

07 支持向量机

最大间隔、软间隔与核方法

■ 感知机已经解决

- 线性可分时找到一个分类超平面
- 用 $f(x) = w^T x + b$ 的符号预测类别

■ 仍未解决

- 多个超平面都能分对训练样本
- 训练正确率无法说明应选择哪一个

■ SVM 的回答

- 比较最近样本到边界的几何距离
- 选择最小距离最大的分类边界

1 从正确分类到最大间隔

怎样把“选择离训练样本更远的边界”变成可训练的模型

■ 同一组训练样本

- 蓝色圆点与黑色方点线性可分

■ 三条候选直线

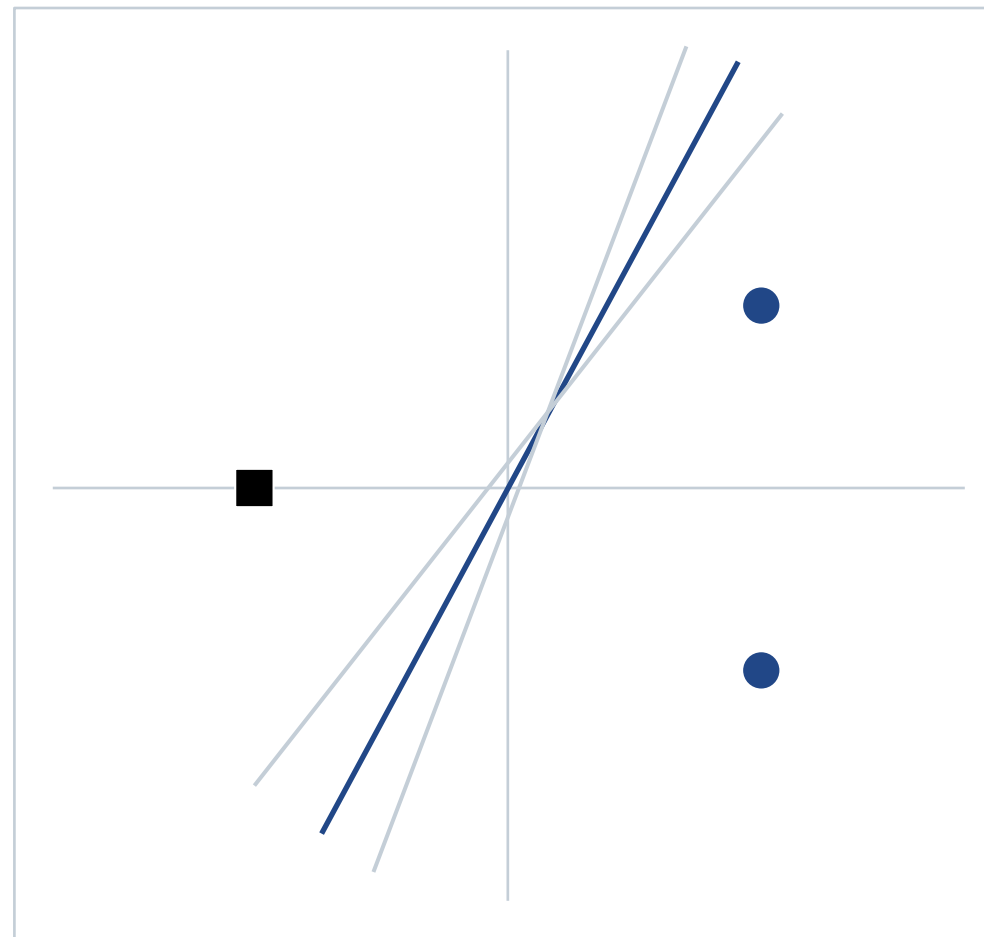
- 都没有训练错误

■ 比较任务

- 找出离最近样本最远的直线

■ 观察重点

- 决定选择的是最小距离，不是平均距离



都对，但离最近样本的距离不同

■ 线性得分

- $f(x) = w^T x + b$
- w 决定边界方向, b 决定边界位置

■ 决策边界

- $f(x) = 0$ 的点组成分类超平面

■ 类别判断

- $f(x) > 0$ 与 $f(x) < 0$ 对应边界两侧

■ 标签约定

- $y_i \in \{-1, +1\}$, 分类正确时 $y_i f(x_i) > 0$

■ 最短方向

- 法向量 w 垂直于超平面 $w^T x + b = 0$

■ 距离公式

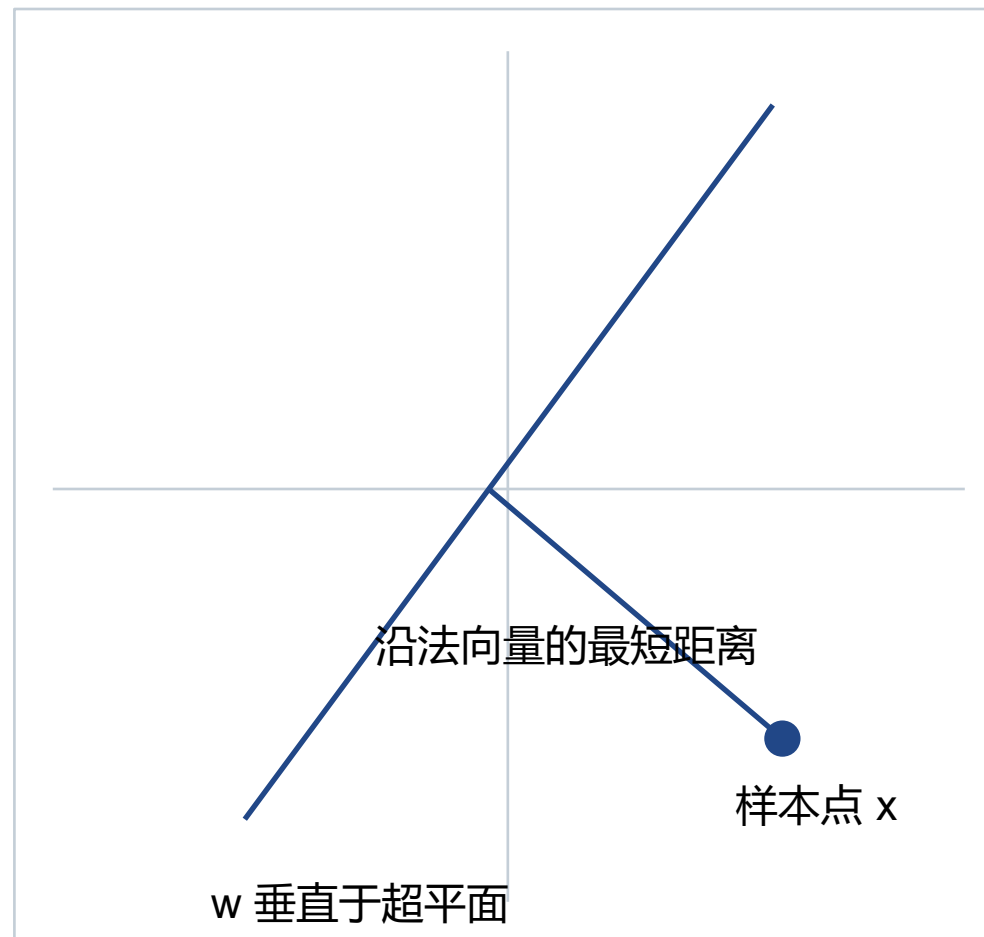
- $$d(x, H) = \frac{|w^T x + b|}{\|w\|}$$

