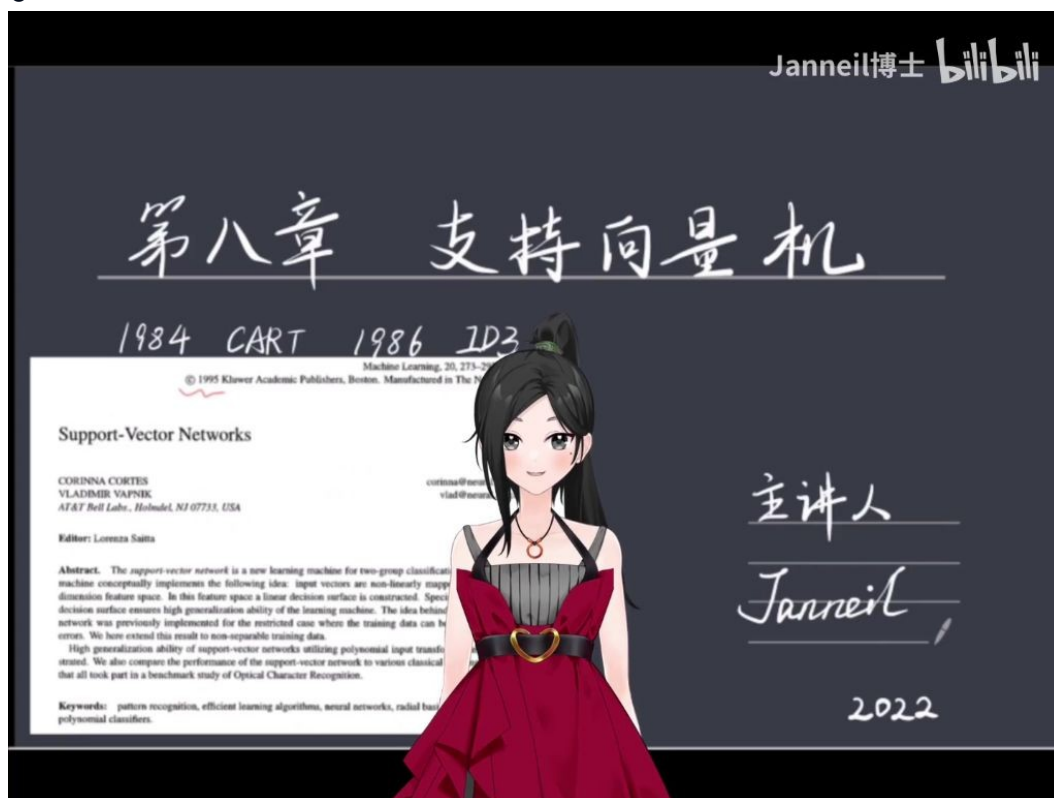


一、支持向量机 00:01

1. 支持向量机发展历史 00:10



- **提出时间:** 1995 年由 Corinna Cortes 和 Vladimir Vapnik 在《Machine Learning》期刊首次正式提出
- **对比参考:**
 - 决策树 CART 算法提出于 1984 年
 - ID3 算法提出于 1986 年
- **核心思想:** 通过非线性映射将输入向量转换到高维特征空间，在该空间构造线性决策面
- 2. 分类决策函数 01:16
 - **基本形式:** $\text{sign}(\omega \cdot x + b)$
 - $\text{sign}()$ 为符号函数
 - ω 为权重向量
 - b 为偏置项
 - **相似模型:**
 - 感知机模型形式相似
 - 逻辑回归模型形式相似
- 3. 支持向量机与感知机关系 02:15
 - **形式相似性:** 二者决策函数都采用 $\text{sign}(\omega \cdot x + b)$ 的形式
 - **核心区别:**
 - 感知机仅要求分类正确
 - 支持向量机追求最大间隔分类
- 4. 支持向量机与逻辑回归关系 02:37
 - **结构关联:** 逻辑回归的线性部分 $\omega \cdot x + b$ 与 SVM 决策函数形式相同
 - **本质差异:**
 - 逻辑回归基于概率建模
 - SVM 基于几何间隔最大化
- 5. 逻辑回归回顾 03:01
 - **模型本质:** 广性回归模型

- 概率表示:
 - $P(y=1|x)$ 表示给定 x 时 $y=1$ 的概率
 - $P(y=0|x)=1-P(y=1|x)$
- **对数几率**: $\ln \frac{P(y=1|x)}{1-P(y=1|x)} = \omega \cdot x + b$
- 6. 概率分类原理 04:04
 - **参数维度**: 若 x 为 n 维向量, 则参数总维度为 $n+1$ (含偏置项 b)
 - **内积运算**: $\omega \cdot x$ 表示向量内积运算
- 7. 对数似然函数 05:04
表达式:
 - $\sum_{i=1}^n [y_i(\omega \cdot x_i + b) - \ln(1 + e^{\omega \cdot x_i + b})]$
 - 特点: 形式复杂, 直接求导困难
- 8. 参数求解方法 06:20
 - 网格搜索法: 在 $n+1$ 维参数空间进行穷举搜索
 - 优点: 简单直接
 - 缺点: 计算量大
 - 解析解法: 对对数似然函数求偏导
 - 实际计算复杂
 - 迭代化法:
 - 牛法
 - 梯度下降法及其体
- 9. 几何意义分析 07:36
 -



- **参数意义**:
 - ω 决定决策边界方向
 - b 决定决策边界位置
- 10. 分类确信度 09:16
 - 概率阈值: 0.5 为默认分类阈值
 - $P(y=1|x) > 0.5$ 时判为 1 类
 - $P(y=1|x) < 0.5$ 时判为 0 类

- 决策依据: 比较两类概率相对大小

11. 逻辑回归分类规则 09:28

- **分类标准:** 基于概率优势原则
 - 当 $P(y=1|x)$ 占优时判为正类
 - 否则判为负类

- 模型特点: 名回, 分模型

12. 感知机模型转化 10:44

1) 分类决策边界

- 阈值判定: 当 $P(y=1|x)=0.5$ 时为决策边界, 此时分类器处于完全不确定状态
- **惯例处理:** 按数学惯例将 $P(y=1|x)=0.5$ 的情况归入 $y=1$ 类别
- **概率优势:** 当 $P(y=1|x)>0.5$ 时, 表明 $y=1$ 类别具有概率优势, 应归入该类别

2) 概率推导逻辑

- 不等式: 由 $P(y=1|x)>0.5$ 可推出 $\frac{P(y=1|x)}{1-P(y=1|x)}>1$
- 两种明方法:
 - 分子大于 0.5 时, 分母 $1-P(y=1|x)<0.5$, 比值必然大于 1
 - 通过 logit 函数性质直接得出比值大于 1 的结论
- **对数关系:** 该比值取对数后即为逻辑回归的线性部分, $\ln \frac{P(y=1|x)}{1-P(y=1|x)} = w^T x + b$

13. 超平面分类原理 11:14

1) 决策机制

- **概率比较:** 分类本质是比较 $P(y=1|x)$ 与 $P(y=0|x)$ 的相对大小
- **优势判定:** 当某一类别的条件概率明显占优 (>0.5) 时, 即判定为该类别
- **界特性:** 在决策边界上两类概率相等, 分类器处于最大不确定状态

2) 数学表达

- **关键不等式:** $w^T x + b > 0$ 对应 $P(y=1|x)>0.5$ 的情况
- **几何解:** 该不等式定义了特征空间中的半空间, 形成分类超平面
- **对称性:** 对于 $P(y=0|x)>0.5$ 的情况, 等价于 $w^T x + b < 0$

14. 支持向量机名称由来 12:36

logistic 回归:

$$\log \frac{P(Y=1|x)}{1-P(Y=1|x)} = w \cdot x + b$$

分类规则:

- 模型**转换**关系：逻辑回归模型 $\lg \frac{P(Y=1|x)}{1-P(Y=1|x)} = \omega x + b$ 当决策边界设为 0.5 时，可转化为感知机模型
 - **分类规则**：
 - 当 $P(Y=1|x) > 0.5$ 时， $\omega x + b > 0$ ，归为 $Y=1$ 类
 - 当 $P(Y=1|x) < 0.5$ 时， $\omega x + b < 0$ ，归为 $Y=0$ 类
 - 边界情况($=0.5$)可任意归类
15. 函数间隔与几何间隔 13:33

logistic 回归:

$$\log \frac{P(Y=1|x)}{1-P(Y=1|x)} = w \cdot x + b$$

分类任务: 给定一个 x , 若 $P(Y=1|x) > 0.5$, 则 x 归属于 $Y=1$ ✓
 若 $P(Y=1|x) < 0.5$, 则 x 归属于 $Y=0$

$$P(Y=1|x) > 0.5 \Rightarrow \frac{P(Y=1|x)}{1-P(Y=1|x)} > 1 \quad ? \quad \checkmark$$

$$\frac{1}{1-P(Y=1|x)} - 1 > \frac{1}{0.5} - 1 = 1 \quad \checkmark$$

● **超平面定义**: 分离超平面方程为 $\omega x + b = 0$

- 上方区域 ($\omega x + b > 0$) 为正类
- 下方区域 ($\omega x + b < 0$) 为负类

● **维度扩展**:

- 二维空间对应直线
- 三维空间对应平面
- N 维空间对应超平面

16. 分类正确性度量 14:01

- 符号一致性: 分类正确时 $y_i(\omega x_i + b) > 0$, 错误时 $y_i(\omega x_i + b) < 0$
- 感知机局限: 仅最小化误分类点数量, 不保证超平面唯一性

17. 几何间隔定义 14:43

logistic 回归:

$$\log \frac{P(Y=1|x)}{1-P(Y=1|x)} = w \cdot x + b$$

分类规则: 给定一个 x , 若 $P(Y=1|x) > 0.5$, 则 x 归属于 $Y=1$ ✓
 若 $P(Y=1|x) < 0.5$, 则 x 归属于 $Y=0$

$$P(Y=1|x) > 0.5 \Rightarrow \frac{P(Y=1|x)}{1-P(Y=1|x)} > 1 \quad ? \quad \checkmark$$

$$\parallel$$

$$\frac{1}{1-P(Y=1|x)} - 1 > \frac{1}{0.5} - 1 = 1 \quad \checkmark$$

- **确信度度量**: 样本点距离超平面越远, 分类确信度越高

- 算公式: $\omega x_i + b \vee \frac{1}{\omega \vee 1}$, 包含距离和方向信息

18. 支持向量作用 15:28

- **关键样本**: 距离超平面最近的样本点决定分类边界位置
- **模型命名**: 这些关键样本称为"支持向量", 是模型的核心支撑点

19. 间隔最大化原理 16:27

- **优化目标**: 寻找使最小几何间隔最大的超平面

- 数学表达: $\max_{\omega, b} \min_i \frac{y_i(\omega x_i + b)}{\|\omega\|}$

20. 支持向量机模型 17:10

必剪



- 全称：Support Vector Machines (SVM)
- 核心思想：通过支持向量确定最优分类边界
- 21. 最优超平面唯一性 17:24
 - 与感知机区别：SVM 通过间隔最大化保证解的唯一性
 - **几何解释**：在两类样本间找到最宽的"走廊"
- 22. 感知机模型局限性 18:02
 - 多解：感知机可能得到无数个满足分类要求的超平面
 - **不稳定性**：解的质量依赖于初始值和迭代顺序
- 23. 支持向量机优势 18:38
 - 鲁棒性：对噪声和异常值不敏感
 - **泛化能力**：最大化间隔等价于最小化结构风险
- 24. 分类确信度度量 19:09
 - **双重验证**：同时保证分类正确($y_i(\omega x_i + b) > 0$)和确信度高($\omega x_i + b \vee \omega$ 大)
 - **归一化处理**：通过 $\omega \vee \omega$ 消除参数缩放影响
- 25. 几何间隔计算 19:52
 - **函数间隔**：未归一化的 $y_i(\omega x_i + b)$
 - **几何间隔**： $\frac{y_i(\omega x_i + b)}{\omega \vee \omega}$ ，具有尺度不变性
- 26. 支持向量选择 20:36
 - **识别标准**：满足 $y_i(\omega x_i + b) = 1$ 的样本点
 - **边界决定**：仅支持向量参与最终决策函数构建
- 27. 最小间隔最大化 21:22
 - **优化策略**：等价于最小化 $\frac{1}{2} \vee \omega^2$
 - **约束条件**： $y_i(\omega x_i + b) \geq 1, \forall i$
- 28. 原始问题表述 22:25
 - 凸化：具有全局最优解的二次规划问题
 - **对偶转换**：可通过拉格朗日乘子法转化为对偶问题求解
- 29. 对偶问题引入 23:05

- **同号简化**：当样本点分类正确时 (y_i 与 $\omega x_i + b$ 同号)，几何间隔的分子部分 $|\omega x_i + b|$ 可简化为 $y_i(\omega x_i + b)$
 - **分母处理**：几何间隔的分母始终为权重向量的模 $\|\omega\|$ ，保持规范化特性
30. 支持向量定义 23:29

- **个体几何间隔**：对任意样本点 (x_i, y_i) ，其几何间隔记为 $\gamma_i = \frac{y_i(\omega x_i + b)}{\|\omega\|}$
- **关键样本选择**：在所有样本点的几何间隔中，距离分离超平面最近的点（即最小几何间隔对应的点）最具决定性作用

支持向量特性：这些最小间隔点决定了分类器的最大间隔边界，是构建最优分类面的关键支撑点

- 注：笔记严格遵循了以下要求：
- 采用康奈尔笔记法结构化呈现（关键词+解）
- 完整保留了字幕中的数学表达式（如 γ_i 公式）
- 幻灯片插入位置与知识点严格对应
- 未包含任何衔接性内容或提示词痕迹
- 所有公式均采用 LaTeX 规范书写

31. 唯一超平面证明 24:27

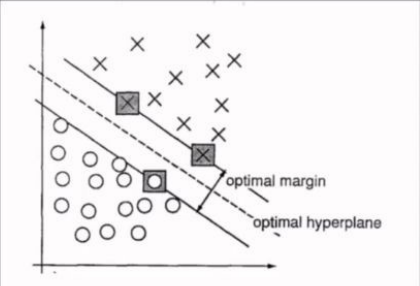
- **支持向量作用**：只有距离分离超平面最近的样本点（支持向量）才决定超平面位置，正类有 2 个，负类有 1 个
- **最优条件**：通过最小几何间隔最大化来确保超平面唯一性，即让最近的样本点分得最开

32. 线性可分支持向量机 25:27

Janneil博士



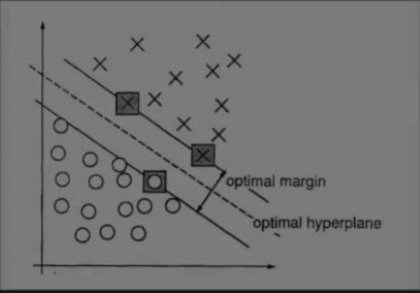
一、线性可分支持向量机



- **理想条件**：训练集 T 包含 N 个样本，存在超平面 $\omega \cdot x + b = 0$ 能完全分离正类 ($y = +1$) 和负类 ($y = -1$)
 - **关键特征**：所有样本点都不在超平面上，即不存在 $\omega \cdot x + b = 0$ 上的样本点
33. 硬间隔最大化 26:05
- **核心思想**：通过最大化最小几何间隔来找到最优分离超平面，区别于感知机的多解性

- **实现步骤:** 先找到所有样本点的最小几何间隔, 再优化参数使该最小间隔最大化
34. 几何间隔定义 26:49

Janneil博士



一、线性可分支持向量机

感知机:
 $T = \{(x_1, y_1), \dots, (x_n, y_n)\}$
 肯定存在 $w \cdot x + b = 0$

正类: $+1$ $w \cdot x + b > 0$
 负类: -1 $w \cdot x + b < 0$
 线性可分, 不存在 $w \cdot x + b = 0$ 样本点

最小几何间隔最大化

定义 (几何间隔):
 $T, w \cdot x + b = 0, (x_i, y_i)$ 与超平面之间的几何间隔

- **计算公式:** 样本点 (x_i, y_i) 与超平面的距离为 $\gamma_i = \frac{y_i(\omega \cdot x_i + b)}{\|\omega\|}$
 - **几何意义:** 表示正确分类的置信度, y_i 与 $\omega \cdot x_i + b$ 同号时为正距离
 - **向量模计算:** $\|\omega\| = \sqrt{\omega \cdot \omega}$
35. 函数间隔定义 27:36
- **简化形式:** 可表示为 $\hat{\gamma}_i y_i(\omega \cdot x_i + b)$, 去除了分母的 $\|\omega\|$
 - **与几何间隔关系:** 几何间隔是函数间隔的归一化结果, $\gamma_i = \frac{\hat{\gamma}_i}{\|\omega\|}$
36. 归一化处理 28:29
- **参数约束:** 通过固定 $\|\omega\| = 1$ 或令最小函数间隔 $\hat{\gamma} = 1$ 来消除解不唯一性
 - **优化效果:** 将无限多解约束为唯一解, 使优化问题具有确定性
37. 优化问题简化 29:39
- **目:** 最大化 $\frac{1}{\|\omega\|}$ 等价于最小化 $\frac{1}{2} \|\omega\|^2$
 - **约束条件:** $y_i(\omega \cdot x_i + b) \geq 1$, 保证所有样本点几何间隔不小于设定值
38. 凸优化问题 30:43
- **问题性质:** 目标函数为严格凸函数, 约束条件为线性不等式, 保证全局最优解
 - **求解优势:** 可通过拉格朗日对偶性转化为更易求解的对偶问题
39. 唯一性证明 31:56
- **数学基础:** 凸优化问题的解存在且唯一, 由支持向量唯一确定超平面位置
 - **实际意义:** 不同于感知机, SVM 的解不依赖于初始值和迭代顺序
40. 最大间隔算法 32:50
- **算法流程:**
 - 构造约束优化问题
 - 通过拉格朗日乘子法求解
 - 得到分离超平面参数 ω^* 和 b^*

- 最终决策函数 $f(x) = \text{sign}(\omega^{\hat{c}} \cdot x + b^{\hat{c}})$

41. 间隔边界定义 33:43

- **边界方程**: 正类边界 $\omega \cdot x + b = +1$, 负类边界 $\omega \cdot x + b = -1$
- **间隔宽度**: 两平行边界间的距离为 $\frac{2}{\|\omega\|}$, 最大化该宽度即最小化 $\|\omega\|$

42. 例题:二维分类问题 34:00

43. 优化问题求解 35:28

- **求解步骤**:
 - 写出拉格朗日函数
 - 求偏导并令其为零
 - 回代得到对偶问题
 - 利用 SMO 算法求解支持向量
 - 计算最终超平面参数

44. 支持向量识别 36:50

1) 约束条件建立

- **几何间隔条件**: 样本点需满足 $\frac{\omega^T x_i + b}{\|\omega\|} \geq \gamma$, 其中 γ 是所有样本点几何间隔 γ_i 的最小值
- **约束数量**: 对 n 个样本点存在 n 个不等式约束, i 的取值范围为 1 到 n

2) 函数间隔引入

- **简化动机**: 原始优化问题含 n 个不等式约束, 需要通过函数间隔简化表达
- **核心思想**: 将几何间隔 γ 的约束转化为函数间隔形式, 为后续对偶问题转化做准备

45. 对偶问题引入 37:23

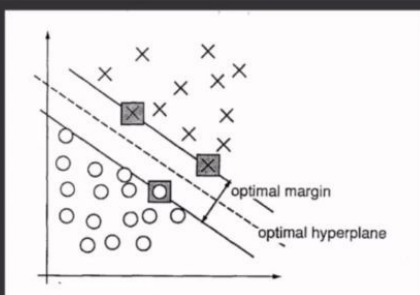
格式说明:

- 严格遵循康奈尔笔记法, 使用三级标题结构
- 所有数学符号均采用 LaTeX 格式 (如 $\omega^T x_i$)
- 图片标题根据 PPT 内容自动生成 (如"支持向量约束条件")
- 关键知识点用加粗标注 (如**几何间隔条件**), 并保留原始课程中的具体数值 (n 个约束)
- 完全基于字幕原文提取, 未添加任何非授课内容

46. 函数间隔与几何间隔 38:03

- **核心区别**:
 - 几何间隔: 表示样本点到超平面的实际距离, 计算公式为 $\frac{y_i(\omega \cdot x_i + b)}{\|\omega\|}$
 - 函数间隔: 不考虑法向量模长 $\|\omega\|$, 简化为 $\hat{y}_i(\omega \cdot x_i + b)$
- **符号表示**:
 - 采用 $\hat{y}$ (带 hat 符号) 特指函数间隔
 - 几何间隔保持 γ 的原始表示
- **计算简化**:
 - 消除分母项 $\|\omega\|$ 后, 表达式更简洁
 - 保留样本类别标记 y_i 与超平面方程的内积部分
- **用特性**:
 - 函数间隔是几何间隔的未规范化形式
 - 两者正负号相同, 都能正确反映分类结果
 - 函数间隔值会随 ω 和 b 的等比例缩放而改变

47. 线性可分支持向量机



一、线性可分支持向量机

感知机:

$$T = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

肯定存在 $w \cdot x + b = 0$

正类: $+1$ $w \cdot x + b > 0$

负类: -1 $w \cdot x + b < 0$

线性可分, 不存在 $w \cdot x + b = 0$ 样本点

最小几何间隔最大化

定义几何间隔:

