

机器学习初步

—非线性分类模型、SVM

李爽

助理教授，特聘副研究员

数据科学与知识工程研究所

E-mail: shuangli@bit.edu.cn

Homepage: shuangli.xyz



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

本章的主要内容

- 非线性分类模型
- 线性可分支持向量机
- 软间隔与正则化
- 核函数与非线性SVM
- 序列最小最优化算法
- 总结

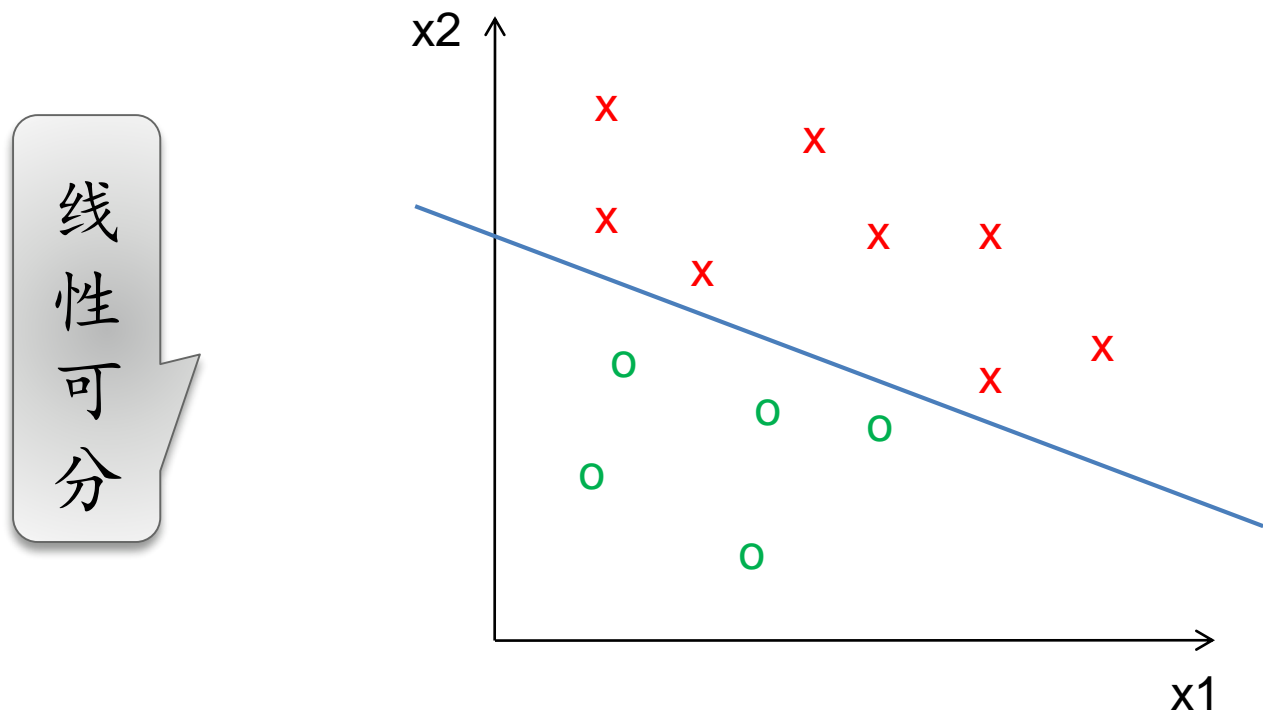
本章的主要内容

- 非线性分类模型
- 线性可分支持向量机
- 软间隔与正则化
- 核函数与非线性SVM
- 序列最小最优化算法
- 总结

一、非线性分类模型：分类模型

- Classification :

Find a *function* f to separate the classes:



f : 线性分类器

在样本空间中寻找
一个将不同类别的
样本分开的超平面

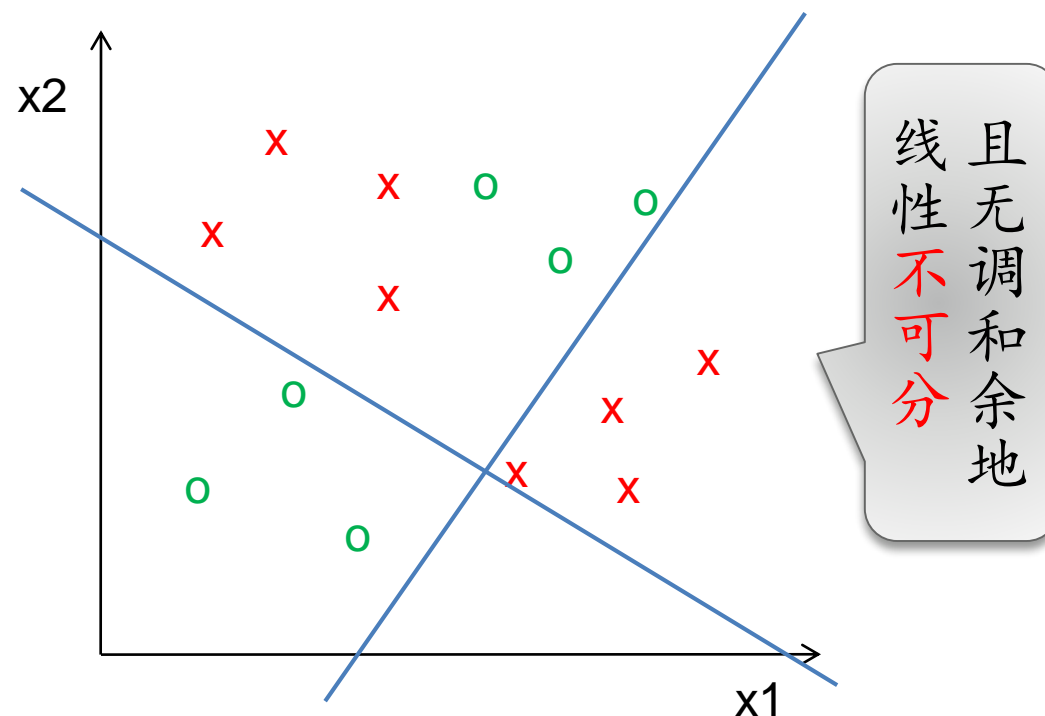
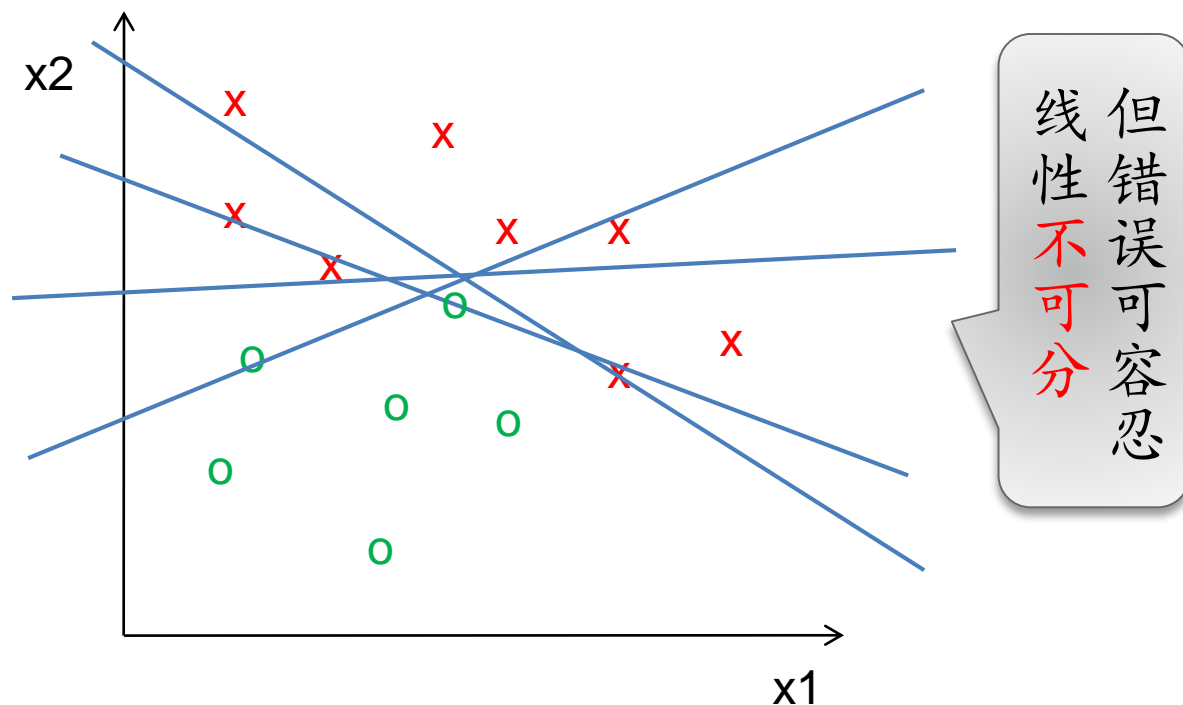
分类决策函数: $f(x) = \text{sgn}(w \cdot x + b)$

一、非线性分类模型：分类模型

- **Classification :**

Find a *function* f to separate the classes:

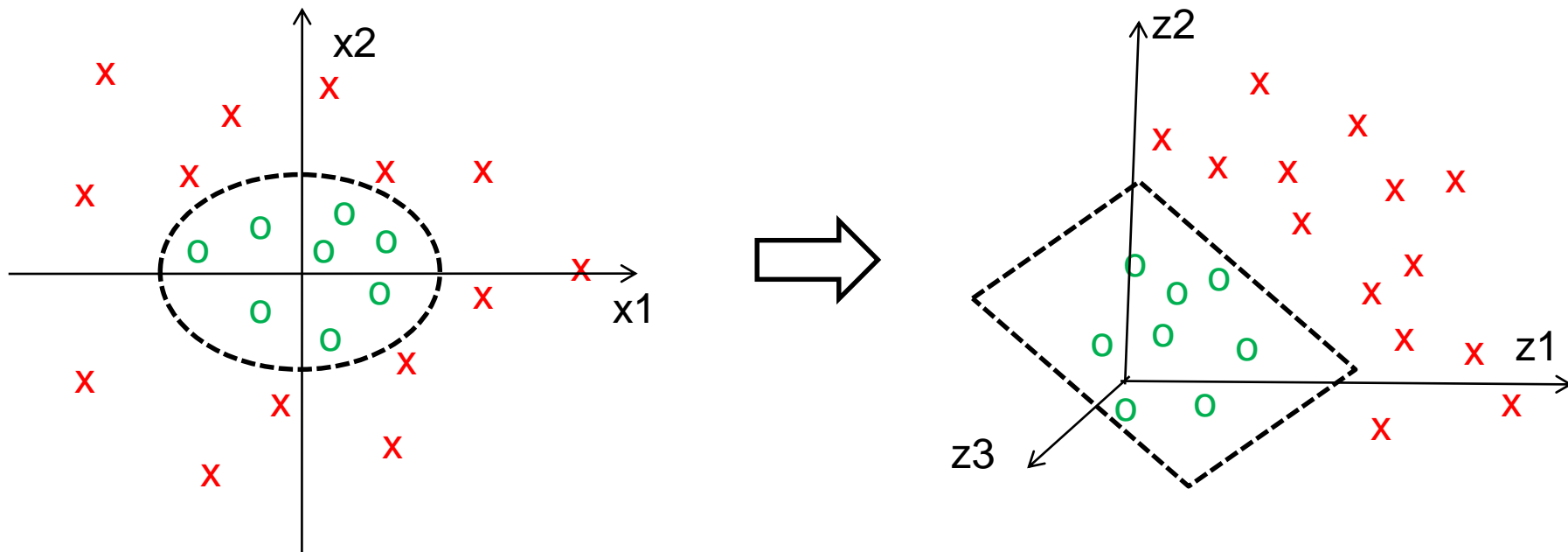
f ? nonlinear classifier



一、非线性分类模型：非线性分类模型

- 非线性分类器

将样本从原空间通过非线性变换映射到新的特征空间，从而在新的特征空间中线性可分（以核方法为代表）；例：kernel SVM



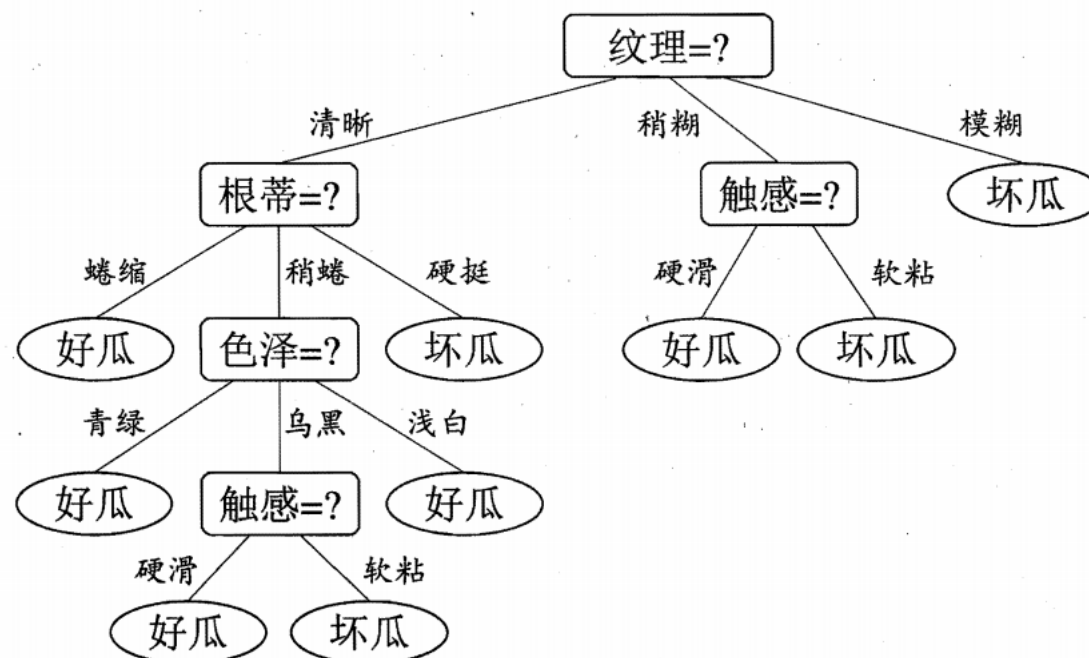
一、非线性分类模型：非线性分类模型

• 非线性分类器

模型本身具有非线性分类能力；例：决策树、MLP、KNN、集成方法等

编号	色泽	根蒂	敲声	纹理	脐部	触感	好瓜
1	青绿	蜷缩	浊响	清晰	凹陷	硬滑	是
2	乌黑	蜷缩	沉闷	清晰	凹陷	硬滑	是
3	乌黑	蜷缩	浊响	清晰	凹陷	硬滑	是
4	青绿	蜷缩	沉闷	清晰	凹陷	硬滑	是
5	浅白	蜷缩	浊响	清晰	凹陷	硬滑	是
6	青绿	稍蜷	浊响	清晰	稍凹	软粘	是
7	乌黑	稍蜷	浊响	稍糊	稍凹	软粘	是
8	乌黑	稍蜷	浊响	清晰	稍凹	硬滑	是
9	乌黑	稍蜷	沉闷	稍糊	稍凹	硬滑	否
10	青绿	硬挺	清脆	清晰	平坦	软粘	否
11	浅白	硬挺	清脆	模糊	平坦	硬滑	否
12	浅白	蜷缩	浊响	模糊	平坦	软粘	否
13	青绿	稍蜷	浊响	稍糊	凹陷	硬滑	否
14	浅白	稍蜷	沉闷	稍糊	凹陷	硬滑	否
15	乌黑	稍蜷	浊响	清晰	稍凹	软粘	否
16	浅白	蜷缩	浊响	模糊	平坦	硬滑	否
17	青绿	蜷缩	沉闷	稍糊	稍凹	硬滑	否

