

$$f(\boldsymbol{x}) = \left(\sum_{i=1}^{m}\lambda_i y_i \boldsymbol{x}_i^{\mathrm{T}}\right)(\boldsymbol{x} - \boldsymbol{x}_j) + y_j$$

例子

假设正例点是 $\boldsymbol{x}_1=(3,3)^{\mathrm{T}}$, $\boldsymbol{x}_2=(4,3)^{\mathrm{T}}$, 负例点是 $\boldsymbol{x}_3=(1,1)^{\mathrm{T}}$。试利用支持向量机算法求出分类超平面。



例子

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

$y_1=1,\quad y_2=1,\quad y_3=-1$。

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，

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j x_i^{\mathrm{T}} x_j$$

$$=18\lambda_1^2+25\lambda_2^2+2\lambda_3^2+42\lambda_1\lambda_2-12\lambda_1\lambda_3-14\lambda_2\lambda_3$$

例子

$$\min_{\lambda}\ 9\lambda_1^2+\frac{25}{2}\lambda_2^2+\lambda_3^2+21\lambda_1\lambda_2-6\lambda_1\lambda_3-7\lambda_2\lambda_3-\sum_{i=1}^{3}\lambda_i$$

$$\text{s.t.}\qquad \lambda_1+\lambda_2-\lambda_3=0,\quad \lambda_i\geq 0\qquad i=1,\dots,3$$

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

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

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

例子

二次规划问题：

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

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

目标函数分别对 λ_1 和 λ_2 求导有

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

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

此时 λ_2 不满足约束，最优解在边界出取得。

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，$\lambda_3 = \lambda_1 + \lambda_2 = \frac{1}{4}$

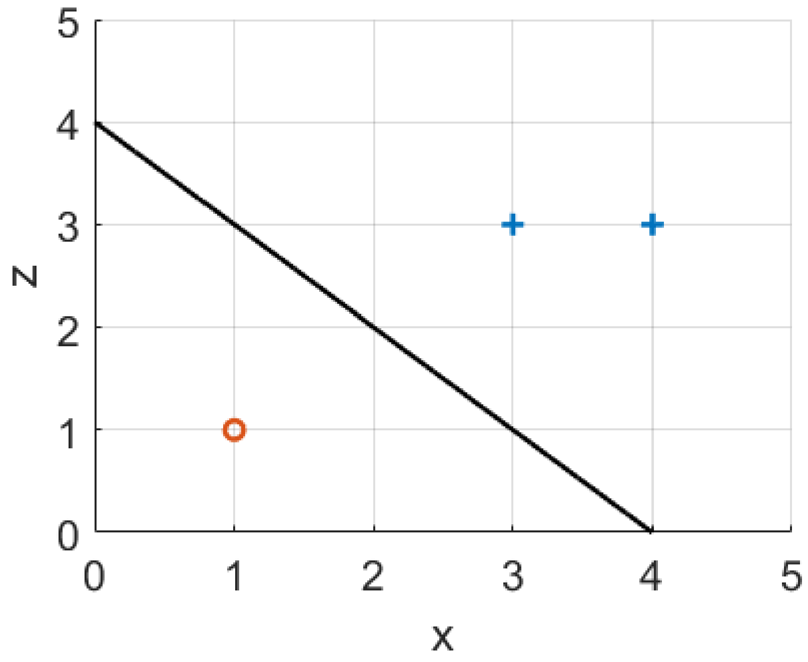
$$w = \sum_{i=1}^{m} \lambda_i y_i x_i = \frac{1}{4} x_1 + 0 - \frac{1}{4} x_3 = \left(\frac{1}{2}, \frac{1}{2}\right)^{\mathrm{T}}$$

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

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

例子

假设正例点是 $\boldsymbol{x}_1 = (3,3)^{\mathrm{T}}$, $\boldsymbol{x}_2 = (4,3)^{\mathrm{T}}$, 负例点是 $\boldsymbol{x}_3 = (1,1)^{\mathrm{T}}$。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 $\boldsymbol{x}_1 = (3,3)^{\mathrm{T}}$, $\boldsymbol{x}_2 = (4,3)^{\mathrm{T}}$, 负例点是 $\boldsymbol{x}_3 = (1,1)^{\mathrm{T}}$。试利用支持向量机算法求出分类超平面。

