第 20 章 软间隔支持向量机

20.1 动机与原问题

命题 20.1.1 (硬间隔 SVM 的缺点与软间隔的引入)

1. 硬间隔过拟合风险

即使采用大间隔，硬间隔 SVM 仍可能过拟合，主要原因：

- 高维/无限维特征映射 Φ 带来强大表达能力；
- 强制要求“完全可分”——若数据含噪声，则模型会强行拟合异常点。

2. 放弃部分样本：软间隔思想

引入松弛变量 $\xi_n \geq 0$，将硬间隔约束

$$y_n(w^\top z_n + b) \geq 1, \quad \forall n$$

放宽为

$$y_n(w^\top z_n + b) \geq 1 - \xi_n, \quad \forall n,$$

并优化目标

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

其中 $C > 0$ 权衡“大间隔”与“噪声容忍”：

- C 大：近似硬间隔，对误分类惩罚高；
- C 小：允许更多误分类或边界违规，提升鲁棒性。

命题 20.1.2 (软间隔 SVM 的完整表述)

引入松弛变量 $\xi_n \geq 0$ 记录第 n 个样本的“间隔违规”程度，软间隔 SVM 可统一写成如下凸二次规划：

原始问题

$$\begin{aligned} \min_{b,w,\xi} \quad & \tfrac{1}{2}\|w\|^2 + C\sum_{n=1}^{N} \xi_n \\ \text{s.t.} \quad & y_n(w^\top z_n + b) \geq 1 - \xi_n, \quad n = 1, \ldots, N, \\ & \xi_n \geq 0, \quad n = 1, \ldots, N. \end{aligned}$$

超参数 $C > 0$ 的作用

- C 大：惩罚间隔违规，近似硬间隔；
- C 小：容忍较大 ξ_n，换取更大间隔与鲁棒性。

规模　该 QP 含 $\tilde{d} + 1 + N$ 个变量 (b, w, ξ) 与 $2N$ 条线性约束。

下一步目标　将原始问题导出对偶形式，从而再次利用核技巧彻底消除对维度 $\tilde{d}$ 的显式依赖。

20.2 对偶问题

命题 20.2.1 (软间隔 SVM 的拉格朗日对偶推导)

1. 原始问题

$$\min_{b,w,\xi} \ \tfrac{1}{2}\|w\|^2 + C\sum_{n=1}^{N}\xi_n \quad \text{s.t.} \quad \begin{cases} y_n(w^\top z_n + b) \ge 1 - \xi_n, \\ \xi_n \ge 0, \quad n = 1, \dots, N. \end{cases}$$

2. 拉格朗日函数

引入乘子 $\alpha_n \ge 0$, $\beta_n \ge 0$,

$$\mathcal{L}(b, w, \xi, \alpha, \beta) = \tfrac{1}{2}\|w\|^2 + C\sum_{n=1}^{N}\xi_n + \sum_{n=1}^{N}\alpha_n\big[1 - \xi_n - y_n(w^\top z_n + b)\big] - \sum_{n=1}^{N}\beta_n\xi_n.$$

3. 消去 ξ_n 与 β_n

对 ξ_n 求导并令为零得

$$\frac{\partial \mathcal{L}(b, w, \xi, \alpha, \beta)}{\partial \xi_n} = C - \alpha_n - \beta_n = 0 \Longrightarrow \beta_n = C - \alpha_n.$$

于是 β_n 可被隐式约束 $0 \le \alpha_n \le C$ 取代，且

$$\mathcal{L}(b, w, \alpha) = \tfrac{1}{2}\|w\|^2 + \sum_{n=1}^{N}\alpha_n\big[1 - y_n(w^\top z_n + b)\big].$$

4. 消去 b 与 w

与硬间隔 SVM 相同，有

$$\sum_{n=1}^{N}\alpha_n y_n = 0, \qquad w = \sum_{n=1}^{N}\alpha_n y_n z_n.$$

5. 标准软间隔对偶

代入化简后得到

$$\max_{\alpha} \left(\sum_{n=1}^{N}\alpha_n - \tfrac{1}{2}\sum_{n,m}\alpha_n\alpha_m y_n y_m\, z_n^\top z_m\right) \quad \text{s.t.} \quad \sum_{n=1}^{N}\alpha_n y_n = 0,\ 0 \le \alpha_n \le C.$$

命题 20.2.2 (标准软间隔 SVM 的对偶形式)

软间隔 SVM 的对偶问题可表示为以下凸二次规划：

$$\min_{\alpha} \ \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m\, z_n^\top z_m - \sum_{n=1}^{N}\alpha_n$$

$$\text{s.t.} \ \sum_{n=1}^{N} y_n\alpha_n = 0, \quad 0 \le \alpha_n \le C, \quad n = 1, \dots, N.$$

