本课件版权所有@LAMDA，为本书教学目的可免费使用，其他目的需征得本书作者同意

周志华 著

MACHINE LEARNING

机器学习

清华大学出版社

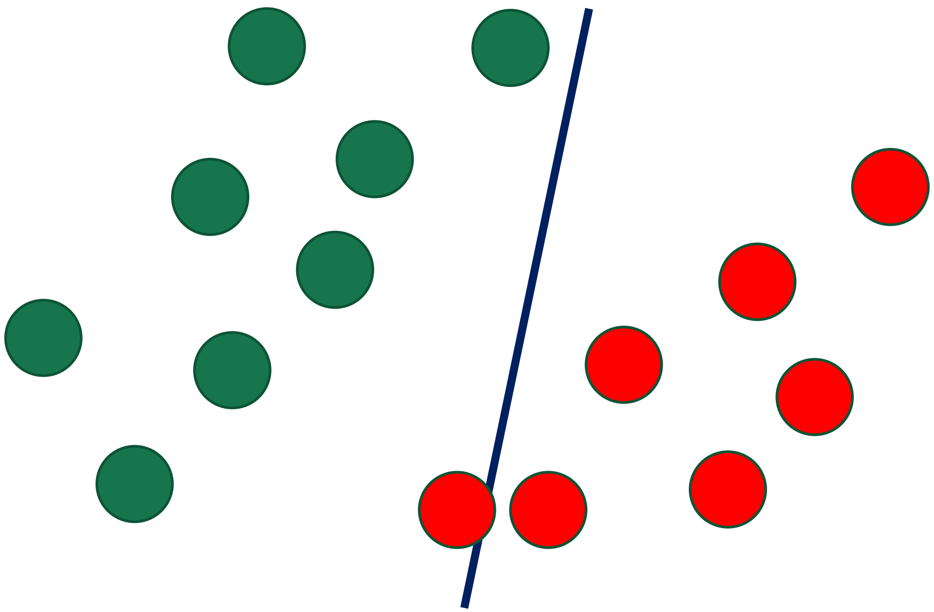
崔磊

QQ：362626744
E-Mail: leicui@nwu.edu.cn
办公室: 信息学院院楼912

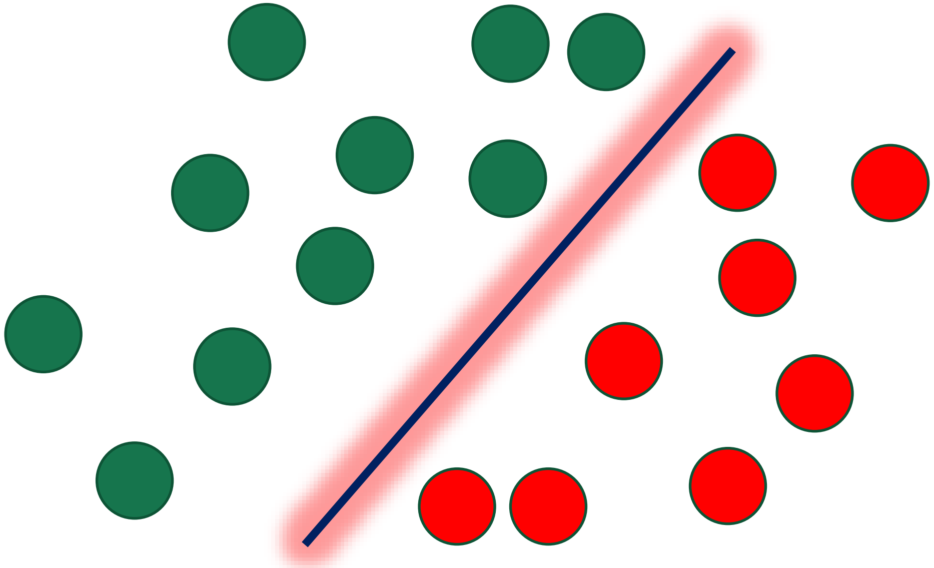