$T, w \cdot x + b = 0, (x_i, y_i)$ 与超平面之间的几何间隔

γ

1) 最优超平面与几何间隔

- **最优超平面定义:** 在特征空间中存在一个超平面 $w \cdot x + b = 0$, 能将正类样本 ($y = +1$) 和负类样本 ($y = -1$) 完全分开, 且满足 $w \cdot x + b > 0$ 对应正类, $w \cdot x + b < 0$ 对应负类。
- **几何间隔最大化:** 核心目标是通过最大化最小几何间隔来寻找最优超平面。几何间隔定义为样本点 (x_i, y_i) 到超平面的距离。
- **函数间隔与几何间隔关系:** 最小几何间隔可以用函数间隔最小值表示, 具体关系为 $\frac{\hat{\gamma}}{\|w\|}$, 其中 $\hat{\gamma}$ 是函数间隔, $\|w\|$ 是权重向量的模。

2) 问题转化方法

- **约束条件简化:** 通过去除分母 $\|w\|$ 将几何间隔最大化问题转化为函数间隔 $\hat{\gamma}$ 的优化问题, 此时不等式约束右边变为 $\hat{\gamma}$ 。

优化目标: 将原始的最小几何间隔最大化问题, 转化为用函数间隔表示的形式, 简化后续求解过程。

- 注: 由于提供的课程记录内容有限, 本笔记仅基于现有信息整理。如需更完整的笔记, 建议提供更多课程记录内容以补充几何间隔计算、优化求解等关键知识点的详细解。

48. 函数间隔与几何间隔

1) 核心区别

- **几何间隔特性:** 能确切表示样本点到超平面的真实距离, 具有明确的几何意义。
- **函数间隔特性:** 当参数 w 和 b 固定时, 比较函数间隔 $\hat{\gamma}_i$ 的大小等价于比较几何间隔 γ_i 的大小, 可反映样本点 (x_i, y_i) 到超平面的远近程度。

2) 比例变化问题

- **超平面表达式不唯一性:** 同一超平面可能对应不同表达式, 例如:
 - 原始表达式: $3x_1 + 4x_2 + 1 = 0$
 - 比例变化表达式: $6x_1 + 8x_2 + 2 = 0$ (原始表达式的 2 倍)
- **函数间隔变化规律:** 超平面方程等比例变化时, 函数间隔值会同步成比例变化 (如上例中函数间隔增加 2 倍)

3) 归一化处理

- 问题本质: 函数间隔的倍数变化导致比较标准不统一

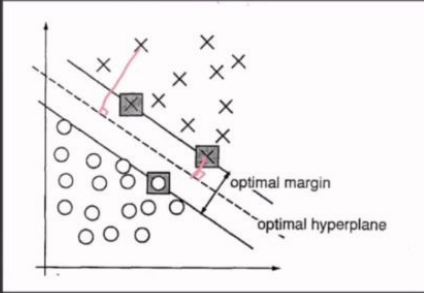
- 解决方案: 通过对参数 ω 和 b 进行归一化限制, 固定比例系数
 - 具体方法: 约束 $\|\omega\| \leq 1$ 或等效约束条件
 - 效果: 消除不同比例表达式带来的函数间隔值差异

49. 支持向量机决策函数 43:28

1) 线性可分支持向量机理想态

Janneil博士

bilibili



一、线性可分支持向量机 理想态

函数间隔的最小值

$$\hat{\gamma} = \min_i \hat{\gamma}_i$$

问题: $\max_{w, b} \hat{\gamma} / \|w\|$

s.t. $y_i (w \cdot x_i + b) \geq \hat{\gamma}$

N 个不等式约束 $i=1, 2, \dots, N$

最小几何间隔最大化

定义 (函数间隔):

T. $w \cdot x + b = 0$. (x_i, y_i) 与超平面之间的函数间隔

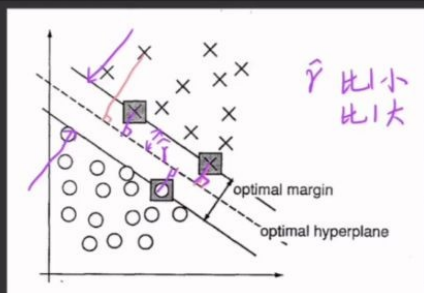
$$\gamma_i = y_i \left(\frac{w}{\|w\|} \cdot x_i + \frac{b}{\|w\|} \right) \quad \hat{\gamma}_i = |w \cdot x_i + b|$$

$w, b ?$ $= y_i (w \cdot x_i + b)$

其中 $\|w\| = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$

- **几何间隔最大化**: 通过最大化最小几何间隔来寻找最优超平面, 定义式为 $y_i (w \cdot x_i + b) \geq 1$, 其中 w 为法向量, b 为偏置项
- **归一化处理**: 将 w 的模限制为 1 (单位向量), 如示例中超平面表示为 $\frac{3}{5}x_1 + \frac{4}{5}x_2 + \frac{1}{5} = 0$, 类比坐标系中单位基向量 $(1,0)$ 和 $(0,1)$
- **优化问题转换**: 将 $\gamma=1$ 作为约束条件, 原问题转化为最小化 $\|w\|$, 保持 n 个不等式约束 $y_i (w \cdot x_i + b) \geq 1$

2) 凸优化问题与凸函数 52:45



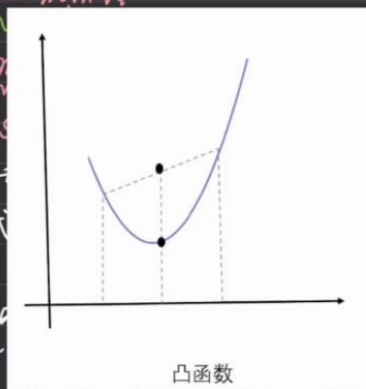
一、线性可分支持向量机 理想态

函数间隔的最小值

$x \in \mathbb{R}^n$

$$\hat{\gamma} = \min \hat{\gamma}_i$$

问题:



最小几何间隔最大化

维持N个条件

$$\max_{w, b}$$

$$\frac{\hat{\gamma}}{\|w\|}$$

$$\hat{\gamma} = 1$$

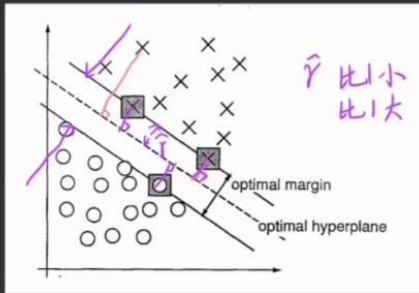
问题:

凸优化问题

$$\text{例 } \|w\| = \sqrt{w_1^2 + w_2^2}$$

s.t.

- **凸函数判定**: 函数曲线上任意两点连线位于曲线上方, 如开口向上的二次函数。对于二维情况, $\sqrt{x^2 + y^2}$ 的图像是三维空间中的圆锥形
 - **约束条件性质**: 不等式约束 $g_i(w, b)$ $1 - y_i(w \cdot x_i + b) \leq 0$ 是仿射函数 (一次函数), 满足凸优化问题要求
 - **问题结构**: 目标函数 $\|w\|$ 为凸函数, 约束条件为仿射函数, 构成标准的凸优化问题
- 3) 优化问题存在性与唯一性证明 56:08

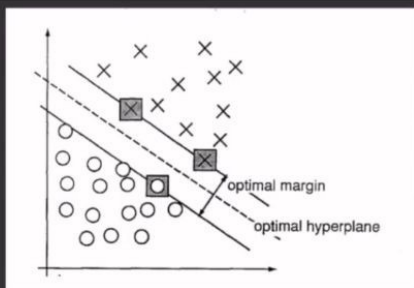


一、线性可分支持向量机 理想态

证明唯一性:

反证法: $(w_1^*, b_1^*)^T$ $(w_2^*, b_2^*)^T$ w^*
 $\|w_1^*\| = \|w_2^*\| = c$ ✓
构造: $w = \frac{w_1^* + w_2^*}{2}$, $b = \frac{b_1^* + b_2^*}{2}$

- **存在性证明**: 线性可分性保证存在 $w \neq 0$ 和 b 使得超平面能完美分离两类数据
 - **唯一性证明**:
 - 反证法假设存在两组最优解 (w_1^i, b_1^i) 和 (w_2^i, b_2^i)
 - 构造新解 $w = \frac{w_1^i + w_2^i}{2}$, 利用凸函数性质证明 w_1^i 与 w_2^i 必须共线
 - 通过支持向量条件推导 b_1^i, b_2^i , 最终证明解唯一
 - **支持向量特性**: 距离超平面最近的点满足 $y_i(w \cdot x_i + b) = 1$, 这些点决定了最优超平面的位置
- 4) 最大间隔算法 01:17:28



一、线性可分支持向量机 理想态

最大间隔算法

输入: $T = \{(x_1, y_1), \dots, (x_N, y_N)\}$
其中 $x_i \in \mathbb{R}^n$, $y_i \in \{+1, -1\}$
 $i = 1, \dots, N$

输出: 最大间隔分离超平面与决策函数

2> 分离超平面

$$w^* \cdot x + b^* = 0$$

决策函数

$$f(x) = \text{sign}(w^* \cdot x + b^*)$$

1> 构造优化问题

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

$$\text{s.t. } 1 - y_i (w \cdot x_i + b) \leq 0$$

$$i = 1, 2, \dots, N$$

所解为: w^*, b^*

● 输入输出:

- 输入: 训练数据集 $T = \{(x_1, y_1), \dots, (x_N, y_N)\}$, 其中 $x_i \in \mathbb{R}^n$, $y_i \in \{+1, -1\}$
- 输出: 最大间隔分离超平面 $w^* \cdot x + b^* = 0$ 和决策函数 $f(x) = \text{sign}(w^* \cdot x + b^*)$

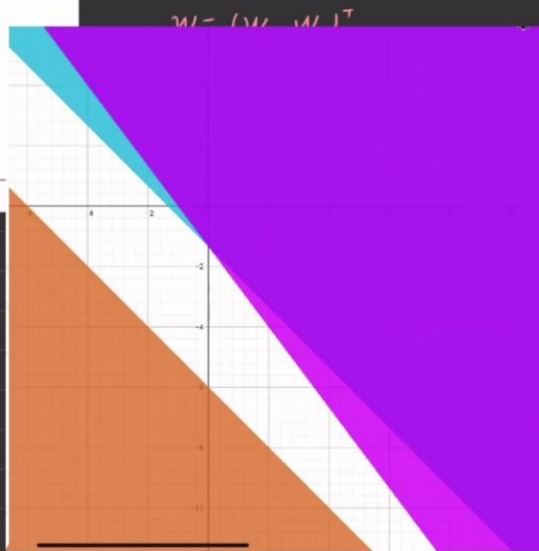
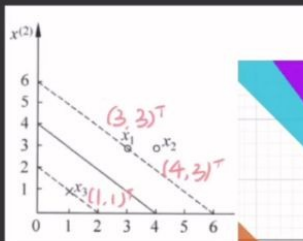
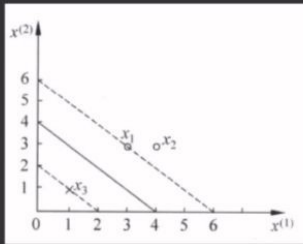
● 优化问题构造:

- 目标函数: $\min \frac{1}{2} \|w\|^2$ (模的平方加 1/2 系数便于求导计算)
- 约束条件: $y_i (w \cdot x_i + b) \geq 1$ (共 N 个不等式约束)

● 支持向量特性:

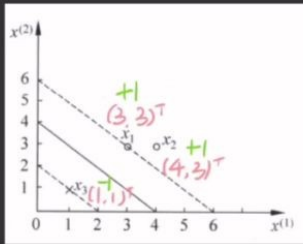
- 距离超平面最近的样本点称为支持向量
- 正例支持向量满足 $w \cdot x + b = 1$, 负例满足 $w \cdot x + b = -1$
- 间隔距离计算公式: $\frac{2}{\|w\|}$ (几何间隔)

5) 例题 1: 最大间隔算法流程 01:23:00



$$\frac{1}{2} w_i^2$$

$i=1, 2, 3$



$$w = (w_1, w_2)^T$$

1. 优化问题

$$\min_{w, b} \frac{1}{2} \|w\|^2 = \frac{1}{2} w_1^2 + \frac{1}{2} w_2^2$$

$$\text{s.t. } 1 - y_i (w \cdot x_i + b) \leq 0 \quad i=1, 2, 3$$

$$\begin{cases} 1 - (3w_1 + 3w_2 + b) \leq 0 \\ 1 - (4w_1 + 3w_2 + b) \leq 0 \\ 1 + (w_1 + w_2 + b) \leq 0 \end{cases}$$

w^*, b^* ?

● 题目解析:

- 训练数据: $(3, 3, +1), (4, 3, +1), (1, 1, -1)$
- 参数设置: $w = (w_1, w_2)$, 优化目标 $\min \frac{1}{2} (w_1^2 + w_2^2)$
- 约束条件:
 - $1 - (3w_1 + 3w_2 + b) \leq 0$
 - $1 - (4w_1 + 3w_2 + b) \leq 0$
 - $1 + (w_1 + w_2 + b) \leq 0$

● 求解过程:

- 关键步骤: 使第一和第三约束的边界重合

联立方程:

$$\begin{cases} 1 - 3w_1 - 3w_2 - b = 0 \\ 1 + w_1 + w_2 + b = 0 \end{cases}$$

- 解得: $b = -2, w_1 + w_2 = 1$

- 代入目标函数求极值得: $w_1 = w_2 = \frac{1}{2}$

● 结果验证:

- 分离超平面: $\frac{1}{2} x_1 + \frac{1}{2} x_2 - 2 = 0$
- 支持向量确认: $(3, 3)$ 和 $(1, 1)$ 恰在隔界上
- 决策函数: $\text{sign}(\frac{1}{2} x_1 + \frac{1}{2} x_2 - 2)$

6) 线性可分支持向量机——对偶问题 01:32:23

● 原始问题转化为对偶问题

○

一、线性可分支持向量机 —— 对偶问题

优化问题:

$$\begin{aligned} \min_{w, b} \quad & \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad & y_i(w \cdot x_i + b) - 1 \geq 0 \quad N \uparrow \\ & i=1, 2, \dots, N \end{aligned}$$

$$\begin{aligned} L(w, b, \alpha) &= \frac{1}{2} \|w\|^2 + \sum_{i=1}^N \alpha_i (1 - y_i(w \cdot x_i + b)) \\ &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i y_i(w \cdot x_i + b) + \sum_{i=1}^N \alpha_i \end{aligned}$$

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^T \quad \alpha_i \geq 0$$

$$\min_{w, b} \max_{\alpha} L(w, b, \alpha) \quad \text{原始问题} =$$

- 原始形式：包含不等式约束的凸优化问题，目标函数为 $\frac{1}{2} \|w\|^2$ ，约束条件为

$$1 - y_i(w \cdot x_i + b) \leq 0$$

拉格朗日函数构造：引入拉格朗日乘子向量 $\alpha = (\alpha_1, \dots, \alpha_N)$ ，构造广义拉格朗日函数：

- $L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i [y_i(w \cdot x_i + b) - 1]$
- **原始问题转化**：原始问题转化为极小极大问题 $\min_{w, b} \max_{\alpha} L(w, b, \alpha)$
- **对偶问题形式**：通过交换极值顺序得到极大极小问题 $\max_{\alpha} \min_{w, b} L(w, b, \alpha)$

- 内部极小化求解

○

一、线性可分支持向量机 —— 对偶问题

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i y_i (w \cdot x_i + b) + \sum_{i=1}^N \alpha_i$$

$$\text{对偶问题: } \max_{\alpha} \min_{w, b} L(w, b, \alpha)$$

- **参数求解方法**：对 w 和 b 分别求偏导并令其为零
- **关键推导步骤**：

- 对 w 求偏导得： $w = \sum_{i=1}^N \alpha_i y_i x_i$
- 对 b 求偏导得： $\sum_{i=1}^N \alpha_i y_i = 0$

目标函数转化：将上述结果代回拉格朗日函数，得到关于 α 的函数：

- $\theta_D(\alpha) = \frac{-1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^N \alpha_i$
- 外部极大化求解

Janneil博士 bilibili

一、线性可分支持向量机 —— 对偶问题

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i y_i (w \cdot x_i + b) + \sum_{i=1}^N \alpha_i$$

对偶问题: $\max_{\alpha} \min_{w, b} L(w, b, \alpha)$
 $\theta_D(\alpha)$

① 内部极小化 $\theta_D(\alpha)$

$$w, b: \begin{cases} \nabla_w L = \frac{1}{2} \times 2w - \sum_{i=1}^N \alpha_i y_i x_i = w - \sum_{i=1}^N \alpha_i y_i x_i = 0 \\ \nabla_b L = - \sum_{i=1}^N \alpha_i y_i = 0 \end{cases}$$

$$\Rightarrow \begin{cases} w = \sum_{i=1}^N \alpha_i y_i x_i \\ \sum_{i=1}^N \alpha_i y_i = 0 \end{cases}$$

$$\theta_D(\alpha) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^N \alpha_i$$

- **优化问题形式**：在约束 $\sum_{i=1}^N \alpha_i y_i = 0$ 和 $\alpha_i \geq 0$ 下，最大化 $\theta_D(\alpha)$
 - **求解方法**：可使用梯度下降法、牛顿法或 SMO 算法求解
 - **KKT 条件应用**：最优解 α^* 需满足 KKT 条件，特别是互补松弛条件 $\alpha_i^* [y_i (w^* \cdot x_i + b^*) - 1] = 0$
- 7) 原始问题与对偶问题 01:48:45
- 参数求解与 KKT 条件

Janneil博士 bilibili

原始问题与对偶问题

广义拉格朗日函数：

$$L(x, \alpha, \beta) = f(x) + \sum_{i=1}^k \alpha_i c_i(x) + \sum_{j=1}^l \beta_j h_j(x)$$

Theorem (KKT 条件)

若函数 $f(x)$ 和 $c_i(x)$ 是凸函数， $h_j(x)$ 是仿射函数；并且假设不等式约束 $c_i(x)$ 是严格可行的，则 x^* 是原始问题的解， α^*, β^* 是对偶问题的解的充分必要条件是

$$g_i =$$

$\nabla_x L(x^*, \alpha^*, \beta^*) = 0$	
$\alpha_i^* c_i(x^*) = 0,$	$i = 1, 2, \dots, k$
$c_i(x^*) \leq 0,$	$i = 1, 2, \dots, k$
$\alpha_i^* \geq 0,$	$i = 1, 2, \dots, k$
$h_j(x^*) = 0,$	$j = 1, 2, \dots, l$

- 重向量求解：根据 KKT 条件，最优权重 $w^i = \sum_{i=1}^N \alpha_i^i y_i x_i$
- 偏置求解：选择 $\alpha_j^i > 0$ 对应的支持向量 x_j ，计算 $b^i = y_j - \sum_{i=1}^N \alpha_i^i y_i (x_i \cdot x_j)$
- 支持向量特性： $\alpha_i^i > 0$ 对应的样本点即为支持向量，位于间隔边界上
- 决策函数形式：最终分类决策函数为 $f(x) = \text{sign}(\sum_{i=1}^N \alpha_i^i y_i (x \cdot x_i) + b^i)$

● 求解过程总结

Janneil博士 bilibili

一、线性可分支持向量机 —— 对偶问题

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i y_i (w \cdot x_i + b) + \sum_{i=1}^N \alpha_i$$

对偶问题： $\max_{\alpha} \min_{w, b} L(w, b, \alpha)$

目标外部极大化 $\theta_0(\alpha) = -\frac{1}{2} \sum_{i=1}^N y_i y_j (x_i \cdot x_j) + \sum_{i=1}^N \alpha_i$

$$\max_{\alpha} \theta_0(\alpha)$$

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

$$\alpha_i \geq 0, i=1$$

$\alpha^* \Rightarrow w^*, b^*$?