■ 分母的作用

- $\|w\|$ 消除参数尺度对得分数值的影响

■ 几何含义

- 距离沿法向量方向测量



■ 单个样本的函数间隔

- $\hat{\gamma}_i = y_i(w^T x_i + b)$

■ 符号

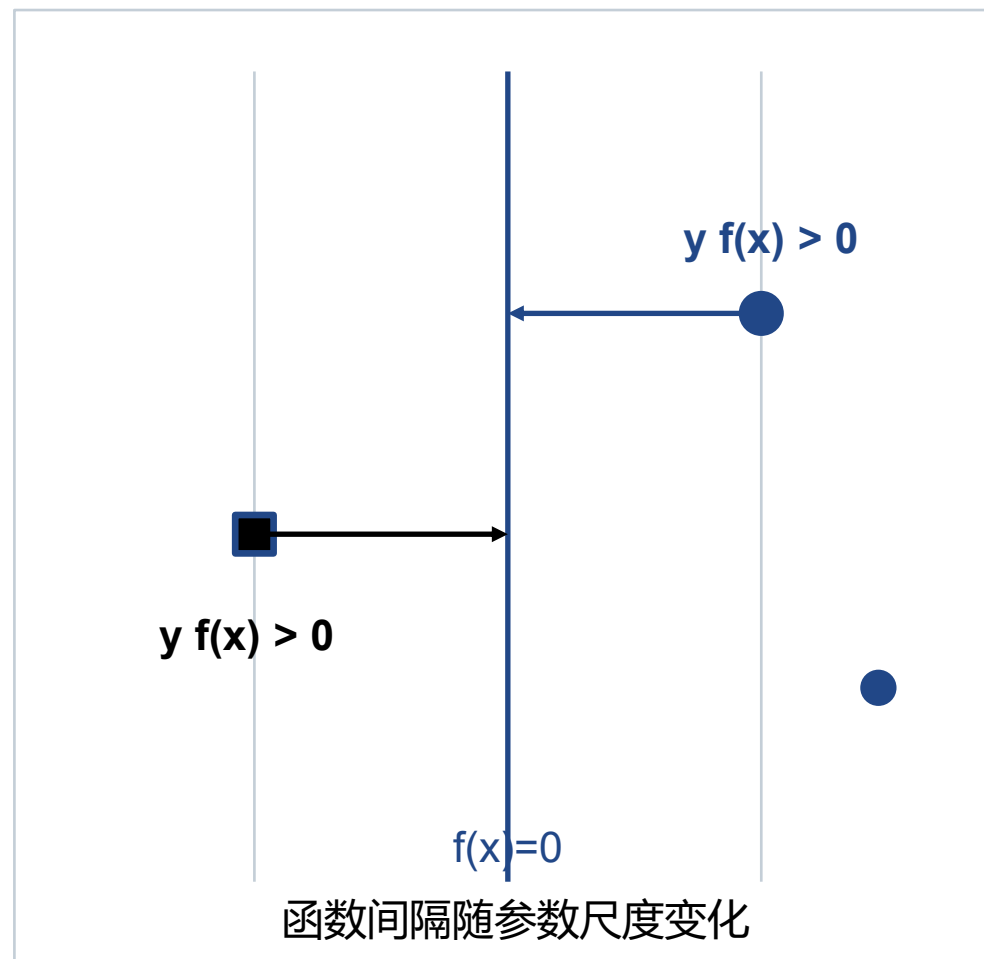
- 正值表示分类正确，负值表示分类错误

■ 数值大小

- 正确分类时，较大数值表示得分离零更远

■ 限制

- 参数同比缩放会同比放大函数间隔



■ 函数间隔

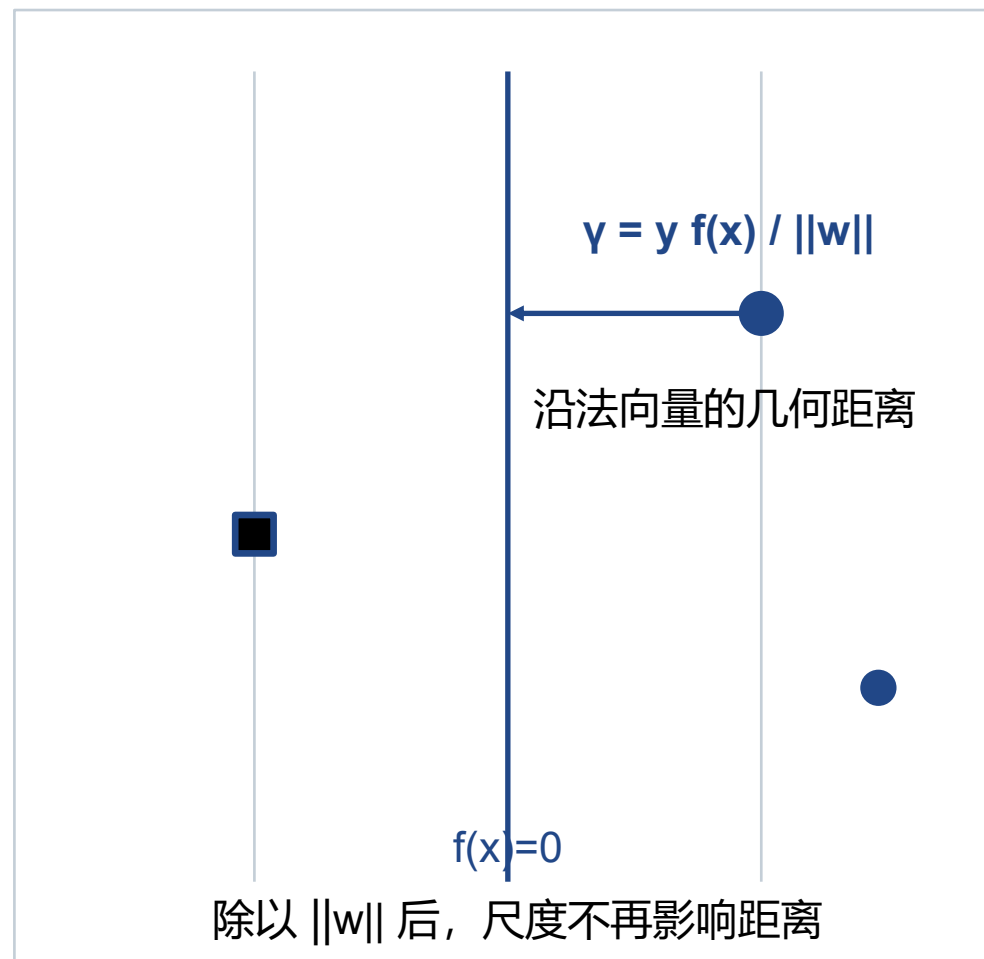
- $\hat{\gamma}_i = y_i(w^T x_i + b)$
- 同时包含分类结果与参数尺度