本章课件致谢：张腾

第六章：支持向量机

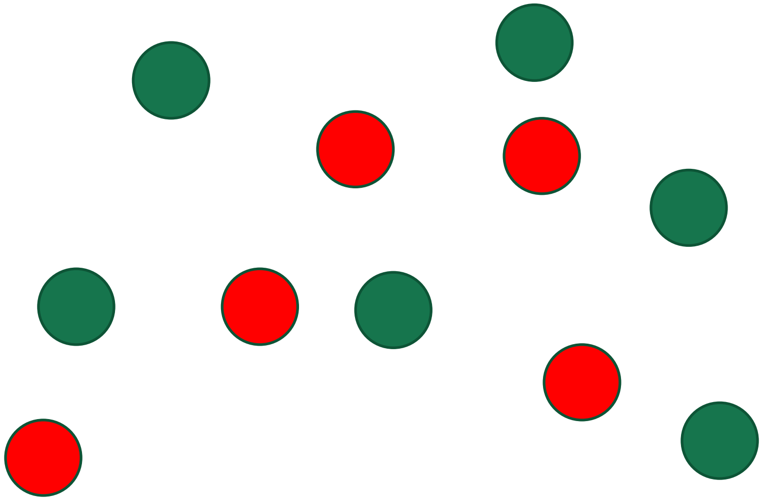
一个例子



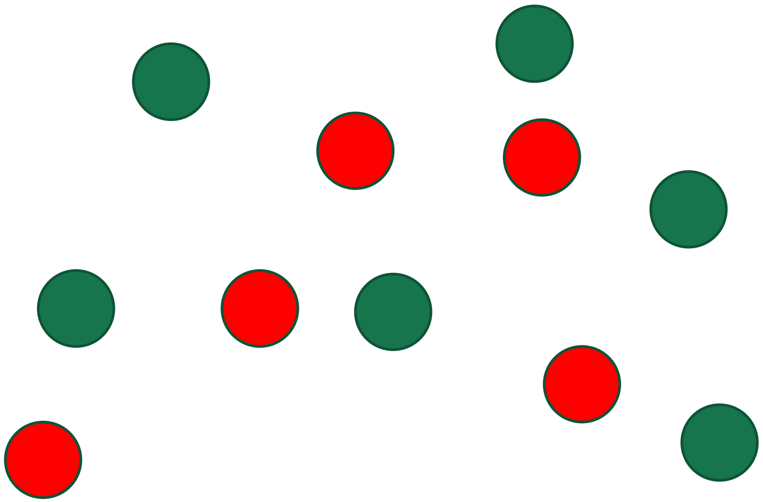
一个例子

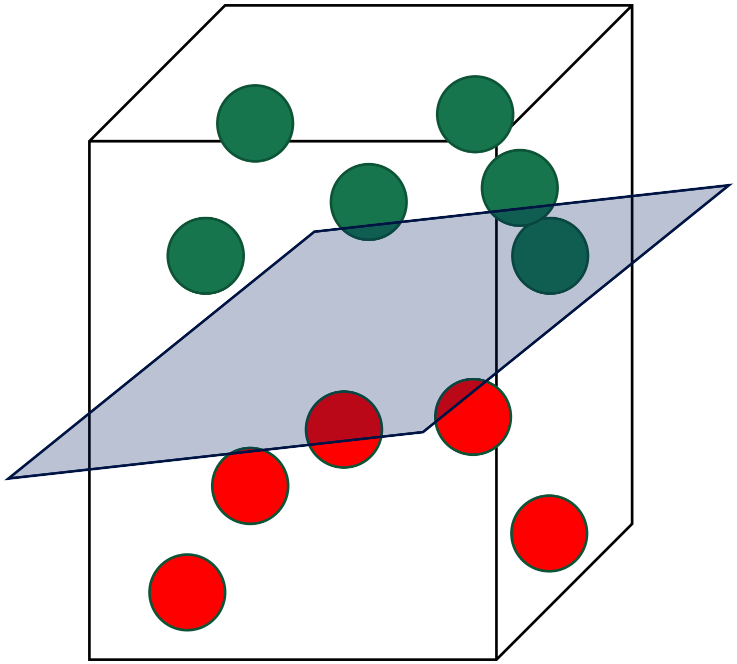


一个例子



魔鬼试炼



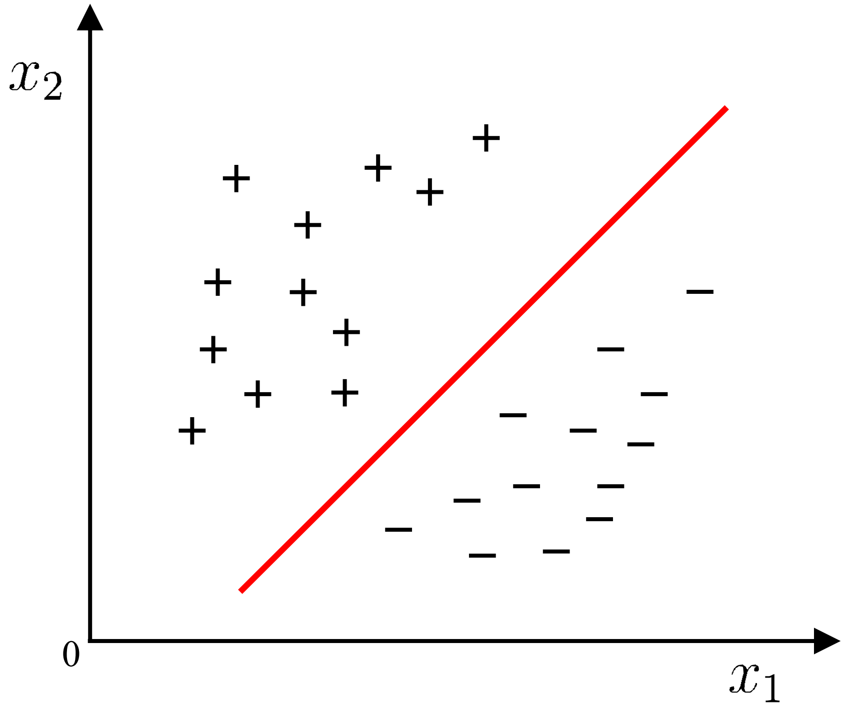