Decision tree

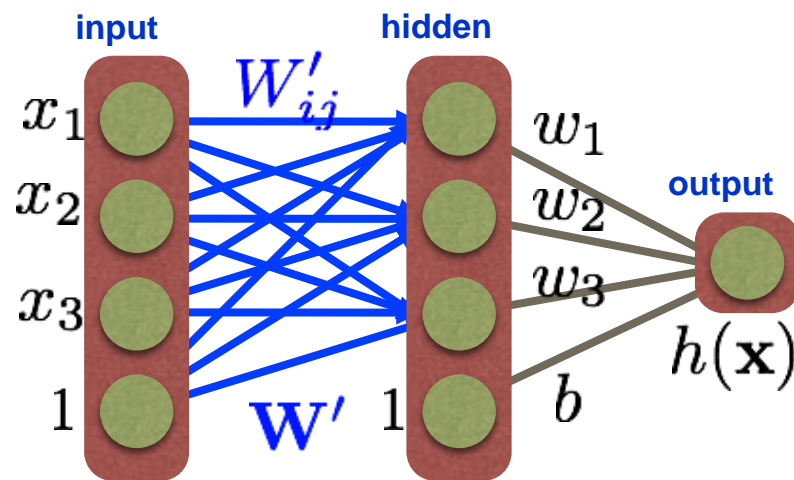
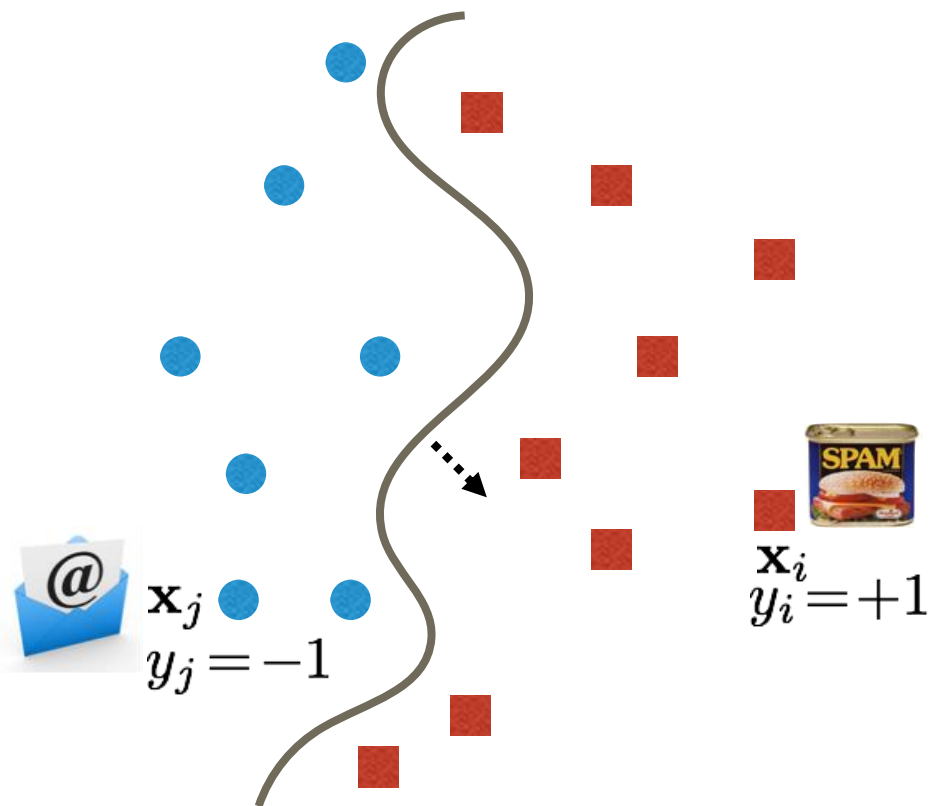


一、非线性分类模型：非线性分类模型

- 非线性分类器

模型本身具有非线性分类能力；例：决策树、MLP、KNN、集成方法等

MLP



$$\sigma(a) = \max(a, 0)$$

Rectified Linear Units (ReLU)

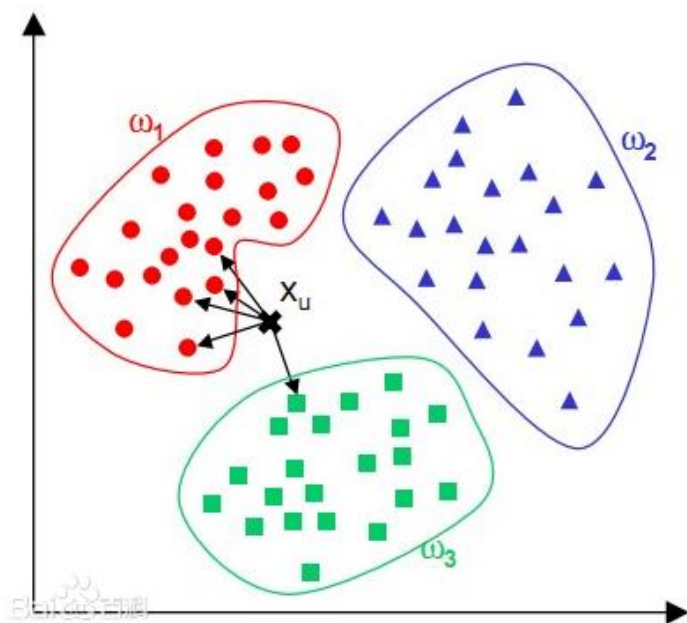
$$h(\mathbf{x}) = \mathbf{w}^\top \sigma(\mathbf{W}'\mathbf{x} + \mathbf{c}) + b$$

$$h(\mathbf{x}) = \mathbf{w}^\top \max(\mathbf{W}'\mathbf{x} + \mathbf{c}, 0) + b$$

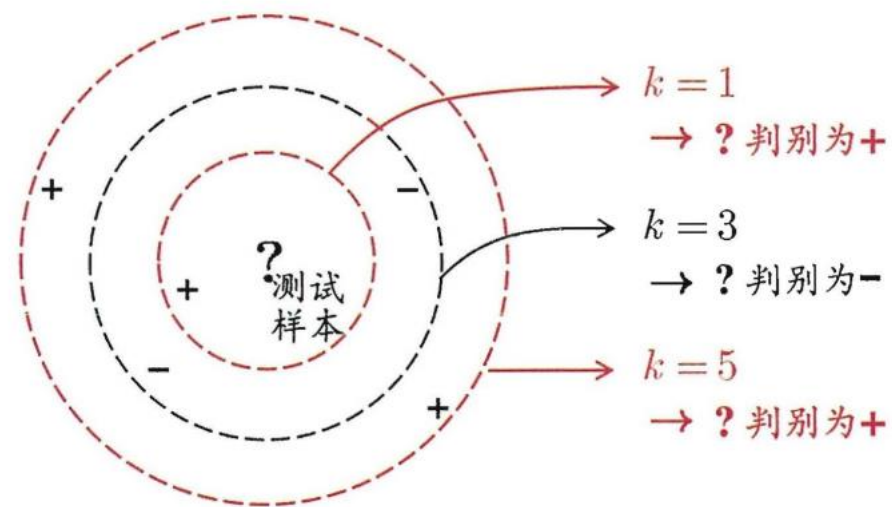
一、非线性分类模型：非线性分类模型

• 非线性分类器

模型本身具有非线性分类能力；例：决策树、MLP、KNN、集成方法等



KNN



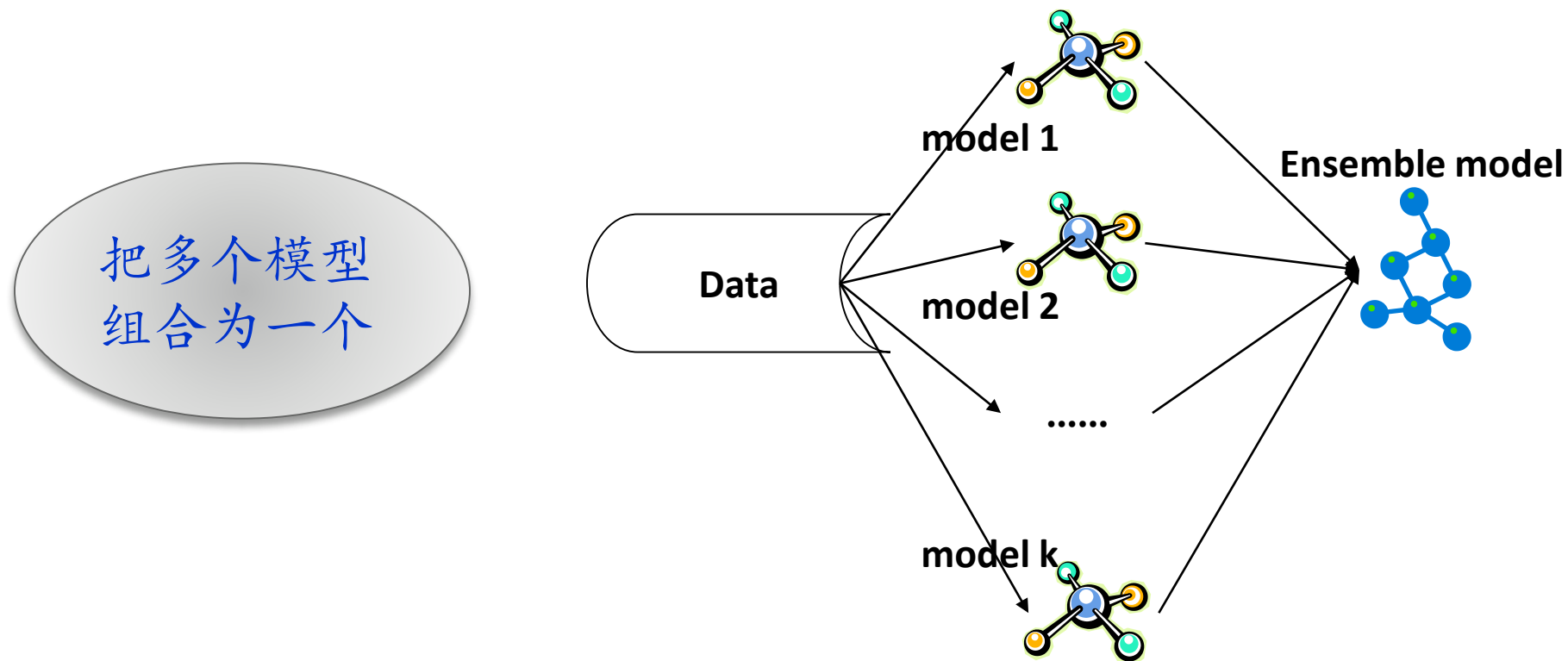
k近邻分类器示意图，虚线显示出等距线；测试样本在 $k=1$ ， $k=5$ 时被判别为正例， $k=3$ 时被判别为反例。

一、非线性分类模型：非线性分类模型

- 非线性分类器

模型本身具有非线性分类能力；例：决策树、MLP、KNN、集成方法等

Ensemble model

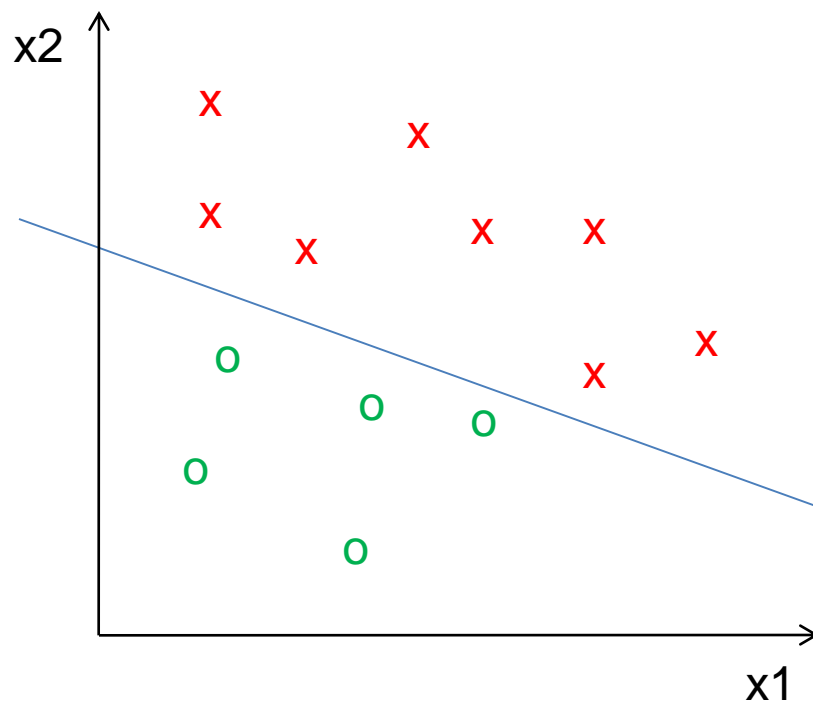


本章的主要内容

- 非线性关系
- 线性可分支持向量机
- 软间隔与正则化
- 核函数与非线性SVM
- 序列最小最优化算法
- 总结

二、线性可分支持向量机：引子

Q: 对于一个在二维特征空间中的二分类问题，你会如何将不同类别的数据分开呢？



× 表示 “正例”

