中國人民大學

RENMIN UNIVERSITY OF CHINA

XXXX

<u>XXXXX</u>

学院 ________________________________

专业 ________________________________

学号 ________________________________

姓名 ________________________________

2025 年 8 月 29 日

摘　要

XXXXX

关键词：XXX

目录

1　问题背景与动机

1.1　硬间隔 SVM 的局限

硬间隔 SVM 追求在训练集上严格线性可分并最大化间隔，但

1. 当数据不可线性可分时无可行解；
2. 对离群点高度敏感：单一点可显著改变最大间隔超平面；
3. 过于强调零训练误差，泛化受损。

1.2　软间隔思想

允许少量样本违反间隔（甚至被误分），用*松弛变量*刻画违约程度，在目标函数中惩罚之。几何上仍保持较宽的间隔，从而提升鲁棒性。

2　软间隔 SVM：约束、几何与优化

2.1　松弛变量与几何解释

给定样本 $\{(x_i, y_i)\}_{i=1}^n, x_i \in \mathbb{R}^d, y_i \in \{\pm 1\}$，判别函数

$$f(x) = \boldsymbol{w}^\top x + \alpha.$$

引入 $\xi_i \geq 0$，把硬间隔约束

$$y_i(\boldsymbol{w}^\top x_i + \alpha) \geq 1$$

放宽为

$$y_i(\boldsymbol{w}^\top x_i + \alpha) \geq 1 - \xi_i, \qquad \xi_i \geq 0.$$

几何分类：

1. $\xi_i = 0$：样本在正确一侧且在间隔之外；
2. $0 < \xi_i < 1$：样本落入间隔区但仍正确分类；
3. $\xi_i \geq 1$：样本被误分。

软间隔的宽度仍定义为 $2/\|\boldsymbol{w}\|$；与硬间隔不同，最近训练点可侵入间隔。

2.2　原始问题（QP）与参数 C

$$\begin{aligned}
\min_{\boldsymbol{w},\alpha,\boldsymbol{\xi}} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^n \xi_i \\
\text{s.t.} \quad & y_i(\boldsymbol{w}^\top x_i + \alpha) \geq 1 - \xi_i,\ i = 1, \ldots, n, \\
& \xi_i \geq 0,\ i = 1, \ldots, n.
\end{aligned} \tag{1}$$

式 (1) 为二次规划：目标二次、约束线性。*正则化超参数* $C>0$ 在“最大化间隔”与“最小化训练误差”之间做权衡。

	C 小	C 大
期望	更大间隔 $1/\|\boldsymbol{w}\|$	使大多数 ξ_i 逼近 0
风险	欠拟合（训练误差偏大）	过拟合（对噪声敏感）
对离群点	不敏感	极敏感

选择 C 通常用验证（交叉验证）完成。

2.3 偏置与同质化

原空间中 $f(x)=\boldsymbol{w}^\top x+\alpha$；同质化升维为

$$x'=(x,1),\quad w'=(\boldsymbol{w},\alpha),\quad f(x)=w'^\top x'.$$

注意 $\boldsymbol{w}^\top(x+\alpha)$ 无意义（维度不匹配）。

3 非线性边界：特征与升维

3.1 总路线

问题：在原空间不可线性可分。策略：构造非线性特征映射 $\Phi:\mathbb{R}^d\to\mathbb{R}^D$，在高维 Φ-空间用*线性*分类器，得到原空间的*非线性*边界。

3.2 抛物提升映射：球面 $\leftrightarrow$ 超平面

取

$$\Phi(x)=\begin{bmatrix}x\\ \|x\|^2\end{bmatrix}\in\mathbb{R}^{d+1}.$$

定理 1 (球 $\leftrightarrow$ 超平面等价)**.** 设原空间球 $\{x:\|x-c\|^2<\rho^2\}$。则

$$\|x-c\|^2<\rho^2 \iff \underbrace{\begin{bmatrix}-2c^\top & 1\end{bmatrix}}_{\text{法向量}}\underbrace{\begin{bmatrix}x\\ \|x\|^2\end{bmatrix}}_{\Phi(x)}<\rho^2-\|c\|^2,$$