大纲

- 间隔与支持向量
- 优化问题求解方法
- 核函数
- 软间隔与正则化

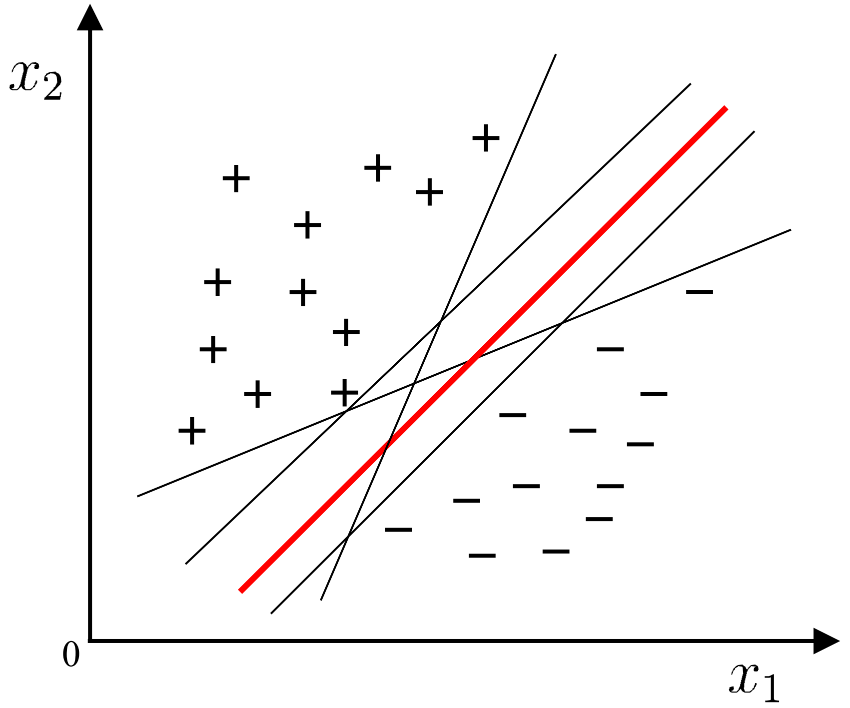
间隔与支持向量

线性模型：在样本空间中寻找一个超平面，将不同类别的样本分开.



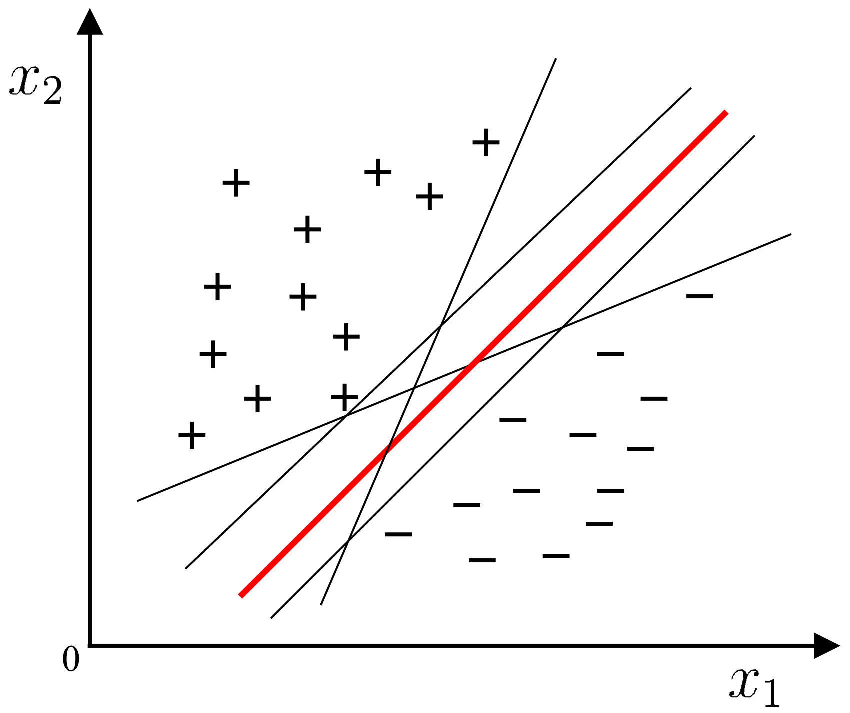
间隔与支持向量

-Q:将训练样本分开的超平面可能有很多，哪一个好呢?