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

$$\begin{aligned}
\text{s.t.} \quad & 1 - (3w_1 + 3w_2 + b) \le 0 \\
& 1 - (4w_1 + 3w_2 + b) \le 0 \\
& 1 + (w_1 + w_2 + b) \le 0
\end{aligned}$$

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。

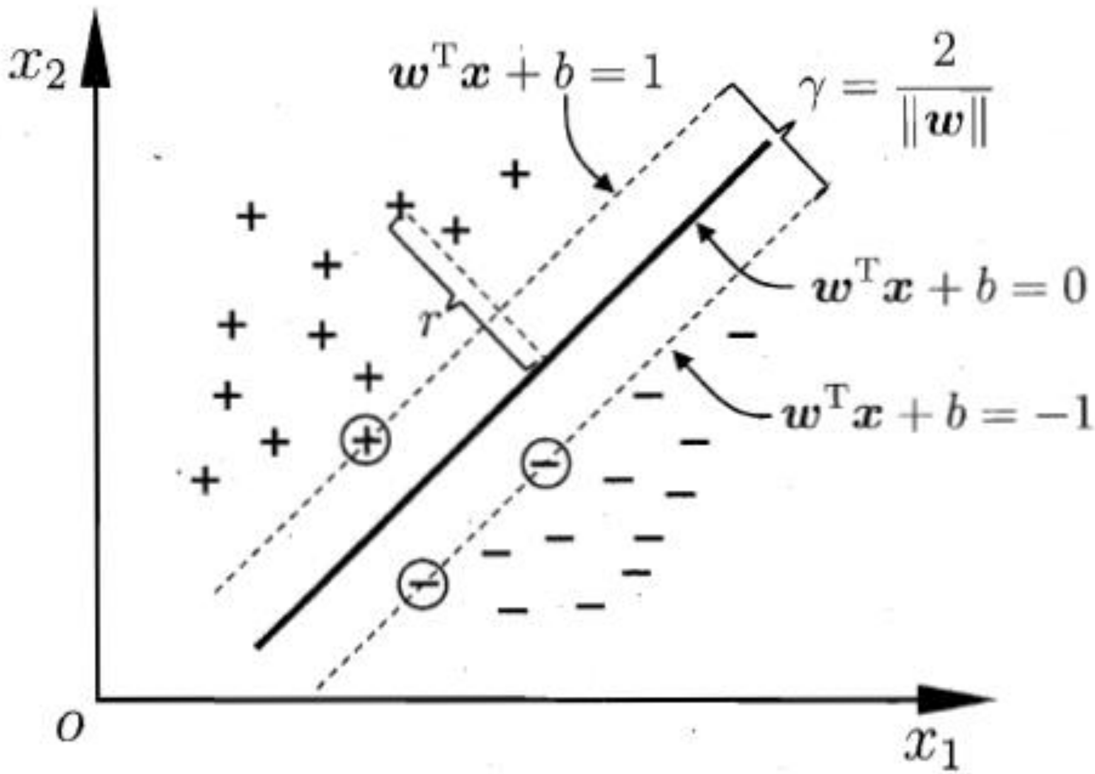
5.4 软间隔SVM

陈飞宇

fchen@cqu.edu.cn

办公室：软件学院529

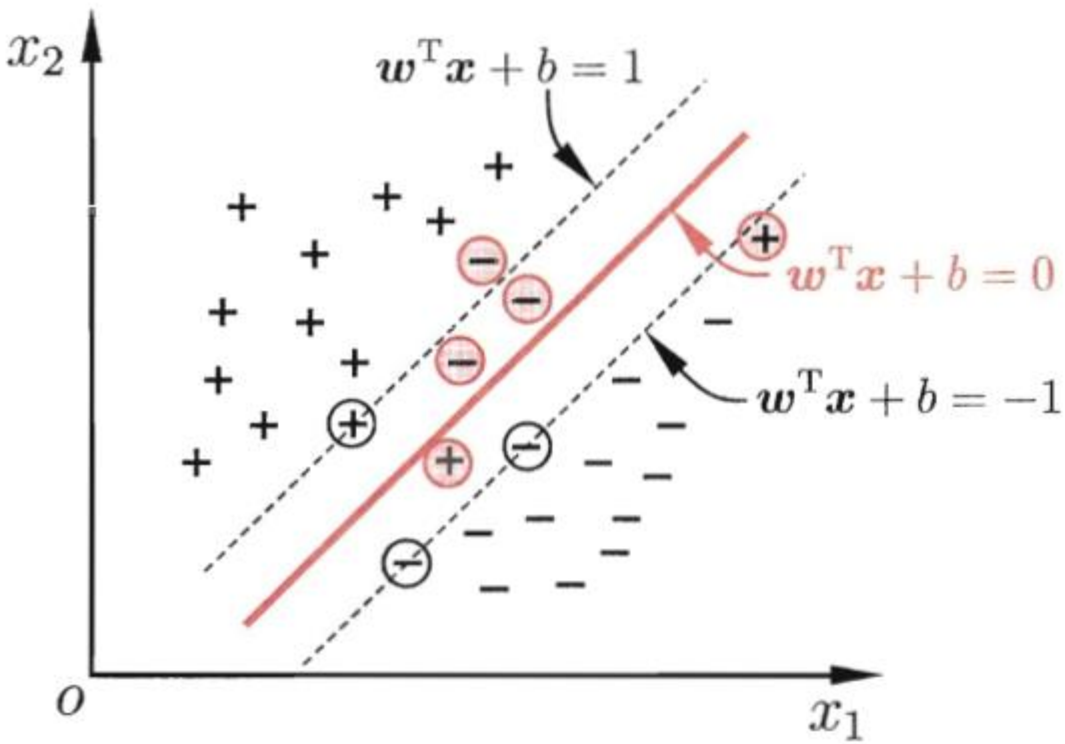
SVM图例



x_2
$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b=1$
$\gamma=\frac{2}{\|\boldsymbol{w}\|}$
r
$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b=0$
$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b=-1$
O
x_1

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）

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

$$\mathbf{s.t.} \quad \boldsymbol{y}_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + \boldsymbol{b}) \geq \mathbf{1} - \xi_i \quad \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{m}$$

$$\xi_i \geq 0 \qquad \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{m}$$

- 变量: $\boldsymbol{w}, \boldsymbol{b}, \xi_i$(松弛变量)
- 约束: 2m 个不等式约束。