即“球内”$\Leftrightarrow$“升维后超平面之下”。超平面是半径 $\to\infty$ 的球的退化情形。

3.3 二次曲面（Quadrics）与二次特征

三维的一般二次曲面

$$Ax_1^2+Bx_2^2+Cx_3^2+Dx_1x_2+Ex_2x_3+Fx_3x_1+Gx_1+Hx_2+Ix_3+\alpha=0.$$

令

$$\Phi(x)=[x_1^2,x_2^2,x_3^2,\ x_1x_2,x_2x_3,x_3x_1,\ x_1,x_2,x_3]^\top,$$

则任意二次边界可写为线性形式 $f(x)=\boldsymbol{w}^\top\Phi(x)+\alpha$。当 d 大时交叉项数量为 $\mathcal{O}(d^2)$，计算代价显著；若仅保留平方项 x_j^2，新增维度线性增长，但边界*只能与坐标轴对齐*（不可任意旋转）。

3.4 一般 p 次多项式边界与升维的普适性

任意次数 $\leq p$ 的多项式

$$f(x) = \sum_{|\alpha| \leq p} c_\alpha \, x^\alpha, \qquad x^\alpha = \prod_{j=1}^{d} x_j^{\alpha_j},$$

通过构造 $\Phi(x)$ 为所有次数 $\leq p$ 的单项式集合（维度 $\binom{d+p}{p}$），都有

$$f(x) = \boldsymbol{w}^\top \Phi(x),$$

从而在 Φ-空间线性化。这一结论说明：*所有多项式形状的边界皆可通过合适升维变为超平面*。但维度随 p 呈 $\mathcal{O}(d^p)$ 爆炸。

3.5 核技巧（Kernel Trick）

对偶形式仅依赖 $\langle \Phi(x_i), \Phi(x) \rangle$。定义核函数

$$K(x,z) = \langle \Phi(x), \Phi(z) \rangle,$$

即可无需显式计算 Φ 而在高维求解。典型示例：多项式核 $K(x,z) = (x^\top z + c)^p$、RBF 核等。核化使高阶多项式在计算上可行。

4 次数、间隔与泛化：复杂度权衡

4.1 提高次数的双重效应

1. 充分高的次数可使原本不可分的数据在高维变为*线性可分*;
2. 次数升高常见到*更宽的有效间隔*（在合适的正则下），但过高会过拟合。

4.2 训练与测试误差随复杂度的典型曲线

随模型复杂度（例如次数 p 或等效自由度）上升，训练误差单调下降；测试误差先降后升。最佳点位于两者平衡处（例如二次决策在示例中给出最小测试误差）。

4.3 与 C 的联动调参

若同时使用多项式特征与软间隔 SVM，需要联合选择

1. 多项式次数 p（控制表示能力）;
2. 正则化超参数 C（控制容错与间隔）。

一般地，不同 p 下的最优 C 不同，改变 p 时应重新验证 C。

5 特征工程：从多项式到任务特定特征

5.1 边缘检测与 HOG

在图像任务中，直接使用原始像素对光照/位置/噪声敏感。可先用 Sobel/方向性高斯导数等算子近似梯度，再在局部网格统计*梯度方向直方图*（如每块 12 个方向 bins），拼接为向量作为特征（HOG），再交由线性 SVM 训练。实证对比：像素 + 线性 SVM 的测试错误率远高于 HOG+ 线性 SVM（错误率显著下降）。

说明 1. 该案例表明：合适的任务特征往往比单纯提升模型复杂度更能改善泛化。

6 实践要点与建议

1. 先做标准化与特征工程，必要时再用核方法提升表示力；
2. 用交叉验证*联合*调参：次数/核参数与 C；
3. 关注鲁棒性：数据含噪/离群点时优先软间隔；
4. 维度-计算量-过拟合三者权衡：去交叉项可显著降维，代价是边界受限（轴对齐）。

附录 A：对偶与核化

A.1 原始问题与拉格朗日函数

软间隔 SVM 的原始（Primal）二次规划为

$$\begin{aligned}
\min_{\boldsymbol{w},\,\alpha,\,\boldsymbol{\xi}} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 \;+\; C\sum_{i=1}^{n}\xi_i \\
\text{s.t.} \quad & y_i(\boldsymbol{w}^\top x_i+\alpha) \;\ge\; 1-\xi_i, \quad i=1,\ldots,n, \\
& \xi_i \;\ge\; 0, \quad i=1,\ldots,n,
\end{aligned} \tag{2}$$