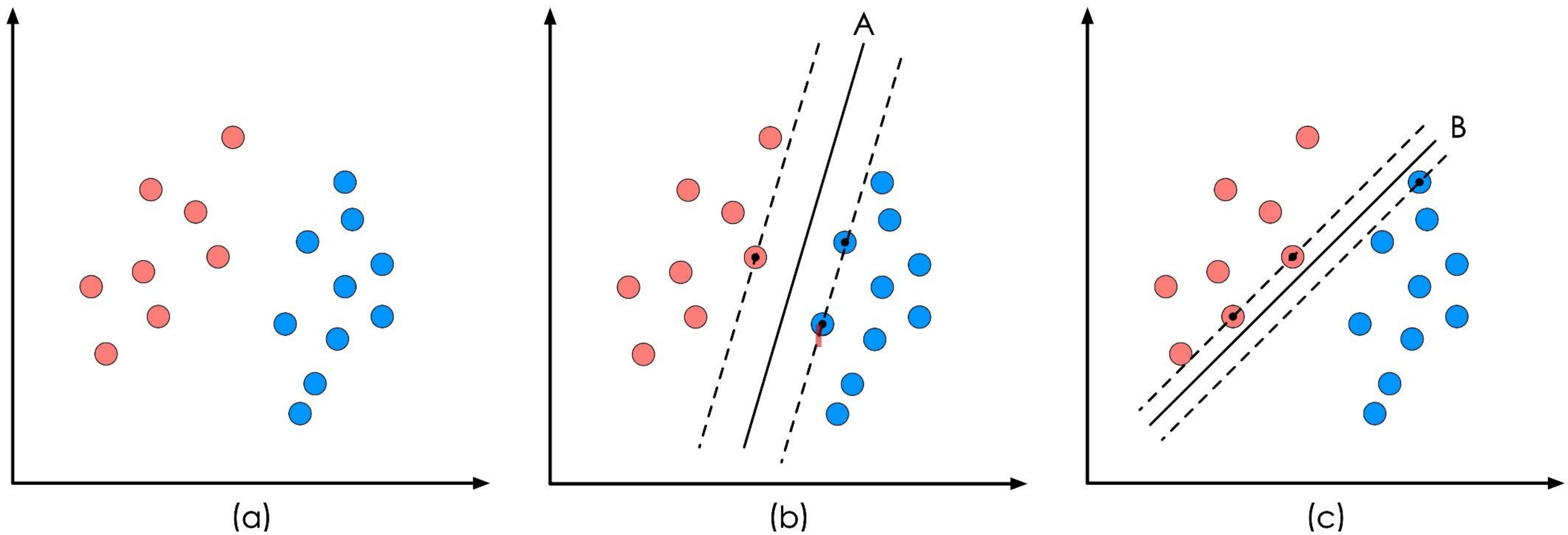
间隔与支持向量

-Q:将训练样本分开的超平面可能有很多，哪一个好呢?



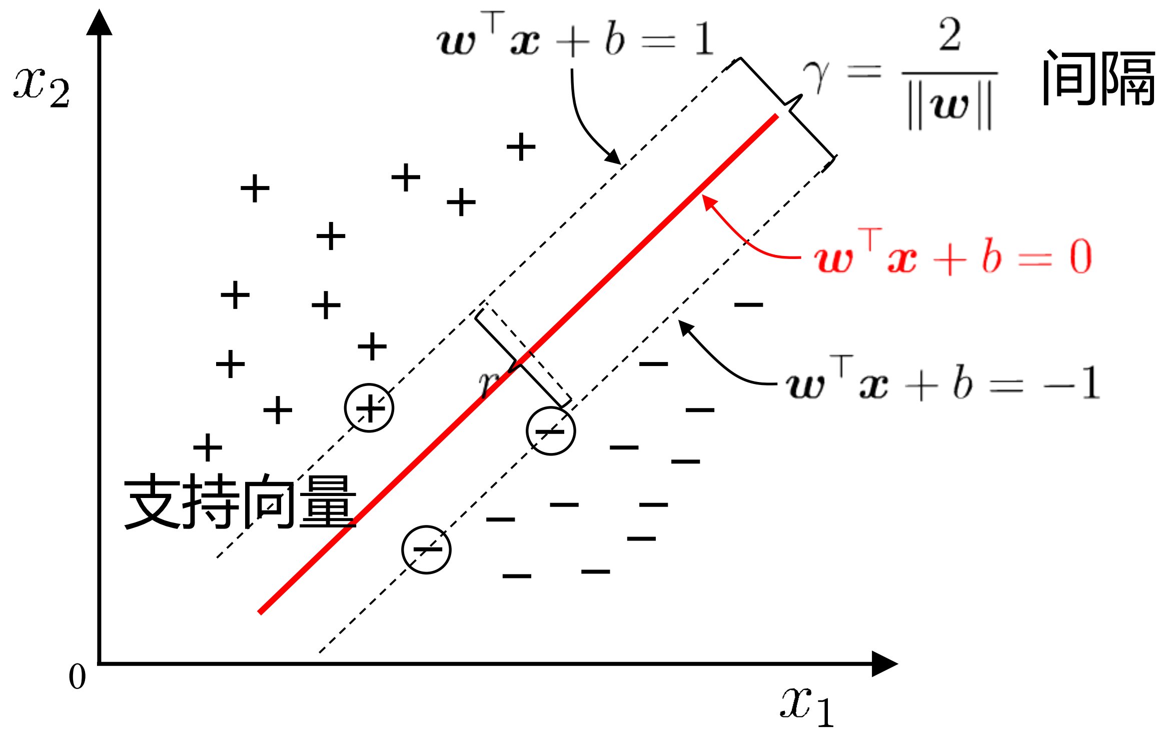
-A:应选择"正中间"，容忍性好，鲁棒性高，泛化能力最强.