■ 几何间隔

- $\gamma_i = \frac{y_i(w^T x_i + b)}{\|w\|}$
- 对正确分类样本等于到超平面的距离

■ 训练集间隔

- $\gamma = \min_i \gamma_i$
- 最近样本限制整条边界的间隔



■ 原参数

- 得分为 $f(x) = w^T x + b$
- 决策边界满足 $f(x) = 0$

■ 缩放参数

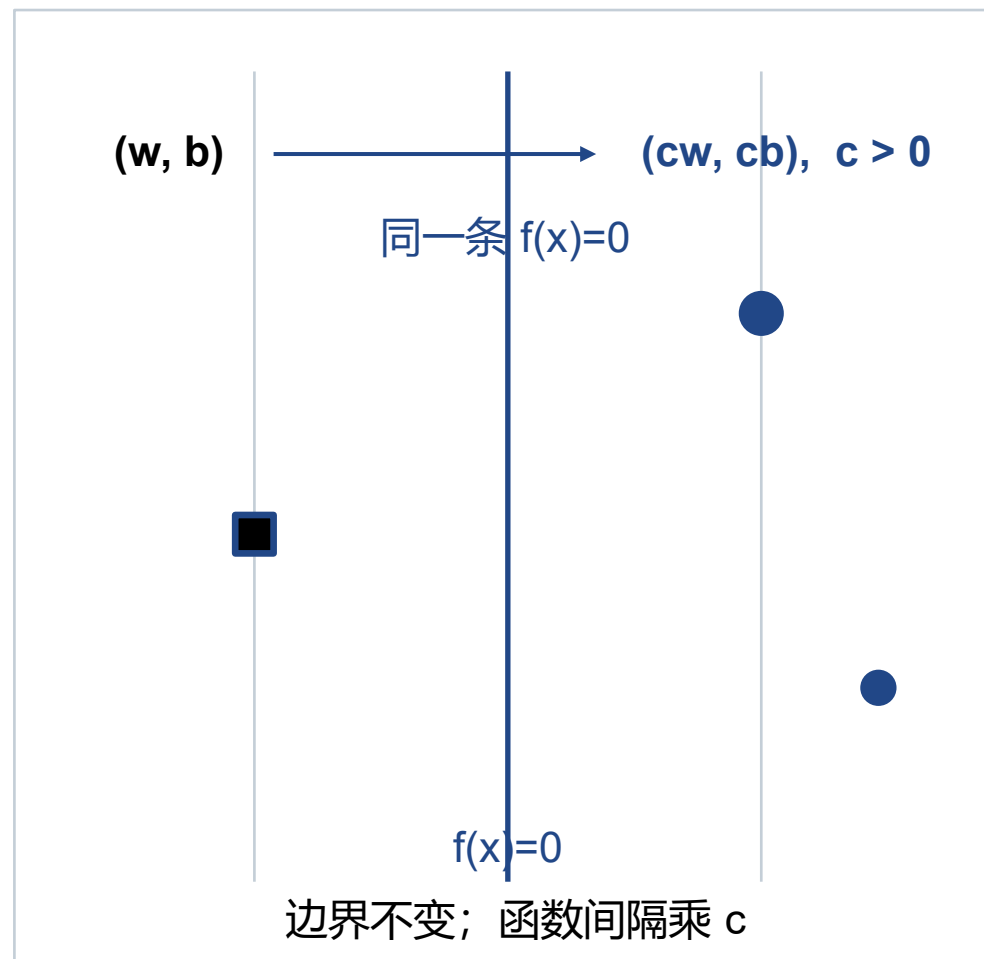
- 取 $c > 0$, 新得分为 $cf(x)$
- 决策边界和分类符号保持不变