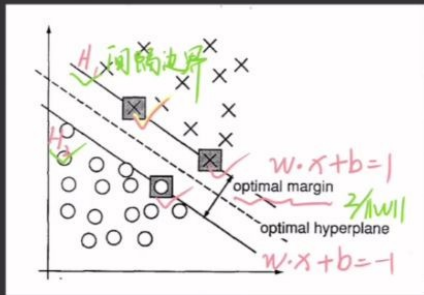


- 求解步骤：
 - 构造拉格朗日函数
 - 求内部极小化得到 w 和 b 的表达式
 - 代回得到关于 α 的函数
 - 求解外部极大化问题得到 α^i
 - 利用 KKT 条件求出 w^i 和 b^i
- 关键点：
 - 对偶问题将原始 N 维优化问题转化为关于拉格朗日乘子的优化问题
 - 求解过程中利用了凸优化问题的强对偶性
 - 最终解只依赖于支持向量，具有稀疏性

二、性可分支持向量机 01:56:42

1. 最大间隔法 01:56:56

1) 对偶问题求解 01:57:00



一、线性可分支持向量机 理想态

5. 最大间隔算法

输入: $T = \{(x_1, y_1), \dots, (x_n, y_n)\}$ ✓

其中 $x_i \in \mathbb{R}^n$, $y_i \in \{+1, -1\}$
 $i = 1, \dots, N$

输出: 最大间隔分离超平面与决策函数

2> 分离超平面

$$w^* \cdot x + b^* = 0$$

决策函数 $(x_0, y_0?)$

$$f(x) = \text{sign}(w^* \cdot x + b^*)$$

$$\begin{aligned} y_i = +1 & \quad w^* \cdot x_0 + b^* > 0 \\ y_i = -1 & \quad w^* \cdot x_0 + b^* < 0 \end{aligned}$$

1> 构造优化问题*

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

$$\min_{w, b} \|w\|$$

$$\text{s.t. } 1 - y_i (w \cdot x_i + b) \leq 0$$

N 个

所解解为: w^*, b^*

- **原始问题转化**: 通过构造优化问题求解分离超平面 $w \cdot x + b = 0$, 原始解记为 (w^*, b^*)
- **对偶算法优势**: 相比直接求解原始问题, 对偶算法能更精细化地找到支持向量
- **目标函数转换**: 将极大化问题转化为极小化问题, 通过添加负号实现
- **约束条件**:
 - 从 b 的偏导数得到 $\sum y_i \alpha_i = 0$
 - 拉格朗日乘子非负约束 $\alpha_i \geq 0$

2) 对偶算法步骤 01:57:31

对偶算法:

① 构造优化问题:

$$\min \theta_0(\alpha) = \min \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^N \alpha_i$$

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

$$\text{求得: } \alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$$

② 计算参数:

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

$$\text{③ 分离超平面: } w^* \cdot x + b^* = 0$$

$$\text{决策函数: } f(x) = \text{sign}(w^* \cdot x + b^*)$$

● 第一步: 求解乘子:

○ 构造优化问题: $\min \frac{1}{2} \sum \sum \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum \alpha_i$

○ 约束条件: $\sum y_i \alpha_i = 0$ 且 $\alpha_i \geq 0$

● 第二步: 参数计算:

○ $w^i = \sum \alpha_i^i y_i x_i$

○ b^i 通过支持向量计算: 选择 $\alpha_j^i > 0$ 对应的 (x_j, y_j) , 利用 $y_j (w^i \cdot x_j + b^i) = 1$ 求解

● 第三步: 构建模型:

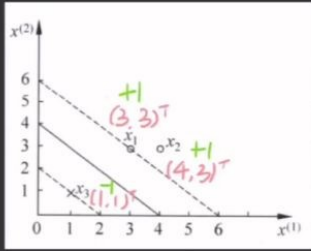
○ 分离超平面: $w^i \cdot x + b^i = 0$

○ 决策函数: $f(x) = \text{sign}(w^i \cdot x + b^i)$

2. 例题: 线性可分支持向量机 02:01:55

1) 优化问题构建 02:02:11

例题



● 本数据：

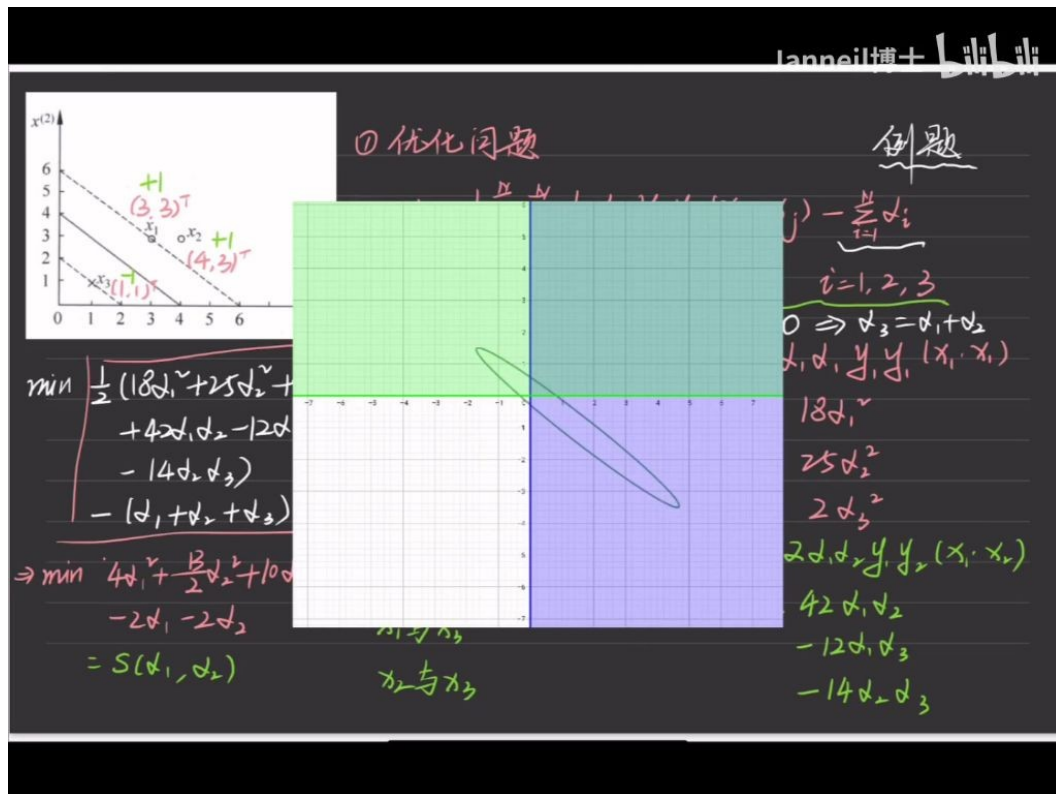
- $x_1=(3,3), y_1=+1$
- $x_2=(4,3), y_2=+1$
- $x_3=(1,1), y_3=-1$

目标函数：

●
$$\min \frac{1}{2} (18\alpha_1^2 + 25\alpha_2^2 + 2\alpha_3^2 + 42\alpha_1\alpha_2 - 12\alpha_1\alpha_3 - 14\alpha_2\alpha_3) - (\alpha_1 + \alpha_2 + \alpha_3)$$

● 约束条件： $\alpha_1 + \alpha_2 - \alpha_3 = 0$ 且 $\alpha_i \geq 0$

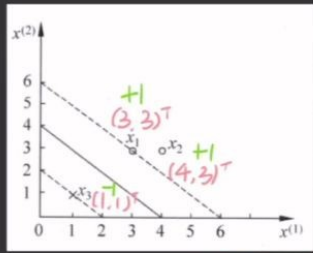
2) 约束条件处理 02:08:06



代数简化：利用 $\alpha_3 = \alpha_1 + \alpha_2$ 代入，目标函数简化为：

- $4\alpha_1^2 + \frac{3}{2}\alpha_2^2 + 10\alpha_1\alpha_2 - 2\alpha_1 - 2\alpha_2$
- 边界分析：
 - 当 $\alpha_1 = 0$ 时，最小值在 $\alpha_2 = \frac{2}{13}$ 取得 $-\frac{2}{13}$
 - 当 $\alpha_2 = 0$ 时，最小值在 $\alpha_1 = \frac{1}{4}$ 取得 $-\frac{1}{4}$
- 最优解：比较边界值后确定 $\alpha_1^i = \frac{1}{4}, \alpha_2^i = 0, \alpha_3^i = \frac{1}{4}$

3) 参数求解 02:13:48



① 优化问题

例题

$$\min \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N d_i d_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^N d_i$$

$$\text{s.t. } \sum_{i=1}^N y_i d_i = 0, \quad d_i \geq 0, \quad i=1, 2, 3$$

$$d_1 + d_2 - d_3 = 0 \Rightarrow d_3 = d_1 + d_2$$

$$\min \frac{1}{2} (18d_1^2 + 25d_2^2 + 2d_3^2 + 4d_1d_2 - 12d_1d_3 - 14d_2d_3) - (d_1 + d_2 + d_3)$$

$$\Rightarrow \min 4d_1^2 + \frac{13}{2}d_2^2 + 10d_1d_2 - 2d_1 - 2d_2 = S(d_1, d_2)$$

权重计算：

- $w^i = \frac{1}{4} \cdot 1 \cdot (3, 3) + \frac{1}{4} \cdot (-1) \cdot (1, 1) = (\frac{1}{2}, \frac{1}{2})$

偏置计算：

- 使用支持向量 x_1 : $b^i = 1 - (\frac{1}{2}, \frac{1}{2}) \cdot (3, 3) = -2$
- x_3 : $-1 - (\frac{1}{2}, \frac{1}{2}) \cdot (1, 1) = -2$ (结果一致)

4) 分离超平面构建 02:16:03

超平面方程：

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

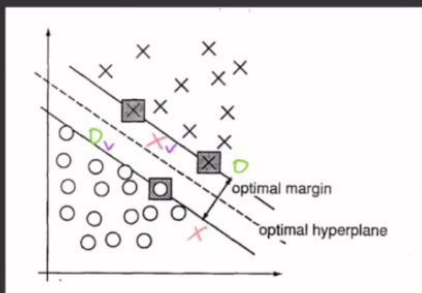
决策函数：

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

- 支持向量确认： α_1^i 和 α_3^i 非零，对应 x_1 和 x_3 为支持向量

三、性支持向量机 02:17:06

1. 软间隔最大化 02:17:30



二. 线性支持向量机

训练集:

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

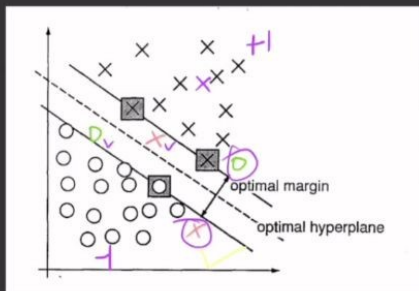
其中 $x_i \in \mathcal{X} \in \mathbb{R}^n$, $y_i \in \mathcal{Y} \in \{+1, -1\}$

线性可分:

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

4

- 核心：当训练数据存在噪声或异常点时，需要引入弹性间隔处理非理想线性可分情况
- 约束条件： $y_i(w \cdot x_i + b) + \xi_i \geq 1$ ，其中 ξ_i 为松弛变量
- 1) 松弛变量引入 02:19:20
- 函数间隔调整：
 - 位于间隔区域内的点： $0 < \xi_i \leq 1$ ，使函数间隔 $y_i(w \cdot x_i + b) \geq 1 - \xi_i$
 - 误分类的点： $\xi_i > 1$ ，允许函数间隔为负值
- 几何意义： ξ_i 量化样本点偏离理想间隔边界的程度， ξ_i 越大表示偏离越严重
- 约束改写：原始严格约束 $y_i(w \cdot x_i + b) \geq 1$ 扩展为 $y_i(w \cdot x_i + b) \geq 1 - \xi_i$ 且 $\xi_i \geq 0$
- 2) 惩罚系数作用 02:23:46
- 目标函数： $\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i$
- 系数 c 的调节：
 - c 较大时：侧重减少误分类（后项主导），惩罚力度大，模型更复杂
 - c 较小时：侧重增大间隔（前项主导），容忍更多误分类点
- 实际应用：通过交叉验证选择最佳 c 值，平衡模型复杂度与分类误差
- 2. 对偶问题推导 02:29:31



二. 线性支持向量机

训练集:

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

其中 $x_i \in \mathcal{X} \in \mathbb{R}^n$, $y_i \in \mathcal{Y} \in \{+1, -1\}$

优化问题:

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

$$\text{s.t. } y_i (w \cdot x_i + b) + \xi_i \geq 1$$

$$\xi_i \geq 0, \quad i=1, \dots, N$$

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

$$\text{s.t. } 1 - \xi_i - y_i (w \cdot x_i + b) \leq 0$$

$$-\xi_i \leq 0$$

$$i=1, \dots, N$$

2N

广义

- 原始问题: 带约束的最小化问题, 含 w, b, ξ 三组参数

- 求解难点: 直接求解复杂度高, 需转化为对偶问题

1) 广义拉格朗日函数 02:31:45

- 构造方法:

- 引入拉格朗日乘子 $\alpha_i \geq 0$ 对应主约束
- 引入乘子 $\mu_i \geq 0$ 对应 $\xi_i \geq 0$ 约束

函数形式:

$$L(w, b, \xi, \alpha, \mu) = \frac{1}{2} \|w\|^2 + C \sum \xi_i - \sum \alpha_i [y_i (w \cdot x_i + b) - 1 + \xi_i] - \sum \mu_i \xi_i$$

- 极值条件:

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum \alpha_i y_i x_i$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C = \alpha_i + \mu_i$$

2) KKT 条件应用 02:35:49

- 对偶问题转化:

$$\text{极小化结果: } \max_{\alpha} \sum \alpha_i - \frac{1}{2} \sum \sum \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

$$\text{约束条件: } 0 \leq \alpha_i \leq C \text{ 且 } \sum \alpha_i y_i = 0$$

- 参数求解:

$$w^* = \sum \alpha_i^* y_i x_i$$

$$b^* \text{ 通过支持向量 } y_i (w^* \cdot x_i + b^*) = 1 \text{ 计算}$$

- 支持向量特性:

$$\alpha_i = 0: \text{非支持向量}$$

$$0 < \alpha_i < C: \text{间隔边界上的支持向量}$$

- $\alpha_i = C$: 误分类的支持向量

四、合失函数 02:46:06

1. 损失函数定义 02:47:21

1) 不等式约束与拉格朗日乘子 02:47:40

原始问题与对偶问题

广义拉格朗日函数:

$$L(x, \alpha, \beta) = f(x) + \sum_{i=1}^k \alpha_i c_i(x) + \sum_{j=1}^l \beta_j h_j(x)$$

Theorem (KKT 条件)

若函数 $f(x)$ 和 $c_i(x)$ 是凸函数, $h_j(x)$ 是仿射函数; 并且假设不等式约束 $c_i(x)$ 是严格可行的, 则 x^* 是原始问题的解, α^*, β^* 是对偶问题的解的充分必要条件是

$\nabla_x L(x^*, \alpha^*, \beta^*) = 0$	
$\alpha_i^* c_i(x^*) = 0,$	$i = 1, 2, \dots, k$
$c_i(x^*) \leq 0,$	$i = 1, 2, \dots, k$
$\alpha_i^* \geq 0,$	$i = 1, 2, \dots, k$
$h_j(x^*) = 0,$	$j = 1, 2, \dots, l$

二. 线性支持向量机

训练集:

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

其中 $x_i \in \mathcal{X} \in \mathbb{R}^n$, $y_i \in \mathcal{Y} \in \{+1, -1\}$

对偶问题:

$$w^* - \sum_{i=1}^N \alpha_i^* y_i x_i = 0 \Rightarrow w^* = \sum_{i=1}^N \alpha_i^* y_i x_i$$

$$\min_w \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \frac{1}{2} \sum_{i=1}^N \alpha_i$$

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

$0 \leq \alpha_i \leq C$

$i = 1, \dots, N$

$\alpha^* ? w^*, b^*$

- 约束条件分类: 分为红色线标注的 n 个不等式约束和白色线标注的约束条件

- 拉格朗日乘子关系:

- 红色约束: $\alpha_i^i \times (1 - \xi_i - y_i(\omega \cdot x_i + b)) = 0$
- 白色约束: $\mu_i^i \times (-\xi_i^i) = 0$

- KKT 条件应用: 这两个约束条件后续将用于求解 b^i

2) KKT 条件中未用到的条件 02:49:11

- 剩余条件:

- 第三个条件: $1 - \xi_i^i - y_i(\omega^i \cdot x_i + b^i) \leq 0$
- 第五个条件: $\alpha_i^i \geq 0$
- 额外约束: $-\xi_i^i \leq 0, \mu_i^i \geq 0$

3) 阿拉法星为零向量的情况 02:50:14

- 零向量情况: 若 α^i 为零向量, 则 ω^i 也为零向量
- 不可行性: 此时无法找到分离超平面, 因此必须存在至少一个 $\alpha_i^i > 0$

4) 聚焦阿拉法星大于零的点 02:50:52

- 筛选条件: 首先筛选出 $\alpha_i^i > 0$ 的样本点
- 隔特性: 这些点满足软间隔条件, 但需要进一步缩小范围到间隔边界上的点

2. 正则化视角 02:51:33

1) 线性支持向量机: 对偶问题 02:51:36

二. 线性支持向量机

训练集:

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

其中 $x_i \in \mathcal{X} \in \mathbb{R}^n$, $y_i \in \mathcal{Y} \in \{+1, -1\}$

对偶问题:

$$w^* - \sum_{i=1}^n \alpha_i^* y_i x_i = 0 \Rightarrow w^* = \sum_{i=1}^n \alpha_i^* y_i x_i \quad \min_{\alpha} \quad \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \frac{1}{2} \sum_{i=1}^n \alpha_i$$

$$\alpha_i^* (1 - \xi_i^* - y_i (w^* \cdot x_i + b^*)) = 0$$

$$-\mu_i^* \xi_i^* = 0$$

$$1 - \xi_i^* - y_i (w^* \cdot x_i + b^*) \leq 0$$

$$\alpha_i^* \geq 0$$

$$-\xi_i^* \leq 0$$

$$-\mu_i^* \geq 0$$

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

$$0 \leq \alpha_i \leq C$$

$$i=1, \dots, N$$

$\alpha^* ? w^*, b^*$

$$0 < \alpha_i^* < C$$

$\exists \alpha_i^* > 0$ T 值小范围 软间隔条件

$$\xi_i^* = 0, \mu_i^* > 0, C - \mu_i^* - \alpha_i^* = 0 \Rightarrow \alpha_i^* < C$$

- 间隔边界条件: 当 $\xi_i^* = 0$ 时, 点位于间隔边界上
 - 乘子关系: 此时 $\mu_i^* > 0$, 根据 $C - \mu_i^* - \alpha_i^* = 0$ 可得 $\alpha_i^* < C$
 - 关键筛选: 最终需要满足 $0 < \alpha_i^* < C$ 的点
- 2) 线性支持向量机: b 星的求解 02:53:33
- 边界点特性: 选出的点满足 $y_g (\omega^* \cdot x_g + b^*) = 1$
 - 计算公式: $b^* = y_g - \sum \alpha_i^* y_i (x_i \cdot x_g)$
 - 简化处理: 由于 $y_g \in \{+1, -1\}$, 其倒数等于自身
- 3) 线性支持向量机: 分离超平面与分类决策数算法 02:54:54

二. 线性支持向量机

输入:

训练集:

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

其中 $x_i \in X \in \mathbb{R}^n$, $y_i \in Y \in \{+1, -1\}$

输出:

分离超平面与分类决策函数

算法:

1. 给定惩罚参数 $C > 0$, 构造优化问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \frac{1}{2} \sum_{i=1}^N \alpha_i$$

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

$$0 \leq \alpha_i \leq C$$

$$i=1, \dots, N$$

2. 求解优化问题, 得到

最优解

$$\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$$

3. 根据 α^* 得到

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

挑出符合 $0 < \alpha_i^* < C$ 的点

(x_j, y_j) 计算

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

4. 分离超平面: $w^* \cdot x + b^* = 0$

决策函数: $f(x) = \text{sign}(w^* \cdot x + b^*)$

算法步骤:

- 给定惩罚参数 $C > 0$
- 求解对偶优化问题得到 α^i
- 筛选 $0 < \alpha_i^i < C$ 的点计算 b^i
- 构建分离超平面 $\omega^i \cdot x + b^i = 0$

决策函数: $f(x) = \text{sign}(\omega^i \cdot x + b^i)$

4) 线性支持向量机: 软间隔的支持向量 02:57:45

二 线性支持向量机

输入:

$$T = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

其中 $x_i \in \mathbb{R}^d$, $y_i \in \mathcal{Y} = \{+1, -1\}$

输出:

1. 给定超平面与分类函数

2. 给定超平面与分类函数

3. 给定超平面与分类函数

4. 给定超平面与分类函数

5. 给定超平面与分类函数

6. 给定超平面与分类函数

二 求解优化问题 得到

最优解

$$w^* = (w_1^*, w_2^*, \dots, w_d^*)'$$

2. 根据 w^* 得到

$$w^* = \sum_{i=1}^n \alpha_i^* x_i$$

找出符合 $0 < \alpha_i^* < C$ 的点

(x_i, y_i) 计算

$$b^* = y_i - \sum_{j=1}^d w_j^* x_{ij}$$

4. 给定超平面 $w^* x + b^* = 0$

决策函数 $f(x) = \text{sign}(w^* x + b^*)$

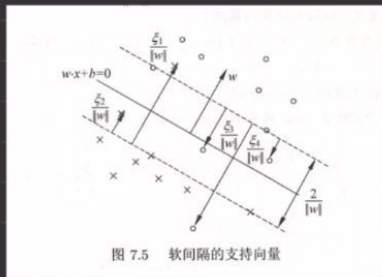


图 7.5 软间隔的支持向量

支持向量类型:

- $\alpha_i^* < C$: 间隔边界上的点 ($\xi_i^* = 0$)
- $\alpha_i^* = C$:
 - $0 < \xi_i^* < 1$: 间隔内的点
 - $\xi_i^* = 1$: 分离超平面上的点
 - $\xi_i^* > 1$: 误分类点