- 数据：$\boldsymbol{x}_i \in \mathbb{R}^n$, $\boldsymbol{y}_i \in \{-\mathbf{1}, \mathbf{1}\}$ $\quad \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{m}$

对偶问题

极小极大问题（Dual Problem）：

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

其中，

$$\mathcal{L}(\boldsymbol{w},\boldsymbol{b},\boldsymbol{\xi},\boldsymbol{\lambda},\boldsymbol{\mu})$$

$$= \frac{1}{2}\|\boldsymbol{w}\|^2 + \boldsymbol{C}\sum_{i=1}^{m}\boldsymbol{\xi}_i - \sum_{i=1}^{m}\boldsymbol{\mu}_i\,\boldsymbol{\xi}_i$$

$$+ \sum_{i=1}^{m}\boldsymbol{\lambda}_i\left(\mathbf{1} - \boldsymbol{\xi}_i - \boldsymbol{y}_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + \boldsymbol{b})\right)$$

极小问题

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

$$+\sum_{i=1}^{m}\lambda_i\left(1-\xi_i-y_i(w^{\mathrm{T}}x_i+b)\right)$$

$$\frac{\partial\mathcal{L}}{\partial w} = w - \sum_{i=1}^{m}\lambda_i\,y_i x_i \qquad \Longrightarrow \qquad w = \sum_{i=1}^{m}\lambda_i\,y_i x_i$$

$$\frac{\partial\mathcal{L}}{\partial b} = -\,\textstyle\sum_{i=1}^{m}\lambda_i\,y_i \qquad \Longrightarrow \qquad 0 = \textstyle\sum_{i=1}^{m}\lambda_i\,y_i$$

$$\frac{\partial\mathcal{L}}{\partial \xi_i} = C-\lambda_i-\mu_i \qquad \Longrightarrow \qquad C = \lambda_i+\mu_i$$