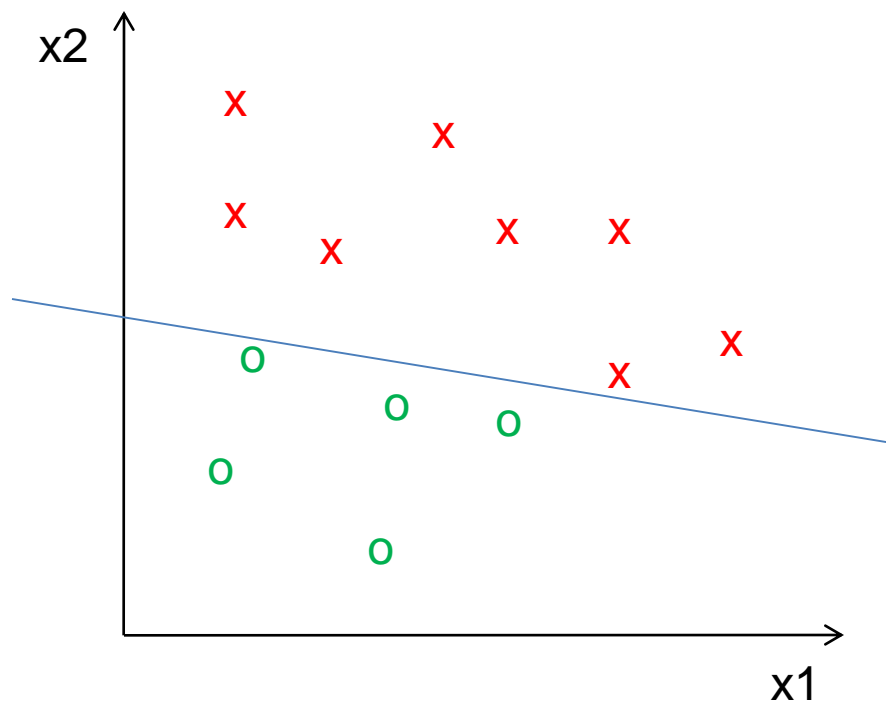
○ 表示 “负例”

$$f(x) = \text{sgn}(w \cdot x + b)$$

超平面(w, b)

二、线性可分支持向量机：引子

Q: 对于一个在二维特征空间中的二分类问题，你会如何将不同类别的数据分开呢？



× 表示 “正例”

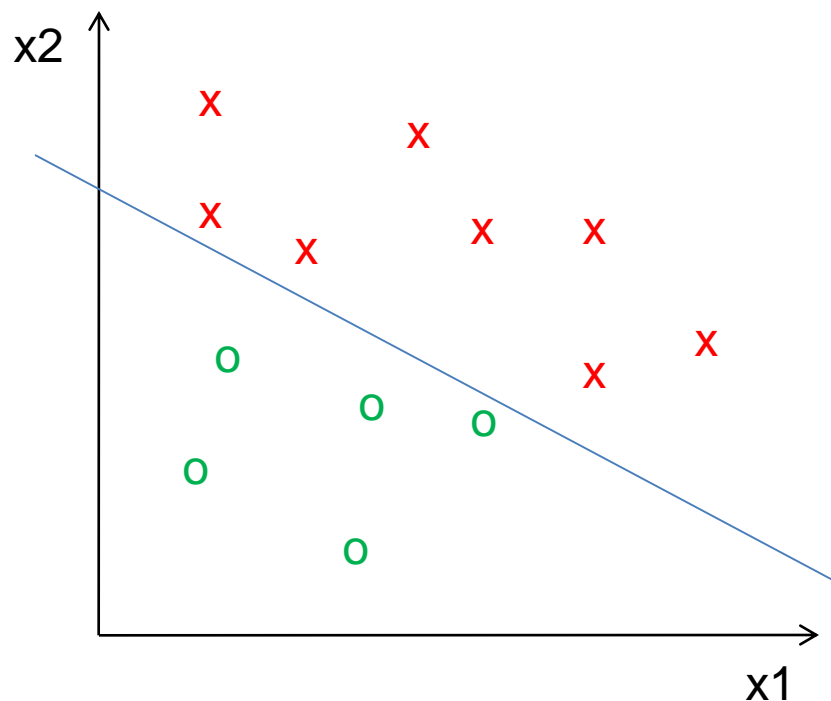
○ 表示 “负例”

$$f(x) = \text{sgn}(w \cdot x + b)$$

超平面(w, b)

二、线性可分支持向量机：引子

Q: 对于一个在二维特征空间中的二分类问题，你会如何将不同类别的数据分开呢？



× 表示 “正例”

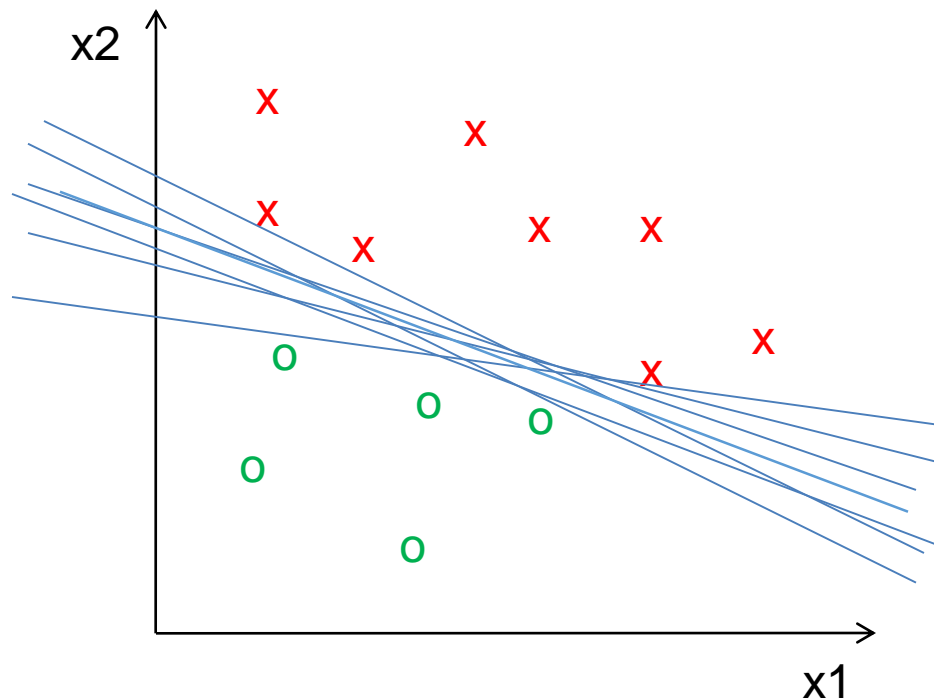
○ 表示 “负例”

$$f(x) = \text{sgn}(w \cdot x + b)$$

超平面(w, b)

二、线性可分支持向量机：引子

Q: 对于一个在二维特征空间中的二分类问题，你会如何将不同类别的数据分开呢？



× 表示 “正例”

○ 表示 “负例”

$$f(x) = \text{sgn}(w \cdot x + b)$$

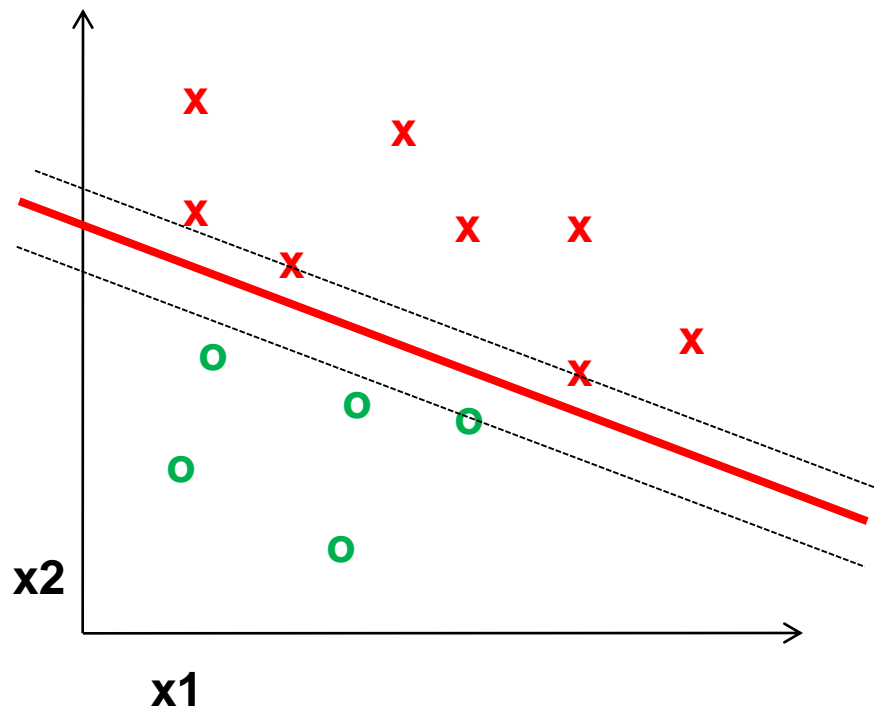
超平面(w, b)

二、线性可分支持向量机：引子

Q: 对于一个在二维特征空间中的二分类问题，你会如何将不同类别的数据分开呢？

Answer:

应选择“**正中间**”，对样本局部扰动的容忍性好，鲁棒性高，对未见实例的泛化能力最强。



× 表示 “正例”

○ 表示 “负例”

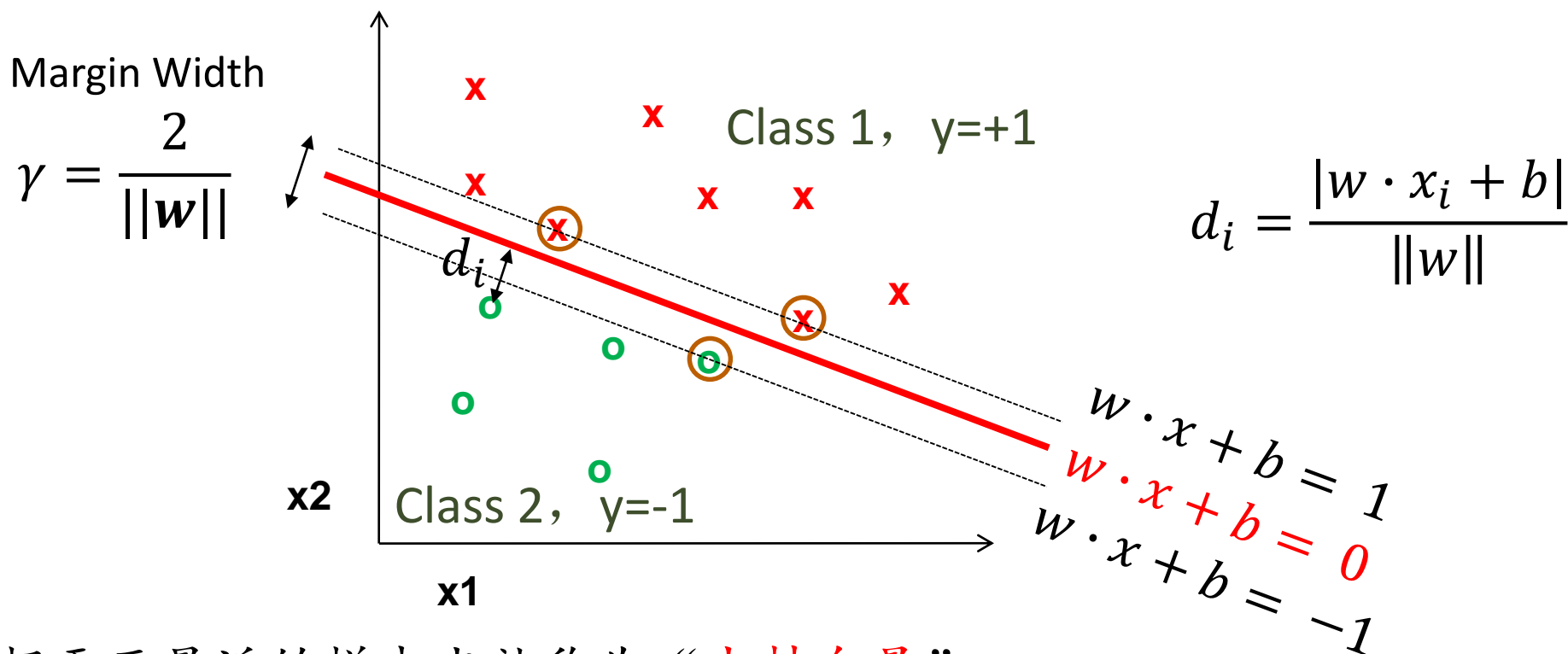
$$f(x) = \text{sgn}(w \cdot x + b)$$

超平面(w, b)

二、线性可分支持向量机：间隔与支持向量

- 最优超平面 (optimal hyperplane)