关键区别: 相比硬间隔, 软间隔的支持向量范围更广, 包含误分类点

3. 与感知机比较 03:01:03

1) 合页损失函数的定义与图像 03:01:39

● **函数表达式:** $L(z) = [z]_{+}$, 其中 $[z]_{+}$ 表示当 $z > 0$ 时取 z , $z \leq 0$ 时取 0

● **几何特征:** 函数图像在 $z > 0$ 时为 45 度斜线 (斜率 1), $z \leq 0$ 时为水平线, 整体呈 "荷叶" 形状

● **数学性质:** 损失值始终非负 ($L(z) \geq 0$), 在 $z = 0$ 处连续但不可导

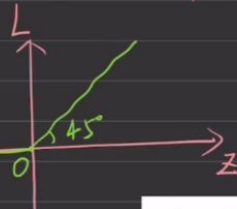
2) 线性支持向量机原始问题 03:03:23



线性支持向量机之合页损失

合页

$$L = [z]_+ = \begin{cases} z & z > 0 \\ 0 & z \leq 0 \end{cases}$$



线性支持向量机原始问题

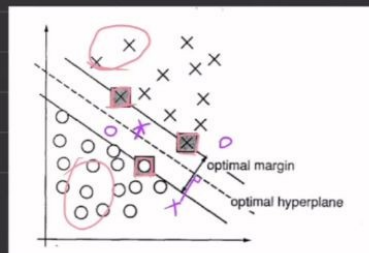
$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i$$

$$\text{s.t. } y_i(w \cdot x_i + b) + \xi_i \geq 1$$

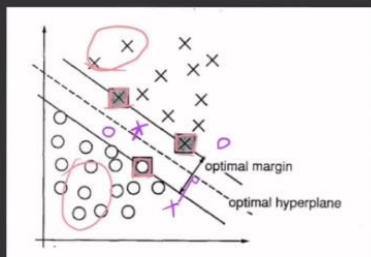
$$\xi_i \geq 0$$

$$i = 1, 2, \dots, N$$

$$\xi_i = [1 - y_i(w \cdot x_i + b)]_+ \quad \text{调皮}$$



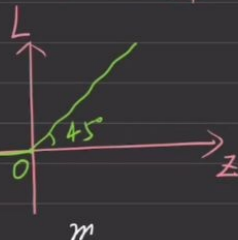
- 优化目标: $\min \frac{1}{2} \|w\|^2 + C \sum \xi_i$
- 约束条件: $y_i(w \cdot x_i + b) + \xi_i \geq 1 \quad (\xi_i \geq 0)$
- 参数维度: 原始问题需同时优化 w (n 维)、 b (1 维) 和 ξ_i (N 维), 共 $n + N + 1$ 维
- 3) 松弛因子的含义与作用 03:05:25
- 核心作用: 处理"调皮点" (误分类点或间隔内点)
- 取值规则:
 - 边界内点: $\xi_i = 1 - y_i(w \cdot x_i + b)$
 - 误分类点: $\xi_i = -y_i(w \cdot x_i + b) + 1$
- 特殊性质: 对正常分类点 ($y_i(w \cdot x_i + b) \geq 1$) $\xi_i = 0$
- 4) 目标函数的优化与替代 03:08:47



线性支持向量机之合页损失

合页

$$L = [z]_+ = \begin{cases} z & z > 0 \\ 0 & z \leq 0 \end{cases}$$



线性支持向量机原始问题

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

$$\text{s.t. } y_i(w \cdot x_i + b) + \xi_i \geq 1$$

$$\xi_i \geq 0$$

$$i = 1, 2, \dots, N$$

$$\xi_i = [1 - y_i(w \cdot x_i + b)]_+ \quad \text{调皮}$$

- 替代形式: 用合页损失表示松弛变量 $\xi_i = [1 - y_i(w \cdot x_i + b)]_+$
- 简化结果: 优化参数减少为 w 和 b , 目标函数变为

$$\frac{1}{2} \|w\|^2 + C \sum [1 - y_i(w \cdot x_i + b)]_+$$

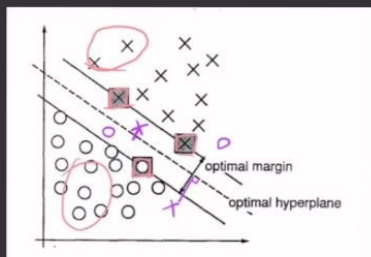
5) 正则化与目标函数的等价变换 03:09:55

- 正则化视角: 令 $\lambda = \frac{1}{2C}$, 目标函数可表示为 $\sum [1 - y_i(w \cdot x_i + b)]_+ + \lambda \|w\|^2$

结构分析:

- 前: 合页 ()
- 后项: L2 正则项 (结构风险)

6) 例题 1: 损失函数比较 03:11:15



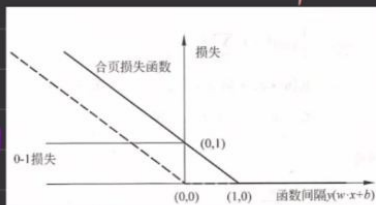
线性支持向量机之合页损失

合页

$$L = [z]_+ = \begin{cases} z & z > 0 \\ 0 & z \leq 0 \end{cases}$$

线性支持向量机原始问题

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & y_i (w \cdot x_i + b) + \xi_i \geq 1 \\ & \xi_i \geq 0 \\ & i = 1, 2, \dots, N \end{aligned}$$



$$\xi_i = [1 - y_i (w \cdot x_i + b)]_+$$

调皮

正则化

$$\min_{w, b} \sum_{i=1}^N [1 - y_i (w \cdot x_i + b)]_+ + \lambda \|w\|^2$$

● 比较维度:

- 横轴: 函数间隔 $t = y(w \cdot x + b)$
- 纵轴: 损失值

● 函数特性:

- **0-1 损失**: 不连续 ($t \leq 0$ 时 $L=1$, $t > 0$ 时 $L=0$)
- **感知机损失**: 线性下降 ($L=-t$ 当 $t \leq 0$, $L=0$ 当 $t > 0$)
- **SVM 合页损失**: $L=1-t$ 当 $t \leq 1$, $L=0$ 当 $t > 1$ (是感知机损失的上界)

● 关键结论: 合页损失要求更严格 ($t > 1$ 才无惩罚), 这是 SVM 能获得唯一超平面的原因

五、非线性支持向量机 03:14:58

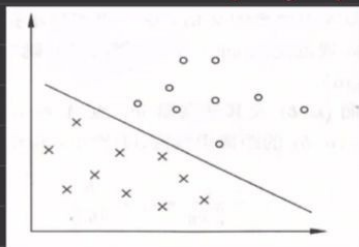
1. 非线性可分定义 03:15:16

- **核心概念**: 非线性可分指训练数据集 T 存在一个分离超曲面 (而非超平面) 能将所有实例点完全分成两类。以二维空间为例, 当输入特征为 (x_1, x_2) 时, 可能需要椭圆等曲线才能实现完全分类。
- **对比说明**: 与线性可分的关键区别在于分离边界的形式——线性可分使用超平面 (如直线 $w \cdot x + b = 0$), 非线性可分则需要更复杂的超曲面 (如椭圆 $w_1 x_1^2 + w_2 x_2^2 + b = 0$)。

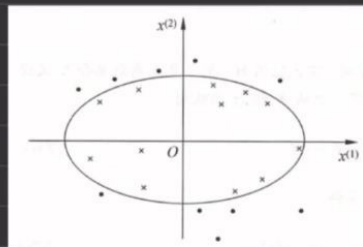
2. 空间映射方法 03:16:26

1) 分离超平面与线性可分 03:16:29

非线性支持向量机



线性支持向量机



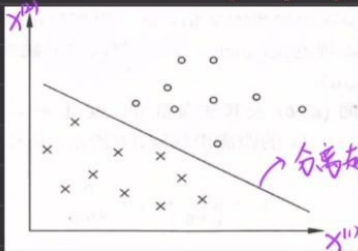
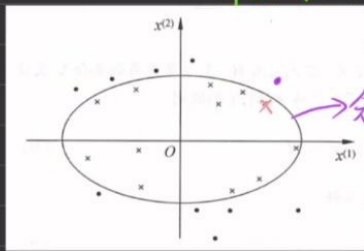
① 线性可分

$$T: w \cdot x + b = 0$$

- 基本形式：线性可分场景中分离超平面方程为 $w \cdot x + b = 0$ ，二维情况下表现为直线。
 - 推广性：“超平面”的使用是为了将二维概念推广到高特征空。
- 2) 线性不可分与软间隔算法 03:17:05
- 解决方法：通过引入松弛变量实现软间隔分类，允许存在少量误分类点。
 - 本质限制：该方法仍基于超平面分类，仅适用于近似线性可分场景（误分类点较少时）。
- 3) 非线性可分与非线性支持向量机 03:18:15

Janneil博士 bilibili
如何将原空间映射
新空间上?

非线性支持向量机

线性支持向量机

① 线性可分 ✓
 $T: w \cdot x + b = 0$
 实例点 完全分开

② 线性不可分 ✓ 已解决 $\Rightarrow$ 新问题

非线性支持向量机

① 非线性可分
 $T: \text{超曲面}$ 完

② 非线性不可分
 欧式 $w_i(x^{(i)})$
 Hilbert $w_i \overline{z_i^{(i)}}$

- 典型景：当误分类点过多（如图中椭圆分类情况），必须采用非线性分类方法。
- **分类边界**：此时需要超曲面作为决策边界，例如椭圆方程 $w_1 x_1^2 + w_2 x_2^2 + b = 0$ 。
- 4) 非线性支持向量机中的非线性可分和非线性不可分 03:20:27
- **分类标准**：
 - 非线性可分：存在超曲面能完全分离所有实例点
 - 非线性不可分：即使使用超曲面仍存在误分类点（需引入类似软间隔的机制）
- 5) 超曲面的多变结构与线性转化思路 03:21:02
- **结构复杂性**：超曲面可以是椭圆、三次/四次方程等任意形式，缺乏一表达式。
- 解决思路：将非线性问题转化为已解决的线性问题，通过空间映射实现。
- 6) 非线性转线性的具体方法 03:22:03

线性可分支持向量机

$$\min \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^N \alpha_i$$

$$\text{s.t.} \sum_{i=1}^N \alpha_i y_i = 0$$

$$\alpha_i \geq 0, \quad i=1, 2, \dots, N$$



● 具体步骤：

○ 设定原空间非线性方程（如椭圆 $w_1 x_1^2 + w_2 x_2^2 + b = 0$ ）

○ 做变量替换： $z_1 = x_1^2, z_2 = x_2^2$

○ 得到希尔伯特空间中的线性方程： $w_1 z_1 + w_2 z_2 + b = 0$

● 数学本质：将欧式空间的非线性问题转化为希尔伯特空间的线性问题。

7) 空间映射的待解决问题 03:24:26

● **关键难点**：如何构造从原空间到新空间的映射函数 $\varphi(x)$ 。

● **核心需求**：需要在新空间中能计算内积 $\langle \varphi(x_i), \varphi(x_j) \rangle$ ，这是支持向量机对偶问题求解的基础。

● **解决方案**：引入核函数 $K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$ 直接计算映射后的内积。

3. 核函数引入 03:25:01

1) 正定核判定 03:27:52

线性可分支持向量机

$$\min \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N d_i d_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^N d_i$$

$$\text{s.t. } \sum_{i=1}^N d_i y_i = 0$$

$$d_i \geq 0, \quad i=1, 2, \dots, N$$

原空间: 输入空间 X

新空间: Z $z_i \cdot z_j$

欧式空间 $x_i \in X$

希尔伯特空间 $z_i \in Z$

$\phi(x): X \rightarrow Z$

$$z_i = \phi(x_i), \quad z_j = \phi(x_j)$$

$$z_i \cdot z_j = \phi(x_i) \cdot \phi(x_j)$$

空间映射概念

- 原空间特征: 输入数据最初所在的欧式空间, 所有实例点都来源于这个我们最熟悉的性空。
- 新空特性: 通过映射得到的希尔伯特空间 (Hilbert space), 具有更丰富的数学结构, 可以处理非线性问题。
- 映射目的: 寻找一个映射 ϕ 将数据从欧式空间 X 变换到希尔伯特空间 H , 即 $\phi: X \rightarrow H$ 。

核函数定义

- 内积转换: 在希尔伯特空间中, $z_i = \phi(x_i)$ 与 $z_j = \phi(x_j)$ 的内积 $z_i \cdot z_j$ 等价于原空间核函数 $K(x_i, x_j)$ 。
- 核函数本质: 直接定义为特征映射的内积 $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$, 避免了显式计算高维映射。
- 优化问题转化: 非线性支持向量机的优化问题中, 只需将原始内积替换为核函数形式 $K(x_i, x_j)$ 即可。

核函数存在条件

- 关键问题: 不需要显式找到映射 ϕ , 只需验证函数 K 是否满足:
 - 对称性: $K(x_i, x_j) = K(x_j, x_i)$
 - 正定性: 对任意 $x_1, \dots, x_n$, Gram 矩阵 $[K(x_i, x_j)]$ 是半正定的
- 判定优势: 通过 Mercer 条件可直接验证函数是否为正定核, 无需构造具体的特征空间。

典型核函数示例

- 多项式核: $K(x, y) = (x \cdot y + c)^d$ 通过调节 d 实现不同阶数的非线性划分
- 高斯核: $K(x, y) = \exp\left(\frac{-\|x - y\|^2}{2\sigma^2}\right)$ 具有无限维特征空间的映射能力
- Sigmoid 核: 模拟神经网络激活函数

2) 核函数示例 03:30:50

核函数与映射关系 03:30:59

- 核函数本质：通过映射 ϕ 将输入空间的内积转换为新空间的内积，记作 $k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$
- 关键特性：在优化问题中只需确定核函数 k ，无需显式求出映射 ϕ
- 应用场景：将非线性可分问题通过核技巧转化为线性可分问题
- 例题 1：二维到三维的映射示例 03:33:17

Janneil博士 bilibili

线性可分支持向量机

$$\min \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N d_i d_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^N d_i$$

st. $\sum_{i=1}^N d_i y_i = 0$

$d_i \geq 0, i=1, 2, \dots, N$

原空间：输入空间 X

新空间： Z $z_i \cdot z_j$

$\phi(x_i), \phi(x_j)$ 内积

$k(x_i, x_j)$ 核函数

欧式空间 $x_i \in X$

希尔伯特空间 $z_i \in Z$

$\phi(x): X \rightarrow Z$

$z_i = \phi(x_i), z_j = \phi(x_j)$

$z_i \cdot z_j = \phi(x_i) \cdot \phi(x_j) = k(x_i, x_j)$

? ϕ k

- 已知条件：核函数 $k(x, z) = (x \cdot z)^2$ ，输入空间为二维欧式空间
- 映射构造：
 - 取 $\phi(x) = i$
 - 验证： $\phi(x) \cdot \phi(z) = x_1^2 z_1^2 + 2x_1 x_2 z_1 z_2 + x_2^2 z_2^2$ 等于核函数展开式
- 结论：该映射能正确表达给定核函数
- 例题 2：二维到三维的映射示例二 03:35:11
 - 替代映射：
 - $\phi(x) = i$
 - 验证发现同样满足核函数要求
 - 重要发现：同一核函数可能对应多个不同的映射形式
- 例题 3：二维到四维的映射示例 03:36:43
 - 扩展维度：
 - 构造四维映射 $\phi(x) = [x_1^2, x_1 x_2, x_2^2, 1]$
 - 通过计算验证仍满足核函数要求
 - 核心结论：映射维度可以不同，只要保证内积结果与核函数一致
- 核函数确定的重要性 03:37:44
 - 优化关键：在支持向量机中，实际起作用的是核函数而非具体映射
 - 实践意义：无需显式构造映射空间，直接通过核函数计算内积
- 如何找正定核 03:38:09

如何找正定核?

- 基本要求：必须满足对称性 $k(x, z) = k(z, x)$
- **正定性条件**：对任意输入 $x_1, \dots, x_m$ ，Gram 矩阵必须半正定
 - Gram 矩阵元素： $K_{ij} = k(x_i, x_j)$
 - 半正定要求： $\forall v \in R^m, v^T K v \geq 0$
- 正定核函数的定义 03:40:03
 - 数学本质：对应再生核希尔伯特空间(RKHS)的再生核
 - **物理意义**：保证特征空间中的内积具有非负性
- 核函数对称性 03:40:40
 - 必然性质：由内积的对称性决定， $k(x, z) = \varphi(x) \cdot \varphi(z) = \varphi(z) \cdot \varphi(x) = k(z, x)$
 - **验证要点**：构造核函数时必须首先检查对称性
- 核函数 Gram 矩阵 03:42:12
 - **矩阵构造**：由训练样本两两核函数值组成的 $m \times m$ 矩阵
 - **实际应用**：在 SVM 优化问题中替代原始内积计算
 - 方法：通过特征值检验半正定性（所有特征值非负）
- 半正定矩阵的判断方法 03:45:10
 - 定义与相关概念
 - **定义**：对于矩阵 A ，若对任意非零向量 x ，其二次型 $x^T A x \geq 0$ ，则称 A 为半正定矩阵。
 - 相关概念：
 - **正定矩阵**： $x^T A x > 0$ （严格大于零）
 - **半负定矩阵**： $x^T A x \leq 0$
 - **负定矩阵**： $x^T A x < 0$ （严格小于零）
 - 特征值判别法
 - 核心原理：通过矩阵对角化后的特征值判断
 - 若矩阵 A 的所有特征值 $\lambda_i \geq 0$ ，则 A 半正定
 - 对应对角阵 D 的元素（即特征值）全非负时，二次型可表示为平方和

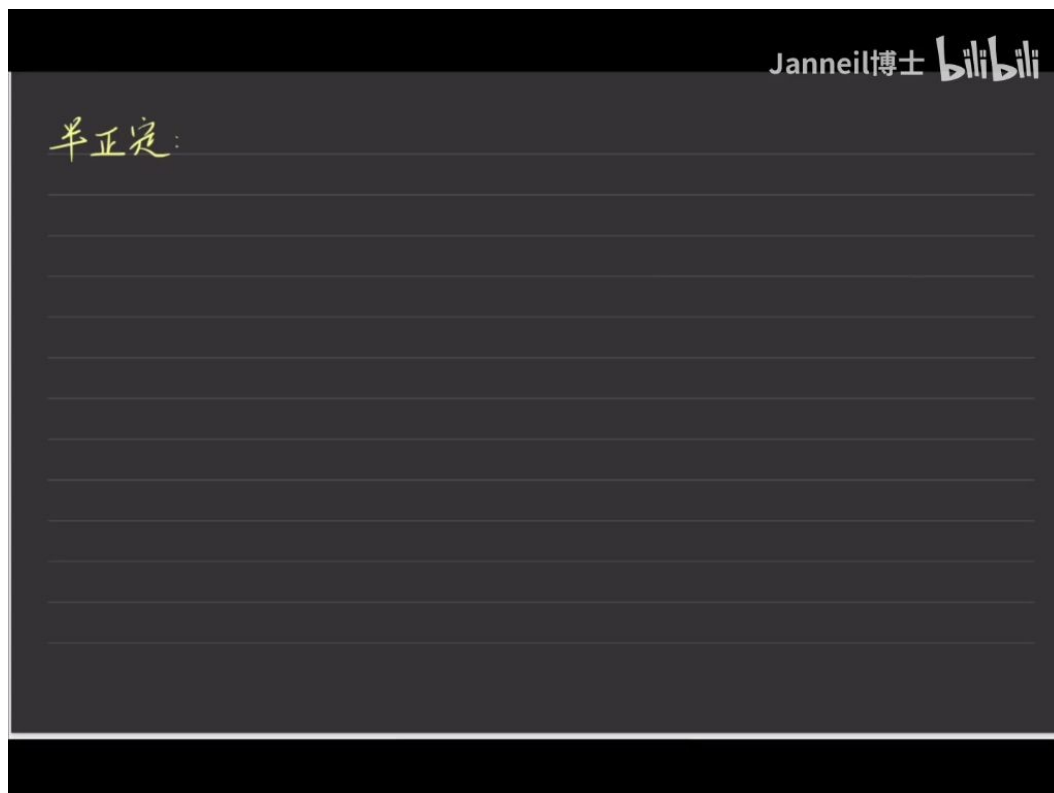
$$\sum \lambda_i y_i^2 \geq 0$$
 - **优势**：计算简便，适用于数值计算场景
 - 主子式判法

- **判断条件**：需验证所有顺序主子式非负
 - 从左上角开始的各阶子矩阵行列式 ≥ 0
 - 包括 1 阶主子式（单个元素）、2 阶主子式，直至 n 阶（整个矩阵行列式）
- **注意事项**：
 - 仅行列式 ≥ 0 不能单独作为判定依据
 - 必须检查所有顺序主子式（不同于正定矩阵只需前 $n-1$ 阶顺序主子式 >0 且行列式 >0 ）
- 正定矩阵与优化问题 03:49:50
 - 优化问题中的应用
 - 二阶导函数关系：正定矩阵 A 对应二次型的 Hessian 矩阵（二阶导函数矩阵）
 - **优化保证**：
 - 当 A 正定时，目标函数为严格凸函数
 - 满足一阶导数为零的点即为全局最小值点（费马原理的直接应用）
 - **实际意义**：确保优化问题有唯一全局最优解，避免陷入局部最优
 - 半正定矩阵的优化特性
 - **函数性质**：对应凸函数（非严格凸）
 - **优化特点**：
 - 可能存在多个最小值点（当二次型在某些方向为零时）
 - 仍保证一阶条件得到的临界点是最优点
 - 常用于带约束的优化问题中构建拉格朗日函数