间隔与支持向量



间隔与支持向量

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



支持向量机基本型

- 最大间隔：寻找参数 $\boldsymbol{w}$ 和 b，使得 γ 最大.

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

$$\Downarrow$$

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

大纲

- 间隔与支持向量
- 优化问题求解方法
- 核函数
- 软间隔与正则化

优化问题求解方法

- 无约束优化问题，可写为：

$$min\ f\ (x)$$

费马引理

导数为0

- 有等式约束的优化问题，可写为：

$$min\ f\ (x)$$
$$s.t.\ h_{i(x)} = 0,\quad i = 1,2,\dots,n$$

$$F(x,\lambda)=f(x)+\sum_{k=1}^{l}\lambda_k h_k(x)$$

拉格朗日乘子法

拉格朗日乘子法

- 拉格朗日乘子法
  - 第一步：引入拉格朗日乘子 $\alpha_i \geq 0$ 得到拉格朗日函数

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

  - 第二步：令 $L(\boldsymbol{w}, b, \boldsymbol{\alpha})$ 对 $\boldsymbol{w}$ 和 b 的偏导为零可得

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

  - 第三步：回代

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

解的稀疏性

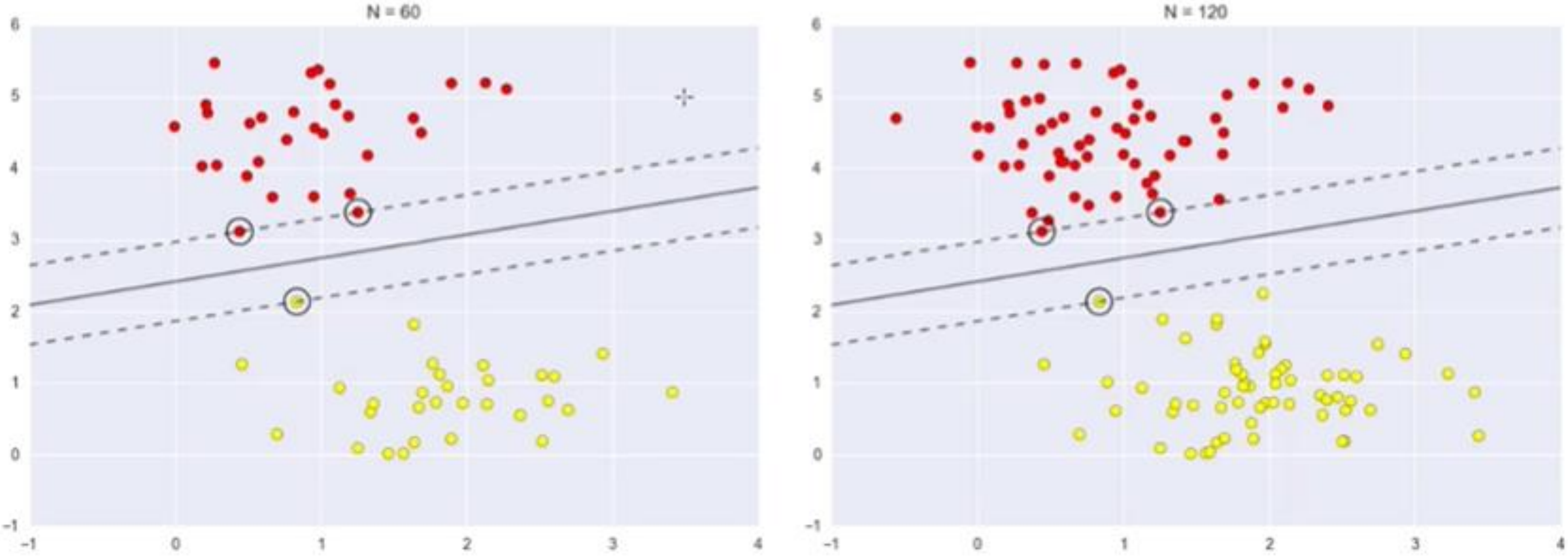
- 最终模型：$f(\boldsymbol{x}) = \boldsymbol{w}^\top \boldsymbol{x} + b = \boxed{\sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i^\top \boldsymbol{x}} + b$

- KKT条件：

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

$$y_i f(\boldsymbol{x}_i) > 1 \quad \Rightarrow \quad \alpha_i = 0$$

支持向量机解的稀疏性：训练完成后，大部分的训练样本都不需保留，最终模型仅与支持向量有关.


N = 60
N = 120

求解方法 - SMO

- 基本思路：不断执行如下两个步骤直至收敛.
  - 第一步：选取一对需更新的变量α_i和α_j.
  - 第二步：固定α_i和α_j以外的参数，求解对偶问题更新α_i和α_j.