将两类数据正确划分且间隔最大的超平面

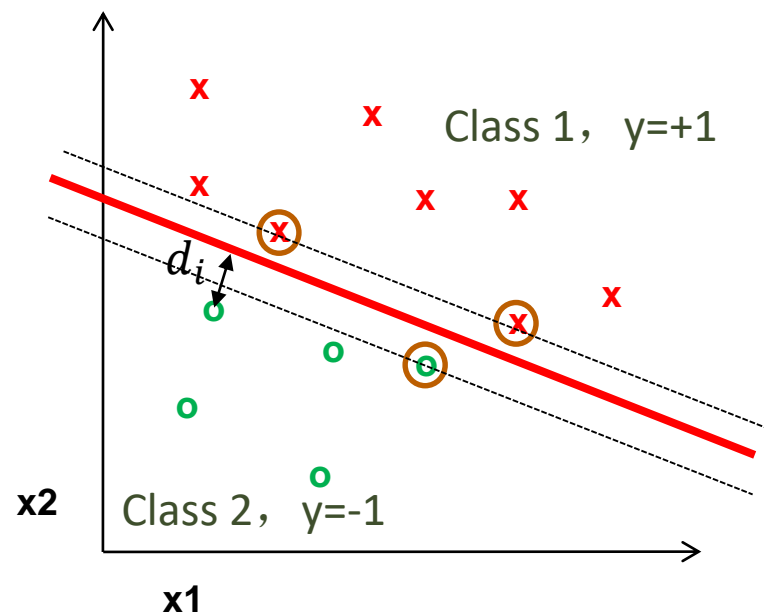


- 距离超平面最近的样本点被称为“支持向量”
- 两个异类支持向量到超平面的距离之和被称为“间隔”

二、线性可分支持向量机：间隔与支持向量

函数间隔 \hat{y} 与几何间隔 \tilde{y}

- 函数间隔 (functional margin) : $\hat{y} = y(w \cdot x + b) = yf(x)$
 $\hat{y} > 0$, 样本分类正确; $\hat{y} < 0$, 样本分类错误
- 找寻超平面 (w, b) 有函数间隔 $\hat{y} = \min \hat{y}_i, \quad (i = 1, 2, \dots, m)$
- 但函数间隔如果成比例的改变 w 和 b (如将它们改成 $2w$ 和 $2b$) , 则函数间隔的值 $f(x)$ 却变成了原来的2倍 (虽然此时超平面没有改变) , 所以只有函数间隔还远远不够。
- 需要真正定义点到超平面的距离——几何间隔 (geometrical margin) : $\tilde{y} = \frac{|w \cdot x_i + b|}{\|w\|} = \frac{|f(x)|}{\|w\|}$
- 于是最大间隔分类器 (maximum margin classifier) 的目标函数可以定义为: $\max \tilde{y}$



二、线性可分支持向量机：间隔与支持向量

- 最大间隔分类器（maximum margin classifier）的目标函数

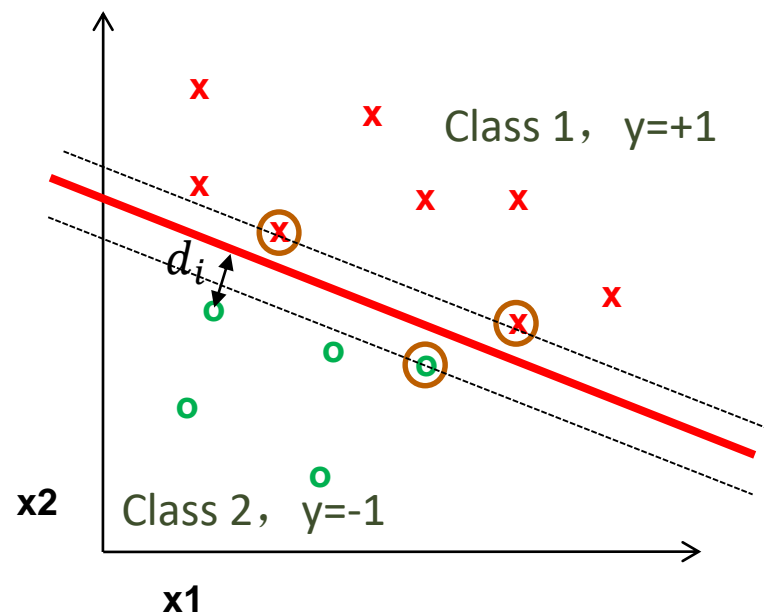
可以定义为：
$$\max \tilde{\gamma} = \max \frac{|f(x)|}{\|w\|} = \max \frac{y(w \cdot x + b)}{\|w\|}$$

- 这里的约束为对所有的样本点，均有：

$$y(w \cdot x_i + b) \geq \hat{\gamma}$$

- 下面做一个变量替换，用 $w' = w/\hat{\gamma}$ ，和 $b' = b/\hat{\gamma}$ 代替上面的 w 和 b ，这样的新变量仍旧是 w 和 b 的函数，所以最大化仍然可以进行。于是，把这两个新的变量代入到原来的约束最大化问题中，就变成了 $y(w \cdot x_i + b) \geq 1$
- 故SVM目标函数变为：

$$\max \frac{1}{\|w\|}, \quad \text{s.t.}, y(w \cdot x_i + b) \geq 1, i = 1, 2, \dots, n$$



二、线性可分支持向量机：线性可分支持向量机

- 支持向量机的基本型

最大化间隔：找到能满足约束的参数 \mathbf{w} 和 b 使 γ 最大，即

$$\begin{aligned} \max_{\mathbf{w}, b} \quad & \frac{2}{\|\mathbf{w}\|} \\ \text{s.t.} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i = 1, 2, \dots, m. \end{aligned}$$

↓ 等价于

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i = 1, 2, \dots, m. \end{aligned}$$

这就是支持向量机（support vector machine，简称SVM）的基本型

二、线性可分支持向量机：线性可分支持向量机

- 上述线性可分支持向量机学习的最优化问题是一个凸二次规划（convex quadratic programming）问题，我们可以直接利用凸二次规划方法求出该约束优化问题的解，从而得到最大间隔分类超平面以及分类决策函数。
- 但是一般我们选择应用拉格朗日对偶性，将原问题转化为对偶问题进行求解。
- **Why?**
 - ✓ 对偶问题的求解更为高效；
 - ✓ 可以更自然地引入核函数，进而推广到非线性分类问题。

二、线性可分支持向量机：线性可分支持向量机

求解以下优化问题：

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) = x_1^2 + x_2^2 \\ \text{s. t.} \quad & h(\mathbf{x}) = x_1 - x_2 - 2 = 0 \\ & g(\mathbf{x}) = (x_1 - 2)^2 + x_2^2 - 1 \leq 0 \end{aligned} \tag{19}$$

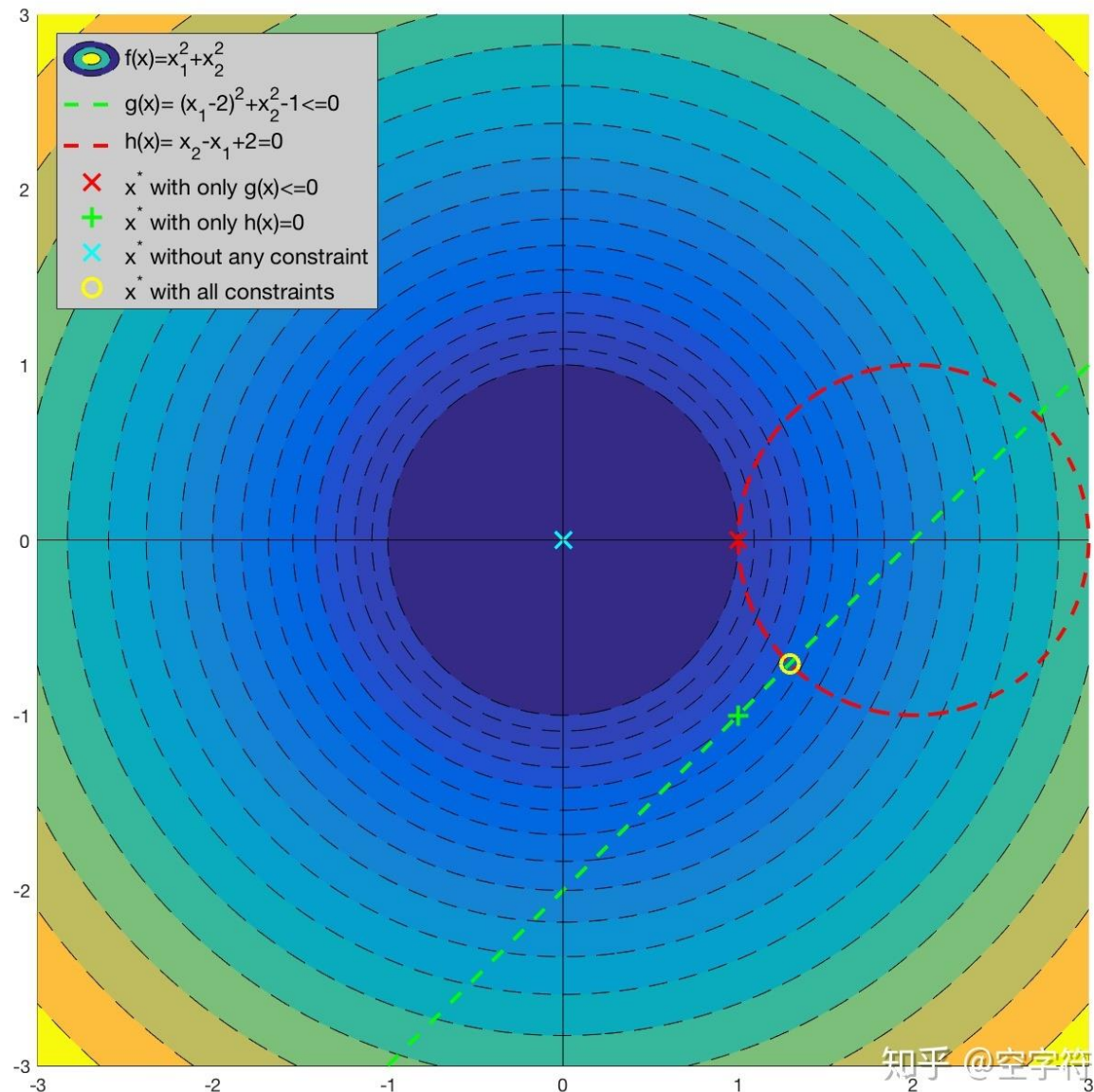
由于优化问题 (19) 相对简单，我们可以先通过作图来直观感受以下：

二、线性可分支持向量机：线性可分支持向量机

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) = x_1^2 + x_2^2 \\ \text{s. t.} \quad & h(\mathbf{x}) = x_1 - x_2 - 2 = 0 \\ & g(\mathbf{x}) = (x_1 - 2)^2 + x_2^2 - 1 \leq 0 \end{aligned}$$

之所以说这个案例比较典型是因为它与线性SVM的数学模型非常相似，且包含了等式和不等式两种不同的约束条件。更重要的是，这两个约束条件在优化问题中都起到了作用。如图所示（左上角所示）：

- ①如果没有任何约束条件，最优解在坐标原点(0, 0)处；
- ②如果只有不等式约束条件 $g(\mathbf{x}) \leq 0$ ，最优解在坐标(1,0)处（红色X）；
- ③如果只有等式约束条件 $h(\mathbf{x}) = 0$ ，最优解在坐标(1,-1)处（绿色+）；
- ④如果两个约束条件都有，最优解在 $(2 - \sqrt{2}/2, -\sqrt{2}/2)$ 处(黄色O)。



二、线性可分支持向量机：线性可分支持向量机

针对这一问题，我们可以设计拉格朗日函数如下：

$$L(\mathbf{x}, \alpha, \beta) = (x_1^2 + x_2^2) + \alpha [(x_1 - 2)^2 + x_2^2 - 1] + \beta(x_1 - x_2 - 2) \quad (20)$$

根据式子(5)可知：

$$\theta_{\mathcal{P}}(x) = \max_{\alpha, \beta: \alpha \geq 0} \mathcal{L}(x, \alpha, \beta) \quad (21)$$

此时，我们依然可以得到，如果 x 不满足上面的两个约束条件，即：

- ①若 $g(x) > 0$ ；则可以任取 α 使得 $\theta_{\mathcal{P}}(x)$ 趋于无穷；
- ②若 $h(x) \neq 0$ ；则只有任取 β ，且 $\beta, h(x)$ 同号，那么 $\theta_{\mathcal{P}}(x)$ 依旧可能趋于无穷；
- ③而只有两个约束条件同时满足， $\theta_{\mathcal{P}}(x)$ 才可能取得极值；

于是同样有：

$$\theta_{\mathcal{P}}(x) = \begin{cases} f(x), & \text{if } x \text{ satisfies primal constraints} \\ \infty, & \text{otherwise} \end{cases} \quad (22)$$

二、线性可分支持向量机：线性可分支持向量机

故，其原始问题为：

$$p^* = \min_x \max_{\alpha, \beta: \alpha > 0} \mathcal{L}(x, \alpha, \beta) \quad (23)$$

接着，那么其对偶问题就应该为：

$$d^* = \max_{\alpha, \beta: \alpha > 0} \min_x \mathcal{L}(x, \alpha, \beta) \quad (24)$$

对于求解对偶问题，一般分为两步：

- 第一步，最小化 $\mathcal{L}(x, \alpha, \beta)$ ；我们将 α, β 视为常数，这时 $L(\mathbf{x}, \alpha, \beta)$ 就只是 \mathbf{x} 的函数。我们可以通过求导等于零的方式寻找其最小值，即 $\theta_{\mathcal{D}}(\alpha, \beta) = \min_{\mathbf{x}} [L(\mathbf{x}, \alpha, \beta)]$

$$\begin{cases} \beta + 2x_1 + \alpha(2x_1 - 4) = 0 \\ 2x_2 - \beta + 2\alpha x_2 = 0 \end{cases} \quad (25)$$