■ 两个间隔的变化

- 函数间隔变为原来的 c 倍
- 几何间隔保持不变

■ 结论

- 最大化函数间隔之前必须固定参数尺度



■ 规范化选择

- 令最近样本满足 $y_i(w^T x_i + b) = 1$

■ 全部样本约束

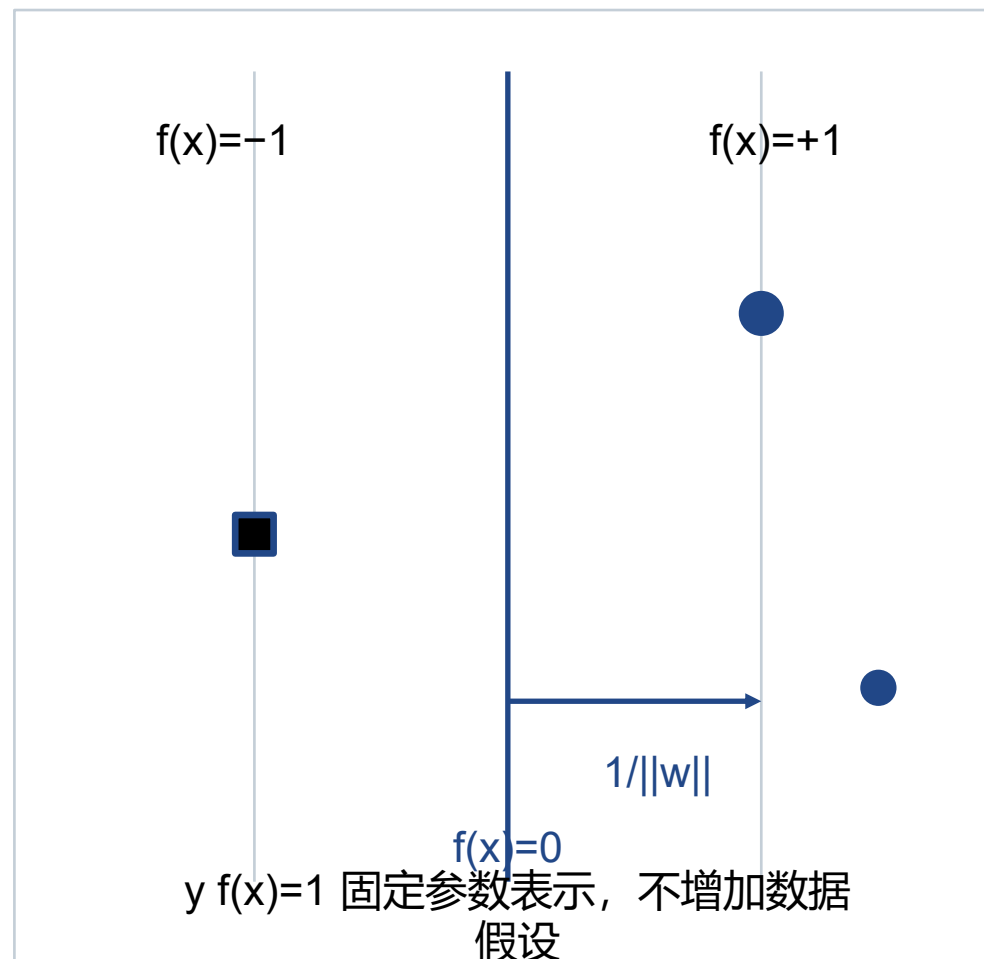
- $y_i(w^T x_i + b) \geq 1$

■ 两条间隔边界

- $w^T x + b = 1$ 与 $w^T x + b = -1$

■ 间隔宽度

- 单侧间隔为 $1/\|w\|$
- 两条间隔边界相距 $2/\|w\|$



■ 优化问题

- $\min_{w,b} \frac{1}{2} \|w\|^2$
- s.t. $y_i(w^T x_i + b) \geq 1, i = 1, \dots, n$

■ 目标函数

- 减小 $\|w\|$ 等价于扩大 $2/\|w\|$

■ 约束

- 每个训练样本分类正确并位于间隔边界外

■ 适用前提

- 训练数据线性可分

三个样本给出三条单位间隔约束

- **正类样本** $(2, 1)$
 - $y = +1$, 约束为 $2w_1 + w_2 + b \geq 1$
- **正类样本** $(2, -1)$
 - $y = +1$, 约束为 $2w_1 - w_2 + b \geq 1$
- **负类样本** $(-2, 0)$
 - $y = -1$, 约束为 $2w_1 - b \geq 1$
- **共同来源**
 - 三式均由 $y_i(w^T x_i + b) \geq 1$ 逐点代入得到