- 仅考虑 α_i和α_j时，对偶问题的约束变为

$$\alpha_i y_i + \alpha_j y_j = -\sum_{k \neq i,j} \alpha_k y_k, \quad \alpha_i \geq 0, \quad \alpha_j \geq 0.$$

用一个变量表示另一个变量，回代入对偶问题可得一个单变量的二次规划，该问题具有闭式解.

- 偏移项b：通过支持向量来确定.

小试牛刀

已知三个数据点：正样本$x_1 = (3, 3)^T$，$x_2 = (4, 3)^T$，负样本$x_3 = (1, 1)^T$，求超平面方程。($y_i \in \{+1, -1\}$)

求解：

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

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

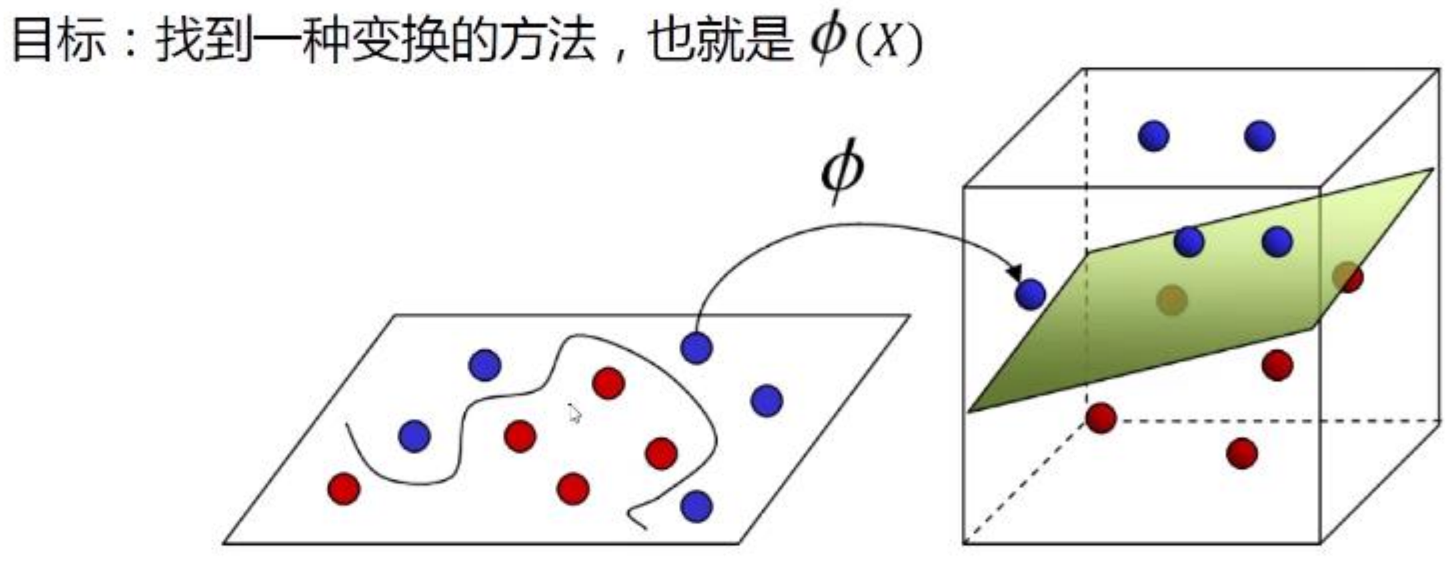
大纲

- 间隔与支持向量
- 对偶问题
- 核函数
- 软间隔与正则化

线性不可分

-Q:若不存在一个能正确划分两类样本的超平面，怎么办?

-A:将样本从原始空间映射到一个更高维的特征空间，使得样本在这个特征空间内线性可分.



核支持向量机

- 设样本 $\boldsymbol{x}$ 映射后的向量为 $\phi(\boldsymbol{x})$，划分超平面为 $f(\boldsymbol{x})=\boldsymbol{w}^\top\phi(\boldsymbol{x})+b$.

原始问题

$$\min_{\boldsymbol{w},b} \ \frac{1}{2}\|\boldsymbol{w}\|^2$$
$$\text{s.t.} \ \ y_i(\boldsymbol{w}^\top\phi(\boldsymbol{x}_i)+b)\geq 1,\ \ i=1,2,\ldots,m.$$

对偶问题

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

只以内积的形式出现

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

预测

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

核函数

- 基本想法：不显式地设计核映射，而是设计核函数.

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^\top \phi(\boldsymbol{x}_j)$$

- Mercer定理(充分非必要)：只要一个对称函数所对应的核矩阵半正定，则它就能作为核函数来使用.