可以解得：

$$\begin{cases} x_1 = \frac{4\alpha - \beta}{2\alpha + 2} \\ x_2 = \frac{\beta}{2\alpha + 2} \end{cases} \quad (26)$$

将 (25) 代入拉格朗日目标函数 (20) 可以得到：

$$\theta_{\mathcal{D}}(\alpha, \beta) = -\frac{\beta^2 + 4\beta + 2\alpha^2 - 6\alpha}{2(\alpha + 1)} \quad (27)$$

二、线性可分支持向量机：线性可分支持向量机

- 第二步，最大化 $\theta_D(\alpha, \beta)$ 此时可以将 $\theta_D(\alpha, \beta)$ 看成是一个二元函数求极值(无条件)的问题，且 $\alpha > 0$ 。用拉格朗日乘数法即可求解。

设 $D = \theta_D(\alpha, \beta)$ ，则 D 分别对 α, β 求偏导并令其为0有：

$$\frac{\partial D}{\partial \alpha} = -\frac{2\alpha^2 + 4\alpha - \beta^2 - 4\beta - 6}{2(\alpha + 1)^2} = 0;$$

$$\frac{\partial D}{\partial \beta} = \frac{2\beta + 4}{2(\alpha + 1)} = 0$$

联立求得：

$$\alpha = \sqrt{2} - 1 (> 0), \beta = -2 \quad (28)$$

再将 (28) 代入 (26) 即可求得 x 。

二、线性可分支持向量机：对偶问题

• 对偶问题

Step1: 引入拉格朗日乘子 $\alpha_i \geq 0$, 构造拉格朗日函数:

$$\begin{aligned} L(\mathbf{w}, b, \alpha) &= \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^m \alpha_i (1 - y_i (\mathbf{w}^T \mathbf{x}_i + b)) \\ &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + \sum_{i=1}^m \alpha_i - \mathbf{w}^T \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i - \left(\sum_{i=1}^m \alpha_i y_i \right) b \end{aligned}$$

则原始问题的对偶问题为: $\max_{\alpha} \min_{\mathbf{w}, b} L(\mathbf{w}, b, \alpha)$

此为关于 \mathbf{w}, b 的无约束优化问题

Step2: 令 $L(\mathbf{w}, b, \alpha)$ 对 \mathbf{w} 和 b 的偏导分别为0可得:

$$\mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i, \quad 0 = \sum_{i=1}^m \alpha_i y_i$$

二、线性可分支持向量机：对偶问题

Step3: 回代可得:

$$\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

$$s.t. \sum_{i=1}^m \alpha_i y_i = 0,$$

$$\alpha_i \geq 0, i = 1, 2 \dots m.$$

上式即为线性可分支持向量机的对偶问题

二、线性可分支持向量机：对偶问题

$$\begin{aligned} L(w, b, \alpha) &= \frac{1}{2} \|w\|^2 + \sum_{i=1}^m \alpha_i (1 - y_i (w^T x_i + b)) \\ &= \frac{1}{2} w^T w + \sum_{i=1}^m \alpha_i - w^T \sum_{i=1}^m \alpha_i y_i x_i - \left(\sum_{i=1}^m \alpha_i y_i \right) b \\ &= \frac{1}{2} w^T \sum_{i=1}^m \alpha_i y_i x_i + \sum_{i=1}^m \alpha_i - w^T \sum_{i=1}^m \alpha_i y_i x_i - \left(\sum_{i=1}^m \alpha_i y_i \right) b \\ &= -\frac{1}{2} w^T \sum_{i=1}^m \alpha_i y_i x_i + \sum_{i=1}^m \alpha_i \\ &= -\frac{1}{2} \left(\sum_{i=1}^m \alpha_i y_i x_i \right)^T \sum_{i=1}^m \alpha_i y_i x_i + \sum_{i=1}^m \alpha_i \\ &= \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j \end{aligned}$$

$$w = \sum_{i=1}^m \alpha_i y_i x_i, \quad 0 = \sum_{i=1}^m \alpha_i y_i$$

二、线性可分支持向量机：对偶问题

- 解出 α^* 后，我们就可以将 \mathbf{w}^* 视为训练样本的一个线性组合：

$$\mathbf{w}^* = \sum_{i=1}^m \alpha_i^* y_i \mathbf{x}_i,$$
$$b^* = y_j - \sum_{i=1}^m \alpha_i^* y_i (\mathbf{x}_i \cdot \mathbf{x}_j)$$

- 从而可得最优超平面：

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

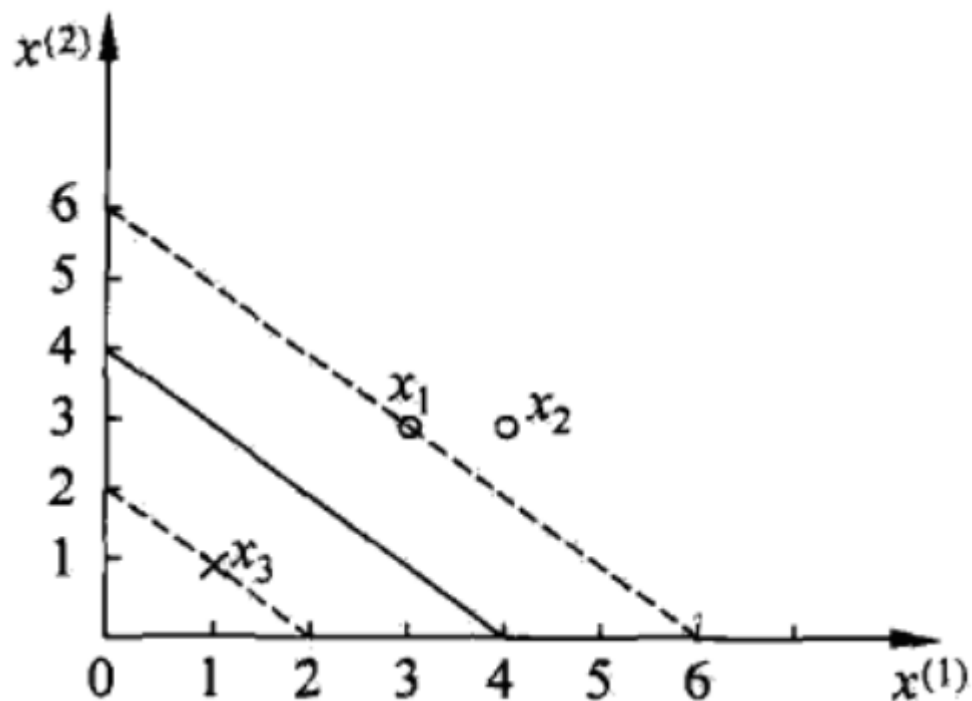
- 最终模型：

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

二、线性可分支持向量机：对偶问题

- **Example**

已知一个如图所示的训练数据集，其正例点为 $x_1 = (3,3)^T$, $x_2 = (4,3)^T$, 负例点是 $x_3 = (1,1)^T$, 试通过求解对偶问题学习线性可分支持向量机



二、线性可分支持向量机：对偶问题

解：

根据所给数据，其对应的对偶问题是

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ & = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2} (18\alpha_1^2 + 25\alpha_2^2 + 2\alpha_3^2 + 42\alpha_1\alpha_2 - 12\alpha_1\alpha_3 - 14\alpha_2\alpha_3) \end{aligned}$$

$$s.t. \quad \alpha_1 + \alpha_2 - \alpha_3 = 0,$$

$$\alpha_i \geq 0, i = 1, 2 \dots m.$$

二、线性可分支持向量机：对偶问题

将 $\alpha_3 = \alpha_1 + \alpha_2$ 带入目标函数并记为

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

对 α_1, α_2 分别求偏导并令其为0，易得 $s(\alpha_1, \alpha_2)$ 在点 $\left(\frac{3}{2}, -1\right)^T$ 处取极值，但该点不满足约束条件 $\alpha_2 \geq 0$ ，因此最大值应在边界上达到。

当 $\alpha_1 = 0$ 时，最大值 $s(0, \frac{2}{13}) = \frac{2}{13}$ ，当 $\alpha_2 = 0$ 时，最大值 $s(\frac{1}{4}, 0) = \frac{1}{4}$ ，因此 $s(\alpha_1, \alpha_2)$ 在 $\alpha_1 = \frac{1}{4}, \alpha_2 = 0$ 处达到最大值，此时 $\alpha_3 = \alpha_1 + \alpha_2 = \frac{1}{4}$ 。

二、线性可分支持向量机：对偶问题

这样， $\alpha_1^* = \alpha_3^* = \frac{1}{4}$ 对应的实例点 x_1, x_3 是支持向量，计算可得

$$w_1^* = w_2^* = \frac{1}{2}, \quad b^* = -2$$

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

分隔超平面为：

$$\frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)} - 2 = 0$$

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

分类决策函数为：

$$f(x) = \text{sgn}\left(\frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)} - 2\right)$$

二、线性可分支持向量机：解的稀疏性

- 需要注意的是，从对偶问题解出的 α_i 是拉格朗日乘子，对应着训练样本 (x_i, y_i) ，于该对偶问题中存在不等式约束，因此上述过程需满足KKT (Karush-Kuhn-Tucker) 条件，即：

$$\begin{cases} \alpha_i \geq 0 \\ y_i f(x_i) - 1 \geq 0 \\ \alpha_i (y_i f(x_i) - 1) = 0 \end{cases} \quad \longrightarrow \quad \begin{aligned} &\text{必有 } \alpha_i = 0 \\ &\text{或 } y_i f(x_i) = 1 \end{aligned}$$

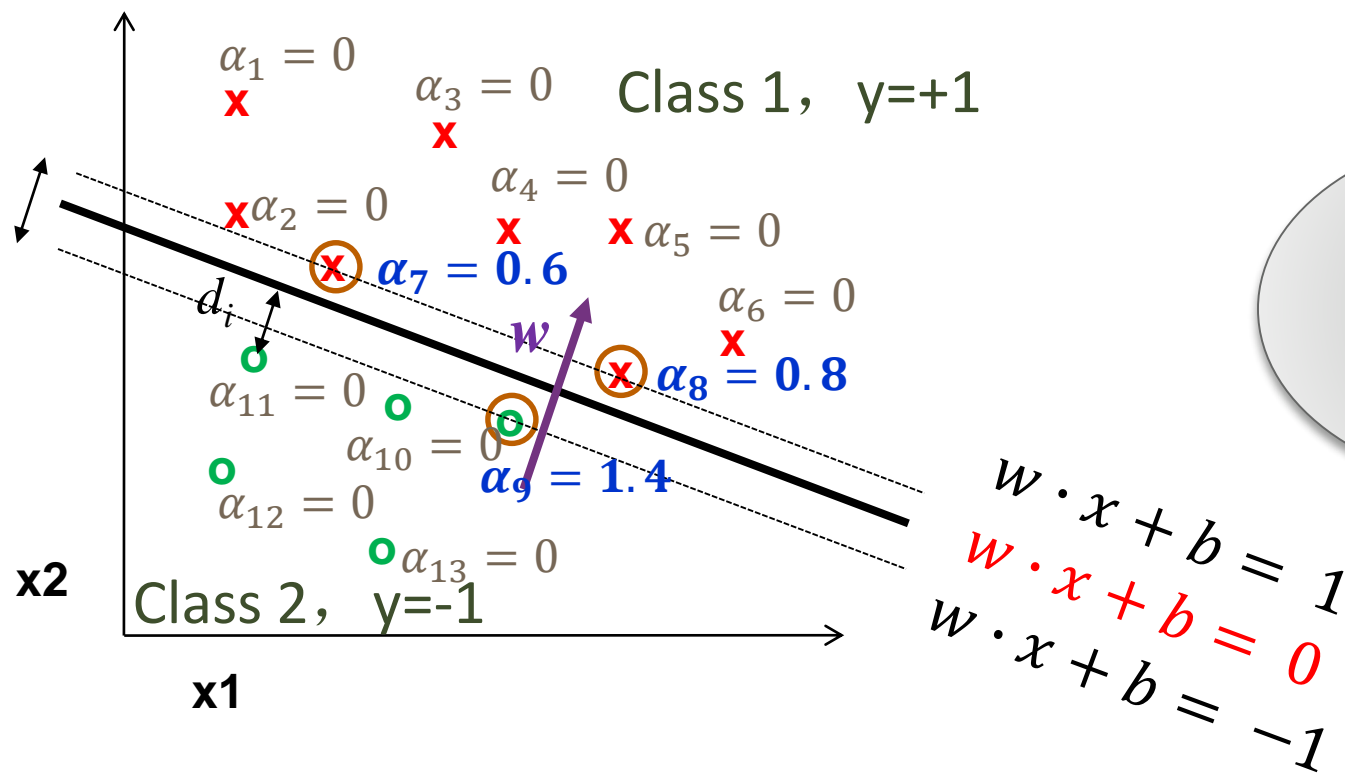
只有很少一部分 α_i 非零！

- 若 $y_i f(x_i) > 1$ ，则有 $\alpha_i = 0$ 。

二、线性可分支持向量机：解的稀疏性

支持向量机解的稀疏性

- 模型训练完成后，大部分的训练样本都无需保留，最终模型仅与
支持向量有关。



$\alpha_i \neq 0$ 所对应的向量为支持向量！

二、线性可分支持向量机：小结

线性可分支持向量机小结

- 原问题：

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i = 1, 2, \dots, m. \end{aligned}$$

- 对偶问题：

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

二、线性可分支持向量机：小结

- 最优超平面：

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

其中，

$$\mathbf{w}^* = \sum_{i=1}^m \alpha_i^* y_i \mathbf{x}_i,$$

$$b^* = y_j - \sum_{i=1}^m \alpha_i^* y_i (\mathbf{x}_i \cdot \mathbf{x}_j)$$

- 分类决策函数：

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^m \alpha_i^* y_i \mathbf{x}_i^T \mathbf{x} + b\right)$$