变量与约束规模

- 变量：N 个 α_n。
- 约束：1 条等式约束 $+\ N$ 条下界约束 $+\ N$ 条上界约束，共 $2N+1$ 条。

与硬间隔 SVM 的差异：硬间隔 SVM 对 α_n 仅有非负约束，软间隔 SVM 额外引入上界 $\alpha_n \le C$。

隐式恢复

$$w=\sum_{n=1}^{N}\alpha_n y_n z_n, \qquad \beta_n = C-\alpha_n, \quad n=1,\dots,N.$$

20.3 软间隔 SVM 的启示

算法 20.3.1: 核软间隔 SVM（Kernel Soft-Margin SVM）

输入: 训练集 $\{(x_n,y_n)\}_{n=1}^N$；核函数 $K(\cdot,\cdot)$；惩罚参数 $C>0$

输出: 支撑向量索引集 $\mathcal{I}_{\text{SV}}$，对应 α_n，偏置 b，决策函数 g_{SVM}

1. 构造 QP 参数

$$\begin{aligned} Q_{n,m} &\leftarrow y_n y_m K(x_n,x_m), \quad n,m=1,\dots,N,\\ p \leftarrow -\mathbf{1}_N, &\quad A_{\text{eq}} \leftarrow y^\top, \quad c_{\text{eq}} \leftarrow 0,\\ \text{lb} \leftarrow 0_N, &\quad \text{ub} \leftarrow C\mathbf{1}_N. \end{aligned}$$

2. 求解二次规划

$$\alpha \leftarrow \text{QP}(Q,p,A_{\text{eq}},c_{\text{eq}},\text{lb},\text{ub}).$$

3. 提取支撑向量

$$\mathcal{I}_{\text{SV}} \leftarrow \{n \mid \alpha_n>0\}, \quad \mathcal{I}_{\text{free}} \leftarrow \{n \mid 0<\alpha_n<C\}.$$

若 $s\in\mathcal{I}_{\text{free}}$，由于 $(C-\alpha_s)\xi_s=0$，则 $\xi_s=0 \quad \forall s$，得出 $b=y_s-y_s\xi_s-w^\top z_s=y_s-w^\top z_s$.

4. 计算偏置 任取 $s\in\mathcal{I}_{\text{free}}$（自由支撑向量），则

$$b\leftarrow y_s-\sum_{n\in\mathcal{I}_{\text{SV}}}\alpha_n y_n K(x_n,x_s).$$

5. 决策函数 对任意新样本 x，

$$g_{\text{SVM}}(x)=\text{sign}\Big(\sum_{n\in\mathcal{I}_{\text{SV}}}\alpha_n y_n K(x_n,x)+b\Big).$$

复杂度

- 训练：$\mathcal{O}(N^3)$（QP）+ $\mathcal{O}(N^2)$（核矩阵）。
- 预测：$\mathcal{O}(|\mathcal{I}_{\text{SV}}|)$（仅支撑向量）。

性质

- 与硬间隔 SVM 步骤几乎一致，仅增加上界约束 $\alpha_n\le C$。
- 原始/对偶问题恒有解，更灵活地容忍噪声。

命题 20.3.1 (软间隔高斯核 SVM 的实际表现与调参)

高斯核

$$K(x,x')=\exp\big(-\gamma\|x-x'\|^2\big)$$

与软间隔参数 C 共同决定决策边界形状：

- $C=1$：较大容忍度，边界光滑，欠拟合风险；

- $C = 10$：中等容忍度，边界与数据分布基本匹配；
- $C = 100$：几乎不容忍违规，边界尖锐，过拟合风险上升。

重要提醒　即使采用软间隔，高斯 SVM 仍可能因 (γ, C) 选择不当而 过拟合。

实践要求　必须通过交叉验证或验证集，对 (γ, C) 进行谨慎联合调参。

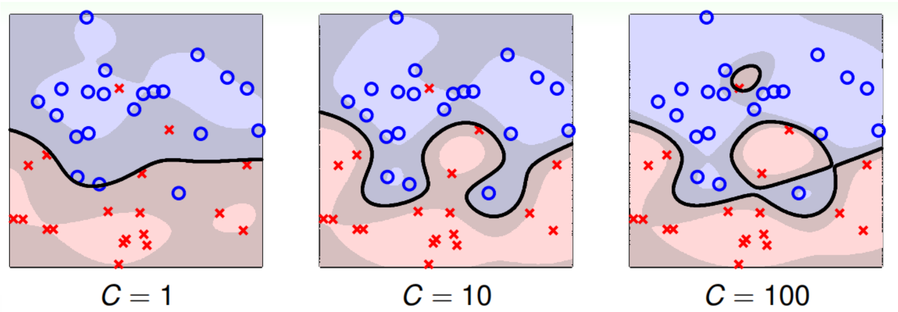

图 20.3.1: 软间隔高斯核函数拟合图像

命题 20.3.2 (α_n 的物理意义)