极小问题

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

$$\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \mathcal{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

$$= \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \lambda_i (1 - \xi_i - y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b)) - \sum_{i=1}^{m} \mu_i \xi_i$$

$$= \sum_{i=1}^{m} \lambda_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j + \sum_{i=1}^{m}(C - \lambda_i - \mu_i)\xi_i$$

$$= \sum_{i=1}^{m} \lambda_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j$$

对偶问题

软间隔SVM 对偶问题：

$$\min_{\boldsymbol{\lambda}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j - \sum_{i=1}^{m}\lambda_i$$

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

$$\lambda_i + \mu_i = C \qquad i = 1, \dots, m$$

$$\lambda_i \geq 0 \qquad i = 1, \dots, m$$

$$\mu_i \geq 0 \qquad i = 1, \dots, m$$

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：

$$\min_{\boldsymbol{\lambda}} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j - \sum_{i=1}^{m}\lambda_i$$

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

$$0 \le \lambda_i \le C \qquad i = 1, \dots, m$$

变量：$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_m)^{\mathrm{T}}$

约束：m 个等式约束，m 个不等式约束。

数据：$\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j \in \mathbb{R}$, $\quad y_i y_j \in \{-1, 1\} \quad i = 1, \dots, m$

KKT 条件

KKT 条件：

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

支持向量

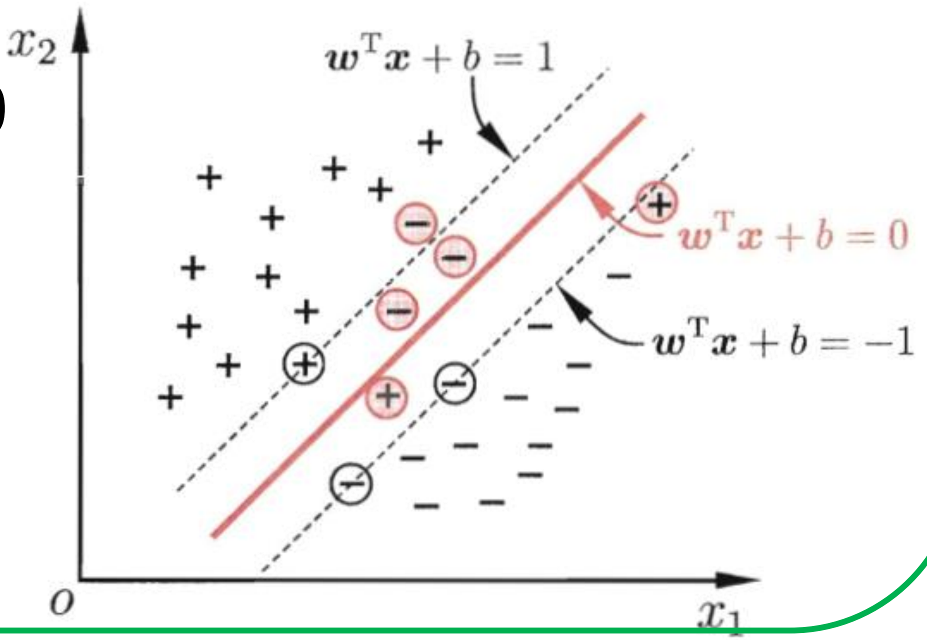
分离超平面：$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b}=\mathbf{0}$

间隔边界： $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b}=\pm\mathbf{1}$

支持向量：

$\boldsymbol{y}_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+\boldsymbol{b})-\mathbf{1}\leq 0$

- 间隔边界上。
- 间隔边界间。
- 分类错误点。



支持向量

- 若 $\lambda_i < C$，则 $\xi_i = 0, \boldsymbol{x_i}$ 落在间隔边界上。
- 若 $\lambda_i = C$， $\mathbf{0} < \xi_i < \mathbf{1}$时， $\boldsymbol{x_i}$ 落在间隔边界和分离超平面之间。
- 若 $\lambda_i = C$， $\xi_i = \mathbf{1}$时, $\boldsymbol{x_i}$ 落在分离超平面上。
- 若 $\lambda_i = C$， $\xi_i > \mathbf{1}$时， $\boldsymbol{x_i}$ 落在分离超平面另外一侧。

软间隔SVM

软间隔支持向量机（SVM）

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

$$\text{s.t.} \quad \boldsymbol{y}_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + \boldsymbol{b}) \geq \mathbf{1} - \xi_i \quad \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{m}$$

$$\xi_i \geq 0 \qquad \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{m}$$