■ 合并正类样本的两条约束

- $2w_1 + b \geq 1 + |w_2|$

■ 负类样本的约束

- $2w_1 - b \geq 1$

■ 合并正负约束

- 两式相加得到 $4w_1 \geq 2 + |w_2|$

- 因此 $w_1 \geq 1/2 + |w_2|/4 \geq 1/2$

■ 目标最小时取等

- 取 $w_2 = 0, w_1 = 1/2$; 此时两类约束共同要求 $b = 0$

- 得到 $\frac{1}{2} \|w\|^2 = 1/8$

■ 逐点核查

- 三个样本均满足 $y_i(w^T x_i + b) = 1$
- 在这个非退化例中，三个边界样本都是支持向量

■ 单侧间隔

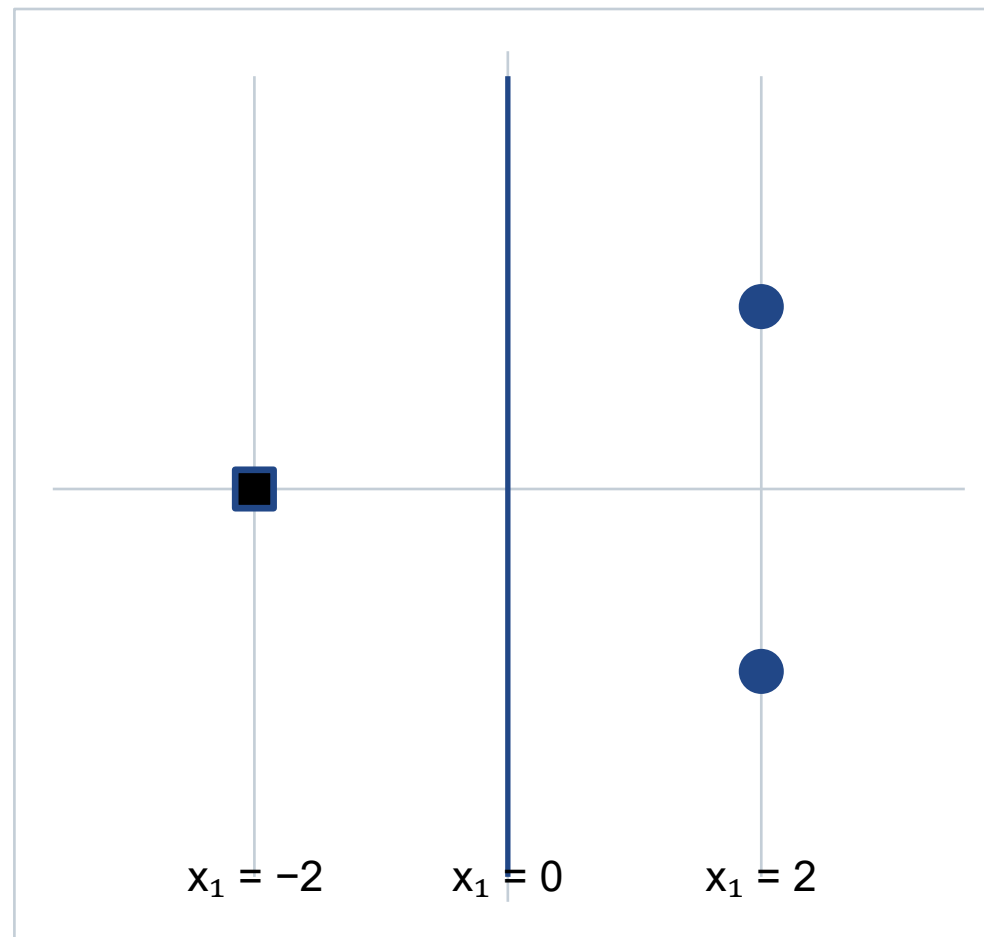
- $\|w\| = 1/2$ ，所以 $1/\|w\| = 2$

■ 间隔边界

- $(1/2)x_1 = \pm 1$ ，即 $x_1 = \pm 2$

■ 总间隔宽度

- $\frac{2}{\|w\|} = 4$



三个样本都在间隔边界上

■ 新输入

- $x = (3, 0)$

■ 决策得分

- $f(x) = w^T x + b = (1/2) \times 3 + 0 = 3/2$

■ 类别

- $f(x) > 0$, 因此预测为正类

■ 边界

- $3/2$ 是决策得分, 不是类别概率

■ 硬间隔中的识别

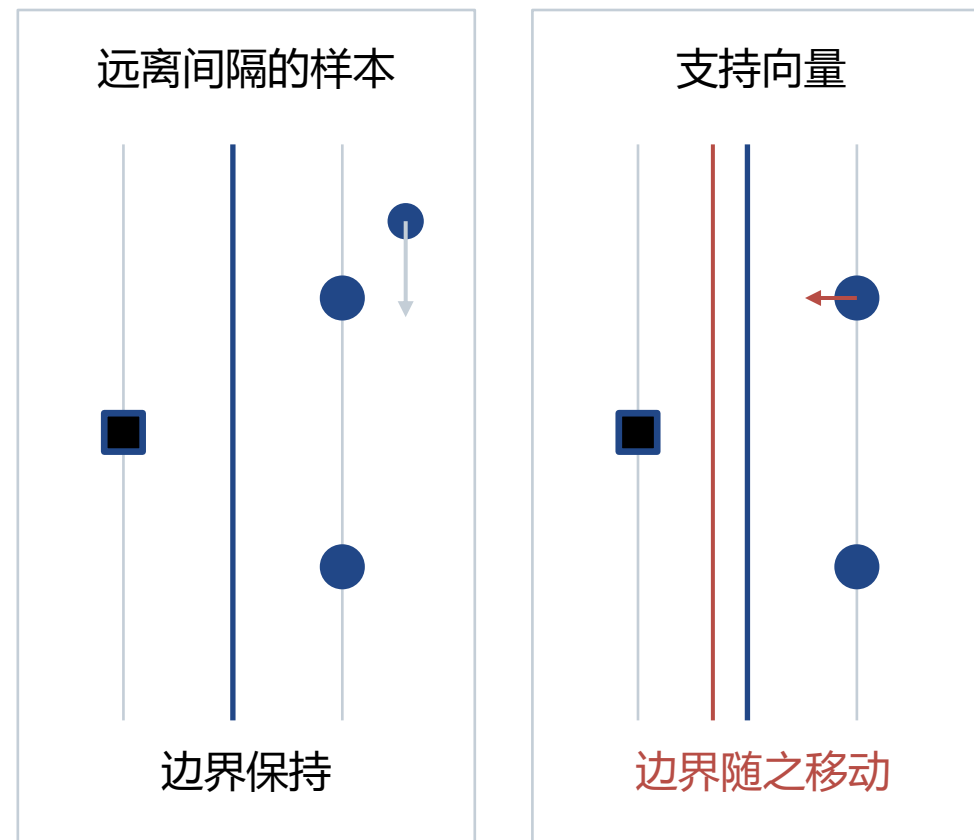
- 满足 $y_i(w^T x_i + b) = 1$ 的样本位于间隔边界上