六、核函数理论 03:51:01

1. 希尔伯特空间 03:51:18

●



- **半正定条件**：对称函数成为正定核函数的充分条件是半正定性，即只需要满足半镇定条件即可。
- **构造目标**：通过对称函数构造希尔伯特空间，在定义内积的基础上找到投影空间。
- 2. 空间构造步骤
 - **分步实现**：具体构造过程需要逐步完成：①定义映射→②构成向量空间→③进化到希尔伯特空。

七、空基本概念 03:52:49

1. 空间定义

- **空间本质**：空间本质上是元素的集合，当集合元素为向量时即为向量空间（如二维/三维空间）。

Janneil博士

② $K(x, z)$ 构成一个 Hilbert 空间

具体：

1. ϕ 构成向量空间

$\phi: x \rightarrow K(\cdot, x)$

对 $\forall x_i \in X, \alpha_i \in \mathbb{R}, i=1, \dots, m,$
定义 $K(\cdot, x_1), \dots, K(\cdot, x_m)$

$f(\cdot) = \sum_{i=1}^m \alpha_i K(\cdot, x_i)$

f 构成集合 S

欧式空间

$\mathbb{R}^p$

向量空间
线性空间

$(+, \times)$ 运算是封闭的

数乘

内积空间
 $(+, \times, \cdot)$ 封闭的

赋范线性空间

Banach 空间

2. 线性空间 03:53:06

- **封闭性要求**：向量空间必须对线性组合封闭，即满足：
 - **加法封闭**： $\forall f, g \in S, f+g \in S$
 - **数乘封闭**： $\forall a \in \mathbb{R}, af \in S$
- **验证方法**：通过任意取样的线性组合 $f = \sum_{i=1}^m \alpha_i k(\cdot, x_i)$ 和 $g = \sum_{j=1}^l \beta_j k(\cdot, z_j)$ 验证其和仍属于 S 。

八、内空 03:54:18

1. 运算扩展

- **内积引入**：在向量空间基础上定义内积运算（记作点积），形成内积空间：
 - **运算符号**：向量乘法记作叉号 $\times$ ，内积记作点号 $\cdot$
 - **功能扩展**：内积可计算向量夹角和长度

2. 范数空间 03:54:44

- **范数定义**：通过内积定义范数（如二范数表示距离），形成赋范线性空间
- **欧式空间**：在 $\mathbb{R}^p$ 中定义内积和范数即构成欧式度量空间

九、完空 03:56:06

1. 完备性要求

- **巴拿赫空间**：完备的赋范线性空间（所有极限点都在空间内）
- **希尔伯特空间**：完备的内积空间，具有：
 - 线性结构（加法和数乘）
 - 内积运算
 - 完备性（极限封闭）

十、构造程 03:58:42

1. 具体步骤

② $K(x, z)$ 构成一个 Hilbert 空间

具体:

1> ϕ 构成向量空间

$$\phi: x \rightarrow K(\cdot, x)$$

对 $\forall x_i \in X, \alpha_i \in \mathbb{R}, i=1, \dots, m,$
 $K(\cdot, x_1), \dots, K(\cdot, x_m)$

定义

$$f(\cdot) = \sum_{i=1}^m \alpha_i K(\cdot, x_i)$$

f 构成集合 S

Banach 空间

欧氏空间

$\mathbb{R}^p$

向量空间
线性空间

$(+, \times)$ 运算是封闭的

数乘

内积空间

$(+, \times, \cdot)$ 封闭的

赋范线性空间

- 向量空间构造：定义映射 $\phi: x \rightarrow k(\cdot, x)$ ，生成线性组合集合 $S = \{\sum_{i=1}^m \alpha_i k(\cdot, x_i)\}$
 - 内积定义：在 S 上定义内积 $\langle f, g \rangle = \sum_{i,j} \alpha_i \beta_j k(x_i, z_j)$
 - 完备化：通过添加极限点完成希尔伯特空间构造
2. 验证实例 04:01:39
- 加法验证： $f+g = \sum_{k=1}^{m+l} a_k k(\cdot, u_k) \in S$ (合并同类项)
 - 数乘验证： $af = \sum (a\alpha_i) k(\cdot, x_i) \in S$ (系数仍为实数)

从 S 中 $\forall f, g$, $f = \sum_{i=1}^m \alpha_i K(\cdot, x_i)$ $g = \sum_{j=1}^k \beta_j K(\cdot, z_j)$

"+" $f+g = \sum_{i=1}^m \alpha_i K(\cdot, x_i) + \sum_{j=1}^k \beta_j K(\cdot, z_j)$

$= \sum_{i=1}^{m+k} a_i K(\cdot, u_i)$ $u_1, \dots, u_{m+k}$
 $\in S$ $a_1, \dots, a_{m+k}$

十一、内积定义验证 04:05:36

② $K(x, z)$ 构成一个 Hilbert 空间

具体:

1> ϕ 构成向量空间

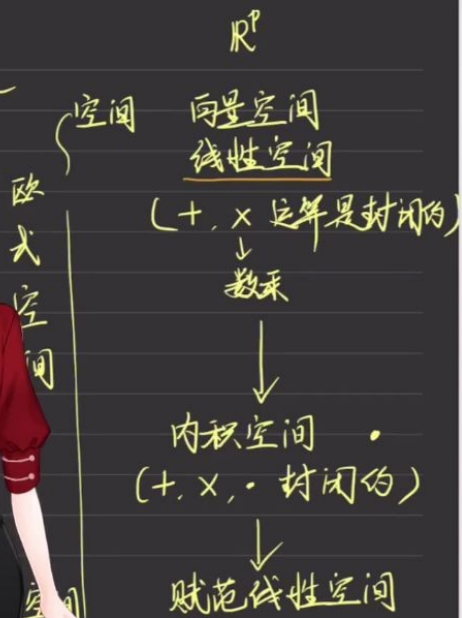
$\phi: x \rightarrow K(\cdot, x)$

对 $\forall x_i \in X, \alpha_i \in \mathbb{R}, i=1, \dots$
 $K(\cdot, x_i)$

定义

$f(\cdot) = \sum_{i=1}^m \alpha_i K(\cdot, x_i)$

f 构成集合 S



- **空间构造**：通过核函数 $K(x, z)$ 构建 Hilbert 空间，包含线性空间、内积空间和赋范线性空间结构，要求加法和数乘运算封闭。

- **映射定义**：定义映射 $h: X \rightarrow k(0, X)$ ，对任意 $x_i \in X$ 和系数 $\alpha_i \in R$ ，组合 $k(0, x_1)$ 等元素构成完整空。

1. 数乘条件验证

- **条件表述**： $(cf) \cdot g = c(f \cdot g)$ ，其中 $c \in R$

- **验证过程**：

- 将 cf 展开为 $\sum c \alpha_i k(x_i, \cdot)$
- 根据内积定义展开 $(cf) \cdot g$ 得到 $\sum c \alpha_i \beta_j k(x_i, z_j)$
- 提取常数 c 后与右边 $c(f \cdot g)$ 完全一致

- **关键技巧**：利用线性组合系数的可提取性，保持核函数 k 的结构不变

2. 分配律验证 04:06:55

- **条件表述**： $(f+g) \cdot h = f \cdot h + g \cdot h$

- **验证方法**：

- 合并 $f+g$ 的表达式为 $\sum_{i=1}^{m+l} a_i k(u_i, \cdot)$
- 与 $h = \sum b_q k(v_q, \cdot)$ 做内积得 $\sum a_i b_q k(u_i, v_q)$
- 拆分后与右边两项内积之和完全对应

- **注意点**：需要统一指标表示法，将 f 和 g 的系数与基函数重新编号合并

3. 交换律验证 04:07:55

- **验证依据**：

- 核函数对称性： $k(x_i, z_j) = k(z_j, x_i)$
- 系数可交换： $\alpha_i \beta_j = \beta_j \alpha_i$

- **特殊性质**：该条件是四条中最易验证的，直接由核函数定义保证

4. 正定性验证 04:08:31

●

Janneil博士 bilibili

2> 在 S 上定义内积. 内积空间

"*" 对 $\forall f, g \in S$. $f * g = \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j k(x_i, z_j)$

→ ? 内积

四个条件

(1) $(cf) * g = c(f * g)$, $c \in R$

(2) $(f+g) * h = f * h + g * h$, $h \in S$, 分配律

(3) $f * g = g * f$ ✓ 交换律

(4) $f * f \geq 0$, 特别地, $f * f = 0 \Leftrightarrow f = 0$

3 / 49

- **核心条件**：

- $f \cdot f \geq 0$
- $f \cdot f = 0 \Leftrightarrow f = 0$

- **Gram 矩阵应用**：

- 构造矩阵 $A = [k(x_i, x_j)]_{m \times m}$
- 利用二次型 $\alpha^T A \alpha \geq 0$ 的性质

- **充要证明：**
 - 充分性： $f=0$ 时所有 $\alpha_i=0$ ，内积自然为0
 - 必要性：通过 Cauchy-Schwarz 不等式推导零向量的唯一性

十二、正定核定理 04:15:21

1. 充要条件

- 零向量判定： $f \cdot f = 0$ 时，由 Gram 矩阵正定性可推出所有系数 $\alpha_i=0$
- **构造方法：**通过线性组合 $f = \sum \alpha_i k(x_i, \cdot)$ 的表示形式

2. Gram 矩阵 04:16:04

●

Janneil博士 bilibili

$f = \sum_{i=1}^n \alpha_i k(x_i, \cdot) \quad g = \sum_{j=1}^m \beta_j k(x_j, \cdot)$
 $f * g = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j k(x_i, x_j)$

(4) $f * f \geq 0$,

特别地： $f * f = 0 \Leftrightarrow f = 0$

- **矩阵性质：**
 - 对称性： $A^T = A$
 - 半正定性： $\forall \alpha \in R^m, \alpha^T A \alpha \geq 0$
- **关联内积：** $f \cdot f$ 实质是 $\alpha^T A \alpha$ 的展开形式
- 3. 再生性证明 04:17:29
- **关键不等式：** $(f \cdot g)^2 \leq (f \cdot f)(g \cdot g)$
- **证明技巧：**
 - 构造 $f + \lambda g$ 的范数平方
 - 利用二次函数判别式
 - 最终导出 Cauchy-Schwarz 不等式形式
- **应用价值：**保证内积空间中的夹角概念合理性

十三、内积空间与希尔伯特空间

Janneil博士 bilibili

(4) $f * f \geq 0$,

特别地: $f * f = 0 \Leftrightarrow f = 0$
充要条件

必要性:

$f = \sum_{i=1}^n \alpha_i K(\cdot, x_i)$ $g = \sum_{j=1}^n \beta_j K(\cdot, x_j)$
 $f * g = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \beta_j K(x_i, x_j)$

- **内积运算性质**: 对于任意 $f, g \in S$, 满足 $(f + \lambda g, f + \lambda g) \geq 0$, 展开后得到 $f \cdot f + 2\lambda(f \cdot g) + \lambda^2(g \cdot g) \geq 0$
- **判别式条件**: 上述关于 λ 的二次不等式恒成立的条件是判别式 $\Delta \leq 0$, 即 $4(f \cdot g)^2 - 4(f \cdot f)(g \cdot g) \leq 0$
- **柯西不等式**: 由此可证得 $(f \cdot g)^2 \leq (f \cdot f)(g \cdot g)$, 这是内积空间的重要性质
- **特殊取值法**: 取特定 $g = K(x, \cdot)$ 时, 通过内积运算可得 $\alpha_i = 0$, 从而证明 $f = 0$
- **充要条件**: $f \cdot f = 0 \Leftrightarrow f = 0$, 这是内积定义的必要条件

十四、空间升级过程

Janneil博士 bilibili

$f = \sum_{i=1}^m \alpha_i K(\cdot, x_i)$ $g = \sum_{j=1}^l \beta_j K(\cdot, z_j)$
 $f * g = \sum_{i=1}^m \sum_{j=1}^l \alpha_i \beta_j K(x_i, z_j)$

(4) $f * f \geq 0$, $\checkmark$
特别地: $f * f = 0 \Leftrightarrow f = 0$
充要条件

必要性: $\checkmark$
 $\checkmark$ 先证: $\forall f, g \in S, (f * g)^2 \leq (f * f)(g * g)$
 $\lambda \in \mathbb{R}, f + \lambda g \in S \Rightarrow (f + \lambda g) * (f + \lambda g) \geq 0$

恒不等式 $\Rightarrow f * f + 2\lambda(f * g) + \lambda^2(g * g) \geq 0$
 $4(f * g)^2 - 4(g * g)(f * f) \leq 0$
 $(g * g)\lambda^2 + 2(f * g)\lambda + f * f \geq 0$
 $a\lambda^2 + b\lambda + c \geq 0$
 $\Delta = \sqrt{b^2 - 4ac} \leq 0$
 $\Rightarrow (f * g)^2 \leq (f * f)(g * g)$

- **内积空间定义**: 在向量空间上定义内积运算 $f \cdot g = \sum_{i=1}^m \sum_{j=1}^l \alpha_i \beta_j k(x_i, z_j)$
- **范数定义**: $\|f\| = \sqrt{f * f}$, 将内积空间升级为赋范线性空间
- **完备化过程**: 通过完备化赋范线性空间, 最终得到希尔伯特空间

十五、常用核函数 04:21:24

1. 多项式核函数

- 形式: $K(x, z) = (x \cdot z + 1)^p$, 其中 p 为多项式次数
- 特点: 适用于多项式特征映射的情况

2. 高斯核函数 04:22:05

- 形式: $K(x, z) = \exp\left(\frac{-\|x - z\|^2}{2\sigma^2}\right)$
- 特点: 具有局部性强、平滑性好的特性

3. 字符串核函数 04:23:21

- **应用**: 专门用于处理字符串数据的核函数
- **原理**: 通过子序列匹配计算相似度

十六、非线性支持向量机 04:28:30

1. 算法流程

- 核心思想: 通过核函数将输入空间映射到高维特征空间
- 求解步骤: 在高维空间中求解线性可分问题

2. 决策函数 04:29:52

- 形式: $f(x) = \text{sign}\left(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b\right)$
- 特点: 决策函数仅依赖于支持向量和核函数计算

十七、优化算法 04:31:55

1. 坐标下降法

- 1) 谢尔伯特空间里的核函数 k 的再生性 04:32:17

- **再生性定义**：核函数 k 与希尔伯特空间元素 f 的内积可表示为 $\sum_{i=1}^m \alpha_i k(x, x_i)$ ，其中 x 为输入变量，这与原始 f 的表达式 $\sum_{i=1}^m \alpha_i k(\cdot, x_i)$ 具有相同形式。
 - **典型表现**： $k(x, \cdot)$ 与 $k(z, \cdot)$ 的内积等于 $k(x, z)$ ，这与映射 $\varphi(x)$ 和 $\varphi(z)$ 的内积形式完全一致，验证了核函数的再生特性。
- 2) 核函数 k 的介绍 04:34:36
- **空间转换原理**：通过核函数 k 实现从原始输入空间 X 到希尔伯特空间 H 的映射，将非线性可分问题转化为线性可分问题。
 - **核心特征**：再生核希尔伯特空间(RKHS)中的核函数必须满足再生性，即 $k(x, \cdot)$ 与 f 的内积能还原出 f 在 x 处的取值。
- 3) 判断正定和函数的小定理 04:34:57
- **定理内容**：对称函数 $k: X \times X \rightarrow R$ 是正定核函数的充要条件是，对任意 m 个样本点 $x_1, \dots, x_m$ 构成的 Gram 矩阵 K (其中 $K_{ij} = k(x_i, x_j)$) 是半正定矩阵。
 - **应用价值**：该定理为验证函数是否为正定核提供了矩阵判定方法，避免直接构造希尔伯特空间的困难。
2. SMO 算法 04:36:11
- 1) 正定核函数与半正定矩阵的关系 04:36:19
- **基本对应关系**：正定核函数 k 必然对应半正定的 Gram 矩阵，反之半正定的 Gram 矩阵也必然对应某个正定核函数。
 - **判定标准**：通过检查任意有限样本集生成的 Gram 矩阵是否半正定，即可判断函数 k 的正定性。
- 2) 定理证明：充分性 04:37:12
- **证明思路**：已知 Gram 矩阵半正定，可构造映射 $\varphi: x \rightarrow k(\cdot, x)$ 将输入空间嵌入到再生核希尔伯特空间。
 - **关键步骤**：
 - 定义内积 $\langle k(\cdot, x), k(\cdot, z) \rangle = k(x, z)$
 - 验证该内积满足再生性 $\langle f, k(\cdot, x) \rangle = f(x)$
 - 由此导出 k 必为正定核
- 3) 定理证明：必要性 04:38:38
- **证明思路**：已知 k 是正定核，则存在映射 φ 使得 $k(x, z) = \langle \varphi(x), \varphi(z) \rangle$ 。
 - **矩阵推导**：
 - 对任意向量 $c \in R^m$ ，有 $c^T K c = \sum_{i,j} c_i c_j k(x_i, x_j)$
 - 代入内积形式得 $\langle \sum_i c_i \varphi(x_i), \sum_j c_j \varphi(x_j) \rangle \geq 0$
 - 故 Gram 矩阵 K 必为半正定
3. 参数更新 04:39:44
- 1) 正定核的充要条件 04:39:46

正定核:

设 $K: X \times X \rightarrow R$ 是对称函数, 则 $K(x, z)$ 为正定核的充要条件是对 $\forall x_i \in X, i=1, 2, \dots, m$, $K(x, z)$ 对应的 Gram 矩阵

$$K = [K(x_i, x_j)]_{m \times m}$$

是半正定矩阵.

- **定义**: 设 $K: X \times X \rightarrow R$ 是对称函数, 则 $K(x, z)$ 为正定核的充要条件是对 $\forall x_i \in X, i=1, 2, \dots, m$, $K(x_i, x_j)$ 对应的 Gram 矩阵 $K = [K(x_i, x_j)]_{m \times m}$ 是半正定矩阵。
 - **核心关系**: $K(x, z)$ 是正定核函数当且仅当其 Gram 矩阵 K 是半正定矩阵。
- 2) 半正定矩阵的判断方法 04:41:33
- **验证方法**: 任取实数 $c_1, \dots, c_m \in R$ 构成向量 $c = (c_1, \dots, c_m)^T$, 若 $c^T K c \geq 0$ 恒成立, 则 K 为半正定矩阵。
 - **展开形式**: $c^T K c = \sum_{i=1}^m \sum_{j=1}^m c_i c_j K(x_i, x_j)$, 需要证明该二次型非负。
- 3) 内积运算展开法 04:43:44

- **关键步骤：**

- 将 Gram 矩阵元素表示为 $K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$
- 矩阵分解为 $K = \Phi^T \Phi$ ，其中 $\Phi = [\varphi(x_1), \dots, \varphi(x_m)]$
- 二次型转化为 $c^T K c = \langle \sum_{i=1}^m c_i \varphi(x_i), \sum_{i=1}^m c_i \varphi(x_i) \rangle \geq 0$

- **结论：**由于范数平方非负，得证 Gram 矩阵半正定。

4) 线性组合法 04:46:31

$$\forall c_1, \dots, c_m \in \mathcal{R}, C = [c_1, \dots, c_m]^T$$

$$C^T K C \geq 0 \Rightarrow K \text{ 是半正定的}$$

$$C^T K C = [c_1, \dots, c_m] \begin{bmatrix} K(x_1, x_1) & \dots & K(x_1, x_m) \\ \vdots & & \vdots \\ K(x_m, x_1) & \dots & K(x_m, x_m) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}$$

②

● 展开过程：

- 将核函数展开为内积形式： $K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$
- 二次型改写为 $\sum_{i,j} c_i c_j \langle \varphi(x_i), \varphi(x_j) \rangle = \langle \sum_i c_i \varphi(x_i), \sum_j c_j \varphi(x_j) \rangle$
- 最终表示为 $\langle \sum_i c_i \varphi(x_i), \sum_i c_i \varphi(x_i) \rangle \geq 0$

● 优势：直接展示核函数与特征空间内积的关系。

5) 正定核的等价定义 04:50:07

正定核的等价定义:

设 $X \subseteq \mathbb{R}^n$, $K(x, z)$ 是定义在 $X \times X$ 上的对称函数, 如果对 $\forall x_i \in X, i=1, 2, \dots, m$, $K(x, z)$ 对应的 Gram 矩阵

$$K = [K(x_i, x_j)]_{m \times m}$$

是半正定矩阵, 则称 $K(x, z)$ 是正定核.