本章的主要内容

- 非线性关系
- 线性可分支持向量机
- 软间隔与正则化
- 核函数与非线性SVM
- 序列最小最优化算法
- 总结

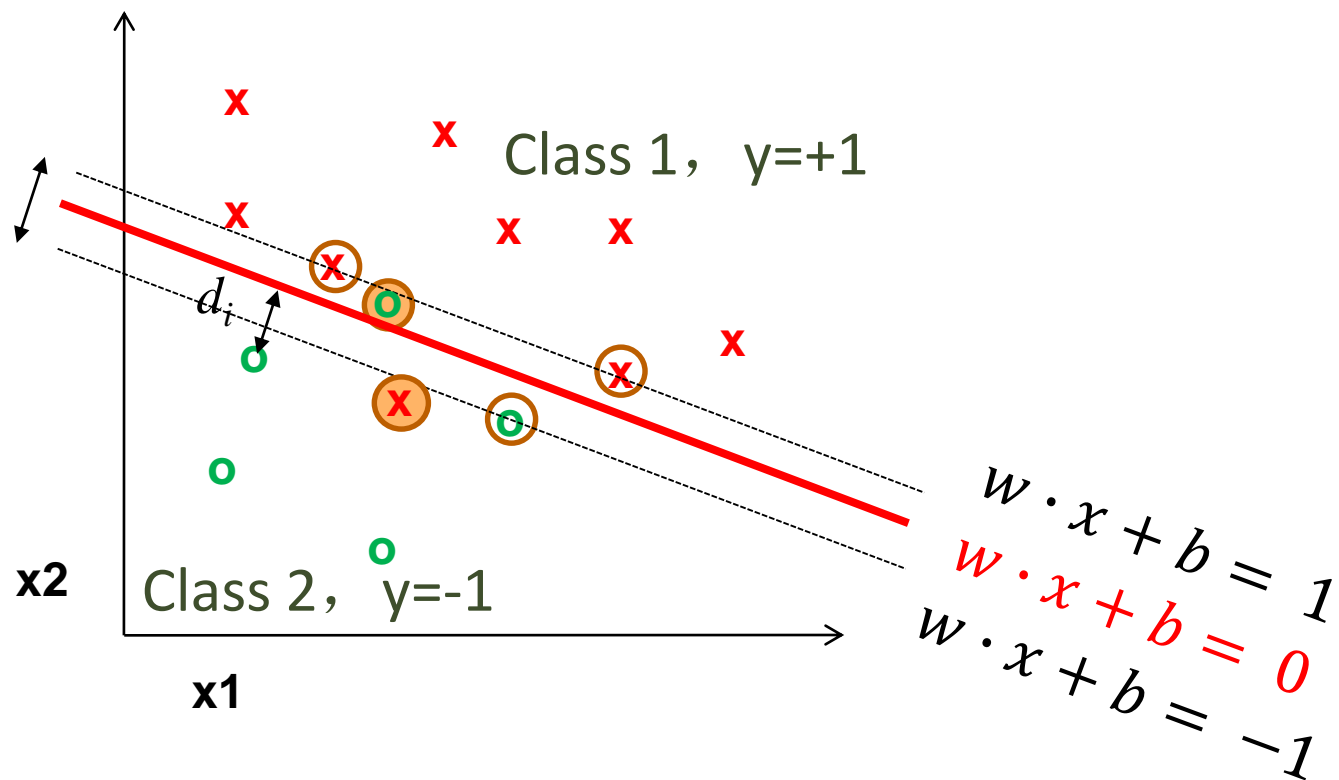
三、软间隔与正则化：软间隔

软间隔

- 在本章前面的讨论中，我们假设训练样本是线性可分的，即存在一个超平面能将不同类的训练样本完全划分开。但是在实际任务中训练样本往往会出现噪声或特异点，此时上述方法中 $y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1$ 的约束条件并不能对所有样本同时成立。
- 因此需要引入“软间隔” (soft margin) 概念。
- 具体来说，前面介绍的线性可分支持向量机要求所有样本均满足约束 $y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1$ ，即所有样本都必须划分正确，这称为“硬间隔”，而“软间隔”则允许支持向量机在一些“困难”或者“噪声”样本上不满足约束。

三、软间隔与正则化：软间隔

如下图所示，其中  表示不满足约束（分类错误）的样本



- 此时支持向量包括两部分：
正确分类且离分类面最近的样本  + 错误分类的样本 

三、软间隔与正则化：软间隔

- 广义最优超平面

将两类数据正确划分且使软间隔最大的超平面即为广义最优超平面。

三、软间隔与正则化：软间隔支持向量机

软间隔支持向量机

- 具体来说，我们对每个样本点 (x_i, y_i) 引入一个松弛变量 $\xi_i \geq 0$ 使下述不等式约束成立：

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, m$$

- 同时，对每个松弛变量 ξ_i 支付一个代价 ξ_i ，目标函数变为：

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

- 其中 $C > 0$ 称为惩罚参数，用于控制对错误的惩罚。

三、软间隔与正则化：软间隔支持向量机

- 软间隔SVM的优化目标函数为：

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

$$s.t. \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, i = 1, 2, \dots, m.$$

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

$$s.t. \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, m$$

- 最小化该目标函数包含两层含义，一是使间隔最大化，二是使错误分类点的个数最小化。 C 可以看为平衡二者的系数。

三、软间隔与正则化：对偶问题

软间隔SVM的对偶问题

- 与线性可分情况下硬间隔SVM的对偶问题相似，可以引入拉格朗日乘子来求解该优化问题。
- 构造拉格朗日函数：

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

s. t. $\alpha_i \geq 0, \mu_i \geq 0$

- 其中 α_i, μ_i 是拉格朗日乘子。

三、软间隔与正则化：对偶问题

- 构造拉格朗日函数：

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

$$\nabla_{\mathbf{w}} L(\mathbf{w}, b, \alpha, \xi, \mu) = \mathbf{w} - \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i = 0 \implies \mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i$$

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

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

- 带入到原式有： $\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$

三、软间隔与正则化：对偶问题

- 令 $L(\mathbf{w}, b, \boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\mu})$ 对 \mathbf{w}, b, ξ_i 的偏导分别为0并带回可得软间隔

SVM的对偶问题：

$$\max_{\alpha} \quad \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

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

$$0 \leq \alpha_i \leq C, i = 1, 2 \dots m.$$

三、软间隔与正则化：解的稀疏性

- 软间隔下支持向量机KKT条件：


$$\left\{ \begin{array}{ll} \alpha_i \geq 0, \mu_i \geq 0 & \text{乘子非负} \\ y_i f(\mathbf{x}_i) - 1 + \xi_i \geq 0 & \text{约束条件} \\ \alpha_i (y_i f(\mathbf{x}_i) - 1 + \xi_i) = 0 & \text{互补条件} \\ \xi_i \geq 0, \mu_i \xi_i = 0 & \end{array} \right.$$

- 可推得最终模型仍仅与支持向量有关，也即引入松弛变量后依然保持了支持向量机解的稀疏性。

三、软间隔与正则化：支持向量机的正则化

支持向量机的正则化

支持向量机学习模型的更一般形式可以写为：

$$\min_f \Omega(f) + C \sum_{i=1}^m l(f(x_i), y_i)$$


结构风险（正则项），描述模型的某些性质

经验风险，描述模型与训练数据的拟合程度

- 正则化可以理解作为一种“罚函数法”，即对不希望得到的结果施以惩罚，从而使得优化过程趋向于希望目标。参数 C 用于实现模型复杂度与经验误差风险间的权衡。

三、软间隔与正则化：支持向量机的正则化

结构风险（正则项）

L_p 范数（norm）是常用的正则化项。

- 其中 L_2 范数 $\|w\|_2$ 较为常用，其倾向于 w 的分量取值尽量均衡，即非零分量个数尽量稠密；
- 在一些稀疏模型中也会采用 L_0 范数 $\|w\|_0$ 和 L_1 范数 $\|w\|_1$ ，倾向于 w 的分量尽量稀疏，即非零分量个数尽量少。

在本章前面的讨论中该项为 $\frac{1}{2} \|w\|^2$

三、软间隔与正则化：支持向量机的正则化

经验风险

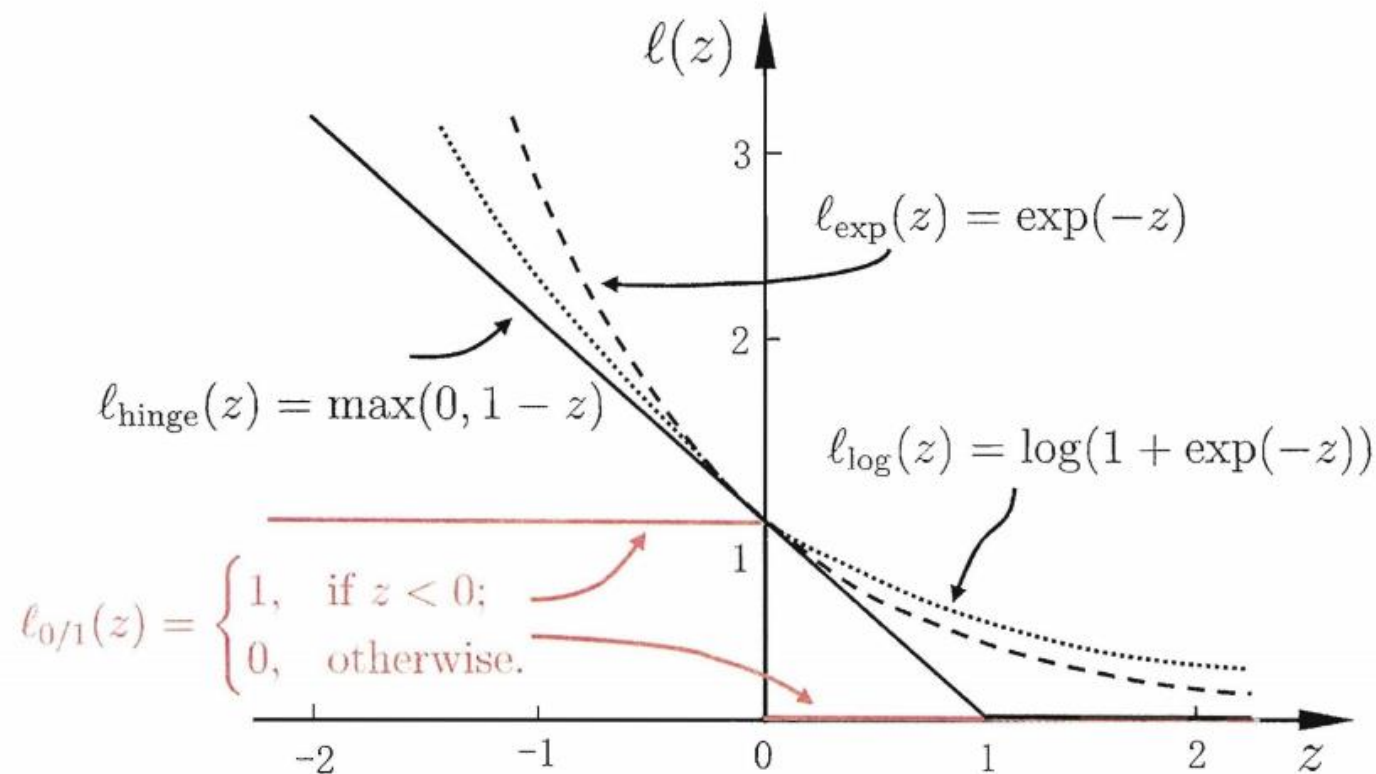
三种常见的替代损失函数：

- **Hinge损失**： $l_{hinge}(z) = \max(0, 1 - z)$
- **指数损失** (exponential loss)： $l_{exp}(z) = \exp(-z)$
- **对率损失** (logistic loss)： $l_{log}(z) = \log(1 + \exp(-z))$

若采用hinge损失，则SVM的目标式可写为：

$$\min_{\mathbf{w}, b, \zeta} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \max(0, 1 - y_i(\mathbf{w}^T \mathbf{x}_i + b))$$

三、软间隔与正则化：支持向量机的正则化



三种常见的替代损失函数: hinge损失、指数损失、对率损失

三、软间隔与正则化：支持向量机的正则化

$$\min_f \Omega(f) + C \sum_{i=1}^m l(f(x_i), y_i)$$



结构风险（正则项），描述模型的某些性质



经验风险，描述模型与训练数据的拟合合度

通过替换上面两个部分，可以得到许多其他学习模型：

- 对数几率回归 (Logistic Regression)
- 最小绝对收缩选择算子 (LASSO)
-

三、软间隔与正则化：小结

软间隔支持向量机小结

- 原问题：

$$\begin{aligned} \min_{\mathbf{w}, b, \zeta} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i \\ \text{s.t.} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, m \end{aligned}$$

- 对偶问题：

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{s.t.} \quad & \sum_{i=1}^m \alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C, i = 1, 2, \dots, m. \end{aligned}$$

三、软间隔与正则化：小结

- 最优超平面：

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

其中，

$$\mathbf{w}^* = \sum_{i=1}^m \alpha_i^* y_i \mathbf{x}_i,$$

$$b^* = y_j - \sum_{i=1}^m \alpha_i^* y_i (\mathbf{x}_i \cdot \mathbf{x}_j)$$

- 分类决策函数：

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^m \alpha_i^* y_i \mathbf{x}_i^T \mathbf{x} + b\right)$$