- $\xi_i \geq \max\left(\mathbf{0}, \mathbf{1} - \boldsymbol{y}_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + \boldsymbol{b})\right)$

$$\min_{\boldsymbol{w},\boldsymbol{b},\xi} \quad \frac{\mathbf{1}}{\mathbf{2}}\|\boldsymbol{w}\|^2 + \boldsymbol{C}\sum_{i=1}^{m}(\mathbf{1} - \boldsymbol{y}_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + \boldsymbol{b})_+$$

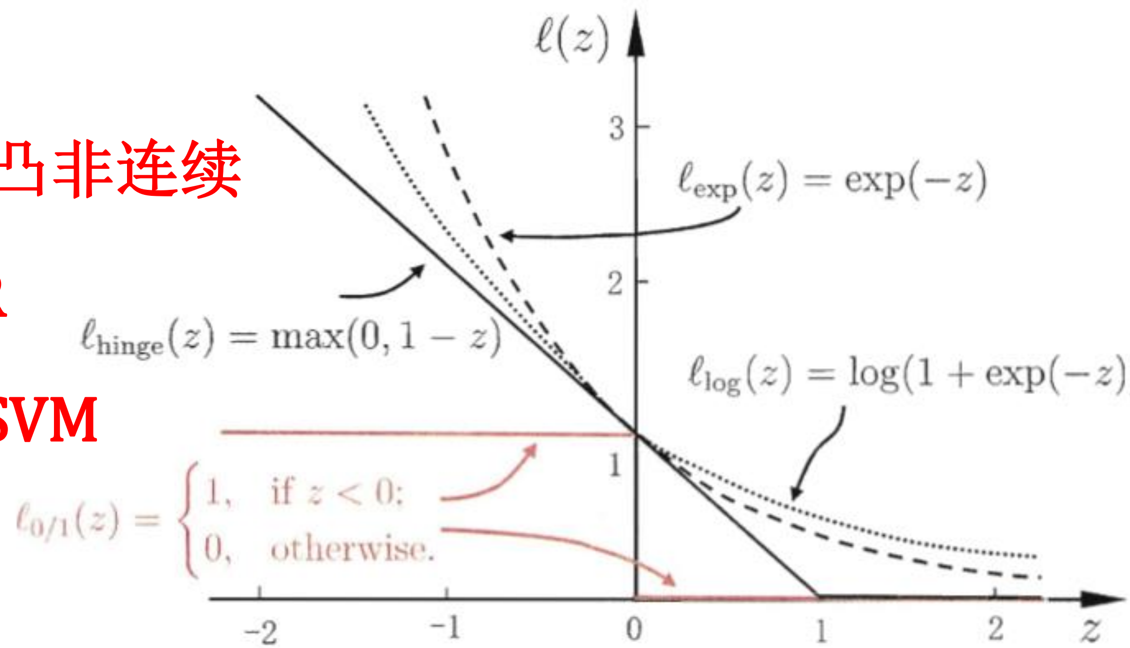
正则化分类模型

$$\min_{w,b} \sum_{i=1}^{m} \ell(y_i(w^{\mathrm{T}}x_i + b)) + \frac{\alpha}{2}\|w\|^2$$

损失函数的选择

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

--- Adaboost



当$C \to +\infty$时

支持向量机（SVM）

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

$$\text{s.t.} \quad y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1 \qquad i = 1,\dots,m$$

软间隔支持向量机（SVM）

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

$$\text{s.t.} \quad y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1 - \xi_i \quad i = 1,\dots,m$$

$$\xi_i \geq 0 \qquad i = 1,\dots,m$$

当$C \rightarrow +\infty$时

软间隔SVM 对偶问题:

$$\min_{\lambda} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j - \sum_{i=1}^{m}\lambda_i$$