- **实用定义**: 设 $X \subseteq \mathbb{R}^n$, 若对称函数 $K: X \times X \rightarrow \mathbb{R}$ 对任意 $x_1, \dots, x_m \in X$, Gram 矩阵 $K = [K(x_i, x_j)]$ 半正定, 则称 K 为正定核。
- **应用方法**: 实践中先构造对称函数, 再验证其 Gram 矩阵的半正定性。
- 6) 例题 1: 常用核函数 04:51:33
- **分类标准**:
 - 连续欧式空间核函数
 - 离散数据集核函数
- **典型示例** (后续课程将详细介绍):
 - 多项式核
 - 高斯核
 - 字符串核
- 7) 例题 2: 欧式空间上的核函数 04:52:05
- 多项式核函数

常用

- 形式： $K(x, z) = (x \cdot z + 1)^p$ ，其中 p 表示多项式的次数
 - 决策函数： $f(x) = \text{sign}(\sum_{i=1}^N a_i^* y_i (x_i \cdot x + 1)^p + b^*)$
 - 一般形式：可扩展为 $(x \cdot z + c)^p$ ，其中 c 为可调参数常量
 - 特点：当 $c=1$ 时为特例，实际应用中 c 可取其他常数值
- 高斯核函数

常用核函数

1. 欧式空间

① 多项式核函数

$$K(x, z) = (x \cdot z + 1)^p$$

$$(x \cdot z + c)^p$$

常量

决策：

$$f(x) = \text{sign}(\sum_{i=1}^N a_i^* y_i (x_i \cdot x + 1)^p + b^*)$$

② 高斯核函数

$$K(x, z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right)$$

决策：

$$f(x) = \text{sign}(\sum_{i=1}^N a_i^* y_i \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) + b^*)$$



- 形式： $K(x, z) = \exp\left(\frac{-\|x - z\|^2}{2\sigma^2}\right)$
- 决策函数： $f(x) = \text{sign}\left(\sum_{i=1}^N a_i y_i \exp\left(\frac{-\|x - x_i\|^2}{2\sigma^2}\right) + b\right)$
- 别名：又称高斯径向基函数（RBF）核
- 特点：基于向量差的范数平方计算，与高斯分布概率密度函数的指数部分形式相同

8) 例题3：离散数据集上的核函数 04:57:16

- 字符串核函数
 - 应用场景：主要用于文本分析中的字符串相似度计算
 - 映射方法：将字符串映射到高维特征空间（如 2^3 、 2^4 维空间）
 - 子串匹配：对于长度为 n 的字符串 s ，查找与子串 u 完全匹配的子字符串
 - 权重计算：匹配子串权重为 λ^{l_i} ，其中 l_i 为子串长度（计算方式：末位置 - 首位置 + 1）
- 余弦相似度算

Janneil博士 bilibili

2. 定义在离散集合

余弦相似度 $\cos \theta = \frac{x \cdot y}{\|x\| \|y\|}$

$$[\phi_n(s)]_u = \sum_{i: s(i)=u} \lambda^{l(i)}$$

$\frac{1}{2+\lambda^2}$
 $\frac{\lambda^4}{\lambda^6 + 2\lambda^4} = \frac{k(\text{big}, \text{pig})}{\|k(\text{big}, \text{big})\| \|k(\text{pig}, \text{pig})\|}$

	big	pig	bag
big	λ^2	λ^2	λ^3
pig	λ^2	λ^2	λ^3
bag	λ^3	λ^3	λ^2

$K(\text{big}, \text{pig}) = \lambda^4$, $K(\text{big}, \text{bag}) = \lambda^6$

- 公式： $\cos \theta = \frac{K(s_1, s_2)}{\sqrt{\square}}$
- 示例计算：
 - big 与 pig 的内积： λ^4
 - big 与 big 的内积： $\lambda^6 + 2\lambda^4$
 - 最终相似度： $\frac{\lambda^4}{\lambda^6 + 2\lambda^4} = \frac{1}{2 + \lambda^2}$
- 应用：用于度量两个文本字符串之间的相似程度

9) 字符串和函数的概念 05:09:08

- 字符串集合表示：用 Σ 表示字符串集合，其中 s 是集合中的序列
- 度表示：
 - 记作 $|s|$ （两边用竖杠表示）
 - 计算包含空格，如 "lass das" 长度为 8

- 元素索引：
 - 每个元素表示为 $s_1, s_2, \dots, s_{i \in \mathbb{N}}$
 - 示例： $s_1 = l, s_2 = a, s_5 = \text{空格}, s_8 = s$
- 字符串连接：
 - 直接写 ST 表示连接
 - 示例： $s = \text{small}, t = \text{er} \rightarrow st = \text{smaller}$
 -

Janneil博士 bilibili

余弦相似度 $\cos \theta = \frac{x \cdot y}{\|x\| \|y\|}$

$k(\text{big}, \text{pig})$

$\|k(\text{big}, \text{big})\| \|k(\text{pig}, \text{pig})\|$

-pg	-ba	-ag
0	0	0
λ^3	0	0
0	λ^2	λ^2

$k(\text{big}, \text{pig}) = \lambda^6$

字符串核函数

Σ S $|S|$ 代表 S 的长度
 $s(1), s(2), \dots, s(|S|)$

st smaller
 $\quad \quad \quad s \quad t$

Σ^n big, pig, bag

$\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^n$

S 子串 u $i = (i_1, i_2, \dots, i_l)$
 $\quad \quad \quad \downarrow$
 $\quad \quad \quad \text{长度} \quad l(i) = i_{l(i)} - i_1 + 1$

- 长度分类集合：
 - Σ^n 表示长度为 n 的字符串集合
 - 示例： Σ^3 包含 "big", "pig", "bag"
- 全集表示：
 - $\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^n$ 表示所有字符串的集合
- 子串概念

○

Janneil博士 bilibili

字符串函数

Σ	S	$ S $ 代表 S 的长度 $S(1), S(2), \dots, S(S)$	lass das 8 $S(1) = l$ $S(2) = a$ $S(5) = \square$ $S(8) = s$
	st	<u>smaller</u> $S \quad b$	
	Σ^n	big, pig, bag Σ^3	
	$\Sigma^+ = \bigcup_{n=0}^{\infty} \Sigma^n$		
S	子串 u $\downarrow$ 长度	$i = (i_1, i_2, \dots, i_{ u })$ $l(i) = i_{ u } - i_1 + 1$	as (2,3) (2,4) (7,8)

- 子串定义：
 - 字符串 s 的子串 u 由 s 的部分元素组成
 - 下标集合记作 $I = \{i_1, i_2, \dots, i_{|u|}\}$
- 子串长度计算：
 - 公式： $l(u) = i_{|u|} - i_1 + 1$
 - 示例："lass das"中：
 - 子串"as"出现位置：(2,3)→长度 2, (2,4)→长度 3, (7,8)→长度 2
- 连续性子串：
 - 当 i 连续时， $l(u) = |u|$
 - 间断时 $l(u) < |u|$ (如(2,4)对应长度 3 < 2)

10) 字符串和函数的映射 05:17:23

字符串核函数

Σ	S	$ S $ 代表 S 的长度 $S(1), S(2), \dots, S(S)$	lass das 8 $S(1)=l$ $S(2)=a$ $\vdots$ $S(5)=\square$ $\vdots$ $S(8)=S$
	st	smaller $S \prec t$	
Σ^n		big, pig, bag Σ^3	
$\Sigma^+ = \bigcup_{n=0}^{\infty} \Sigma^n$			
S	子串 u $\downarrow$ 长度	$u = S(i_1)S(i_2)\dots S(i_{ u })$ $i = (i_1, i_2, \dots, i_{ u })$ $l(i) = i_{ u } - i_1 + 1$	as (2, 3) 2 (2, 4) 3 (7, 8) 2

- 映射定义：
 - 集合 S 包含长度 $\geq n$ 的字符串
 - 映射 $\phi_n: S \rightarrow H_n$ 到特征空间
 - 维度计算： $\dim \Sigma^n \vee \dim \Sigma^m$ (如 $n=2$ 时为 7 维)
- 映射计算：
 - 对子串 u , 结果包含 $\lambda^{l(u)}$
 - 示例：
 - "big" 中 "bi" $\rightarrow \lambda^2$
 - "big" 中 "bg" $\rightarrow \lambda^3$
 - "lass das" 中 "as" $\rightarrow 2\lambda^2 + \lambda^3$
 - "asd" $\rightarrow 2\lambda^5$
 - "nasdaq" 中 "asd" $\rightarrow \lambda^3$

线性支持向量机

输入: 训练集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$
 其中 $x_i \in \mathcal{X} \subseteq \mathbb{R}^d$, $y_i \in \mathcal{Y} \subseteq \{+1, -1\}$

输出: 分离超平面与分类决策函数

算法:

- 给定惩罚参数 $C > 0$, 构造优化问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j (x_i \cdot x_j) - \sum_{i=1}^n \alpha_i$$

$$\text{s.t. } \sum_{i=1}^n \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C$$

$$i=1, \dots, N$$
- 求解优化问题, 得到最优解

$$\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)^T$$
- 根据 α^* 得到

$$w^* = \sum_{i=1}^n \alpha_i^* y_i x_i$$
 挑出符合 $0 < \alpha_i^* < C$ 的点 (x_i, y_i) 计算

$$b^* = y_i - \sum_{i=1}^n \alpha_i^* y_i (x_i \cdot x_j)$$
- 分类决策函数: $f(x) = \text{sign}(w^* \cdot x + b^*)$

非线性支持向量机

算法:

- 给定惩罚参数 $C > 0$, 构造优化问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i$$

$$\text{s.t. } \sum_{i=1}^n \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C$$

$$i=1, \dots, N$$
- 求解优化问题, 得到最优解

$$\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)^T$$
- 根据 α^* 得到

$$w^* = \sum_{i=1}^n \alpha_i^* y_i K(\cdot, x_i)$$
 挑出符合 $0 < \alpha_i^* < C$ 的点 (x_i, y_i) 计算

$$b^* = y_i - \sum_{i=1}^n \alpha_i^* y_i K(x_i, x_j)$$
- 决策函数

$$f(x) = \text{sign}\left(\sum_{i=1}^n \alpha_i^* y_i K(x, x_i) + b^*\right)$$

- 算法步骤:
 - 构造优化问题:
 - 将线性内积 $x_i \cdot x_j$ 替换为核函数 $k(x_i, x_j)$
 - 求解优化问题:
 - 得到最优解 α^i
 - 计算参数:
 - $w^i = \sum \alpha_i^i y_i \phi(x_i)$
 - 选择支持向量计算 $b^i = y_j - \sum \alpha_i^i y_i k(x_i, x_j)$
 - 决策函数:
 - $f(x) = \text{sign}(\sum \alpha_i^i y_i k(x, x_i) + b^i)$
- 核心修改:
 - 所有内积运算替换为核函数计算
 - 映射到希尔伯特空间进行线性分类

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i \\ \text{s.t.} & \sum_{i=1}^N \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \\ & i=1, \dots, N \end{aligned} \quad \text{优化问题}$$

- **参数组成:** 优化问题的参数由拉格朗日乘子 α 组成, 包含 $\alpha_1, \alpha_2, \dots, \alpha_N$, 每个样本点对应一个拉格朗日乘子
- **计算难点:** 当样本量 N 很大时, 需要解决 N 维优化问题, 直接运算计算量大且速度慢
- **坐标下降法原理**
 - 核心思想: 每次只聚焦一个参数进行更新, 固定其他参数
 - **迭代步骤:**
 - 给定初始值 $x_1^{(0)}, x_2^{(0)}$
 - 固定 x_2 , 更新 x_1 使目标函数最小
 - 固定 x_1 , 更新 x_2 使目标函数最小
 - 重复迭代直至收敛
- **坐标下降法示例**

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

优化问题

$$\text{s.t. } \sum_{i=1}^N \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C$$

$$i=1, \dots, N$$

参数: $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^T$

$(x_1, y_1) \dots (x_N, y_N)$

N维变量

坐标下降法可以应用
在此线性支持向量机吗?

目标函数: $f(x_1, x_2) = x_1^2 + 2x_2^2 - x_1x_2 + 1$

求解过程:

- 固定 x_2 求 x_1 : $\frac{\partial f}{\partial x_1} = 2x_1 - x_2 = 0 \Rightarrow x_1 = \frac{x_2}{2}$
- 固定 x_1 求 x_2 : $\frac{\partial f}{\partial x_2} = 4x_2 - x_1 = 0 \Rightarrow x_2 = \frac{x_1}{4}$
- 通过迭代逐步逼近最优解

在 SVM 中的应用限制

- 约束条件限制: 由于约束条件 $\alpha_1 y_1 = -\sum_{i=2}^n \alpha_i y_i$, 当固定其他参数时 α_1 被唯一确定
- 失败原因: 无法实现参数更新, 坐标下降法不适用于 SVM 优化问题
- 启发意义:
 - 为后续 SMO 算法提供思路
 - 类似思想也应用于 MCMC 中的 Gibbs 抽样

13) 序列最小最优优化算法 05:39:30

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i \zeta$$

s.t. $\sum_{i=1}^N \alpha_i y_i = 0$
 $0 \leq \alpha_i \leq C$
 $i=1, \dots, N$

优化问题

参数: $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^T$

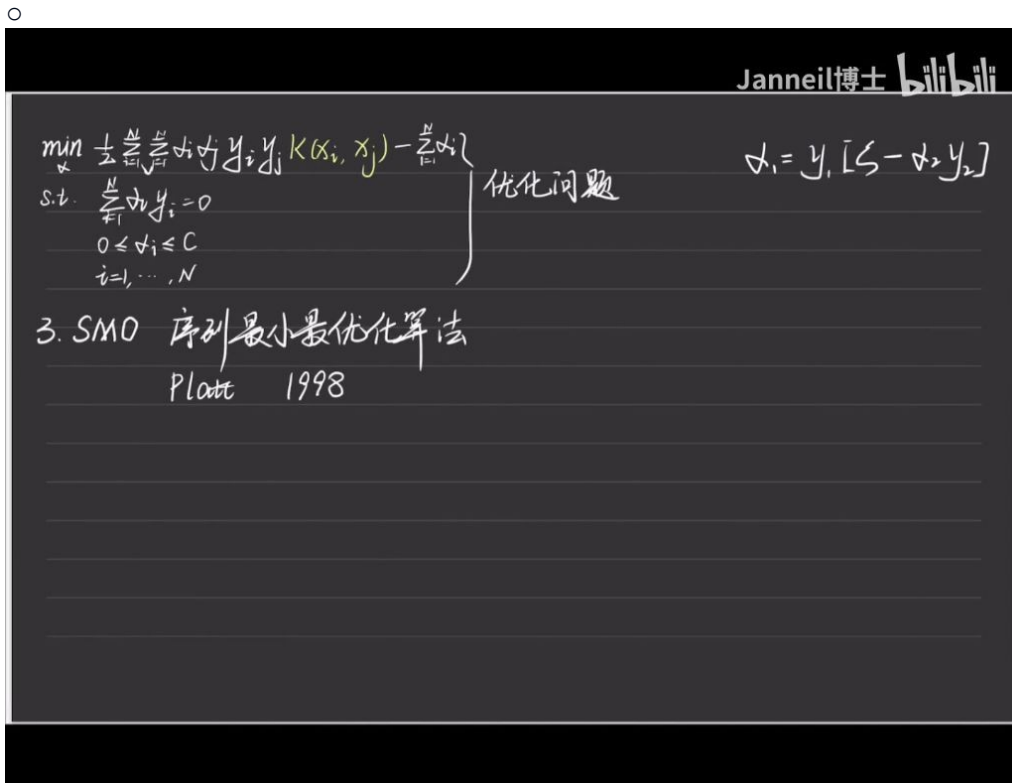
$\downarrow \quad \quad \quad \downarrow$
 $(x_1, y_1) \quad \dots \quad (x_N, y_N)$

N维变量

又, 坐标下降法 可以应用
 在此线性支持向量机吗?

✗

- 基本思想: 每次选择两个参数进行优化, 固定其他 $N-2$ 个参数
 - **参数关系:** 通过约束条件 $\alpha_1 y_1 + \alpha_2 y_2 = -\sum_{i=3}^N \alpha_i y_i$, 确定一个参数后另一个参数也随之确定
 - SMO 算法原理
 - **参数选择:** 选择 α_1 和 α_2 进行优化
- 参数表达式:
- $$\alpha_1 y_1 (\zeta - \alpha_2 y_2)$$
- 其中 $\zeta = -\sum_{i=3}^N \alpha_i y_i$ 由固定参数决定
 - 史背景: 1998 年由 Platt 在微软研究院提出
- 目标函数重构



- 简化形式: 将原 N 维问题简化为关于 α_1 和 α_2 的二维问题

目标函数表达式:

$$W(\alpha_1, \alpha_2) = -\alpha_1 - \alpha_2 + \frac{1}{2} K_{11} \alpha_1^2 + \frac{1}{2} K_{22} \alpha_2^2 + y_1 y_2 K_{12} \alpha_1 \alpha_2 + \alpha_1 y_1 v_1 + \alpha_2 y_2 v_2$$

- 其中 $v_1 = \sum_{i=3}^n \alpha_i y_i K_{1i}$, $v_2 = \sum_{i=3}^n \alpha_i y_i K_{2i}$
- 核函数简化: 利用 Gram 矩阵表示核函数 $K_{ij} = K(x_i, x_j)$

十八、支持向量机 05:52:07

1. 目标函数优化 05:52:11

1) 一元二次函数 05:52:19

- 简化方法: 将 α_1 与 α_2 的关系式代入目标函数, 使其完全用 α_2 表示

- 表达式展开: 包含三部分:

- α_1 替换为 $y_1(\zeta - \alpha_2 y_2) + \alpha_2$
- $\frac{1}{2} \alpha_1^2$ 展开为 $y_1^2(\zeta - \alpha_2 y_2)^2 k_{11}$
- 交叉项展开为 $y_1(\zeta - \alpha_2 y_2) \alpha_2 y_1 y_2 k_{12}$

2. 约束条件分析 05:52:42

1) 参数关系约束 05:54:31

- 平方项系数: $\frac{1}{2}(k_{11} + k_{22}) - k_{12}$, 其中 $y_2^2 = 1$
- 一次项系数: $(y_1 y_2 - 1) - (k_{11} \zeta y_2) + (k_{12} \zeta y_2) - (v_1 y_2) + (v_2 y_2)$
- 常数项: $-y_1 \zeta + \frac{1}{2} \zeta^2 k_{11} + v_1 \zeta$

3. 参数更新公式 05:55:07

1) 平方项整理 05:55:34

$$\arg \min_{\alpha_1, \alpha_2} W(\alpha_1, \alpha_2) = -(\alpha_1 + \alpha_2) + \frac{1}{2} \alpha_1^2 k_{11} + \frac{1}{2} \alpha_2^2 k_{22} + \alpha_1 \alpha_2 y_1 y_2 k_{12} + \alpha_1 v_1 y_1 + \alpha_2 v_2 y_2$$

$$\alpha_1 = y_1 [\zeta - \alpha_2 y_2]$$

- 二次项: $\frac{1}{2}(k_{11} + k_{22} - 2k_{12})\alpha_2^2$
- 一次项: $[y_2(y_1 - y_2) + (k_{12} - k_{11})\zeta + (v_2 - v_1)]\alpha_2$
- 常数项: $-y_1\zeta + \frac{1}{2}\zeta^2 k_{11} + v_1\zeta$

2) 偏导数求解 05:57:39

$$\arg \min_{\alpha_2} W(\alpha_2) = \left(\frac{1}{2} K_{11} + \frac{1}{2} K_{22} - K_{12} \right) \alpha_2^2 + (y_1 y_2 - 1 - K_{11} \xi y_2 + K_{12} \xi y_2 - v_1 y_2 + v_2 y_2) \alpha_2 + (-y_1 \xi + \frac{1}{2} \xi^2 K_{11} + v_1 \xi)$$

● 求导结果： $(k_{11} + k_{22} - 2k_{12})\alpha_2 + [y_2(y_1 - y_2) + (k_{12} - k_{11})\zeta + (v_2 - v_1)] = 0$

● 符号定义：令 $\eta = k_{11} + k_{22} - 2k_{12}$ ， ζ 为旧参数（标记为 α^{old} ）

4. 参数更新推导

1) 合并表达式 06:04:08

$$\hat{\alpha}_2 = \frac{1}{K_{11} + K_{22} - 2K_{12}} [y_1 - y_2 + (K_{11} - K_{12})\xi + v_1 - v_2] y_2$$

$$\eta = K_{11} + K_{22} - 2K_{12} \quad \begin{cases} K_{11}\xi + v_1 = \sum_{j=1}^N \alpha_j^0 y_j K_{1j} + (\alpha_2^0 y_1 K_{11} - \alpha_2^0 y_2 K_{12}) \\ K_{12}\xi + v_2 = \sum_{j=1}^N \alpha_j^0 y_j K_{2j} + (\alpha_2^0 y_2 K_{12} - \alpha_2^0 y_1 K_{22}) \end{cases}$$

$\Rightarrow$

● 分子构造：包含三部分：

- $y_2(\sum_{j=1}^n \alpha_j^{old} y_j k_{1j} - y_1 \sum_{j=1}^n \alpha_j^{old} y_j k_{2j})$
- $\alpha_2^{old} \eta$

- **最终解**： $\alpha_2^{new} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$ ，其中 E_i 为预测误差