■ 本例的结果

- 三个训练样本都达到等号，并对应非零对偶系数

■ 移动样本的影响

- 移动远离间隔的样本通常不会立即改变边界
- 移动支持向量可能改变可达到的最大间隔

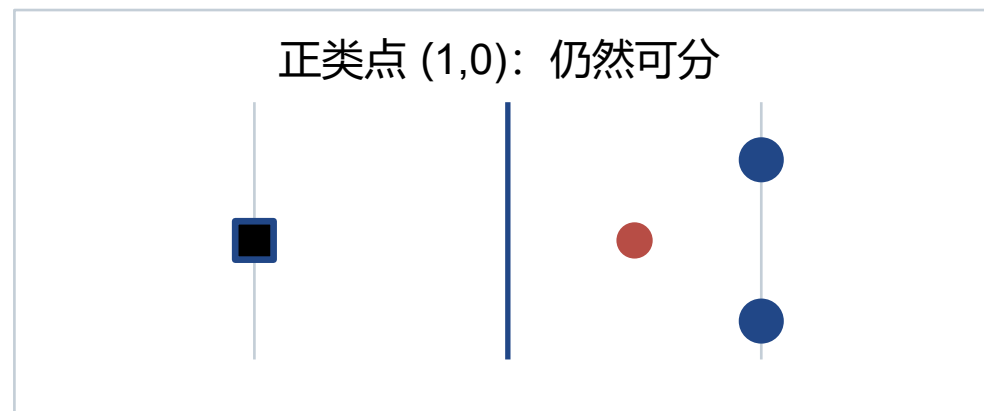


2 从硬约束到软间隔

怎样在约束违反或数据不可分时保留最大间隔原则

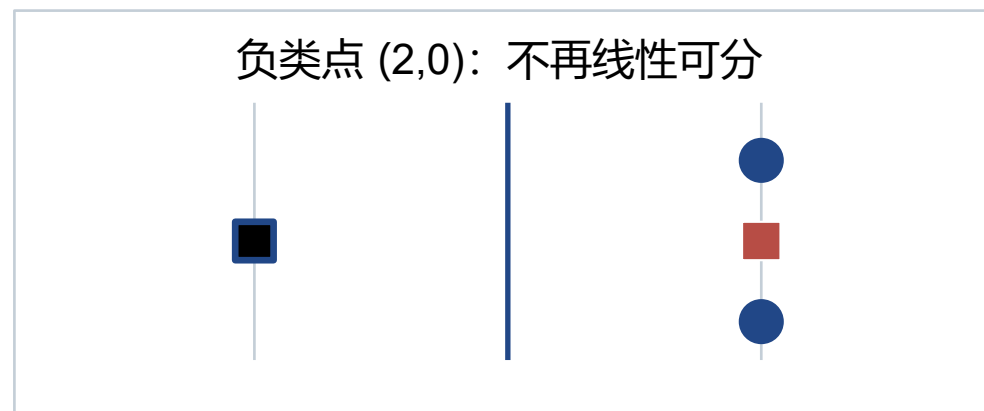
■ 仍然线性可分

- 新增正类点 $(1, 0)$
- 违反当前边界的单位间隔约束
- 重新训练后数据仍然线性可分



■ 不再线性可分

- 新增负类点 $(2, 0)$
- 位于两个正类点之间
- 数据不再线性可分，硬间隔问题不可行



■ 需要的扩展

- 允许并惩罚单位间隔违反

■ 松弛变量

- $\xi_i \geq 0$ 记录第 i 个样本对单位间隔要求的违反量

■ 放宽后的约束

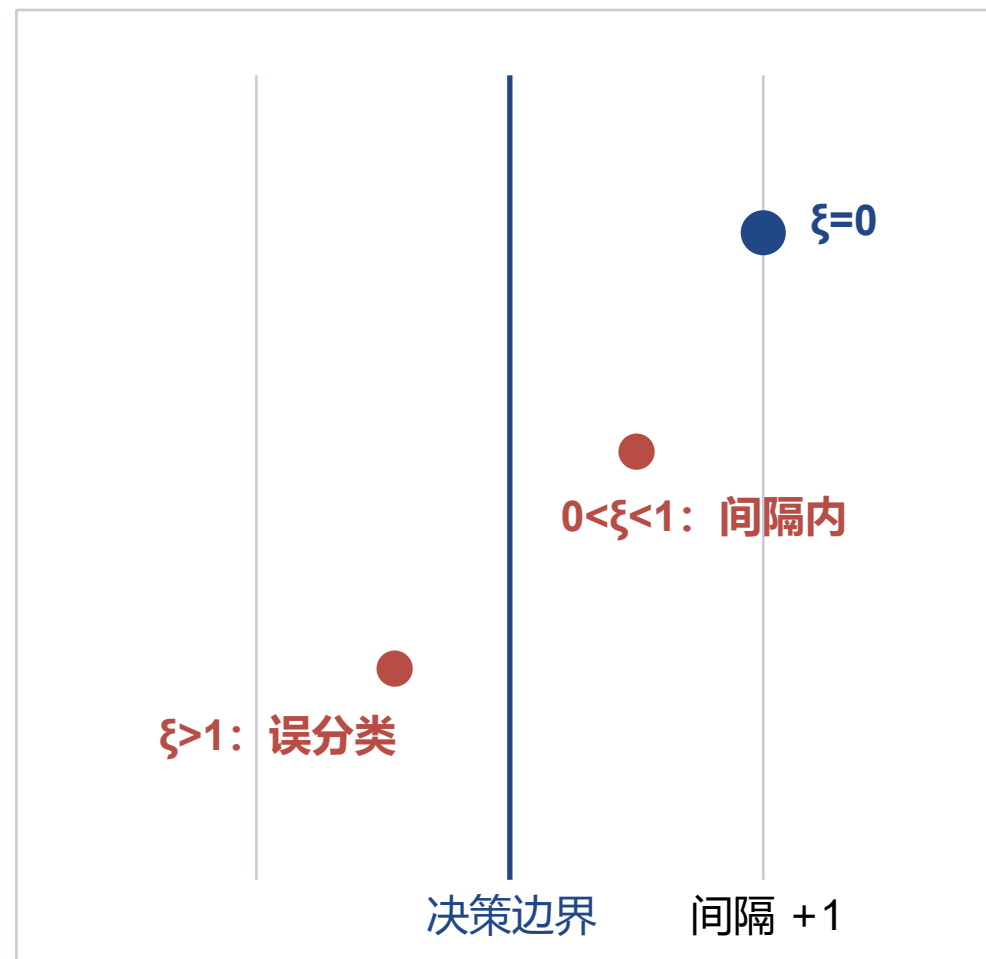
- $y_i(w^T x_i + b) \geq 1 - \xi_i$

■ 约束含义

- 较大的 ξ_i 允许更大的间隔违反

■ 概念边界

- ξ_i 是训练变量，不是误差次数或预测概率



■ 优化问题

- $\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_i \xi_i$
- s.t. $y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$