其中 $x_i\in\mathbb{R}^d$, $y_i\in\{\pm1\}$，$C>0$ 为正则化超参数。引入拉格朗日乘子

$$\lambda_i \;\ge\; 0 \quad \text{对应约束} y_i(\boldsymbol{w}^\top x_i+\alpha)\ge 1-\xi_i, \qquad \mu_i \;\ge\; 0 \quad \text{对应约束}\xi_i\ge 0.$$

拉格朗日函数

$$\mathcal{L}(\boldsymbol{w},\alpha,\boldsymbol{\xi};\boldsymbol{\lambda},\boldsymbol{\mu})=\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\lambda_i\big[y_i(\boldsymbol{w}^\top x_i+\alpha)-1+\xi_i\big]-\sum_{i=1}^{n}\mu_i\xi_i. \tag{3}$$

A.2 KKT 条件与驻点

对 $\boldsymbol{w}, \alpha, \xi_i$ 的一阶驻点条件（Stationarity）：

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} &= \boldsymbol{w} - \sum_{i=1}^{n} \lambda_i y_i x_i = \mathbf{0} \;\;\Rightarrow\;\; \boldsymbol{w}^{\star} = \sum_{i=1}^{n} \lambda_i y_i x_i, && (4)\\
\frac{\partial \mathcal{L}}{\partial \alpha} &= -\sum_{i=1}^{n} \lambda_i y_i = 0 \;\;\Rightarrow\;\; \sum_{i=1}^{n} \lambda_i y_i = 0, && (5)\\
\frac{\partial \mathcal{L}}{\partial \xi_i} &= C - \lambda_i - \mu_i = 0 \;\;\Rightarrow\;\; 0 \le \lambda_i \le C,\;\; \mu_i = C - \lambda_i. && (6)
\end{aligned}
$$

再加上原始可行性（Primal feasibility）

$$
y_i(\boldsymbol{w}^{\top} x_i + \alpha) \ge 1 - \xi_i, \quad \xi_i \ge 0,
$$

对偶可行性（Dual feasibility）$\lambda_i \ge 0,\ \mu_i \ge 0\ (\Rightarrow 0 \le \lambda_i \le C)$, 以及互补松弛（Complementary slackness）：

$$
\begin{aligned}
\lambda_i\big[y_i(\boldsymbol{w}^{\top} x_i + \alpha) - 1 + \xi_i\big] &= 0, && (7)\\
\mu_i \xi_i &= 0. && (8)
\end{aligned}
$$

由于问题满足 Slater 条件（取 $\boldsymbol{w} = \mathbf{0},\ \alpha = 0,\ \xi_i = 2$ 即严格可行），强对偶成立，KKT 为充要最优条件。

A.3 消元得到对偶问题

将 (4) 代入 (3) 并最小化掉原始变量，得到*对偶*（*Dual*）：

$$
\begin{aligned}
\max_{\boldsymbol{\lambda}} \quad & \sum_{i=1}^{n} \lambda_i \;-\; \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j\, y_i y_j\, (x_i^{\top} x_j) \\
\text{s.t.} \quad & \sum_{i=1}^{n} \lambda_i y_i = 0, \qquad 0 \le \lambda_i \le C,\;\; i = 1, \ldots, n.
\end{aligned} \tag{9}
$$

该目标函数关于 $\boldsymbol{\lambda}$ 为严格凹函数（负半定二次型），带线性约束的凹优化，标准 QP 可用 SMO 等算法求解。

A.4 偏置项的恢复与判别函数

求得最优 $\boldsymbol{\lambda}^{\star}$ 后，由 (4) 有

$$
\boldsymbol{w}^{\star} = \sum_{i=1}^{n} \lambda_i^{\star} y_i x_i.
$$

对任一*边界支持向量* k（满足 $0 < \lambda_k^{\star} < C$，此时 (7) 给出 $y_k(\boldsymbol{w}^{\star\top} x_k + \alpha^{\star}) = 1$），可恢复偏置

$$
\alpha^{\star} = y_k - \sum_{i=1}^{n} \lambda_i^{\star} y_i\, (x_i^{\top} x_k). \tag{10}
$$

实践中可对所有 $0<\lambda_k^\star<C$ 的样本求得的 (10) 取平均以减小数值误差。最终判别函数为

$$f(x)=\operatorname{sign}\left(\sum_{i=1}^{n}\lambda_i^\star y_i\,(x_i^\top x)+\alpha^\star\right). \tag{11}$$