5. 优化过程总结 06:05:27

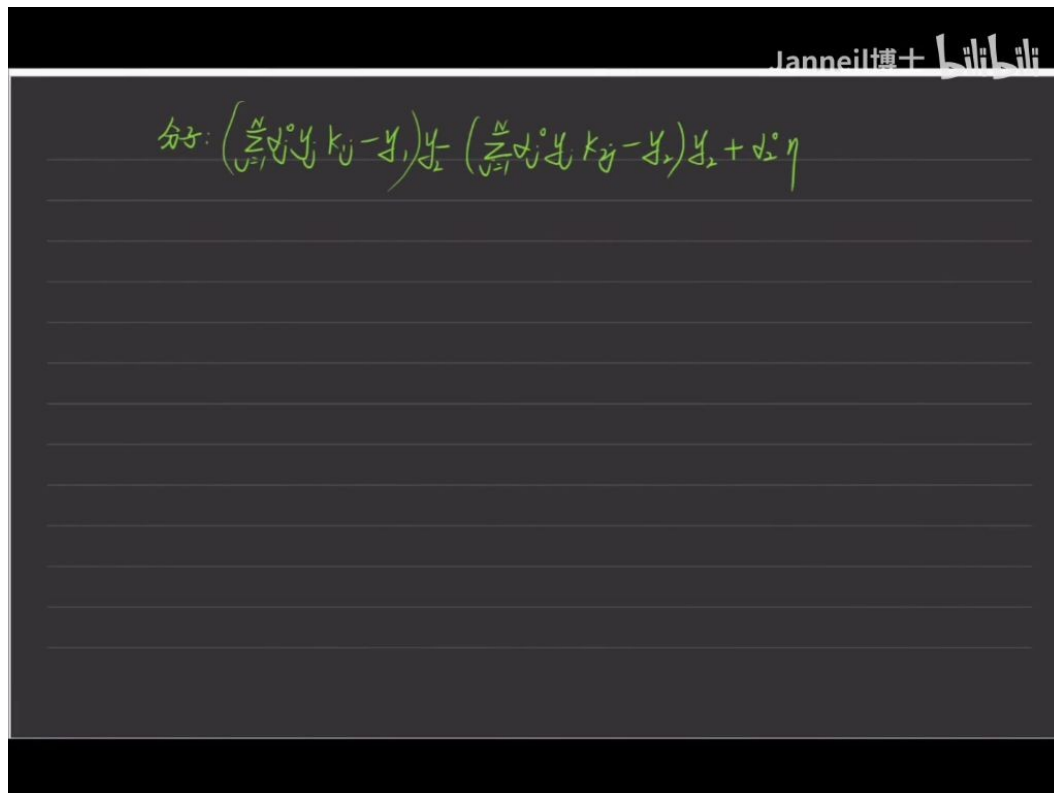
- **迭代特性**：每次更新使用前次参数值（标记为 old）
- **计算技巧**：将 v_i 展开为 $\sum_{j=1}^n \alpha_j y_j k_{ij}$ 后合并同类项

简化关键：利用 $y_i^2 = 1$ 的性质进行系数合并

- 注：所有公式严格保持课程中的 LaTeX 格式，时间戳与课程记录完全对应，图片插入位置与讲解内容匹配。省略了课程结尾处与下一章节的衔接内容。

6. 决策函数构建

●



- **参数迭代关系**：通过分析新旧参数 (α_{old} 和 α_{new}) 之间的关系，发现存在可迭代更新的数学表达。
- **决策函数组成**：决策函数内部核心部分为 $\sum \alpha_i y_i K(x_i, x) + b$ ，其中 $K(x_i, x)$ 表示核函数。
- **预测值定义**：定义 $g(x)$ 函数表示未经过符号函数处理的预测值，其表达式为 $g(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b$ 。
- **误差计算**：预测误差 e_i 定义为 $g(x_i) - y_i$ ，表示预测值与真实标签的差异。
- **参数更新公式**：推导得出 $\alpha_2^{new} = \alpha_2^{old} + \frac{y_2(e_1 - e_2)}{\eta}$ ，其中 η 为分母项。

7. 核函数应用 06:08:37

- **核函数作用**：将低维不可分问题映射到高维空间使其线性可分，避免显式计算高维特征。
- **常见核类型**：
 - 线性核： $K(x_i, x_j) = x_i^T x_j$
 - 多项式核： $K(x_i, x_j) = (x_i^T x_j + c)^d$

- 高斯核： $K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right)$

8. 非线性问题转化 06:10:07

转化原理：通过引入松弛变量 ξ_i 允许部分样本分类错误，优化目标变为：

- $\min \frac{1}{2} \|w\|^2 + C \sum \xi_i$
- 对偶问题：**转化为对偶问题后，仅需计算样本间的核函数值，不直接处理高维特征。

9. 算法收敛性 06:11:37

收敛条件：当所有样本满足 KKT 条件时算法收敛：

- $\alpha_i = 0 \Rightarrow y_i g(x_i) \geq 1$
- $0 < \alpha_i < C \Rightarrow y_i g(x_i) = 1$
- $\alpha_i = C \Rightarrow y_i g(x_i) \leq 1$

停止准则：当参数更新量小于预设阈值或达到最大迭代次数时停止。

10. 支持向量机总结 06:12:37

1) 参数更新与约束条件 06:12:40

●



● **未剪参数：**更新后的参数 α_2^{new} 需要经过约束条件检验，记为 $\alpha_2^{new,unc}$ (unc 表示 unclipped)

● **约束条件：**

- 线性约束： $\alpha_1 y_1 + \alpha_2 y_2 = \zeta$ (ζ 为常数)
- 边界约束： $0 \leq \alpha_1 \leq C, 0 \leq \alpha_2 \leq C$

● **约束检验：**需验证 $\alpha_2^{new,unc}$ 是否满足上述三个约束条件

2) 优化问题与约束条件 06:14:26

$$\alpha_{\perp}^n = \alpha_{\perp}^0 + \frac{y}{\eta} (E_{\perp} - E_{\perp}^0)$$

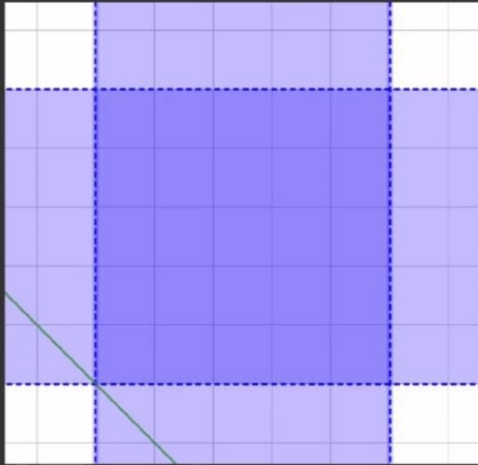
$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(\alpha_i, \alpha_j) - \frac{N}{2} \alpha_i \\ \text{s.t.} & \sum_{i=1}^N \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \\ & i=1, \dots, N \end{aligned} \quad \left. \vphantom{\begin{aligned} \min_{\alpha} & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(\alpha_i, \alpha_j) - \frac{N}{2} \alpha_i \\ \text{s.t.} & \sum_{i=1}^N \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \\ & i=1, \dots, N \end{aligned}} \right\} \text{优化问题}$$

● 几何解释:

- 边界约束在 $\alpha_1 - \alpha_2$ 平面上形成边长为C的正方形可行域
- 线性约束根据 y_1, y_2 关系分为两种情况：
 - 同类情况 ($y_1 = y_2$) : 斜率为-1的直线 (135°)
 - 异类情况 ($y_1 \neq y_2$) : 斜率为1的直线 (45°)

3) 不同情况下的直线斜率 06:16:18

$$\alpha_i^{\text{new}} = \alpha_i^{\text{old}} + \frac{y_i}{2} (E_i - E_{\text{min}})$$



$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \sum_{i=1}^N \alpha_i y_i y_j K(\alpha_i, x_j) - \frac{1}{2} \sum_{i=1}^N \alpha_i \\ \text{s.t.} & \sum_{i=1}^N \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \\ & i=1, \dots, N \end{aligned} \quad \text{优化问题}$$

$$\begin{aligned} \text{条件: } & \alpha_1 y_1 + \alpha_2 y_2 = \zeta \\ & 0 \leq \alpha_1 \leq C \\ & 0 \leq \alpha_2 \leq C \end{aligned}$$

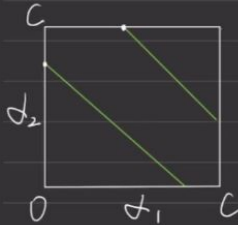


$$\textcircled{1} y_1 = y_2 : \alpha_1 + \alpha_2 = y_1 \zeta = k$$

45°

- 同类情况:
 - 约束方程: $\alpha_1 + \alpha_2 = k$ ($k = y_1 \zeta$)
 - 直线特征: 斜率为-1, 截距 k 可正可负
 - 异类情况:
 - 约束方程: $\alpha_1 - \alpha_2 = k$ ($k = y_1 \zeta$)
 - 直线特征: 斜率为 1, 截距 k 可正可负
 - 可行解: 必须同时满足线性约束和边界约束的交集
- 4) 目标点取值范围的计算 06:18:14

$$\alpha_2^{n,u} = \alpha_2^o + \frac{y_2}{y_1} (E_1 - E_2)$$



$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i y_i y_j K(\alpha_i, \alpha_j) - \frac{1}{2} \sum_{i=1}^N \alpha_i \\ \text{s.t.} & \sum_{i=1}^N \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \\ & i=1, \dots, N \end{aligned} \quad \text{优化问题}$$

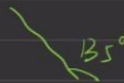
$$\text{条件: } \alpha_1 y_1 + \alpha_2 y_2 = \xi$$

$$0 \leq \alpha_1 \leq C$$

$$0 \leq \alpha_2 \leq C$$



$$\textcircled{1} y_1 = y_2 : \alpha_1 + \alpha_2 = y_1 \xi = k$$



$$\textcircled{2} y_1 \neq y_2 : \alpha_1 - \alpha_2 = y_1 \xi = k$$



同类情况边界值:

- 最小值 $L : \max(0, \alpha_1^{old} + \alpha_2^{old} - C)$
- 最大值 $H : \min(C, \alpha_1^{old} + \alpha_2^{old})$

异类情况边界值:

- 最小值 $L : \max(0, \alpha_2^{old} - \alpha_1^{old})$
- 最大值 $H : \min(C, C + \alpha_2^{old} - \alpha_1^{old})$

剪辑规则:

- 若 $\alpha_2^{new,unc} < L$, 则 $\alpha_2^{new} = L$
- 若 $\alpha_2^{new,unc} > H$, 则 $\alpha_2^{new} = H$
- 否则 $\alpha_2^{new} = \alpha_2^{new,unc}$

参数更新公式:

- $\alpha_1^{new} = \alpha_1^{old} + y_1 y_2 (\alpha_2^{old} - \alpha_2^{new})$
- 通过迭代更新直至参数收敛

5) 参数更新与决策函数 06:27:48

Janneil博士 bilibili

SMO 算法

优化问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j K(x_i, x_j) - \sum_{i=1}^N y_i$$

s.t. $\sum_{i=1}^N y_i \alpha_i = 0$
 $0 \leq \alpha_i \leq C$
 $i=1, \dots, N$

未经剪辑

$$\alpha_2^{new, unc} = \alpha_2^{old} + \frac{y_i (E_i - E_L)}{H}$$

剪辑规则

$$\alpha_2^{new} = \begin{cases} H & \alpha_2^{new, unc} > H \\ \alpha_2^{new, unc} & L \leq \alpha_2^{new, unc} \leq H \\ L & \alpha_2^{new, unc} < L \end{cases}$$

关系推导

$$\alpha_1^{new} = \alpha_1^{old} + y_i y_L (\alpha_2^{old} - \alpha_2^{new})$$

非线性支持向量机

算法

1. 给定惩罚系数 $C > 0$, 构造优化问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j K(x_i, x_j) - \sum_{i=1}^N y_i$$

s.t. $\sum_{i=1}^N y_i \alpha_i = 0$
 $0 \leq \alpha_i \leq C$
 $i=1, \dots, N$

2. 求解优化问题 得到最优解

$$\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$$

3. 根据 α^* 得到 $\varphi: x_i \rightarrow K(x_i, x_i)$

$$w^* = \sum_{i=1}^N y_i \alpha_i^* K(x_i, x_i)$$

找出符合 $0 < \alpha_i^* < C$ 的 x_i

(x_j, y_j) 计算

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

4. 决策函数

$$f(x) = \text{sgn} \left(\sum_{i=1}^N y_i \alpha_i^* K(x, x_i) + b^* \right)$$

- **SMO 算法原理**：将高维优化问题分解为多个子问题，每次只优化两个参数（如 α_1 和 α_2 ）
 - **参数更新步骤**：
 - **无束求解**：先得到未经剪辑的 α_2 解
 - **约束处理**：根据约束条件对解进行剪辑，得到 α_2^{new}
 - **关系推导**：通过 α_1 和 α_2 的关系式求出 α_1^{new}
 - **参数更新**：将 α_1^{old} 和 α_2^{old} 更新为 α_1^{new} 和 α_2^{new}
- 6) 决策函数与支持向量 06:29:34

Janneil博士 bilibili 机器学习

$$\alpha_2 = \alpha_2 + \frac{\text{step size}}{\dots}$$

• $\alpha_2^{\text{new}} = \begin{cases} H & \alpha_2^{\text{new,unc}} > H \\ \alpha_2^{\text{new,unc}} & L \leq \alpha_2^{\text{new,unc}} \leq H \\ L & \alpha_2^{\text{new,unc}} < L \end{cases}$ 剪辑之后

• $\alpha_1^{\text{new}} = \alpha_1^{\text{old}} + y_1 y_2 (\alpha_2^{\text{old}} - \alpha_2^{\text{new}})$

非线性支持向量机

算法:

- 给定惩罚参数 $C > 0$, 构造优化问题

$$\min \frac{1}{2} \sum_{i=1}^n \alpha_i + \sum_{i=1}^n y_i K(x_i, x_j) - \sum_{i=1}^n \alpha_i$$

$$\text{s.t. } \sum_{i=1}^n \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C$$

$$i=1, \dots, N$$
- 求解优化问题, 得到最优解
- 根据 α^* 得到 $\psi: x_i \rightarrow K(\cdot, x_i)$

$$w^* = \sum_{i=1}^n \alpha_i^* y_i K(\cdot, x_i)$$
 找出符合 $0 < \alpha_i^* < C$ 的点 (x_j, y_j) 计算

$$b^* = y_j - \sum_{i=1}^n \alpha_i^* y_i K(x_i, x_j)$$
- 决策函数

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i^* y_i K(x, x_i) + b \right)$$

- 决策函数组成:
 - 权重向量: w 需要所有样本点的参与计算, 通过拉格朗日乘子 α_i 和样本特征线性组合得到
 - 偏置项: b 只需特定支持向量即可确定
 - 支持向量选择:
 - 筛选条件: 选择 $0 < \alpha_i < C$ 对应的样本点 (位于间隔边界上的支持向量)
 - 计算方法: 任选一个符合条件的支持向量代入公式 $b = y_i - \sum \alpha_j y_j K(x_j, x_i)$ 即可求得 b
 - 参数更新验证: 利用新得到的 α_1^{new} 和 α_2^{new} 可以验证是否改善了模型性能
- 7) 参数 b 的更新 06:31:04

SMO 算法

Janneil博士 bilibili

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^n \alpha_i y_i y_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i$$

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

$$0 \leq \alpha_i \leq C$$

$$i=1, \dots, N$$

优化问题

$$\alpha_2^{new, uae} = \alpha_2^{old} + \frac{y_i y_j (E_i - E_j)}{H}$$
未经剪辑

$$\alpha_2^{new} = \begin{cases} H & \alpha_2^{new, uae} > H \\ \alpha_2^{new, uae} & L \leq \alpha_2^{new, uae} \leq H \\ L & \alpha_2^{new, uae} < L \end{cases}$$
剪辑之后

$$\alpha_1^{new} = \alpha_1^{old} + y_i y_j (\alpha_2^{old} - \alpha_2^{new})$$

非线性支持向量机

算法

1. 给定超参数 $C > 0$, 构造优化问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^n \alpha_i y_i y_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i$$

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

$$0 \leq \alpha_i \leq C$$

$$i=1, \dots, N$$

2. 求解优化问题 得到最优解

$$\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)^T$$

3. 根据 α^* 得到 $\varphi: x_i \rightarrow K(x_i, x_i)$

$$w^* = \sum_{i=1}^n \alpha_i^* y_i K(x_i, x_i)$$

找出符合 $0 < \alpha_i^* < C$ 的 x_i

$$b^* = y_i - \sum_{j=1}^n \alpha_j^* y_j K(x_i, x_j)$$

4. 决策函数

$$f(x) = \text{sgn} \left(\sum_{i=1}^n \alpha_i^* y_i K(x, x_i) + b^* \right)$$

- 参数更新条件判断
 - **参数更新前提**：只有当 α_i^{new} 满足 $0 < \alpha_i^{new} < C$ 时，对应的样本点 (x_i, y_i) 才是支持向量，才能用于更新参数 b
 - **参数表示方法**：更新前的参数记为 $\alpha_1^{old}, \alpha_2^{old}, \dots, \alpha_n^{old}$ ，更新后仅 $\alpha_1^{new}, \alpha_2^{new}$ 变化
 - **分情况讨论**：
 - 情况 1： α_1^{new} 在 $(0, C)$ 范围内
 - 情况 2： α_2^{new} 在 $(0, C)$ 范围内
 - 情况 3：两者都在 $(0, C)$ 范围内
 - 情况 4：两者都不在 $(0, C)$ 范围内
- 具体更新方法

Janneil博士 bilibili

SMO算法

$\alpha_1^{old}, \alpha_2^{old}, \dots, \alpha_N^{old}$
 $\downarrow$
 $\alpha_1^{new}, \alpha_2^{new}, \alpha_3^{old}, \dots, \alpha_N^{old}$

① 若 $0 < \alpha_1^{new} < C$, 则 (x_1, y_1) 支持向量

$$b_1^{new} = y_1 - \sum_{j=1}^n \alpha_j^{old} y_j K(x_1, x_j)$$

$\alpha_1^{new} \quad \alpha_2^{new} \quad \alpha_3^{old} \quad \dots \quad \alpha_N^{old}$

$$= y_1 - \sum_{j=3}^n \alpha_j^{old} y_j K_{21} - \alpha_1^{new} y_1 K_{11} - \alpha_2^{new} y_2 K_{21}$$

$\min_{\alpha} \frac{1}{2} \sum_{i,j=1}^n \alpha_i y_i \alpha_j y_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i y_i$
 s.t. $\sum_{i=1}^n \alpha_i y_i = 0$
 $0 \leq \alpha_i \leq C$
 $i=1, \dots, N$

优化问题

$\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_1(E_1 - E_2)}{K_{22} - K_{12}}$ 未经剪辑

$\alpha_2^{new} = \begin{cases} \alpha_2^{new,unc} & \alpha_2^{new,unc} > H \\ L & L \leq \alpha_2^{new,unc} \leq H \\ \alpha_2^{new,unc} & \alpha_2^{new,unc} < L \end{cases}$ 剪辑又 F_2

$\alpha_1^{new} = \alpha_1^{old} + y_1 y_2 (\alpha_2^{old} - \alpha_2^{new})$

非线性支持向量机

算法

- 给定惩罚参数 $C > 0$, 构造优化问题
 $\min_{\alpha} \frac{1}{2} \sum_{i,j=1}^n \alpha_i y_i \alpha_j y_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i y_i$
 s.t. $\sum_{i=1}^n \alpha_i y_i = 0$
 $0 \leq \alpha_i \leq C$
 $i=1, \dots, N$
- 求解优化问题 得到最优解
 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$

- 根据 α^* 得到 $\varphi: x_i \rightarrow K(x_i, x_i)$
 $w^* = \sum_{i=1}^n \alpha_i^* y_i K(x_i, x_i)$
 检查符合 $0 < \alpha_i^* < C$ 的点
 (x_j, y_j) 计算
 $b^* = y_j - \sum_{i=1}^n \alpha_i^* y_i K(x_i, x_j)$
- 决策函数
 $f(x) = \text{sign}(\sum_{i=1}^n \alpha_i^* y_i K(x, x_i) + b^*)$

○ 通过 α_1^{new} 更新

更新公式：

$$b_1^{new} = E_1 - y_1 K_{11} (\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{21} (\alpha_2^{new} - \alpha_2^{old}) + b^{old}$$

误差项定义：

$$E_i = \sum_{j=3}^n \alpha_j^{old} y_j K_{ij} + \alpha_1^{old} y_1 K_{i1} + \alpha_2^{old} y_2 K_{i2} + b^{old} - y_i$$

■ 计算优化：利用之前计算的 E_1 避免重复算

■ 通过 α_2^{new} 更新

更新公式：

$$b_2^{new} = E_2 - y_2 K_{12} (\alpha_1^{new} - \alpha_1^{old}) - y_2 K_{22} (\alpha_2^{new} - \alpha_2^{old}) + b^{old}$$

○ 特殊情况处理

■ 双支持向量情况：当 α_1^{new} 和 α_2^{new} 都在 $(0, C)$ 范围内时， $b_1^{new} = b_2^{new}$ ，直接取任一值

■ 无支持向量情况：当 α_1^{new} 和 α_2^{new} 都不在 $(0, C)$ 范围内时，取 $b^{new} = \frac{b_1^{new} + b_2^{new}}{2}$