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

$$0 \le \lambda_i \le C \qquad i = 1, \dots, m$$

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。

5.5 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):

$$\min_{\lambda} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j - \sum_{i=1}^{m}\lambda_i$$

$$\mathbf{s.t.} \quad \sum_{i=1}^{m}\lambda_i y_i = 0 \qquad i = 1, \ldots, m$$

$$0 \le \lambda_i \le C \qquad i = 1, \ldots, m$$

SMO的两个部分：

- 解析地求解两个变量二次规化问题。
- 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：

$$\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j x_i^{\mathrm{T}} x_j - \sum_{i=1}^{m}\lambda_i$$

$$= \frac{1}{2}\left(x_i^{\mathrm{T}} x_i\right)\lambda_i^2 + \frac{1}{2}\left(x_j^{\mathrm{T}} x_j\right)\lambda_j^2 + \left(y_i y_j x_i^{\mathrm{T}} x_j\right)\lambda_i\lambda_j - \lambda_i$$

$$- \lambda_j + \left(\sum_{k\neq i,j}^{m} y_i y_k x_i^{\mathrm{T}} x_k \lambda_k\right)\lambda_i + \left(\sum_{k\neq i,j}^{m} y_j y_k x_j^{\mathrm{T}} x_k \lambda_k\right)\lambda_j$$

约束：

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

SMO算法

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

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

其中:

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

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

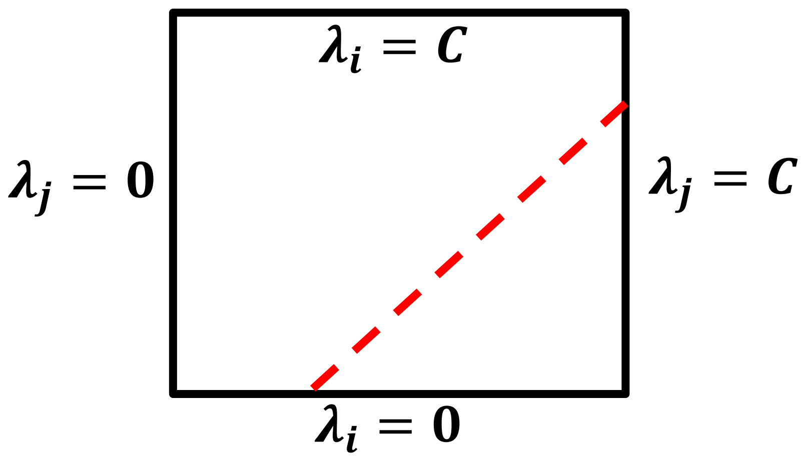
$$c_i = \sum_{k\neq i,j}^{m} y_i y_k x_i^{\mathrm{T}} x_k \lambda_k - 1 \quad c_j = \sum_{k\neq i,j}^{m} y_j y_k x_j^{\mathrm{T}} x_k \lambda_k - 1$$

SMO算法

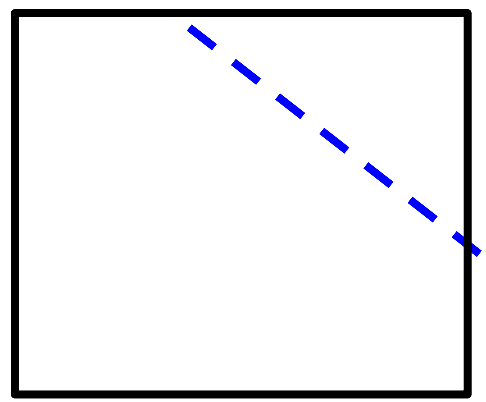
$$\min_{\lambda_i,\lambda_j} \quad \frac{1}{2}K_{ii}\lambda_i^2 + \frac{1}{2}K_{jj}\lambda_j^2 + c_{ij}\lambda_i\lambda_j + c_i\lambda_i + c_j\lambda_j$$