由互补松弛条件

$$\alpha_n\left[1 - y_n(w^\top z_n + b)\right] = 0, \qquad (C - \alpha_n)\xi_n = 0,$$

可将样本按 α_n 分为三类：

- 非支持向量（$\alpha_n = 0$）
  $\xi_n = 0$，样本落在“胖边界”之外或之上，对决策边界无影响。
- □ 自由支撑向量（$0 < \alpha_n < C$）
  $\xi_n = 0$，样本恰好落在“胖边界”上，用于唯一确定偏置 b。
- △ 边界支撑向量（$\alpha_n = C$）
  $\xi_n > 0$ 记录违规量，样本已越过“胖边界”，但仍在容忍范围内。

结论　系数 α_n 不仅决定支撑向量，还能用于后续数据分析，如异常检测、样本重要性评估等。

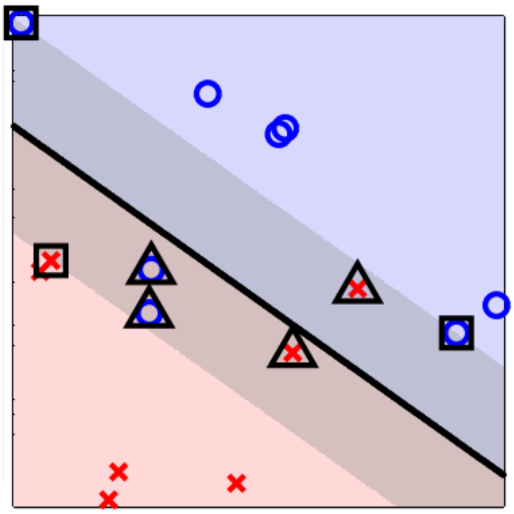

图 20.3.2: α_n 的物理意义

例题 20.1 选择题：SVM 训练误差的范围

对于大小为 10000 的数据集，假设 SVM 的支持向量数为 1126，其中 1000 个支持向量被“有界”（bounded）。则 0/1 误差下 $E_{\text{in}}(g_{\text{SVM}})$ 的可能范围是：

1) $0.0000 \le E_{\text{in}}(g_{\text{SVM}}) \le 0.1000$
2) $0.1000 \le E_{\text{in}}(g_{\text{SVM}}) \le 0.1126$
3) $0.1126 \le E_{\text{in}}(g_{\text{SVM}}) \le 0.5000$
4) $0.1126 \le E_{\text{in}}(g_{\text{SVM}}) \le 1.0000$

解答 正确选项为 $\boxed{1}$。有界支持向量是唯一可能违反“胖边界”的样本。当 $\xi_n \ge 1$ 时，违反会导致 0/1 错误；当 $\xi_n < 1$ 时，违反不导致 0/1 错误。因此，训练误差范围为 $0.0000 \le E_{\text{in}}(g_{\text{SVM}}) \le 0.1000$。■

20.4 模型选择

命题 20.4.1 (SVM 的留一交叉验证误差与支撑向量数的关系)

1. 留一交叉验证误差上界

对含 N 个样本的训练集，留一交叉验证误差满足

$$E_{\text{LOOCV}} \le \frac{|\mathcal{I}_{\text{SV}}|}{N},$$

其中 $\mathcal{I}_{\text{SV}}$ 为全部支撑向量的索引集。

直观解释 若最优解中 $\alpha_N = 0$（即第 N 个样本为非支撑向量），则去掉该样本后剩余的最优 $\alpha_1, \dots, \alpha_{N-1}$ 仍保持不变，因此预测误差为 0；唯有支撑向量才可能导致留一误差。

2. 用支撑向量数调参

- $n_{\text{SV}}(C,\gamma)$ 是 (C,γ) 的非光滑函数，直接优化困难；
- 可作为计算 E_{CV} 过于耗时时的“安全检查”：在少量网格点上排除 n_{SV} 过大的危险模型；
- 仅为上界，实际误差可能远低于该比例。

20.5 总结

笔记 [软间隔支持向量机]

- 动机与原问题：引入松弛变量 ξ_n 以允许间隔违规。
- 对偶问题：拉格朗日乘子 α_n 被常数 C 上界约束。
- 软间隔 SVM 的启示：通过“有界/自由”支持向量进行数据分析。
- 模型选择：使用交叉验证，或近似地依据支持向量数目 n_{SV}。