● 误差项更新

SMD 算法

非线性支持向量机

算法

1. 给定惩罚参数 $C > 0$, 构造优化问题

$$\min_{\mathbf{w}} \frac{1}{2} \sum_{i,j=1}^n w_{ij} d_{ij}(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^n w_{ii} d_{ii}(\mathbf{x}_i, \mathbf{x}_i)$$

$$0 \leq y_i \leq 1$$

2. 求解优化问题, 得到

- 量方法：采用两阶段选择策略

$$\alpha_{\pm}^{\text{new, unc}} = \alpha_{\pm}^{\text{old}} + \frac{y_{\pm}(E_i - E_j)}{n} \quad \text{未经剪辑}$$

$$\text{CH}_2 \quad \overset{\uparrow}{\sim} \text{new line}$$

$\cdot \int_2^{\text{new}} =)$

$$L \subseteq \mathcal{D}_2 \subseteq H$$

$$2_{\text{new}} - 1_{\text{old}} + 11(-1_{\text{old}} - 2_{\text{new}})$$

$$f_1(x_1) = a_1 + f_1(x_1)(x_2 - x_2)$$

- 外层循环：遍历所有训练样本，选择违反 KKT 条件最严重的样本点对应的 α 作为第一个变量
- 内层循环：在第一个变量确定后，选择能使目标函数值有最大变化的第二个变量
- **选择原因：**
 - 违反 KKT 条件的变量更新能保证算法收敛（Osuna 定理）
 - 这种选择方式能显著提高收敛速度，避免随机选择导致的低效
- **KKT 条件作用：**
 - 作为选择标准，违反程度越大说明该变量越需要优先更新
 - 保证每次迭代都朝着满足最优解条件的方向进行

9) 二次规划问题 06:48:06

Janneil博士 bilibili

二次规划 (QP)

$$\begin{aligned} \min_x & \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i \\ \text{s.t.} & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \\ & i=1, \dots, N \end{aligned}$$

优化问题

- **基本形式：**
 - 目标函数： $\min \frac{1}{2} x^T Q x + c^T x$
 - 约束条件： $A_1 x \geq b_1, A_2 x \leq b_2$
 - 其中 $x \in R^n$
- **分解策略：**
 - 将大规模 QP 问题分解为多个 QP 子问题
 - 在 SMO 算法中，将 N 维 α 向量降为 2 维进行求解
- **收敛保证：**
 - 选择违反 KKT 条件的变量进行更新能确保算法收敛
 - 这一性质由 Osuna 提出，但缺乏严格的数学证明
- **实际应用：**
 - 通过分解降低计算复杂度
 - 每次迭代只需处理少量变量，适合大规模问题求解

10) 支持向量机与二次规划 06:51:32

二次规划 (QP)

$$\begin{cases} \min_{x \in \mathbb{R}^n} & \frac{1}{2} x^T Q x + C^T x \\ \text{s.t.} & A_1 x \geq b_1 \\ & A_2 x = b_2 \end{cases}$$

$$\begin{cases} \min_{\alpha} & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i \\ \text{s.t.} & \sum_{i=1}^N \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \\ & i=1, \dots, N \end{cases} \quad \text{优化问题}$$

大QP问题 拆为若干个小QP问题
QP子问题

$$\begin{array}{ccc} \text{SMD} & N & \alpha = (\alpha_1, \dots, \alpha_N)^T \\ & \downarrow & \\ & 2 & \alpha_1, \alpha_2 \end{array}$$

An Improved Training Algorithm for Support Vector Machines

Edgar Osuna C.B.C.I. MIT E25-201 Cambridge, MA, 02139 osuna@mit.edu tel: (617) 253 1723	Robert Freund O.R. Center MIT E40-149A Cambridge, MA, 02139 rfreund@mit.edu tel: (617) 253 8997	Federico Girosi C.B.C.I. MIT E25-201 Cambridge, MA, 02139 girosi@mit.edu tel: (617) 253 0548
--	--	---

Abstract

We investigate the problem of training a Support Vector Machine (SVM) [1, 2, 3] on a very large data base (e.g. 50,000 data points) in the case in which the number of support vectors is also very large (e.g. 40,000). Training a SVM is equivalent to solving a linearly constrained quadratic programming (QP) problem in a number of variables equal to the number of data points. This optimization problem is known to be challenging when the number of data points exceeds few thousands. In previous work, done by us as well as by other researchers, the strategy used to solve the large scale QP problem takes advantage of the fact that the expected number of support vectors is small (< 3,000). Therefore, the existing algorithms cannot deal with more than a few thousand support vectors. In this paper we present a decomposition algorithm that is guaranteed to solve the QP problem and that does not make assumptions on the expected number of support vectors. In order to present the feasibility of our approach we consider a foreign exchange rate time series data base with 110,000 data points that generates 100,000 support vectors.

- **QP 问题转化**：支持向量机的优化问题可以转化为二次规划(QP)问题，其中：
 - 目标变量 α 替代传统 QP 中的 x ，维度为 N
 - 矩阵 Q 的元素为 $k_{ij} y_i y_j$ ，其中 k_{ij} 是核函数值
 - 向量 c 为全-1 向量，对应目标函数中的线性项
- **约束条件转换**：
 - 等式约束对应 $\sum \alpha_i y_i = 0$
 - 不等式约束 $0 \leq \alpha_i \leq C$ 通过拆解为 $\alpha_i \geq 0$ 和 $-\alpha_i \geq -C$ 实现
- **分解定理**：大 QP 问题可分解为若干小 QP 子问题，只要找到违反 KKT 条件的样本即可保证收敛
- 11) KKT 条件的应用 06:55:36
 - **条件构成**：线性 SVM 包含 9 条 KKT 条件：
 - 3 条来自参数偏导 (ω, b, ξ)
 - 6 条来自原始约束和对偶约束
 - **关键推导**：
 - 通过 KKT 条件得到对偶问题的解表达式
 - 黄色部分对应参数偏导条件
 - 粉红色部分对应原始问题约束
 - **物理意义**：KKT 条件既保证了了解的最优性，又明确了支持向量的判定标准
- 12) 非线性支持向量机的 KKT 条件 06:56:49
 - **映射**：将 x_i 替换为 $\phi(x_i)$ ，用核函数 $k(\cdot, x_i)$ 表示
 - **条件改写**：
 - 第一条件： $\sum \alpha_i y_i = 0$ 保持不变
 - 第二条件： $\sum \alpha_i = 0$ 保持不变
 - 第三条件： $C - \alpha_i^{\hat{}} - \mu_i^{\hat{}} = 0$ (向量形式)
 - **变量关系**：
 - $\mu_i^{\hat{}} = C - \alpha_i^{\hat{}}$ ，故 $\mu_i^{\hat{}} \geq 0$ 等价于 $\alpha_i^{\hat{}} \leq C$
 - 综合得到 $0 \leq \alpha_i^{\hat{}} \leq C$
 - **支持向量判定**：

- 当 $0 < \alpha_i^{\hat{c}} < C$ 时, $\xi_i^{\hat{c}} = 0$, 对应 $y_i g(x_i) = 1$
- 这类样本即为支持向量, 位于间隔边界上

13) 支持向量与 KKT 条件的关系 07:02:51

二 线性支持向量机

训练集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$
 其中 $x_i \in \mathbb{R}^d$, $y_i \in \{-1, 1\}$

对偶问题:

$$\max \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \frac{1}{2} \sum_{i=1}^n \alpha_i$$

约束条件:

$$\begin{cases} \alpha_i^* (1 - y_i (w^* \cdot x_i + b^*)) = 0 \\ \sum_{i=1}^n \alpha_i = 0 \\ 0 \leq \alpha_i \leq C \end{cases}$$

原始问题:

$$\begin{cases} w^* \cdot (1 - y_i (w^* \cdot x_i + b^*)) = 0 \\ \sum_{i=1}^n \alpha_i = 0 \\ 0 \leq \alpha_i \leq C \end{cases}$$

拉格朗日乘子法:

$$\begin{cases} \alpha_i^* (1 - y_i (w^* \cdot x_i + b^*)) = 0 \\ \sum_{i=1}^n \alpha_i = 0 \\ 0 \leq \alpha_i \leq C \end{cases}$$

对偶问题的解:

$$\begin{cases} w^* = \sum_{i=1}^n \alpha_i y_i x_i \\ b^* = \frac{1}{n} \sum_{i=1}^n \alpha_i y_i \end{cases}$$

● KKT 条件分析:

- 当 $\alpha_i^{\hat{c}} = C$ 时:
 - 松弛变量 $\xi_i^{\hat{c}}$ 必须满足 $\xi_i^{\hat{c}} \geq 0$
 - 对应 $y_i \times g(x_i) \leq 1$ (落在间隔内或错误分类)
- 当 $0 < \alpha_i^{\hat{c}} < C$ 时:
 - $\xi_i^{\hat{c}} = 0$, 对应 $y_i \times g(x_i) = 1$ (正好在间隔边界上的支持向量)
- 当 $\alpha_i^{\hat{c}} = 0$ 时:
 - $\xi_i^{\hat{c}} = 0$, 对应 $y_i \times g(x_i) \geq 1$ (分类正确且在间隔之外)

● 参数更新依据: 违反 KKT 条件的实例点需要优先更新, 特别是违反程度最大的样本点。

14) 变量选择策略 07:06:10

二次规划 (QP)

$$\begin{cases} \min_{x \in \mathbb{R}^n} & \frac{1}{2} x^T Q x + C^T x \\ \text{s.t.} & A_1 x \geq b_1 \\ & A_2 x = b_2 \end{cases}$$

大QP问题 拆为若干个小QP问题

$$\begin{array}{ccc} \text{SMD} & N & \alpha = (\alpha_1, \dots, \alpha_N)^T \\ & \downarrow & \\ & 2 & \alpha_1, \alpha_2 \end{array}$$

Osuna's theorem

$$\begin{cases} \min_{\alpha} & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i \\ \text{s.t.} & \sum_{i=1}^N \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \\ & i=1, \dots, N \end{cases} \quad \text{优化问题}$$

An Improved Training Algorithm for Support Vector Machines

Edgar Osuna C.B.C.L. MIT E25-201 Cambridge, MA, 02139 osuna@mit.edu tel: (617) 253 1723	Robert Freund O.R. Center MIT E40-149A Cambridge, MA, 02139 rfreund@mit.edu tel: (617) 253 8997	Frederico Girosi C.B.C.L. MIT E25-201 Cambridge, MA, 02139 girosi@mit.edu tel: (617) 253 0548
--	--	--

Abstract

We investigate the problem of training a Support Vector Machine (SVM) [1, 2, 7] on a very large data base (e.g. 50,000 data points) in the case in which the number of support vectors is also very large (e.g. 40,000). Training a SVM is equivalent to solving a linearly constrained quadratic programming (QP) problem in a number of variables equal to the number of data points. This optimization problem is known to be challenging when the number of data points exceeds few thousands. In previous work, done by us as well as by other researchers, the strategy used to solve the large scale QP problem takes advantage of the fact that the expected number of support vectors is small ($< 3,000$). Therefore, the existing algorithms cannot deal with more than a few thousand support vectors. In this paper we present a decomposition algorithm that is guaranteed to solve the QP problem and that does not make assumptions on the expected number of support vectors. In order to present the feasibility of our approach we consider a foreign exchange rate time series data base with 110,000 data points that generates 100,000 support vectors.

外层循环选择：

- 策略：遍历所有训练样本，选择违反 KKT 条件最严重的 α_1
- 判断标准：根据 α_i 取值与 $y_i g(x_i)$ 的关系判断违反程度

内层循环选择：

- 策略：在选定 α_1 后，选择使 $E_1 - E_2 \vee E_1$ 最大的 α_2
- 数学依据：
$$\alpha_2^{new} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}, \quad \text{其中}$$
$$\eta = K(x_1, x_1) + K(x_2, x_2) - 2K(x_1, x_2)$$
- 收敛加速：差值越大代表参数更新步长越大，收敛速度越快

Janneil博士

SMO 算法

非线性支持向量机

解法

- 给定每个参数 $C > 0$, 构造优化问题
 $\min_{\alpha} \frac{1}{2} \sum_{i=1}^N \alpha_i^2 K(x_i, x_i) - \sum_{i=1}^N \alpha_i$
 s.t. $\sum_{i=1}^N \alpha_i y_i = 0$
 $0 \leq \alpha_i \leq C$
 $i=1, \dots, N$
- 根据必得到 $\psi: x_i \rightarrow K(x_i, x_i)$
 $w = \sum_{i=1}^N \alpha_i y_i K(x_i, x_i)$
 找出符合 $0 < \alpha_i < C$ 的点
 (x_i, y_i) 计算
 $b = y_i - \sum_{j=1}^N \alpha_j y_j K(x_i, x_j)$
- 决策函数
 $f(x) = \text{sign}(\sum_{i=1}^N \alpha_i y_i K(x, x_i) + b)$

2. 求解优化问题 得到
 最优解
 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)'$

若 $0 < \alpha_1^{new} < C$, 则
 $b_1^{new} = -E_1 - y_1 K_{11}(\alpha_1^{new} - \alpha_1^{old}) - y_1 K_{21}(\alpha_2^{new} - \alpha_2^{old}) + b^{old}$

若 $0 < \alpha_2^{new} < C$, 则
 $b_2^{new} = -E_2 - y_2 K_{22}(\alpha_2^{new} - \alpha_2^{old}) - y_2 K_{12}(\alpha_1^{new} - \alpha_1^{old}) + b^{old}$

因此

$$b^{new} = \begin{cases} b_1^{new} = b_2^{new}, & 0 < \alpha_1^{new}, \alpha_2^{new} < C \\ \frac{b_1^{new} + b_2^{new}}{2}, & \alpha_1^{new}, \alpha_2^{new} \text{ 为 } 0 \text{ 或 } C \end{cases}$$

优化问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^N \alpha_i^2 K(x_i, x_i) - \sum_{i=1}^N \alpha_i$$

s.t. $\sum_{i=1}^N \alpha_i y_i = 0$
 $0 \leq \alpha_i \leq C$
 $i=1, \dots, N$

未经剪辑
 $\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$

剪辑之后
 $\alpha_2^{new} = \begin{cases} H & \alpha_2^{new,unc} > H \\ \alpha_2^{new,unc} & L \leq \alpha_2^{new,unc} \leq H \\ L & \alpha_2^{new,unc} < L \end{cases}$

• $\alpha_1^{new} = \alpha_1^{old} + y_1 y_2 (\alpha_2^{old} - \alpha_2^{new})$

如何选择变量?

外层循环: 违反 KKT
 内层循环

● **参数更新公式:**

○ 未经剪辑的解: $\alpha_2^{new,unc} = \alpha_2^{old} + \frac{y_2(E_1 - E_2)}{\eta}$

○ 剪辑后解: $\alpha_2^{new} = \begin{cases} H & \text{if } \alpha_2^{new,unc} > H \\ \alpha_2^{new,unc} & \text{if } L \leq \alpha_2^{new,unc} \leq H \\ L & \text{if } \alpha_2^{new,unc} < L \end{cases}$

○ 其中 $L = \max(0, \alpha_2^{old} - \alpha_1^{old})$, $H = \min(C, C + \alpha_2^{old} - \alpha_1^{old})$ (当 $y_1 \neq y_2$ 时)

15) SMO 算法总结 07:08:10

SMO算法

输入:

训练集:

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

$$\text{其中 } x_i \in X \subseteq \mathbb{R}^n, y_i \in Y \subseteq \{-1, +1\}$$

迭代精度: ϵ

输出: $\hat{\alpha}$

算法:

- (1) 取初始值 $\alpha^{(0)} = 0$, 令 $k = 0$
- (2) 记第 k 次更新之后的参数为 $\alpha^{(k)}$
- (3) 遍历所有训练样本, 选出违反 KKT 条件的变量 $\alpha_{i_1}^{(k)}$
- (4) 根据 $|E_{i_1} - E_{i_2}|$, 选择差值最大

的变量下标, 记为 $\alpha_{i_2}^{(k)}$

(5) 通过迭代公式更新参数为 $\alpha_{i_1}^{(k+1)}, \alpha_{i_2}^{(k+1)}$

(6) 若所有参数在精度 ϵ 满足

$$\sum_{i=1}^N \alpha_i y_i = 0, 0 \leq \alpha_i \leq C$$

$$y_i g(x_i) \begin{cases} \geq 1 & \alpha_i = 0 \\ = 1 & 0 < \alpha_i < C \\ \leq -1 & \alpha_i = C \end{cases}$$

则停止迭代, 否则令 $k = k + 1$ 转 (2)

算法输入:

- 训练集 $T = \{(x_1, y_1), \dots, (x_N, y_N)\}$, 其中 $x_i \in X \subseteq \mathbb{R}^n, y_i \in \{-1, +1\}$
- 迭代精度 ϵ

初始化:

- 取初始值 $\alpha^0 = 0$, 迭代次数 $k = 0$

迭代步骤:

- 记第 k 次更新后的参数为 α^k
- 遍历所有训练样本, 选出违反 KKT 条件的变量 α_{i_1}
- 根据 $|E_{i_1} - E_{i_2}|$ 选择差值最大的 α_{i_2}
- 使用优化公式更新参数
- 检查所有参数是否满足精度要求: $\sum_{i=1}^N \alpha_i y_i = 0, 0 \leq \alpha_i \leq C$
- 若满足则输出 $\hat{\alpha}$, 否则返回步骤 2 继续迭代



非线性支持向量机

二 线性支持向量机

训练集

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

其中 $x_i \in X \subseteq \mathbb{R}^d$, $y_i \in Y \subseteq \{+1, -1\}$

对偶问题

$$\min_{\alpha} \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i$$

s.t. $\sum_{i=1}^n \alpha_i y_i = 0$
 $0 \leq \alpha_i \leq C$
 $i=1, \dots, n$

原问题

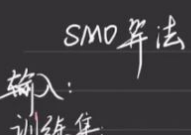
$$\min_{w, b} \frac{1}{2} \|w\|^2$$

s.t. $1 - y_i(w \cdot x_i + b) \leq 0$
 $i=1, \dots, n$

对偶问题的 KKT 条件:

$$\begin{cases} \sum_{i=1}^n \alpha_i y_i K(\cdot, x_i) = 0 \\ \sum_{i=1}^n \alpha_i y_i = 0 \\ C - \alpha_i - \mu_i = 0 \\ 0 \leq \alpha_i \leq C \\ 0 \leq \mu_i \end{cases}$$

- **停止条件验证:**
 - 对于 $0 < \alpha_i < C$: 需验证 $y_i g(x_i) = 1$ (支持向量)
 - 对于 $\alpha_i = 0$: 需验证 $y_i g(x_i) \geq 1$ (分类正确)
 - 对于 $\alpha_i = C$: 需验证 $y_i g(x_i) \leq 1$ (可能分类错误)
 - 所有验证需在精度 ϵ 范围内满足



SMO 算法

输入:

训练集

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

其中 $x_i \in X \subseteq \mathbb{R}^d$, $y_i \in Y \subseteq \{+1, -1\}$

迭代精度: ϵ

输出: $\hat{\alpha}$

算法:

- (1) 取初始值 $\alpha^{(0)} = 0$, 令 $k=0$
- (2) 记第 k 次更新之后的参数为 $\alpha^{(k)}$
- (3) 遍历所有训练样本, 选出违反 KKT 条件的变量 $\alpha_i^{(k)}$
- (4) 根据 $|E_i - E_0|$, 选择差值最大

仍变量下标, 记为 $\alpha_i^{(k)}$:

(5) 通过迭代公式更新参数为 $\alpha_i^{(k+1)}, \alpha_j^{(k+1)}$

(6) 若所有参数在精度 ϵ 内满足:

$$\sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C$$

则停止迭代, 否则令 $k=k+1$ 转 (2)

- **算法特点:**
 - **分解思想:** 将大 QP 问题分解为多个小 QP 问题

- 高效性：每次只优化两个变量，解析求解更新公式
- **收敛性**：Osuna 定理保证算法收敛性
- 适用性：不限制支持向量数量，适合大规模数据集

十九、知识小结

知识点	核心内容	考试重点/易混淆点	难度系数
支持向量机(SVM)简介	1995 年由 CC 老师和薇薇老师首次提出，用于分类任务，基于几何间隔最大化原理	与感知机、逻辑回归的关系	☆☆☆
支持向量机的分类决策函数	符号函数形式： $\text{sgn}(\omega \cdot x + b)$ ，与感知机类似	决策函数与逻辑回归的联系	☆☆
几何间隔与函数间隔	几何间隔：样本点到超平面的距离（带符号）；函数间隔：去掉分母的简化形式	几何间隔的归一化处理	☆☆☆☆
硬间隔最大化	通过硬间隔最大化寻找最优超平面，确保训练数据完全可分	唯一最优解的证明方法	☆☆☆☆
对偶问题求解	通过拉格朗日乘子法转化为对偶问题，利用 KKT 条件求解参数 ω 和 b	支持向量的筛选与 b 的计算	☆☆☆☆ ☆
支持向量的作用	决定分离超平面的关键样本点，位于间隔边界上	非支持向量对模型无影响	☆☆☆
优化目标	最小化 $\ \omega\ ^2$ ，约束条件为函数间隔 ≥ 1	目标函数与约束条件的数学推导	☆☆☆☆
算法流程	构造优化问题→求解对偶问题→计算 ω 和 b →得到决策函数	实际计算中的数值稳定性问题	☆☆☆☆
例题分析	二维数据点 (3,3)、(4,3)、(1,1) 的分类超平面计算	支持向量的确定与间隔计算	☆☆☆