三、软间隔与正则化：小结

- 正则化：

$$\min_f \Omega(f) + C \sum_{i=1}^m l(f(x_i), y_i)$$



结构风险（正则项），描述模型的某些性质



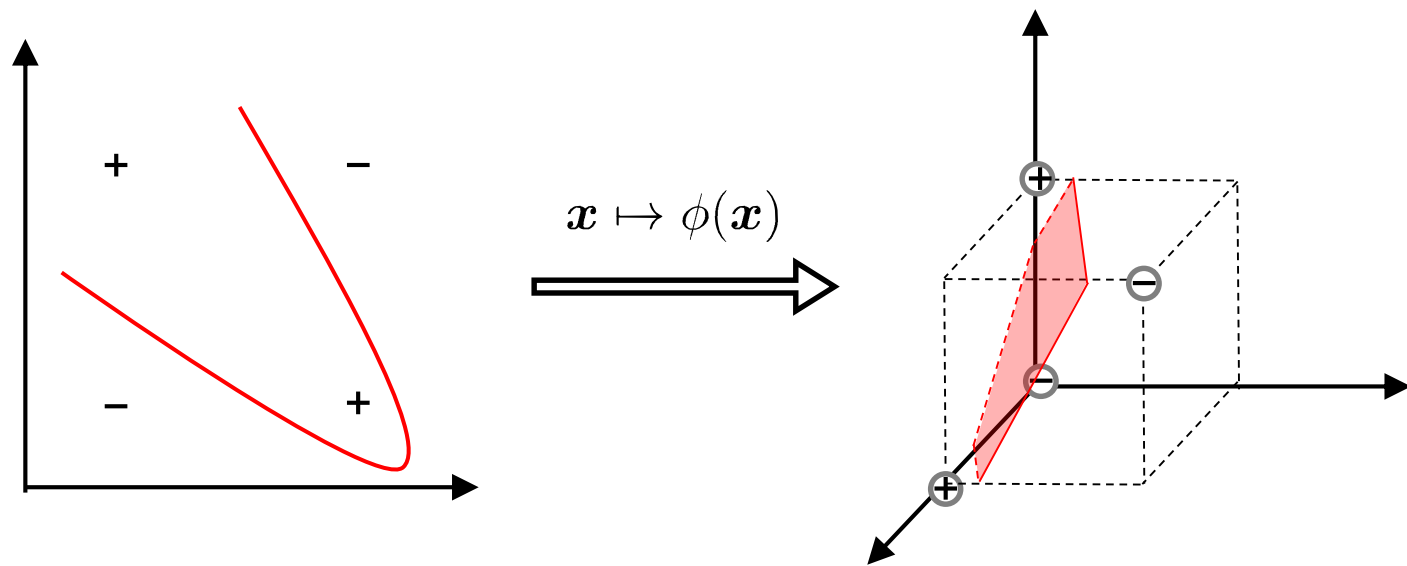
经验风险，描述模型与训练数据的拟合合度

本章的主要内容

- 非线性关系
- 线性可分支持向量机
- 软间隔与正则化
- 核函数与非线性SVM
- 序列最小最优化算法
- 总结

四、核函数与非线性SVM：非线性可分

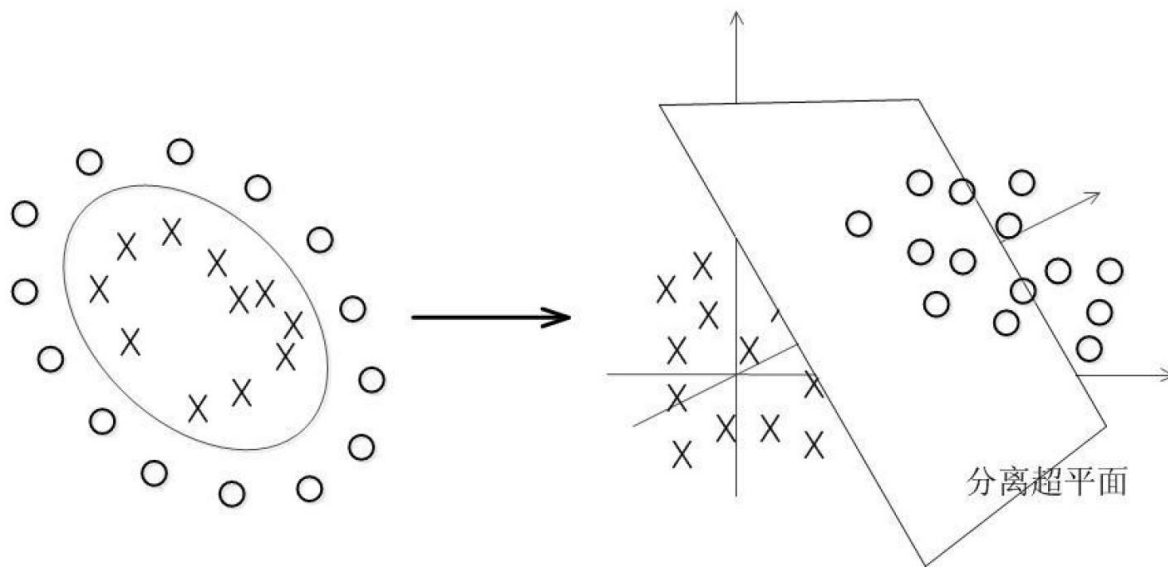
Q: 如果训练样本完全线性不可分怎么办？



- 最优超平面 **非线性化** 的思想：把样本从原空间通过 **非线性变换** 映射到新的特征空间；在特征空间中 **构造（广义）超平面**

四、核函数与非线性SVM：非线性可分

- 在**线性不可分**的情况下，支持向量机首先在**低维空间**中完成计算，然后通过**核函数**将输入空间**映射到高维特征空间**，最终在**高维特征空间**中构造出最优分离超平面，从而把平面上本身不好分的非线性数据分开。



四、核函数与非线性SVM：非线性可分

- 在我们遇到核函数之前，如果用原始的方法，那么在用线性学习器学习一个非线性关系，需要选择一个非线性特征集，并且将数据写成新的表达形式，这等价于应用一个固定的非线性映射，将数据映射到特征空间，在特征空间中使用线性学习器。
- 考虑函数： $f(x) = \sum_{i=1}^m w_i \phi_i(x) + b$ ，其中 $\phi(\cdot)$ 是从输入空间到某个特征空间的映射，这意味着建立非线性学习器分为两步：
 - ① 首先使用一个非线性映射将数据变换到一个特征空间；
 - ② 然后在高维特征空间使用线性学习器分类。

四、核函数与非线性SVM：核技巧

核技巧 (kernel trick)

- 思考：如何应对映射到高维特征空间后计算量激增问题？ **核技巧**
- 回顾最优超平面（对偶形式）

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

只涉及样本之间的**内积运算**！

四、核函数与非线性SVM：核技巧

$$x \rightarrow \phi(x) \quad \Longrightarrow \quad x_i^T x_j \rightarrow \phi(x_i)^T \phi(x_j)$$

- 实际上，特征空间中的内积运算仍然对应着原空间中某种运算，因此，无需显式地进行映射，我们可以定义合适的核函数来在原空间中计算特征空间的内积，即：

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

四、核函数与非线性SVM：核技巧

Example

- 给定输入 $\mathbf{x} = [x_1, x_2]$
- 假设 $\phi(\mathbf{x}) = [x_1^2, \sqrt{2x_1x_2}, x_2^2]$

$$\begin{aligned} \text{• 则 } \phi(\mathbf{x}) \cdot \phi(\mathbf{y}) &= \begin{pmatrix} x_1^2 \\ \sqrt{2x_1x_2} \\ x_2^2 \end{pmatrix} \begin{pmatrix} y_1^2 \\ \sqrt{2y_1y_2} \\ y_2^2 \end{pmatrix} \\ &= \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right)^2 = (\mathbf{x} \cdot \mathbf{y})^2 \end{aligned}$$

通过核函数实现
隐式映射

✓ Kernel function
 $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^2$

四、核函数与非线性SVM：非线性SVM

Kernel SVM

- 新的对偶问题：

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i^T, \mathbf{x}_j) \\ \text{s.t.} \quad & \sum_{i=1}^m \alpha_i y_i = 0, \\ & \alpha_i \geq 0, i = 1, 2 \dots m. \end{aligned}$$

- 分类模型：

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^m \alpha_i y_i K(\mathbf{x}_i^T, \mathbf{x}) + b\right)$$

四、核函数与非线性SVM：非线性SVM

- 理论上，如果原始空间维数有限，那么一定存在一个高维特征空间使样本线性可分。
- 但是在实际任务中，我们往往很难确定合适的核函数使训练样本在特征空间中恰好线性可分；即使找到了合适的核函数，也很难确定该线性可分的结果是不是由过拟合所造成的。
- 因此，一般会允许支持向量机在某些训练样本上出错，即在特征空间中使用“广义最优超平面”。

四、核函数与非线性SVM：核函数

- 思考：如何选择合适的核函数？

它必须是某个特征空间中的内积！

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，则可以作为核函数来使用。

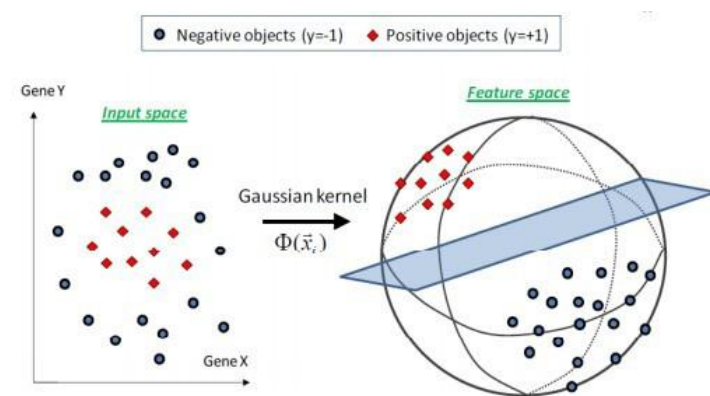
核矩阵

$$K = \begin{bmatrix} \kappa(x_1, x_1), \kappa(x_1, x_2) \cdots \kappa(x_1, x_m) \\ \kappa(x_2, x_1), \kappa(x_2, x_2) \cdots \kappa(x_2, x_m) \\ \cdots \\ \kappa(x_m, x_1), \kappa(x_m, x_2) \cdots \kappa(x_m, x_m) \end{bmatrix}$$

非线性分类模型、SVM

四、核函数与非线性SVM：核函数

常见的核函数



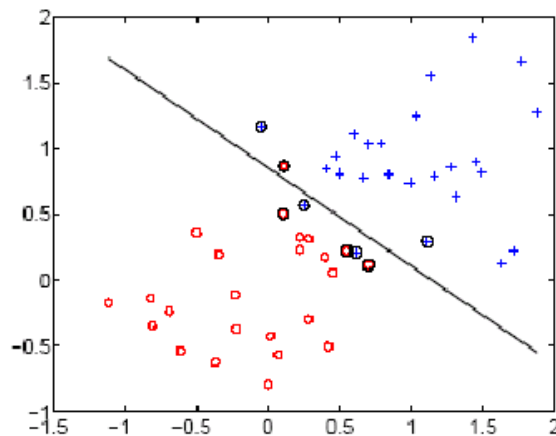
名称	表达式	Kernel SVM
线性核	$K(x, x') = (x \cdot x')$	线性SVM
多项式核	$K(x, x') = [(x \cdot x') + 1]^q$	用SVM实现的多项式判别函数
径向基函数 (RBF) 核 (高斯核)	$K(x, x') = \exp\left\{-\frac{\ x - x'\ ^2}{2\sigma^2}\right\}$	用SVM实现的RBF网络
Sigmoid核	$K(x, x') = \tanh(v(x \cdot x') + c)$	用SVM实现的三层MLP

四、核函数与非线性SVM：非线性SVM

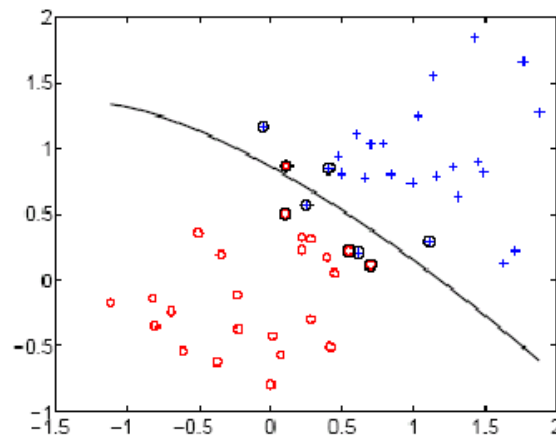
- 概括来说，使用核函数
 - ① 实际中，我们会经常遇到线性不可分的样例，此时，我们的常用做法是把样例特征映射到高维空间中去，映射到高维空间后，相关特征便被分开了，也就达到了分类的目的；
 - ② 但进一步，如果凡是遇到线性不可分的样例，一律映射到高维空间，那么这个维度大小是会高到可怕的(如无穷维)，会产生维数爆炸，那咋办呢？
 - ③ 此时，核函数就隆重登场了，核函数的价值在于它虽然也是将特征进行从低维到高维的转换，但核函数绝就绝在它事先在低维上进行计算，而将实质上的分类效果表现在了高维上，也就如上文所说的避免了直接在高维空间中的复杂计算。