■ 间隔项

- 较小的 $\|w\|$ 对应较宽的间隔

■ 违反项

- $\sum_i \xi_i$ 汇总训练样本的间隔违反

■ 结论

- 软间隔保留最大间隔原则，同时允许有限违反

■ 较大的 C

- 单位违反付出的代价更高
- 边界通常更重视减少训练违反

■ 较小的 C

- 允许用更多违反换取更宽的间隔
- 部分难分样本对边界的影响减弱

■ 选择条件

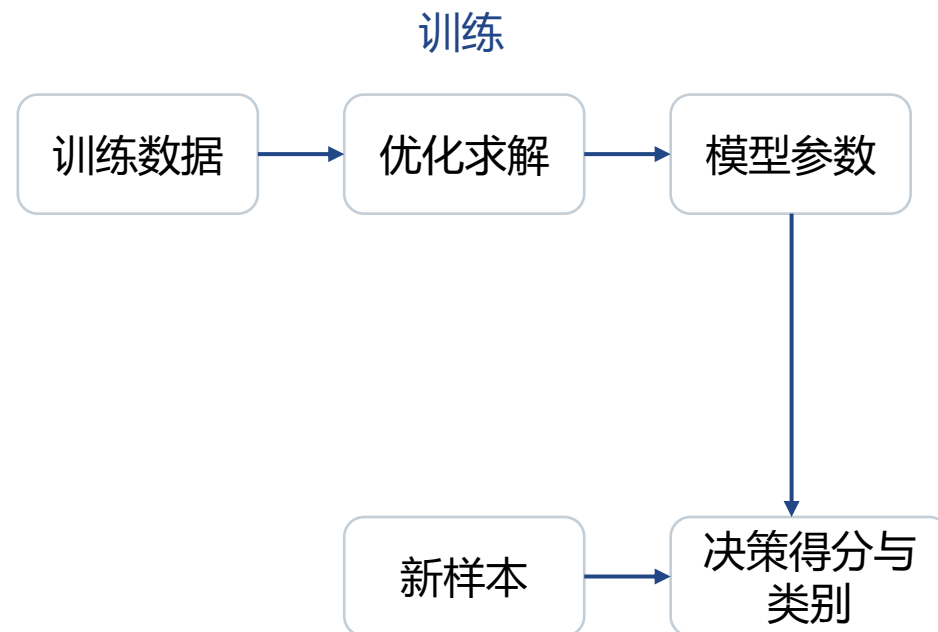
- 结果受数据、特征尺度和核参数共同影响
- 使用验证数据比较候选设置

■ 训练

- 输入带标签样本、特征表示、 C 与核参数，优化器由此求解模型
- 输出偏置、支持向量及其系数；线性情形还可恢复 w

■ 推理

- 输入训练所得模型与新样本，输出决策得分及其符号
- 推理直接使用训练输出，不重新训练模型



推理：直接使用训练输出

3 从表示限制到核机制

原表示不足时，怎样改变表示并只计算模型真正需要的内积

- **正类**

- $(1, 1)$ 与 $(-1, -1)$

- **负类**

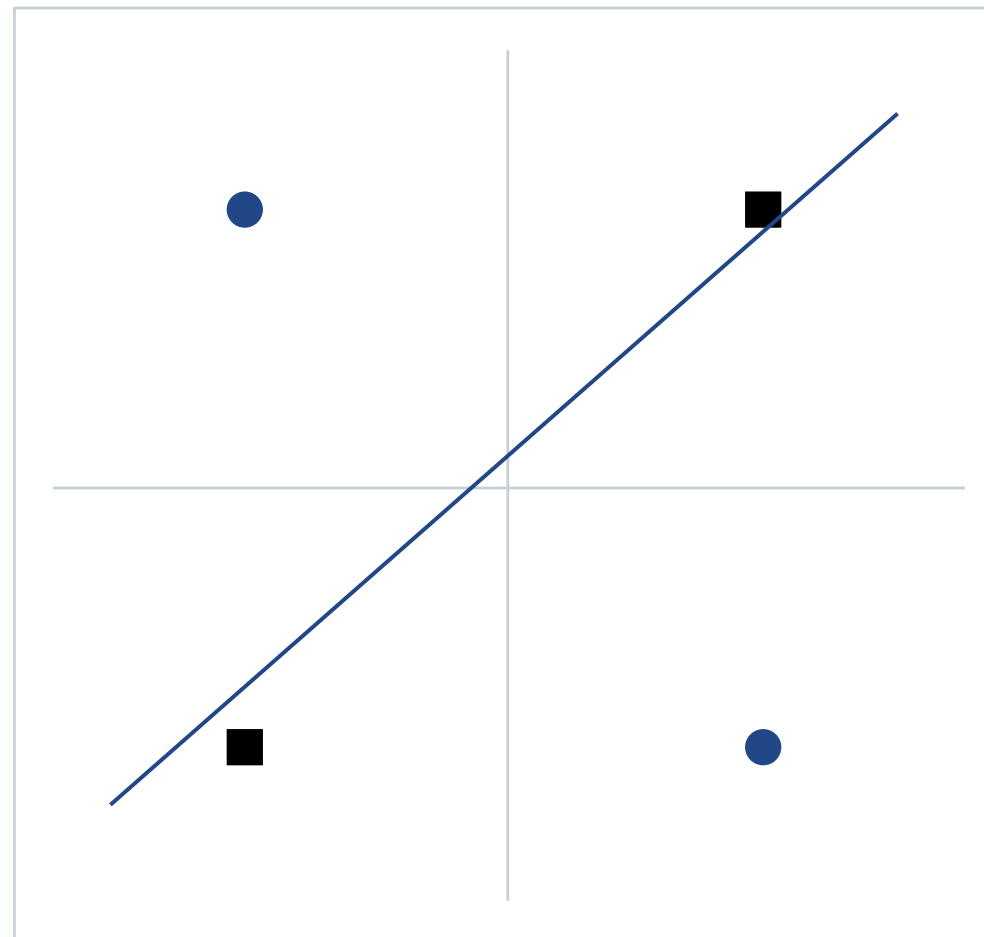
- $(1, -1)$ 与 $(-1, 1)$

- **原空间限制**

- 同类位于对角位置，单一直线无法分开两类

- **问题类型**

- 这里需要改变表示，而不是继续放宽间隔



任一直线都会混合两类

■ 构造新特征

- $u = x_1 x_2$

■ 正类映射

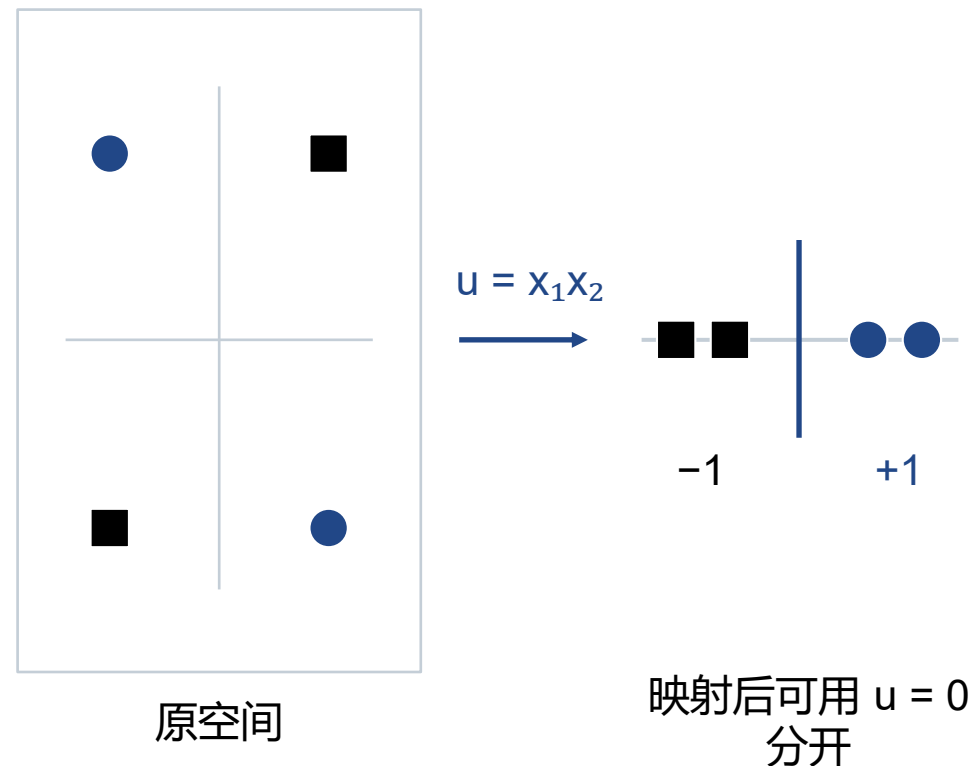
- 同号坐标得到 $u = 1$

■ 负类映射

- 异号坐标得到 $u = -1$

■ 映射后的边界

- 一维特征空间用 $u = 0$ 分开两类



■ 当前问题

- 新表示可能很高维，需要确认训练真正使用哪些特征运算

■ 对偶训练目标

- $\max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$

■ 系数约束

- 硬间隔: $\alpha_i \geq 0$; 软间隔: $0 \leq \alpha_i \leq C$; 两者均满足 $\sum_i \alpha_i y_i = 0$

■ 关键结论与范围

- 训练目标只通过 $\langle x_i, x_j \rangle$ 使用特征; 本讲采用该结果, 不推导完整对偶

■ 支持向量表示

- $w = \sum_i \alpha_i y_i x_i$
- $f(x) = \sum_i \alpha_i y_i \langle x_i, x \rangle + b$

■ 式中对象

- α_i 由训练得到, y_i 给出类别方向
- $\langle x_i, x \rangle$ 是训练样本与新输入的内积

■ 支持向量联系

- KKT 条件把 $\alpha_i > 0$ 与支持向量联系起来
- 推理只需累加非零系数对应的样本
- 固定三点例的 α_i 可取 $1/16, 1/16, 1/8$