- 常用核函数：

名称	表达式	参数
线性核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^\top \boldsymbol{x}_j$	
多项式核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^\top \boldsymbol{x}_j)^d$	$d \geq 1$为多项式的次数
高斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\lVert\boldsymbol{x}_i-\boldsymbol{x}_j\rVert^2}{2\delta^2}\right)$	$\delta > 0$为高斯核的带宽(width)
拉普拉斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\lVert\boldsymbol{x}_i-\boldsymbol{x}_j\rVert}{\delta}\right)$	$\delta > 0$
Sigmoid核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh(\beta \boldsymbol{x}_i^\top \boldsymbol{x}_j + \theta)$	tanh为双曲正切函数，$\beta > 0$, $\theta < 0$

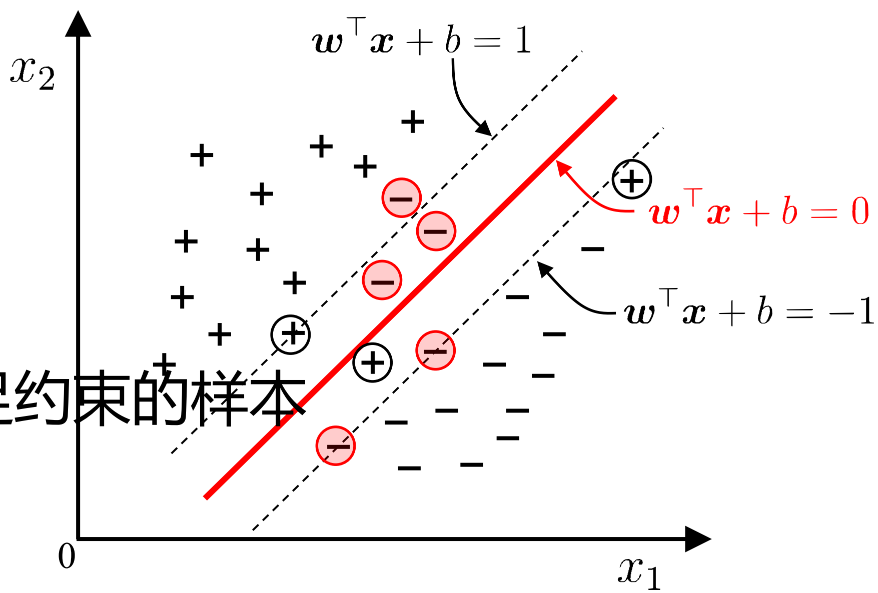
大纲

- 间隔与支持向量
- 对偶问题
- 核函数
- 软间隔与正则化

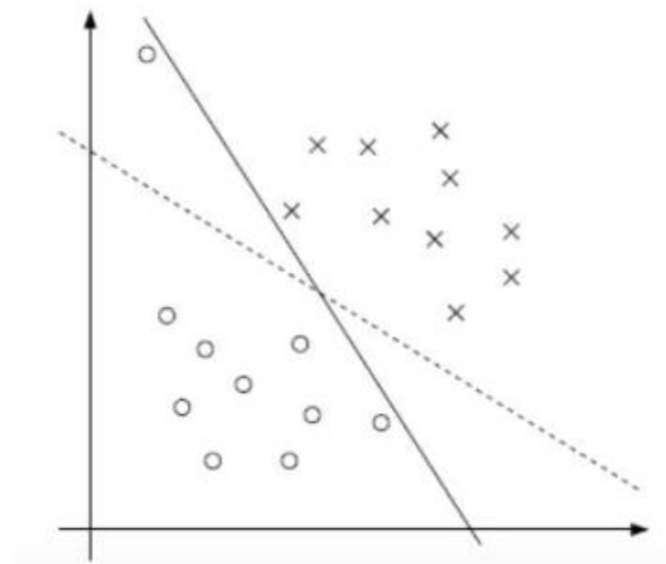
软间隔

-Q:现实中，很难确定合适的核函数使得训练样本在特征空间中线性可分；同时一个线性可分的结果也很难断定是否是有过拟合造成的.

-A:引入"软间隔"的概念，允许支持向量机在一些样本上不满足约束.



不满足约束的样本



软间隔

为了解决该问题，引入松弛因子

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i$$

新的目标函数： $\min \frac{1}{2} ||w||^2 + C \sum_{i=1}^{n} \xi_i$

当C趋近于很大时：意味着分类严格不能有错误

当C趋近于很小时：意味着可以有更大的错误容忍

C是我们需要指定的一个参数！

0/1损失函数

- 基本想法：最大化间隔的同时，让不满足约束的样本应尽可能少.