四、核函数与非线性SVM：实例与应用

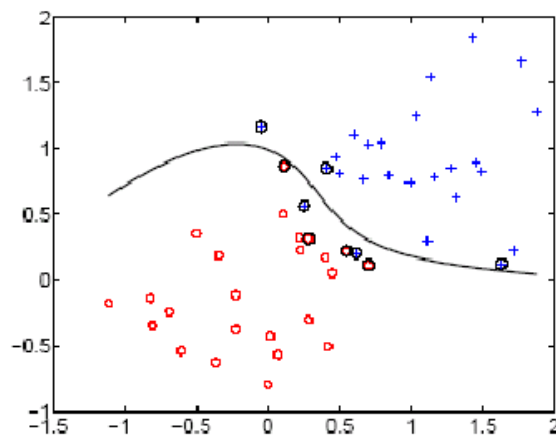
SVM的实例



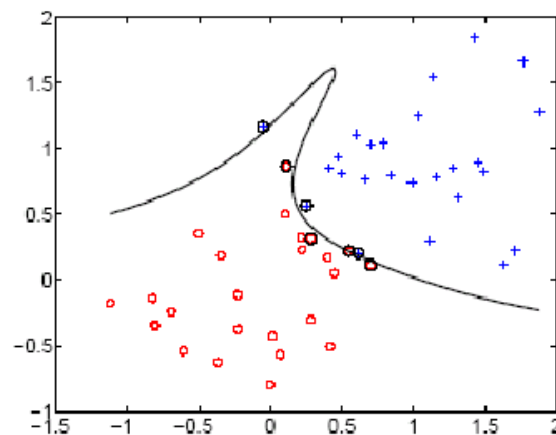
linear



2nd order polynomial



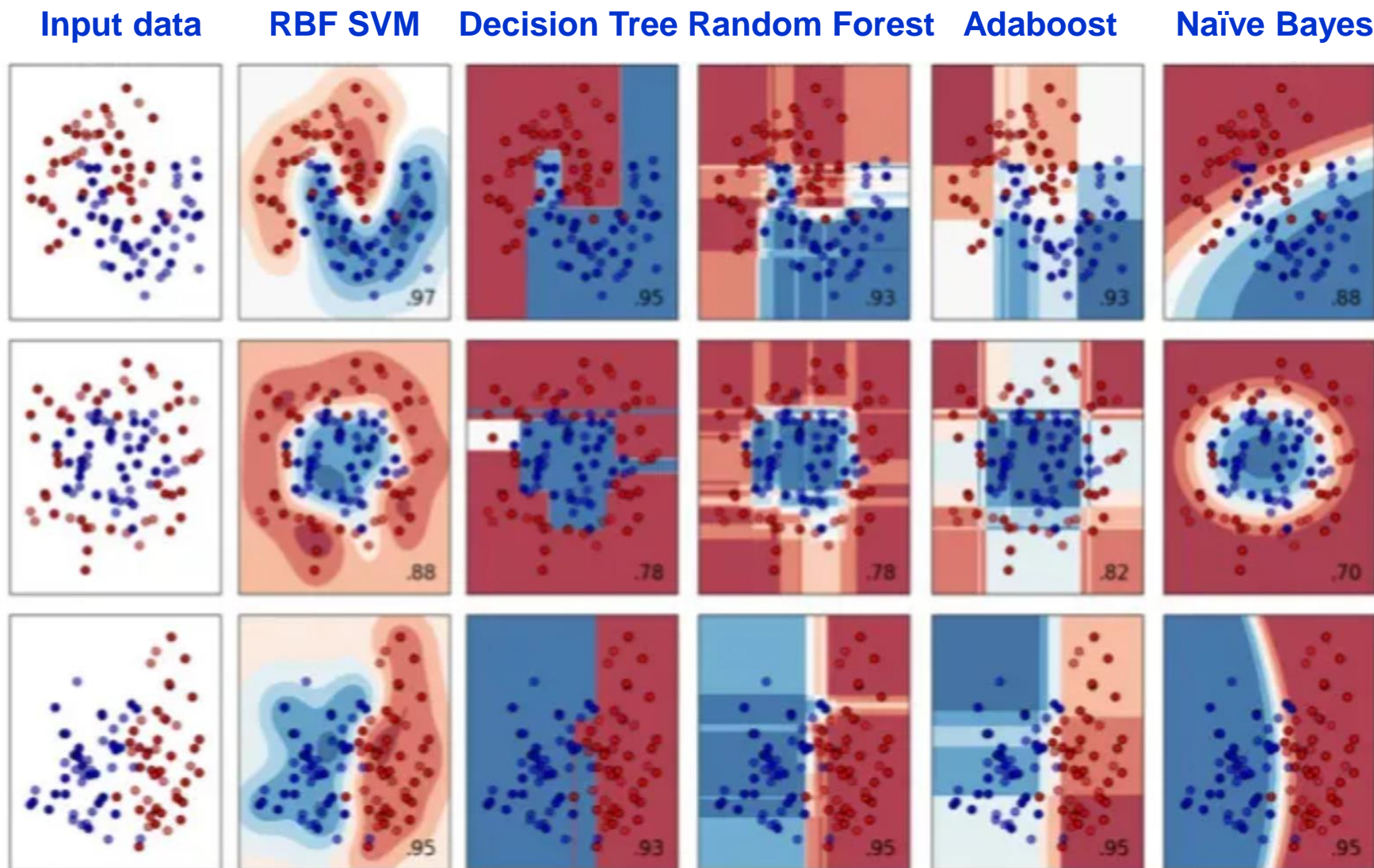
4th order polynomial



8th order polynomial

四、核函数与非线性SVM：实例与应用

SVM的实例



四、核函数与非线性SVM：实例与应用

Former bell Lab's experiments on USPS data



Method	Test Error
Human	2.5%
Decision Tree	16.2%
2-layer MLP	5.9%
5-layer MLP	5.1%
SVM with 3 types of kernels	4.0% 4.1% 4.1%

- 多项式核
- 高斯核
- Sigmoid核

四、核函数与非线性SVM：实例与应用

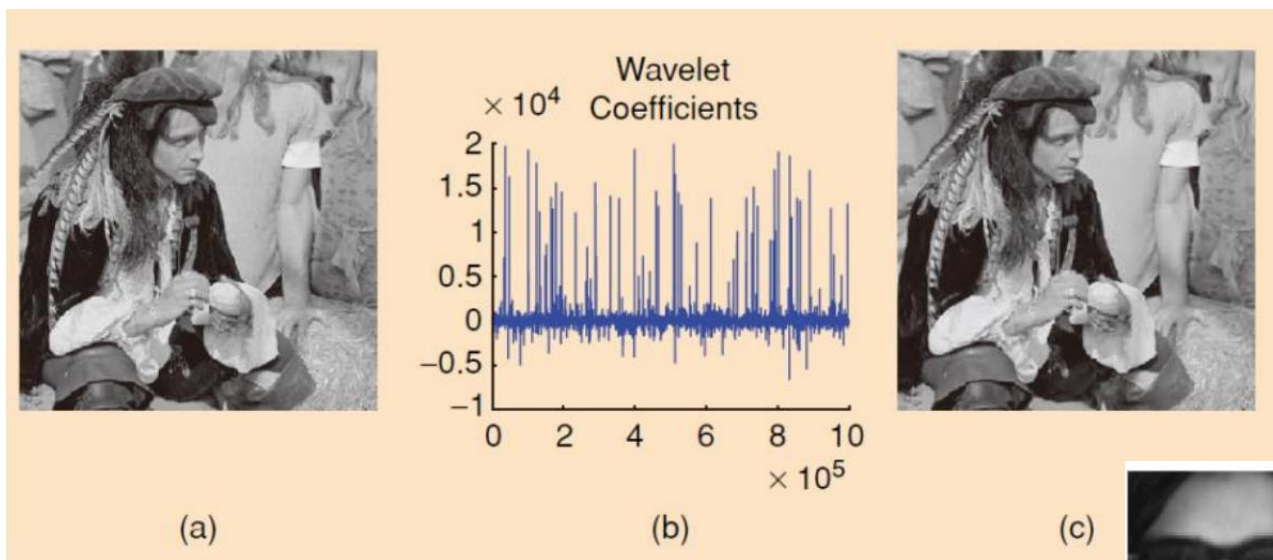
- SVM的应用

文本分析

account	stock	politics	day	health	movie	video	market	farm	car	account	stock	politics	day	health	movie	video	market	farm	car
pay	trading	music	index	image	bank	imports	exports	dollars	surplus	pay	trading	music	index	image	bank	imports	exports	dollars	surplus
club	game	bar	star	football	brain	sensor	study	effort	town	club	game	bar	star	football	brain	sensor	study	effort	town
sell	people	oil	good	walk	city	billion	value	China	normal	sell	people	oil	good	walk	city	billion	value	China	normal
top	money	card	job	start	run	beat	dance	wine	goal	top	money	card	job	start	run	beat	dance	wine	goal
arsenal	tax	large	salary	asset	China	federal	crisis	beer	small	arsenal	tax	large	salary	asset	China	federal	crisis	beer	small
price	house	Obama	wall	hubel	match	cell	state	card	bad	price	house	Obama	wall	hubel	match	cell	state	card	bad
rate	school	manage	season	system	nose	league	play	union	climate	rate	school	manage	season	system	nose	league	play	union	climate
capital	people	results	credit	plan	time	USA	street	down	year	capital	people	results	credit	plan	time	USA	street	down	year
Bush	team	free	invest	make	win	set	news	NBA	wealth	Bush	team	free	invest	make	win	set	news	NBA	wealth

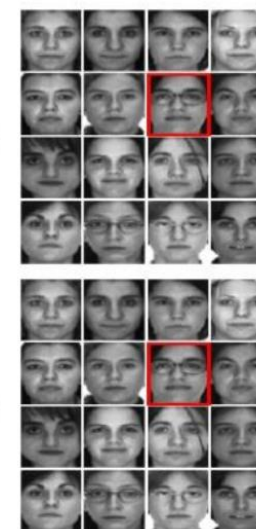
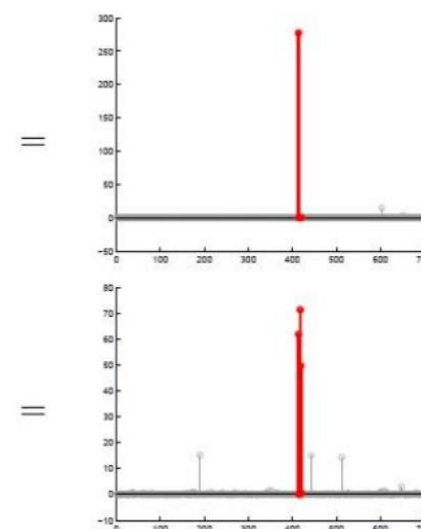
四、核函数与非线性SVM：实例与应用

- SVM的应用



信号处理

人脸识别



本章的主要内容

- 非线性关系
- 线性可分支持向量机
- 软间隔与正则化
- 核函数与非线性SVM
- 序列最小最优化算法
- 总结

五、序列最小最优化算法：SMO

- 经过上述章节的讨论我们了解到，支持向量机的学习问题可以形式化为求解凸二次规划问题，并可以引用拉格朗日算子进行求解。但是当训练样本容量很大时，这些算法往往会变得十分低效，以致无法使用。所以，如何快速高效地实现支持向量机学习就成为一个重要的问题。
- 本节介绍一种实现支持向量机学习的高效算法：序列最小最优化（Sequential Minimal Optimization, SMO）算法。（该算法1998年由Platt提出）

五、序列最小最优化算法：SMO

- SMO算法要解决如下凸二次规划的对偶问题：

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

此时变量为拉格朗日乘子 α_i ，一个变量 α_i 对应于一个样本点 (x_i, y_i) ，变量的总数对应于训练样本的容量

□ 基本思路：

- Step1: 选取一对需更新的变量 α_i 和 α_j
- Step2: 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

五、序列最小最优化算法：SMO

仅考虑 α_i 和 α_j 时，对偶问题的约束可重写为：

$$\alpha_i y_i + \alpha_j y_j = c, \quad \alpha_i \geq 0, \alpha_j \geq 0.$$

其中

$$c = - \sum_{k \neq i, j} \alpha_k y_k$$

是使 $\sum_{i=1} \alpha_i y_i = 0$ 成立的常数。

可以认为 α_i 和 α_j 是两个变量，其余为固定值；但当 α_i 确定， α_j 也会相应的确定

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问题，这样的二次规划问题具有闭式解，因此不必调用数值优化算法即可高效地计算出更新后的 α_i 和 α_j

本章的主要内容

- 非线性关系
- 线性可分支持向量机
- 软间隔与正则化
- 核函数与非线性SVM
- 序列最小最优化算法
- 总结

六、总结

- 支持向量机（SVM）是一种二类分类模型，其基本模型是定义在特征空间上的间隔最大的分类器，间隔使其有别于感知机。
- 本章按照由简至繁的思路构建了三类支持向量机模型：线性可分支持向量机（硬间隔支持向量机）、软间隔支持向量机及非线性支持向量机。
- 支持向量机的学习策略是间隔最大化，可以形式化为一个求解凸二次规划的问题，支持向量机的学习算法就是求解凸二次规划的最优化算法，通过引入拉格朗日算子可转化为相应的对偶问题。
- 通过使用核函数可以学习非线性支持向量机，等价于隐式地高维特征空间中学习线性支持向量机。

六、总结

思考

- 支持向量机的“最大间隔”思想
- 对偶问题及其解的稀疏性
- 引入“软间隔”缓解线性不可分问题
- 支持向量机的正则化
- 通过向高维特征空间映射解决线性不可分问题（核技巧）
- 高效求解算法：SMO

六、总结

一些成熟的SVM软件包

- LIBSVM <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- LIBLINEAR <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>
- SVM^{light}、SVM^{perf}、SVM^{struct}
http://svmlight.joachims.org/svm_struct.html
- Pegasos <http://www.cs.huji.ac.il/~shais/code/index.html>

六、总结

- 本章作业

(必做) 1. 参考线性可分支持向量机的对偶问题，试给出软间隔支持向量机对偶问题的推导过程。

(必做) 2. 已知正例点 $x_1 = (1,2)^T, x_2 = (2,3)^T, x_3 = (3,3)^T$, 负例点 $x_4 = (2,1)^T, x_5 = (3,2)^T$, 试求最大间隔分类超平面和分类决策函数。

六、总结

- 本章作业

(选做) 利用SVM分类MNIST手写数字集

- 可以利用LIBSVM工具包，MNIST从网上自行下载，每个数据展开为256维向量来处理；
- 选取20%样本点为测试集，剩下80%为训练集；
- 训练有soft margin的linear SVM，利用交叉验证选取margin大小，并将训好的模型测试test data，给出模型分类错误率；
- 训练有soft margin的RBF SVM，利用交叉验证选取margin大小与kernel bandwidth，并将训好的模型测试test data，给出模型分类错误率；
- 写出作业报告，交纸质版报告。

谢谢！

李爽

E-mail: shuangli@bit.edu.cn



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY