07 支持向量机

最大间隔、软间隔与核方法

学习目标

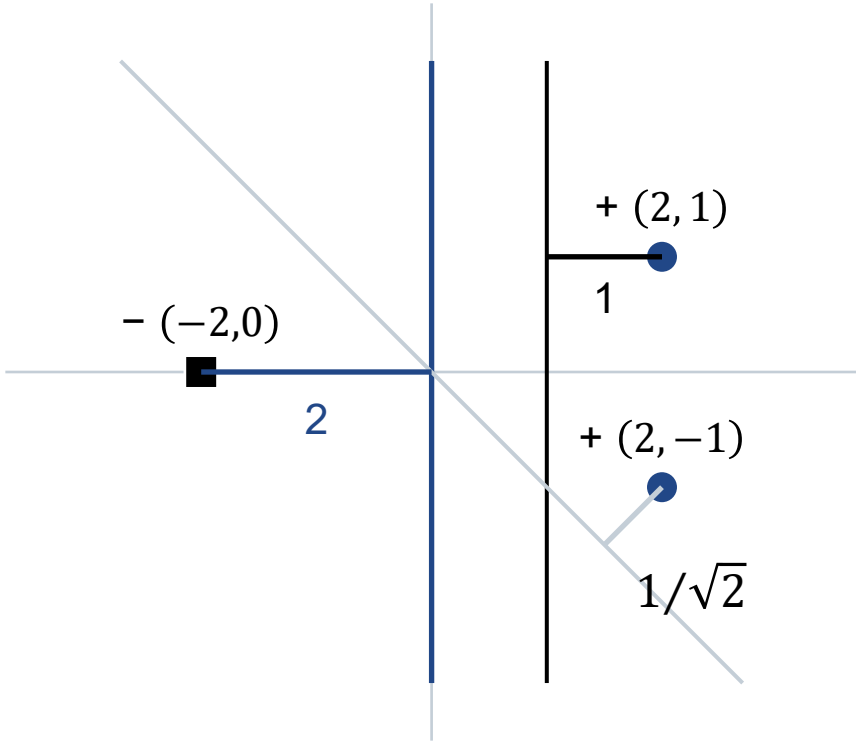
- 建立硬间隔模型
  - 解释最大间隔原则，区分函数间隔与几何间隔，说明目标与约束的形成
  - 完成线性可分例子的距离、间隔、支持向量和分类得分计算
- 理解软间隔与核扩展
  - 解释松弛变量与 C 的职责和边界
  - 解释特征映射、对偶内积与核替换，完成核值及预测贡献核查
- 比较并验证核方案
  - 结合表示假设、参数、缩放、验证证据和资源成本，初步选择与检查核方案

1 从正确分类到最大间隔

最大间隔原则怎样形成模型，并用于训练与预测

最大间隔提供边界选择依据

- **正确分类仍需选择边界**
  - 线性可分时，感知机可找到分隔面；多个正确边界无法用训练正确率区分
- **最大间隔原则**
  - 固定边界时，输入扰动的欧氏长度小于样本距离，分类符号保持不变
  - 最近样本限制共同的扰动余量；平均距离不能反映最薄弱位置
  - 选择最小余量最大的边界；不构成未知数据的性能保证



$x_1 = 0$: 2; $x_1 = 1$: 1

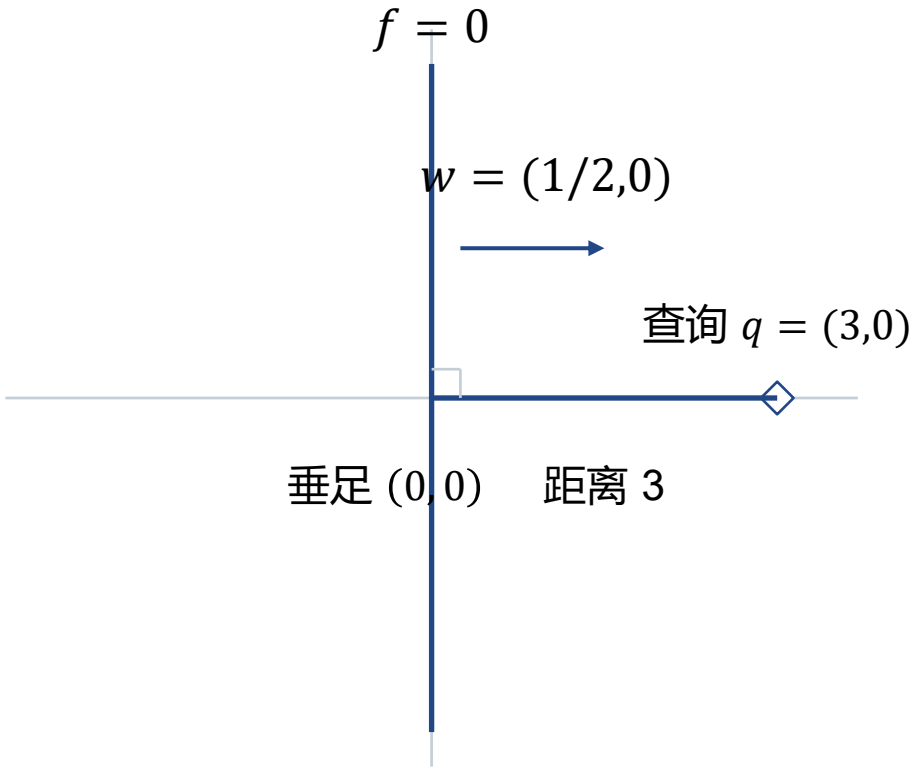
$x_1 + x_2 = 0$: $1/\sqrt{2}$

线性得分与超平面距离

线性得分表示边界，距离用于比较。

线性得分、边界与分类

- $f(x) = w^T x + b, H: f(x) = 0$
- x 为输入特征；非零法向量 w 决定方向，偏置 b 决定位置
- 标签 $y_i \in \{-1, +1\}$，得分正负对应边界两侧
- 正确分类时 $y_i f(x_i) > 0$；零得分位于边界，得分不是概率

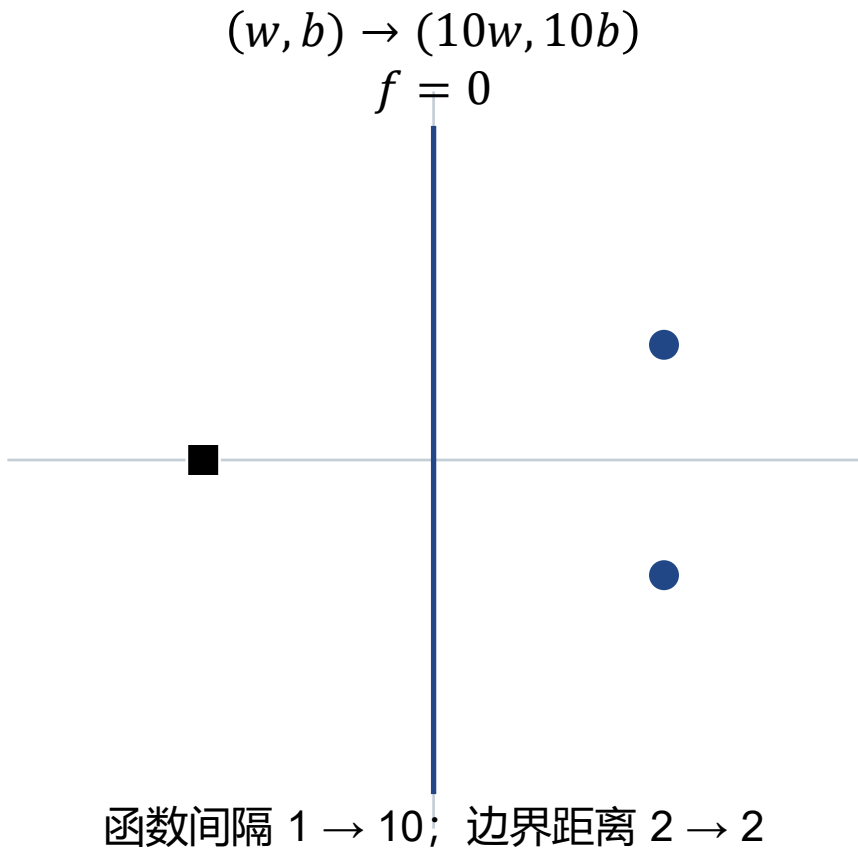


点到超平面的距离

- $d(x, H) = \frac{|w^T x + b|}{\|w\|}$
- 垂直距离沿法向量测量；分母消除参数尺度影响

函数间隔与参数尺度

- **函数间隔统一分类正误**
  - $\hat{\gamma}_i = y_i(w^T x_i + b)$
  - 正值表示分对，负值表示分错，零值表示在边界上
- **同一边界可以有不同的得分余量**
  - 将 w, b 同乘 $c > 0$，边界与分类符号不变
  - 函数间隔变为 c 倍，点到边界的距离保持不变
- **选择边界须消除尺度影响**
  - 任意放大参数即可增大函数间隔，不能据此判断边界更合理



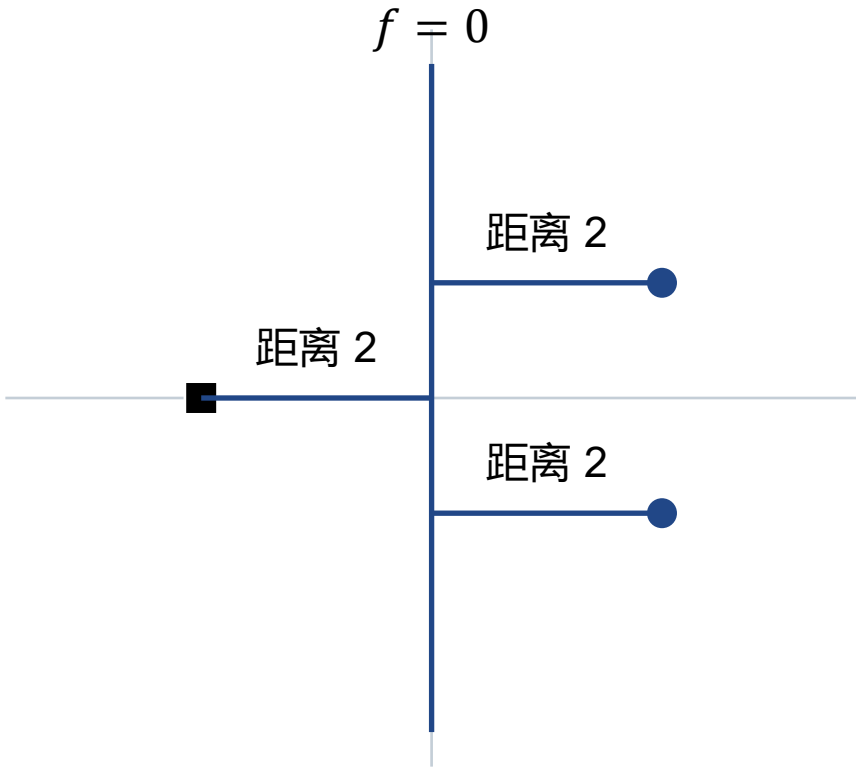
样本与决策边界均未移动

几何间隔消除尺度影响

函数间隔的尺度影响可由法向量范数消除，得到不随参数正比例缩放而改变的几何间隔。

从函数间隔到几何间隔

- 函数间隔 $\hat{\gamma}_i = y_i(w^T x_i + b)$ 包含参数尺度
- $\gamma_i = \frac{y_i(w^T x_i + b)}{\|w\|}$
- 几何间隔是按标签方向取符号的距离：正确分类为正，错误分类为负
- 正确分类时，几何间隔等于样本到超平面的距离



$\|w\| = 1/2$，几何间隔均为 2

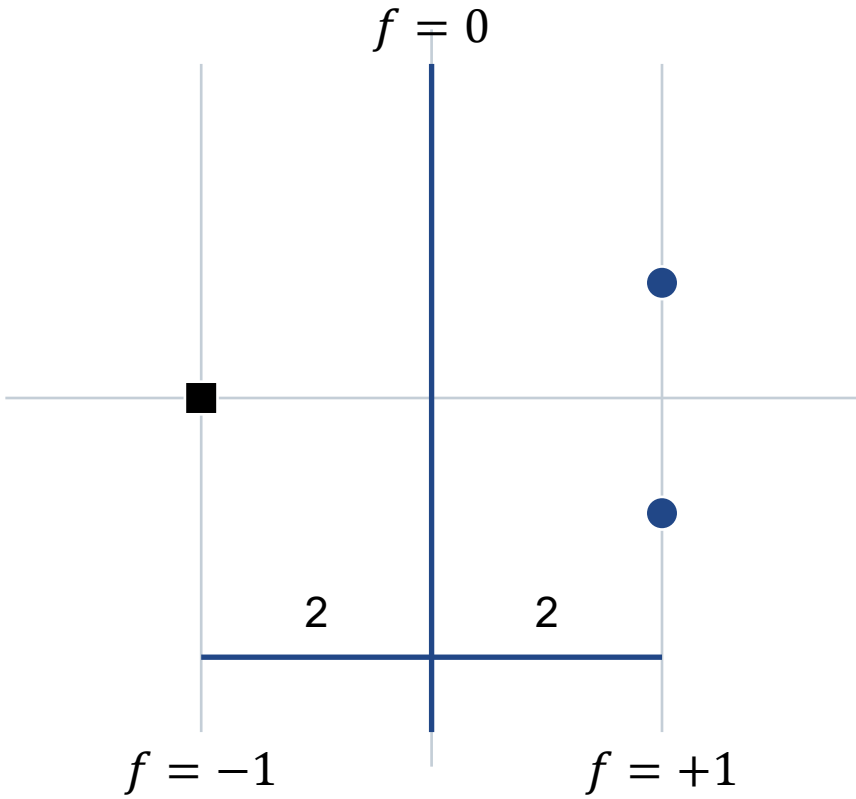
训练集的几何间隔

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

单位函数间隔实现参数规范化

固定最小函数间隔为 1，可为同一边界选定参数尺度，并将几何间隔直接写成参数范数的函数。

- **规范化依据**
  - 严格分开的有限训练集满足 $m = min_i y_i(w^T x_i + b) > 0$
  - 将 w, b 同除以 m，最小函数间隔变为 1；边界与几何间隔保持不变
- **单位间隔约束**
  - $y_i(w^T x_i + b) \geq 1$
  - 最近样本取等号，两条间隔边界为 $w^T x + b = \pm 1$
- 单侧间隔与总宽度
  - 单侧间隔为 $1/\|w\|$，总宽度为 $2/\|w\|$



单侧间隔 2；总宽度 4

最大间隔转化为约束优化

规范化后，最大化间隔宽度 $2/\|w\|$，等价于在单位间隔约束下最小化 $\frac{1}{2}\|w\|^2$。

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

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

- **优化对象**
  - 从可行的 w, b 中选择使目标最小的一组参数
- **目标与约束的职责**
  - 目标减小参数范数，扩大几何间隔
  - 约束要求每个样本分类正确，并位于间隔边界上或外侧
- **适用条件**
  - 训练数据线性可分

硬间隔模型的训练与预测

- **准备输入**
  - 固定特征表示，给出带标签的线性可分训练集
  - 若使用特征缩放，保存训练时得到的缩放规则
- **求解模型**
  - 目标为凸二次函数、约束为线性不等式，可交给二次规划求解器
  - 求得满足约束且使目标最小的 w, b
- **检查与输出**
  - 核查可行性与求解器的最优性结果，保存 w, b
- **新样本预测**
  - 使用同一特征处理，再计算 $f(x) = w^T x + b$ 并取符号

三点训练例的目标与逐点约束

正类在 $x_1 = 2$，负类在 $x_1 = -2$。边界居中且与两侧平行，预期可平衡最近距离；下面用约束与下界检验这一几何判断。

优化对象

- $w = (w_1, w_2)$，目标为最小化 $\frac{1}{2}(w_1^2 + w_2^2)$，同时求偏置 b

逐点产生的约束

- 正类 $(2, 1)$：$2w_1 + w_2 + b \geq 1$
- 正类 $(2, -1)$：$2w_1 - w_2 + b \geq 1$
- 负类 $(-2, 0)$：$2w_1 - b \geq 1$

可行性的含义

- 三条约束均由 $y_i(w^T x_i + b) \geq 1$ 代入得到，须同时满足
- 同时满足约束只说明参数可行，尚未确定目标是否最小

可行性与目标下界共同确定最优解

可行参数未必最优。由约束推出目标下界，再找到达到下界的可行参数，就能确定最优解。

- **合并正类样本的两条约束**
  - $2w_1 + b \geq 1 + |w_2|$
  - 同时满足两条约束，等价于满足 $2w_1 + b \geq max(1 - w_2, 1 + w_2)$
- **合并正负约束**
  - 与负类约束 $2w_1 - b \geq 1$ 相加，得到 $4w_1 \geq 2 + |w_2|$
  - 因此 $w_1 \geq 1/2 + |w_2|/4 \geq 1/2$
- **下界与可行解共同确定最优**
  - 任意可行解有 $\frac{1}{2}(w_1^2 + w_2^2) \geq \frac{1}{2}w_1^2 \geq 1/8$
  - 取 $w_2 = 0, w_1 = 1/2$，正类约束要求 $b \geq 0$，负类约束要求 $b \leq 0$，因此 $b = 0$
  - 该参数满足全部约束，且 $\frac{1}{2}\|w\|^2 = 1/8$，达到下界，因此最优

最优参数确定间隔与预测

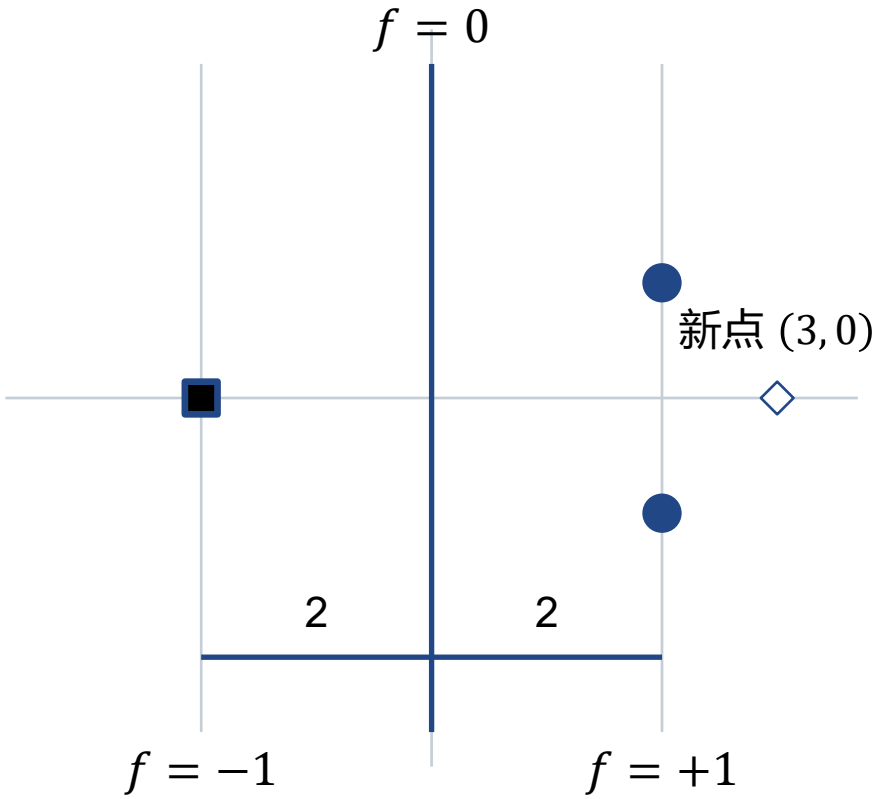
最优参数 $w = (1/2, 0), b = 0$ 同时确定模型的几何性质与新样本预测。

间隔与支持向量

- **支持向量**是训练后参与决定边界的关键样本
- 本例三点均满足 $y_i f(x_i) = 1$，位于间隔边界且均为支持向量
- 一般识别不能仅依据间隔边界位置
- $\|w\| = 1/2$，单侧间隔为 2，总宽度为 4；间隔边界 $(1/2)x_1 = \pm 1$，即 $x_1 = \pm 2$

固定模型的新样本预测

- 对未参与训练的 $x = (3, 0)$
- $f(x) = w^T x + b = (1/2) \times 3 + 0 = 3/2$
- 得分为正，预测为正类；得分不是类别概率



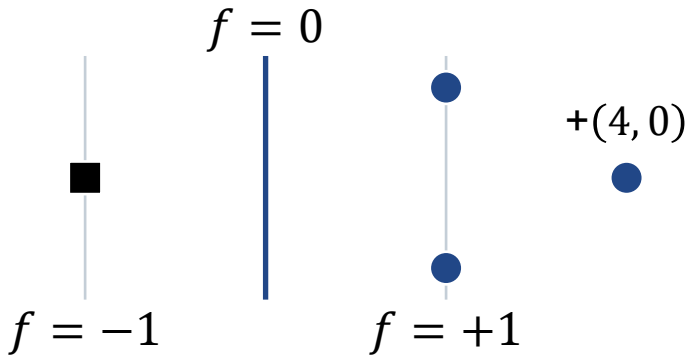
单侧间隔 2；总宽度 4

远点与支持点变化对原解的影响

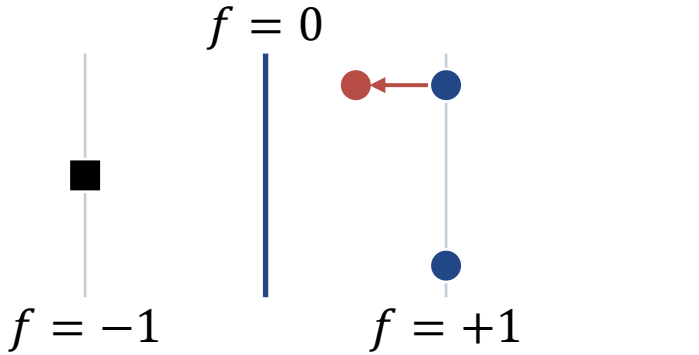
训练样本变化后，原解可能失效。下面分别增加远点与移动支持点。

- **增加远点：原解保持**
  - 独立加入正类 $(4,0)$，原解满足新增约束
  - 新约束只缩小可行域，原最优解仍可行，因此仍然最优
- **移动支持点：原解失效**
  - 独立把 $(2,1)$ 移到 $(1,1)$，原解违反约束，须重新训练
- **原解失效与问题不可行**
  - 原解失效不等于训练集不可分
  - 仍可分时可以重新求解硬间隔；不可分时，硬间隔问题才不可行

独立扩展一：增加远点，原最优解保持



独立扩展二：只移动一个正类支持点



$(2,1) \rightarrow (1,1)$：原解违反约束，须重求

2 从硬约束到软间隔

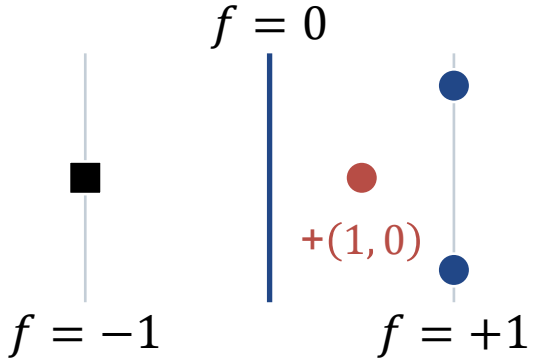
松弛变量与违反惩罚怎样放宽严格间隔要求

严格间隔约束的适用局限

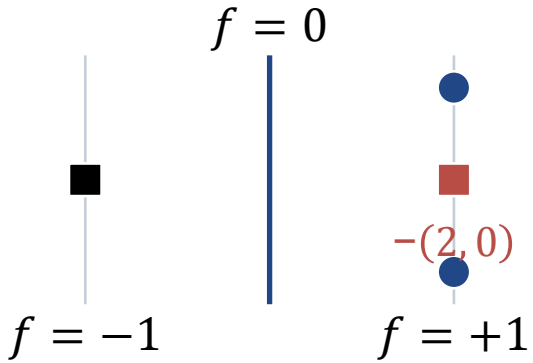
原解失效时，仍可分的数据可以重新训练硬间隔；
不可分的数据则无法满足全部严格约束。

- **仍可分，但原解失效**
  - 独立新增正类点 $(1,0)$，违反原模型的单位间隔约束
  - 数据仍可分，可以重新训练硬间隔
- **不可分，硬间隔不可行**
  - 独立新增负类点 $(2,0)$，位于两正类点的中点
  - 无法线性分开两类，硬间隔约束整体不可行
- 软间隔的作用
  - 不可分数据以及噪声或异常样本，暴露严格约束的局限
  - 软间隔允许单位间隔违反，并对违反施加惩罚

独立扩展一：违反原单位间隔，仍然可分



独立扩展二：负类处于两个正类的中点



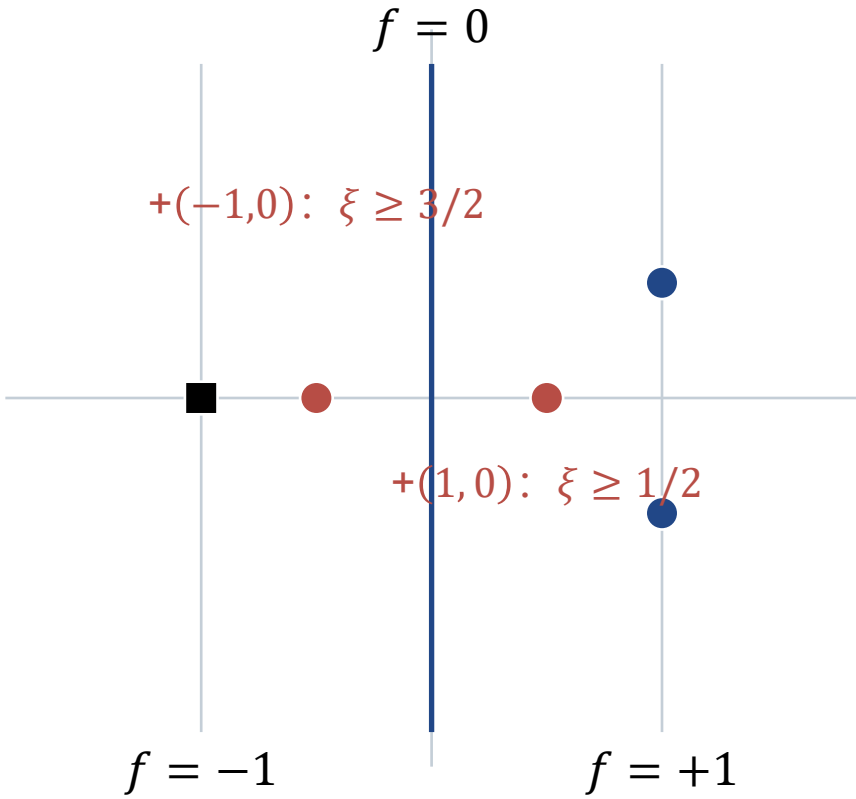
第二组数据无法被线性分隔

松弛变量放宽单位间隔约束

非负松弛变量表示各样本的单位间隔约束放宽量。

- **松弛变量与软约束**
  - **松弛变量** $\xi_i \geq 0$ 是第 i 个样本的约束放宽量
  - $y_i(w^T x_i + b) \geq 1 - \xi_i$
- 固定模型下的松弛量
  - 松弛量至少弥补 $1 - y_i f(x_i)$ 的正缺口；较大松弛允许较大违反
- 松弛变量的训练职责
  - ξ_i 与模型参数共同参与训练，由目标和约束决定
  - 松弛量不是误差次数或预测概率

图示是固定参照模型，不是新的最优解。



参照 $f(x) = x_1/2$，不是新训练解

软间隔目标权衡间隔与违反

松弛允许违反，惩罚控制违反量；C 调节惩罚与间隔项的相对权重。

$$min_{w,b,\xi}\ \frac{1}{2}\|w\|^2 + C\sum_i \xi_i$$

$$\text{s.t.} y_i f(x_i) \geq 1 - \xi_i, \xi_i \geq 0$$

- **消去松弛变量**
  - 固定 $C > 0$ 与 w, b，惩罚随 ξ_i 增大，最优取 $\xi_i^* = max(0, 1 - y_i f(x_i))$，代回目标

$$min_{w,b}\ \frac{1}{2}\|w\|^2 + C\sum_i max(0, 1 - y_i f(x_i))$$

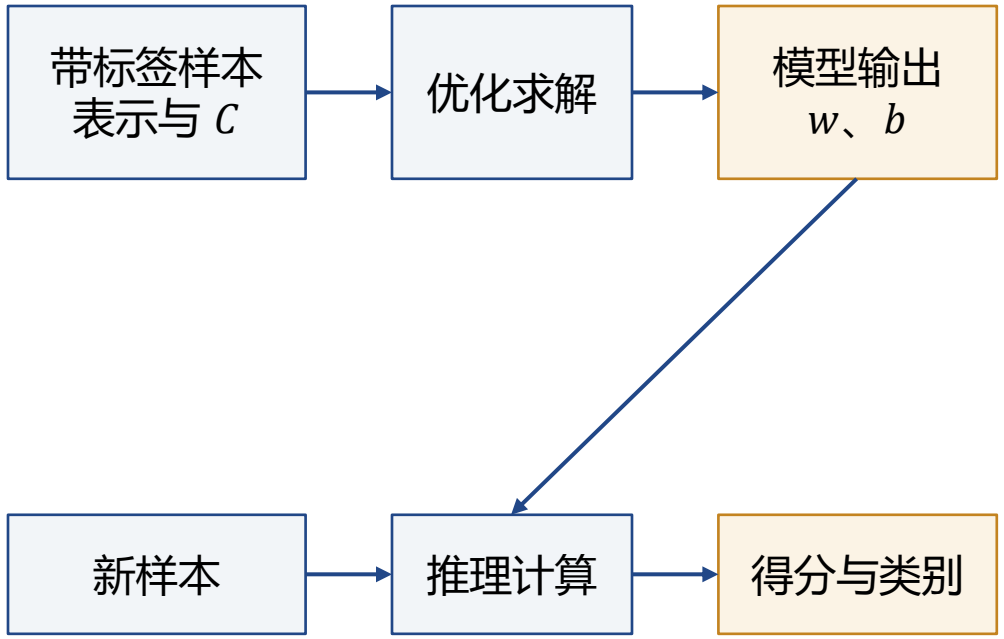
- **目标与 C 的职责**
  - 范数正则项中，较小范数对应较宽间隔；合页损失（hinge loss）汇总间隔违反
  - 固定数据与表示，较大 C 提高单位违反的代价，通常更重视减少违反
  - 较小 C 允许更多违反换取更宽间隔，部分难分样本对边界的影响减弱
  - 结果受数据与特征尺度影响，C 没有固定优势；用验证数据比较设置

软间隔模型的训练与预测

预测使用固定训练输出，无需重新优化松弛变量或调整 C。

- **训练**
  - 输入带标签样本、固定表示与 C，优化 w, b 和松弛变量
  - 输出预测模型 w, b；松弛变量用于训练约束与惩罚
- **推理**
  - 新样本与训练模型共同确定得分 $f(x) = w^T x + b$，符号给出类别
  - 使用固定训练输出，不重新求解模型

软间隔放宽约束并惩罚违反，原特征空间中的决策边界仍为线性边界。



训练输出和新样本共同进入推理

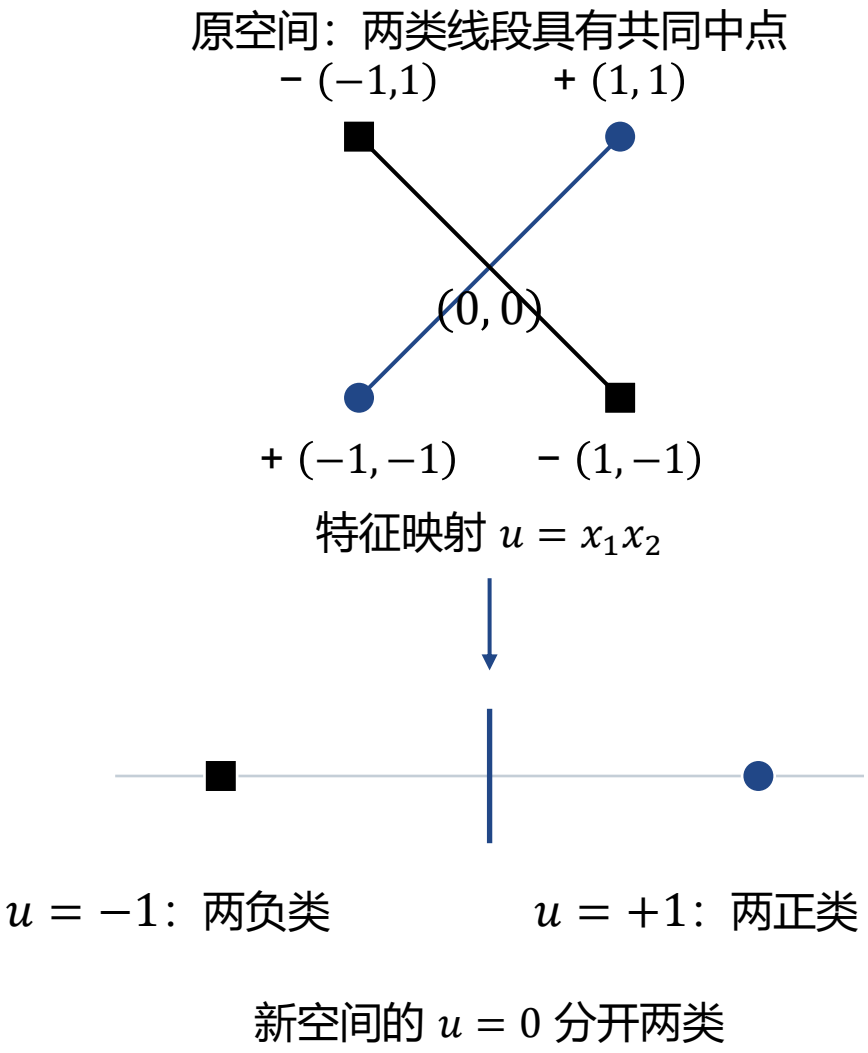
3 从表示限制到核机制

特征空间内积怎样连接表示改变与核化训练、预测

特征映射使 XOR 在新空间线性可分

软间隔仍采用原表示的线性边界。XOR 的类别关系需要改变特征表示才能线性分开。

- **原空间的限制**
  - 正类 $(1,1),(-1,-1)$，负类 $(1,-1),(-1,1)$；同类位于相对角点
  - 两类连线具有共同中点 $(0,0)$，无法用一条直线严格分开
- **特征映射**
  - **特征映射** ϕ 将输入转换为新的特征表示；对这四点取 $u = x_1 x_2$
  - 正类得到 $u = 1$，负类得到 $u = -1$，新空间中的 $u = 0$ 分开两类



对偶形式以样本系数与内积表示模型

高维映射可能昂贵。要省去显式坐标，先用对偶形式检查训练是否只需内积。

每个样本约束引入非负拉格朗日乘子 α_i，构造拉格朗日函数

$$L(w,b,\alpha)=\frac{1}{2}\|w\|^2+\sum_i \alpha_i\,[1-y_i(w^T x_i+b)]$$

可行点使约束项非正；固定乘子最小化L 给出原问题的下界。驻点把法向量写成样本贡献之和

$$\nabla_w L=w-\sum_i \alpha_i\, y_i x_i=0 \Longrightarrow w=\sum_i \alpha_i\, y_i x_i$$

另有 $\partial_b L=-\sum_i \alpha_i\, y_i=0$。驻点关系代回 L 产生内积项，再最大化该下界

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

- **系数与内积接口**
  - 硬间隔：$\alpha_i\geq 0$；软间隔：$0\leq\alpha_i\leq C$；两者均满足 $\sum_i \alpha_i\, y_i=0$
  - 训练目标只通过样本内积使用特征；共同映射下改用 $\langle\phi(x_i),\phi(x_j)\rangle$

支持向量与内积预测得分

将样本加权表示代入得分式，预测也只需内积值。训练与预测都不必读取映射坐标。

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

- **得分组成**
  - 训练系数 α_i、类别标签 y_i 与内积 $\langle x_i, x \rangle$ 形成单项贡献，最后加偏置
  - x_i 为训练样本，x 为新输入
- **支持向量联系**
  - 对偶系数 $\alpha_i > 0$ 的训练样本称为支持向量
  - 推理只需累加非零系数对应的样本；位于间隔边界不保证系数非零
  - 换用特征映射后，预测中的内积相应变为 $\langle \phi(x_i), \phi(x) \rangle$
- **三点训练例的系数核查**
  - 按两个正类、一个负类的顺序，$\alpha = (1/16, 1/16, 1/8)$，满足 $\sum_i \alpha_i \, y_i = 0$ 且恢复 $w = (1/2, 0)$

核函数替代特征空间内积计算

若能从原输入直接得到同一内积，就能省去显式构造映射坐标；核函数实现这一替换。

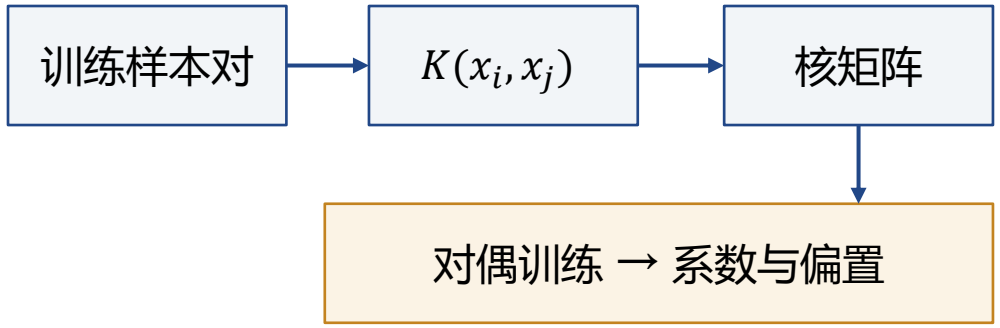
核函数的定义

- 存在将所有输入映至同一内积空间的共同映射 ϕ，使任意 x, z 满足
- $K(x,z) = \langle \phi(x), \phi(z) \rangle$
- 输入为原空间对象，输出为映射后的内积；核值不是距离、概率或预测结果

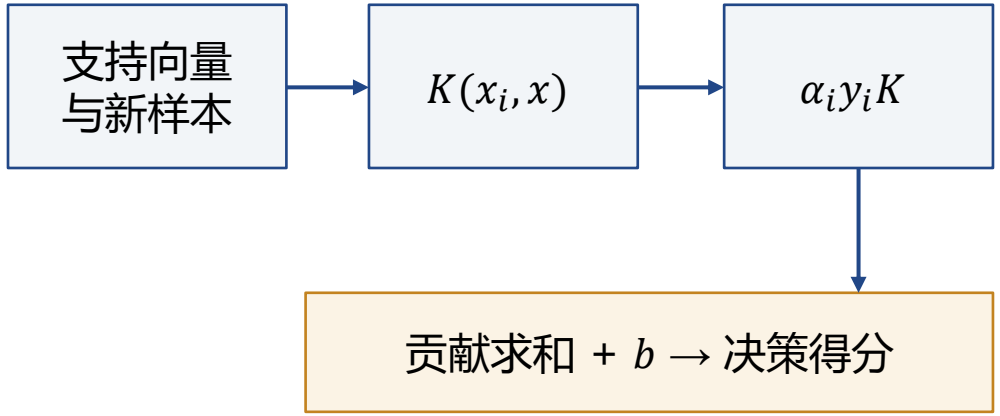
训练与预测中的替换

- 训练以 $K(x_i, x_j)$ 替代映射内积，汇集为核矩阵
- 预测使用同一映射，两种表达等价
- $f(x) = \sum_i \alpha_i y_i \langle \phi(x_i), \phi(x) \rangle + b$
- $f(x) = \sum_i \alpha_i y_i K(x_i, x) + b$

训练：核矩阵进入对偶求解



预测：核值形成贡献，再累加偏置



核矩阵条件界定核的合法性

内积替换要求核对应共同特征空间。任意有限样本的核矩阵都需满足对称与半正定条件。

核矩阵及其条件

- $G_{ij} = K(x_i, x_j)$ 汇集样本两两核值
- 对称：$G = G^T$
- 半正定：对任意与样本数同维的实向量 c，都有 $c^T Gc \geq 0$

核合法性的边界

- 条件必须对任意有限样本集成立，任意相似度函数不能直接作为核
- 单个数据集通过矩阵检查，不能证明函数的一般合法性

核化训练确定系数与预测模型

选定合法核后，核矩阵进入软间隔对偶训练，得到系数与偏置以及预测所需模型。

训练输入

- 带标签样本、合法核 K、核参数与惩罚系数 $C > 0$

$$max_{\alpha} \sum_{i} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

$$0 \le \alpha_i \le C, \sum_{i} \alpha_i y_i = 0$$

模型输出与预测

- 二次规划求解系数，再确定偏置；保存支持向量、标签、非零系数、偏置及核设置
- $f(x) = \sum_{i:\alpha_i>0} \alpha_i y_i K(x_i, x) + b$，由符号给出类别

核化与软间隔的关系

- 核化保持特征空间中的线性模型，可在原空间表达非线性关系
- 核化改变表示与内积计算；软间隔允许并惩罚单位间隔约束的违反，两者可结合

KKT 连接对偶系数、约束与偏置

求得系数后，还需从原约束确定偏置。KKT（Karush–Kuhn–Tucker）最优性条件连接对偶系数与约束状态，在线性与核化 SVM 中承担同一作用。

$$0 < \alpha_s < C \Longrightarrow \xi_s = 0, y_s f(x_s) = 1$$

$$b = y_s - \sum_i \alpha_i \, y_i K(x_i, x_s)$$

- **偏置的条件与结果**
  - s 是满足该条件的训练样本下标，ξ_s 是其松弛变量
  - 该样本的单位间隔取等号，代入预测式即可解出偏置
- KKT 互补关系
  - 两条互补关系连接训练系数与原约束状态
  - $\alpha_i[y_i f(x_i) - 1 + \xi_i] = 0, (C - \alpha_i)\xi_i = 0$
- 偏置公式的适用限制
  - 若没有 $0 < \alpha_s < C$ 的样本，按最优性条件确定允许的偏置；不能直接套用上式

数学工具支撑 SVM 的模型与计算

SVM 的模型与计算依赖不同数学工具：约束优化建立学习目标，对偶形式提供内积表达，KKT 条件连接最优解与约束，半正定条件界定核合法性。

数学工具	在 SVM 中的作用	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	把训练和预测都写成样本内积形式	CS229 Section 6；Cornell Kernels continued
KKT 条件	由互补关系识别关键约束，并确定偏置	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

4 核计算、选择与验证

核查计算接口，并用验证证据与资源成本支持方案选择

核计算与显式映射得到相同内积

二次多项式核提供一个内积核查实例：比较同一输入的直接核计算与显式映射内积。

- **核函数与共同映射**
  - $K(x,z) = (1 + x^T z)^2$
  - $\phi(x) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, \sqrt{2}x_1x_2, x_2^2)$
  - 取 $x = (1,1), z = (1,0)$
- **两条计算路径**
  - 直接核计算：$x^T z = 1$，$K(x,z) = (1+1)^2 = 4$
  - 显式映射：$\phi(x) = (1, \sqrt{2}, \sqrt{2}, 1, \sqrt{2}, 1)$，$\phi(z) = (1, \sqrt{2}, 0,1,0,0)$
  - 映射内积：$\phi(x)^T \phi(z) = 1 + 2 + 1 = 4$
- 两条计算路径得到相同内积
  - 直接核计算返回同一个特征空间内积；这组数值核查不代替一般恒等式证明

支持向量加权核值形成预测贡献

核值乘以系数和标签形成单项预测贡献；完整得分还需累加其他贡献与偏置。

- 给定的支持向量项
  - $\alpha_i = 1/2,\ y_i = +1,\ K(x_i, x) = 4$
- 单项预测贡献
  - $\alpha_i y_i K(x_i, x) = \frac{1}{2} \times 1 \times 4 = 2$
- 完整预测得分
  - $f(x) = \sum_{i \in SV} \alpha_i\, y_i K(x_i, x) + b$
  - SV 是非零系数对应的支持向量下标集合
  - 最终得分还要累加其他支持向量并加偏置

数值 2 只是给定支持向量项的贡献。系数来自单项演算设定，与三点线性训练解不同；本例未完成完整核 SVM 训练。

表示假设提供核方案的选择依据

核	表达式	表示与候选依据	主要参数
线性核	$x^T z$	原表示适合线性关系；样本多或维度高时关注计算成本	通常无核参数
多项式核	$(\gamma x^T z + r)^d$	有限阶特征交互；次数与尺度具有任务解释	d, γ, r
RBF 核	$\exp(-\gamma \|x - z\|^2)$	局部平滑非线性；缺少明确有限阶交互假设	γ

常用合法参数约定：多项式次数 d 为正整数，$\gamma \geq 0, r \geq 0$；RBF 采用 $\gamma > 0$。这里的 γ 是核参数，与几何间隔同名但用途不同。

选择候选核时，同时考虑解释需要、参数敏感性与计算成本。核族没有固定优劣顺序，表示依据用于形成候选，不能替代验证。

核方案的验证与计算成本

表示假设形成候选后，共同比较核、核参数和 C，并检查数据处理及资源成本。

- **预处理核查**
  - 对依赖距离的核记录特征缩放方式
  - 缩放参数只用各训练折拟合，再应用于对应验证折
- **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
  - 最终测试数据不参与核与参数选择
- **结果检查**
  - 观察总体指标、重要错误类型和结果稳定性
- **训练与存储成本**
  - 同时检查训练时间与核矩阵存储成本；若显式保存完整核矩阵，条目数随样本数平方增长

SVM 的表示空间与间隔约束

常用形式	表示空间	间隔约束	预测得分
硬间隔线性 SVM	原特征空间	线性可分；所有样本满足单位间隔	$w^T x + b$
软间隔线性 SVM	原特征空间	松弛允许违反，C 权衡惩罚	$w^T x + b$
核 SVM	共同特征空间，以核计算内积	通常采用软间隔，仍由 C 权衡	$\sum_i \alpha_i y_i K(x_i, x) + b$

这些形式都在所选表示与约束下建立最大间隔模型；得分符号给出类别，非零系数样本参与预测。

软间隔改变约束处理，核化改变表示与内积计算；这两个扩展维度可以结合。

本讲总结

- **边界选择形成硬间隔模型**
  - 多个正确边界用最小几何间隔比较；几何间隔消除函数间隔的尺度影响
  - 将最小函数间隔规范化为 1，形成目标函数与单位间隔约束
- **训练结果通过支持向量用于预测**
  - 对偶系数把模型写成样本加权表示，非零系数对应支持向量
  - 固定训练输出后，累加样本贡献与偏置形成得分；得分符号给出类别，得分不是类别概率
- **不同限制对应不同扩展**
  - 松弛变量与 C 管理约束违反；原表示不足时，可用特征映射与核内积改变模型使用的表示
  - 对偶形式使训练与预测依赖内积，核函数据此替代显式映射后的内积计算
- **方案选择需要验证证据**
  - 结合表示假设、特征缩放、C 与核参数比较候选，同时检查训练与存储成本
  - 合法核提供使用前提，实际效果仍需验证；最大间隔不构成无条件泛化保证