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

约束：



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



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

SMO算法

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

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

- 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。

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

- 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

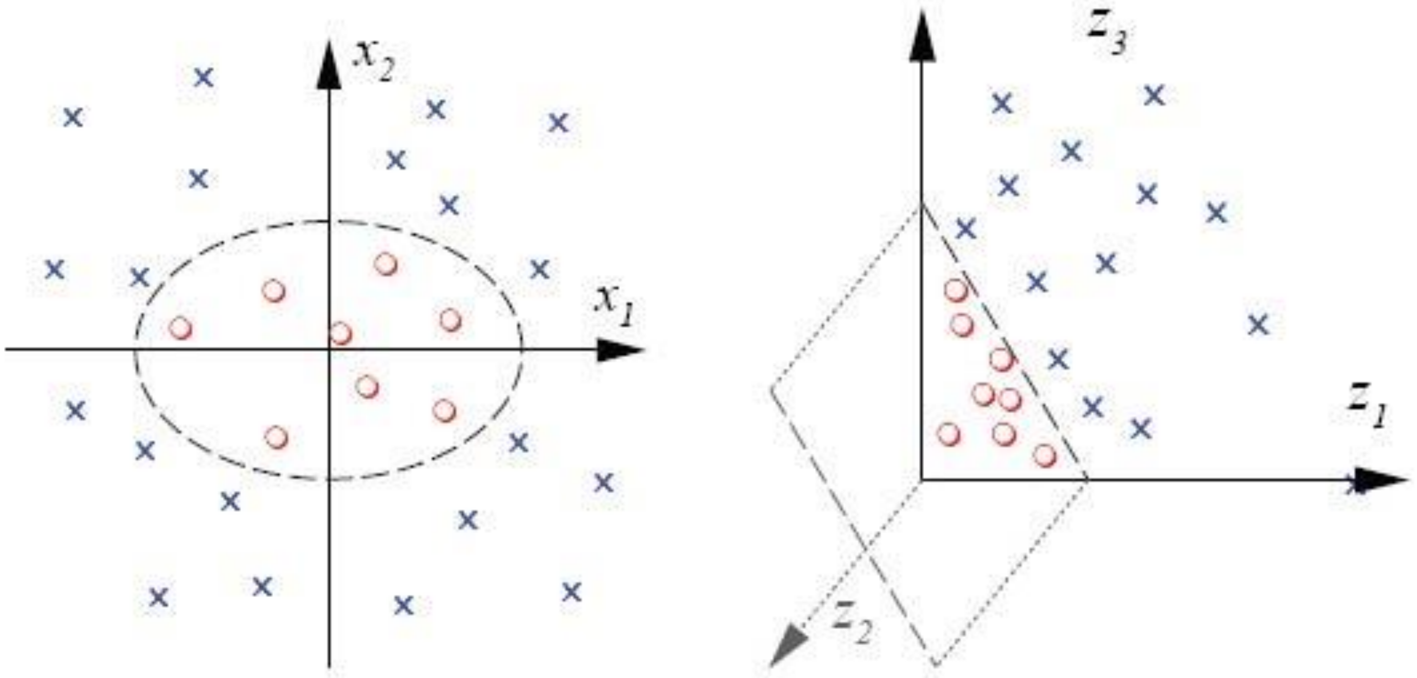
- 序列极小化优化算法（SMO）是求解SVM模型最高效的算法。
- SMO每次迭代解两个变量的二次优化问题，其最优解可以显式表达。
- SMO利用了启发式算法根据数据点违反KKT条件的大小选取需要迭代的变量。

5.5 非线性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$\{x_i\}_{i=1}^m$，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： $w_1x^{(1)}+w_2x^{(2)}+b=0$

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 $\boldsymbol{\phi}$ 是不现实的。

定义核函数（**Kernel Function**）：

$$\boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{\phi}(\boldsymbol{x}_i)^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x}_i)$$

这样，$\boldsymbol{\phi}$ 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。

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	$K(x_i, x_j) = x_i^{\mathrm{T}} x_j$
多项式核	$K(x_i, x_j) = (x_i^{\mathrm{T}} x_j)^d$
高斯核	$K(x_i, x_j) = \exp(-\frac{\lVert x_i - x_j \rVert^2}{2\sigma^2})$
拉普拉斯核	$K(x_i, x_j) = \exp(-\frac{\lVert x_i - x_j \rVert}{\sigma})$
Sigmoid 核	$K(x_i, x_j) = \tanh(\beta x_i^T x_j + \theta)$

非线性SVM

对偶问题：

$$\min_{\lambda} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j - \sum_{i=1}^{m}\lambda_i$$

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

$$0 \le \lambda_i \le C \qquad i = 1, \dots, m$$

非线性SVM

对偶问题:

$$\min_{\lambda} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j K_{ij} - \sum_{i=1}^{m}\lambda_i$$

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

$$0 \le \lambda_i \le C \qquad i = 1, \dots, m$$

- 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。

练习题

- 求解P35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。