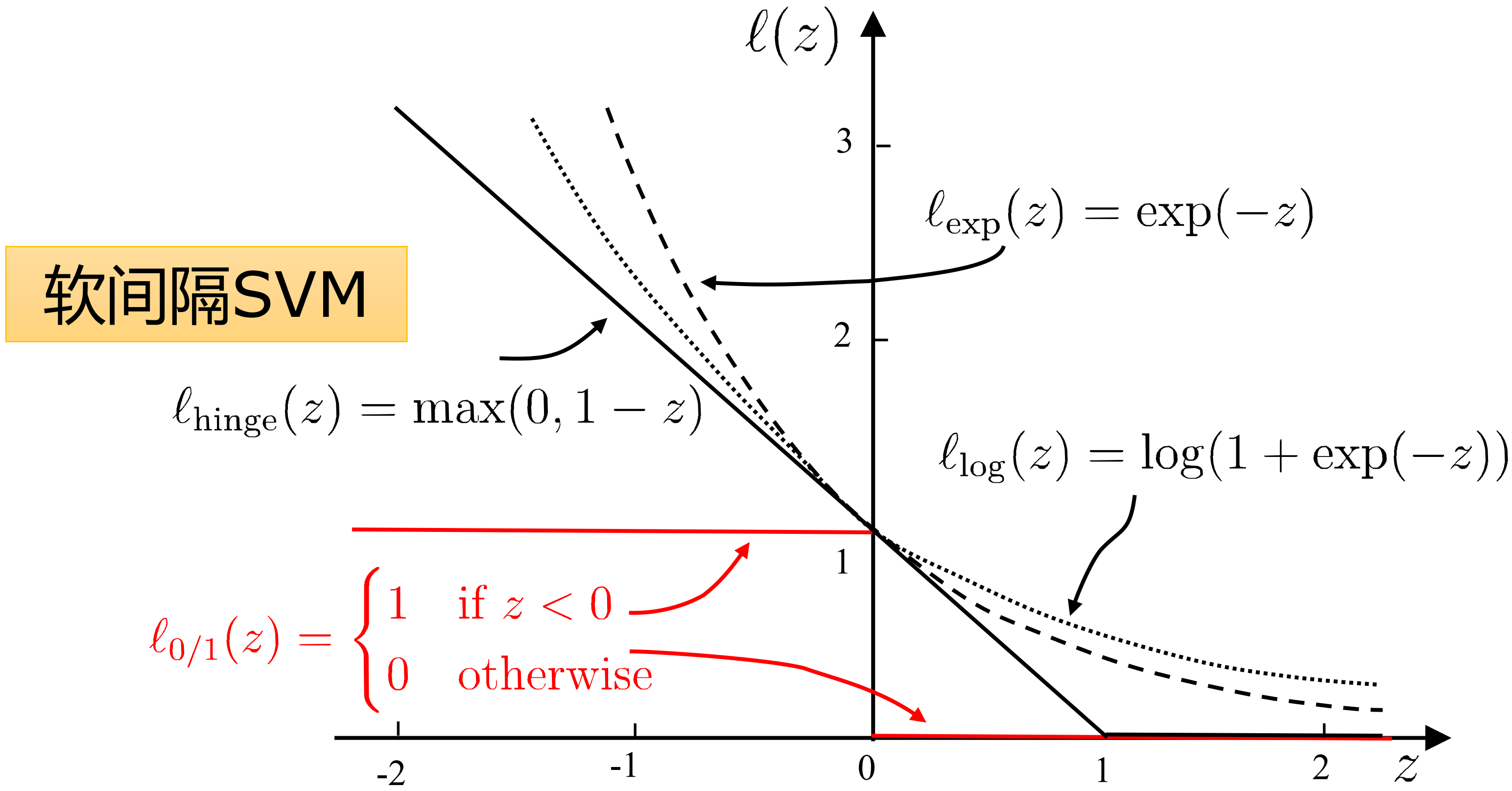
$$\min_{\boldsymbol{w},b} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} l_{0/1}\left(y_i(\boldsymbol{w}^\top\phi(\boldsymbol{x}_i)+b)-1\right)$$

其中 $l_{0/1}$ 是"0/1损失函数"

$$l_{0/1} = \begin{cases} 1 & z < 0 \\ 0 & otherwise \end{cases}$$

- 存在的问题：0/1损失函数非凸、非连续，不易优化!

替代损失



替代损失函数数学性质较好，一般是0/1损失函数的上界

软间隔支持向量机

原始问题

$$\min_{\boldsymbol{w},b} \ \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \max\left(0, 1 - y_i(\boldsymbol{w}^\top \phi(\boldsymbol{x}_i) + b)\right)$$

对偶问题

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

$$\text{s.t.} \ \sum_{i=1}^{m}\alpha_i y_i = 0, \ 0 \le \alpha_i \le C, \ i = 1, 2, \dots, m.$$

根据KKT条件可推得最终模型仅与支持向量有关，也即hinge损失函数依然保持了支持向量机解的稀疏性.

正则化

- 支持向量机学习模型的更一般形式

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

结构风险，描述模型的某些性质

经验风险，描述模型与训练数据的契合程度

- 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)
  - ……

大纲

- 间隔与支持向量
- 对偶问题
- 核函数
- 软间隔与正则化

Take Home Message

- 支持向量机的"最大间隔"思想
- 优化问题求解方法
- 通过向高维空间映射解决线性不可分的问题
- 引入"软间隔"缓解特征空间中线性不可分的问题

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/

- $\text{SVM}^{\text{light}}$、$\text{SVM}^{\text{perf}}$、$\text{SVM}^{\text{struct}}$
  http://svmlight.joachims.org/svm_struct.html

- Pegasos
  http://www.cs.huji.ac.il/~shais/code/index.html