■ 特征映射

- $\phi(x)$ 把输入表示到新的特征空间

■ 定义

- $K(x, z) = \langle \phi(x), \phi(z) \rangle$

■ 输入与输出

- 输入是两个原空间对象 x, z
- 输出是它们在特征空间中的内积标量

■ 必要边界

- 核值不是距离、类别概率或最终预测

■ 训练中的替换

- 将 $\langle \phi(x_i), \phi(x_j) \rangle$ 替换为 $K(x_i, x_j)$

■ 特征空间中的对偶预测

- $f(x) = \sum_i \alpha_i y_i \langle \phi(x_i), \phi(x) \rangle + b$

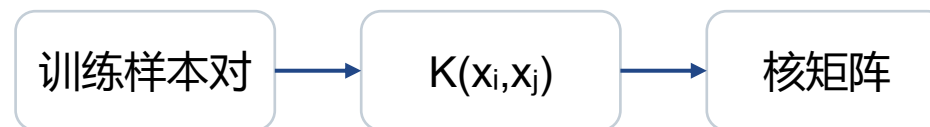
■ 核替换

- $\langle \phi(x_i), \phi(x) \rangle = K(x_i, x)$

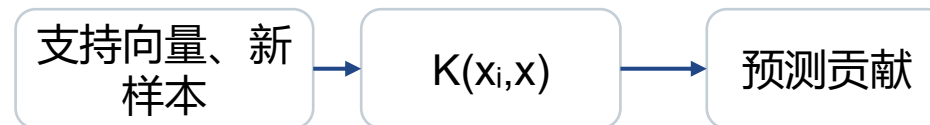
■ 核化后的预测式

- $f(x) = \sum_i \alpha_i y_i K(x_i, x) + b$

训练



预测



两条路径调用同一个特征空间内积

■ 核矩阵

- 对任意有限样本集定义 $G_{ij} = K(x_i, x_j)$

■ 对称性

- $G = G^T$

■ 半正定性

- 对任意向量 c , 都有 $c^T G c \geq 0$

■ 使用边界

- 该条件排除不能对应内积的任意相似度函数
- 采用已知合法核, 不证明 Mercer 定理

4 核计算、选择与验证

先核查核值怎样对应特征空间内积，再用验证证据选择具体方案

■ 核函数

- $K(x, z) = (1 + x^T z)^2$

■ 对应的显式映射

- $\phi(x) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, \sqrt{2}x_1x_2, x_2^2)$

■ 固定输入

- $x = (1, 1), \quad z = (1, 0)$

■ 核查任务

- 分别计算 $K(x, z)$ 与 $\phi(x)^T \phi(z)$

■ 直接计算核函数

- $x^T z = 1$
- $K(x, z) = (1 + 1)^2 = 4$

■ 计算显式映射内积

- $\phi(x) = (1, \sqrt{2}, \sqrt{2}, 1, \sqrt{2}, 1)$
- $\phi(z) = (1, \sqrt{2}, 0, 1, 0, 0)$
- $\phi(x)^T \phi(z) = 1 + 2 + 1 = 4$

■ 核查结论

- 直接核计算返回同一个特征空间内积

- 给定一个支持向量项

- $\alpha_i = 1/2, y_i = +1, K(x_i, x) = 4$

- 计算该项贡献

- $\alpha_i y_i K(x_i, x) = \frac{1}{2} \times 1 \times 4 = 2$

- 完整预测式

- $f(x) = \sum_{i \in SV} \alpha_i y_i K(x_i, x) + b$

- 结果边界

- 数值 2 只是一个支持向量的贡献
 - 最终得分还要累加其他支持向量并加偏置

核	表达式	表示倾向	主要参数
线性核	$x^T z$	原特征的线性关系	通常无核参数
多项式核	$(\gamma x^T z + r)^d$	有限阶特征交互	d, γ, r
RBF 核	$\exp(-\gamma \ x - z\ ^2)$	局部平滑的非线性	γ

■ 线性核

- 原特征已经提供合适表示
- 样本多或特征维度高，需要较低计算成本

■ 多项式核

- 有理由关注有限阶特征交互
- 次数和尺度能够获得任务解释

■ RBF 核

- 需要局部、平滑的非线性相似性
- 没有明确的有限阶交互假设

■ 共同要求

- 同时考虑解释需要、参数敏感性与计算成本

■ 预处理核查

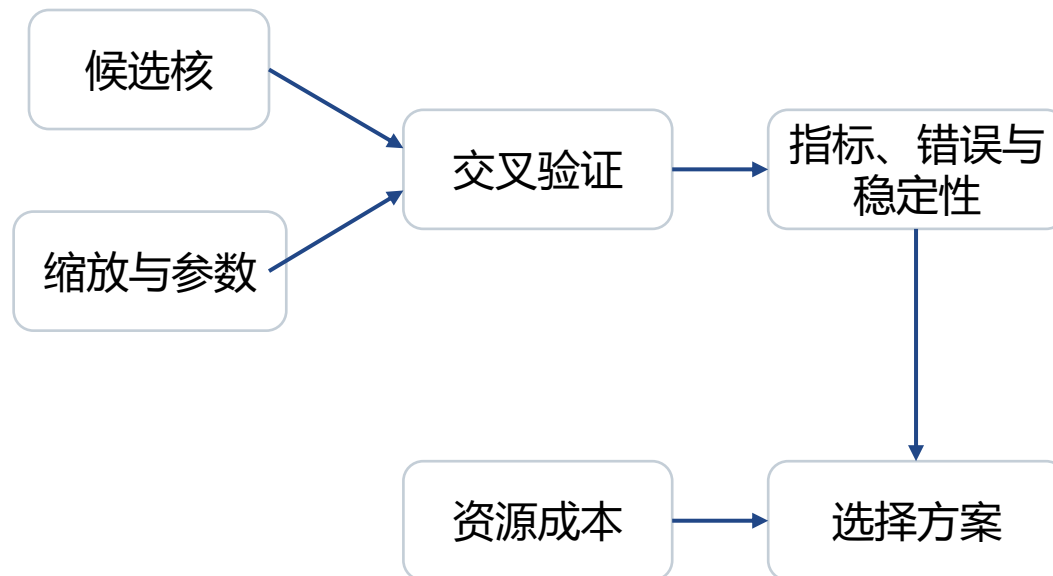
- 对依赖距离的核记录特征缩放方式

■ 候选比较

- 使用独立验证集或交叉验证比较核函数及参数
- C 与核参数需要共同考察

■ 结果检查

- 观察总体指标、重要错误类型和结果稳定性



合法性是前提，验证证据决定选择

数学工具	本讲用它得到什么	继续学习位置
约束优化	把最大间隔写成目标与约束	CS229 Section 6; Cornell Lecture 13
拉格朗日对偶	得到只依赖样本内积的预测表示	CS229 Section 6; Cornell Kernels continued
KKT 条件	连接非零对偶系数与支持向量	CS229 Section 6
半正定矩阵	给出合法核的核矩阵边界	CS229 Section 5.4; Cornell Kernels

- **硬间隔线性 SVM：选择可分边界**

- 原特征中要求所有样本满足单位间隔；推理使用 $f(x) = w^T x + b$

- **软间隔线性 SVM：容纳间隔违反**

- 原特征中由 ξ_i 与 C 管理违反；推理仍使用 $f(x) = w^T x + b$

- **核 SVM：改变模型使用的表示空间**

- 核定义特征空间，通常仍采用软间隔；推理使用 $f(x) = \sum_i \alpha_i y_i K(x_i, x) + b$

- **共同框架**

- 三者都在模型实际使用的表示空间中选择最大间隔边界

■ 选择

- 多个正确边界仍需比较最近样本的几何间隔

■ 建模

- 规范化消除尺度歧义，形成目标函数与单位间隔约束

■ 容错

- 松弛变量放宽约束， C 为间隔违反定价

■ 表示

- 原表示不足时改变特征空间，用核函数计算所需内积

■ 验证

- 合法性只给出可用边界；具体方案仍需验证证据与成本支持