分类几何解释：间隔宽度为 $2/\|\boldsymbol{w}^\star\|$。样本类型与 KKT 的对应关系：

1. $\lambda_i^\star=0\Rightarrow\xi_i^\star=0,\ y_i(\boldsymbol{w}^{\star\top}x_i+\alpha^\star)\geq 1$：在间隔外，*非*支持向量；
2. $0<\lambda_i^\star<C\Rightarrow\xi_i^\star=0,\ y_i(\boldsymbol{w}^{\star\top}x_i+\alpha^\star)=1$：*边界*支持向量；
3. $\lambda_i^\star=C\Rightarrow y_i(\boldsymbol{w}^{\star\top}x_i+\alpha^\star)=1-\xi_i^\star\leq 1$：在间隔内或被误分的*错误/软*支持向量。

A.5 与合页损失（Hinge Loss）的等价

原始问题 (2) 可改写为正则化风险最小化：

$$\min_{\boldsymbol{w},\alpha}\ \frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{n}\ell_{\text{hinge}}\big(y_i,\ \boldsymbol{w}^\top x_i+\alpha\big),\quad \ell_{\text{hinge}}(y,t)=\max\{0,\ 1-y\,t\}. \tag{12}$$

证明要点：对给定 $(\boldsymbol{w},\alpha)$，对每个 i 的最优 ξ_i 来自

$$\min_{\xi_i\geq 0}\ C\xi_i\quad \text{s.t. }\xi_i\geq 1-y_i(\boldsymbol{w}^\top x_i+\alpha),$$

其最优解为 $\xi_i^\star=\max\{0,\ 1-y_i(\boldsymbol{w}^\top x_i+\alpha)\}$，代回得到 (12)。这说明 SVM 是“二范数正则化 + 合页损失”的最大间隔分类器。

A.6 核化（Kernelization）

在对偶 (9) 与判别 (11) 中，数据仅以内积 $x_i^\top x_j$、$x_i^\top x$ 出现。定义*核函数*

$$K(x,z)=\langle\Phi(x),\Phi(z)\rangle,$$

其中 $\Phi(\cdot)$ 是（可能高维/无限维的）特征映射，且 K 对称且正定。将所有内积替换为核函数：

$$\begin{aligned}\max_{\boldsymbol{\lambda}}\quad & \sum_{i=1}^{n}\lambda_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\,y_iy_j\,K(x_i,x_j)\\ \text{s.t.}\quad & \sum_{i=1}^{n}\lambda_iy_i=0,\qquad 0\leq\lambda_i\leq C.\end{aligned} \tag{13}$$

$$f(x)=\operatorname{sign}\left(\sum_{i=1}^{n}\lambda_i^\star y_i\,K(x_i,x)+\alpha^\star\right). \tag{14}$$

常见核：线性核 $K(x,z)=x^\top z$、多项式核 $K(x,z)=(x^\top z+c)^p$、RBF 核 $K(x,z)=\exp\big(-\|x-z\|^2/(2\sigma^2)\big)$ 等。核化避免了显式构造 Φ 及其维度爆炸。

A.7 硬间隔与极限情形

当数据严格可分且取 $C\to\infty$ 时，软间隔对偶 (13) 退化到硬间隔的对偶，最优解强制 $\xi_i^\star=0$，得到经典最大间隔分离超平面。

A.8 计算与实现要点

1）对偶问题是带盒约束与一个等式约束的凹 QP，SMO 类分解方法高效稳定；2）偏置 $\alpha^\star$ 建议在所有 $0 < \lambda_i^\star < C$ 的边界支持向量上取平均；3）特征缩放与核参数（如多项式次数、RBF 的 σ）必须与 C 一并经验证优化。

小结： 软间隔 SVM 的对偶仅依赖样本的成对相似度（内积/核），最优解由少量支持向量决定；核技巧在不显式升维的情形下实现了高维线性分割，从而在原空间产生非线性决策边界。

附录 B：常见核与特征维度增长

1. 多项式核 $K(x,z) = (x^\top z + c)^p$，对应显式特征维度 $\binom{d+p}{p}$；
2. RBF 核 $K(x,z) = \exp(-\frac{\|x-z\|^2}{2\sigma^2})$，对应无限维特征；
3. 仅平方项时新增维度 $\mathcal{O}(d)$，包含交叉项时为 $\mathcal{O}(d^2)$。