第 19 章 核支持向量机

19.1 核技巧

命题 19.1.1 (对偶 SVM 再审视：脱离维度依赖的关键)

当前已将对偶 SVM 化为

$$\min_{\alpha}\ \tfrac{1}{2}\alpha^{\top}Q\alpha-\mathbf{1}^{\top}\alpha\quad \text{s.t.}\quad y^{\top}\alpha=0,\ \alpha_n\geq 0,$$

其中

$$Q_{n,m}=y_ny_m\,\Phi(x_n)^{\top}\Phi(x_m).$$

目标　避免显式依赖维度 d，需快速计算内积 $\Phi(x_n)^{\top}\Phi(x_m)$。

示例：二阶多项式核定义

$$\Phi_2(x)=\Big(1,\ x_1,\dots,x_d,\ x_1^2,\ x_1x_2,\ \dots,\ x_dx_1,\ x_d^2\Big)^{\top}\in\mathbb{R}^{\frac{(d+1)(d+2)}{2}}.$$

对任意 $x,x'\in\mathbb{R}^d$，其内积可化简为

$$\Phi_2(x)^{\top}\Phi_2(x')=1+\sum_{i=1}^{d}x_ix_i'+\sum_{i=1}^{d}\sum_{j=1}^{d}x_ix_jx_i'x_j'=1+x^{\top}x'+(x^{\top}x')^2,$$

计算复杂度仅 $O(d)$。

命题 19.1.2 (核技巧：变换 + 内积一步完成)

定义核函数

$$K(x,x')\triangleq\Phi(x)^{\top}\Phi(x').$$

示例核：二阶多项式

$$K_2(x,x')=1+x^{\top}x'+(x^{\top}x')^2.$$

1. 二次规划矩阵

$$Q_{n,m}=y_ny_m\,K(x_n,x_m).$$

2. 最优偏置 b 任取支撑向量 (x_s,y_s)（$\alpha_s>0$），则

$$b=y_s-\sum_{n=1}^{N}\alpha_ny_n\,K(x_n,x_s).$$

3. 最优假设 $g_{\mathbf{SVM}}$ 对测试点 x，

$$g_{\mathrm{SVM}}(x)=\operatorname{sign}\Big(\sum_{n=1}^{N}\alpha_ny_n\,K(x_n,x)+b\Big).$$

核技巧　用高效核函数 $K(\cdot,\cdot)$ 代替显式 $\Phi(\cdot)$，彻底摆脱对维度 $\tilde{d}$ 的依赖。

算法 19.1.1: 核硬间隔支持向量机（Kernel Hard-Margin SVM）

输入: 训练集 $\{(x_n, y_n)\}_{n=1}^{N} \subset \mathbb{R}^d \times \{-1, +1\}$；核函数 $K(\cdot,\cdot)$

输出: 支撑向量 $\{(x_n, y_n)\}_{n\in\mathcal{I}_{\text{SV}}}$，对应 α_n，偏置 b，决策函数 g_{SVM}

1. 构造 QP 参数：

$$
\begin{gathered}
Q_{n,m} \leftarrow y_n y_m K(x_n, x_m), \quad \text{for } n, m = 1, \ldots, N \\
p \leftarrow -\mathbf{1}_N, \quad A \leftarrow y^\top, \quad c \leftarrow 0 \\
\text{下界 lb} \leftarrow \mathbf{0}_N, \quad \text{上界 ub} \leftarrow +\infty
\end{gathered}
$$

2. 求解 QP：

$$
\alpha \leftarrow \text{QP}(Q, p, A, c)
$$

3. 提取支撑向量与偏置：

$$
\begin{gathered}
\mathcal{I}_{\text{SV}} \leftarrow \{n \mid \alpha_n > \varepsilon\}, \quad \varepsilon = 10^{-6} \\
b \leftarrow y_s - \sum_{n\in\mathcal{I}_{\text{SV}}} \alpha_n y_n K(x_n, x_s), \quad \text{for any } s \in \mathcal{I}_{\text{SV}}
\end{gathered}
$$

4. 定义决策函数：

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

时间复杂度说明

- $\mathcal{O}(N^2)$：计算核矩阵 Q（N^2 次核函数计算）；
- 二次规划变量数为 N，等式约束 1 个，N 个不等式；
- $\mathcal{O}(\#\text{SV})$：预测阶段仅依赖支撑向量核值。

备注 核 SVM 利用核技巧（kernel trick）避免显式映射到高维特征空间，训练和预测均仅依赖于支撑向量。

例题 19.1 选择题：二次核函数的计算

考虑两个样本 $\mathbf{x}$ 和 $\mathbf{x}'$，满足 $\mathbf{x}^T\mathbf{x}' = 10$。则二次核函数 $K_{\Phi_2}(\mathbf{x}, \mathbf{x}')$ 的值为：

1) 1
2) 11
3) 111
4) 1111

解答 正确选项为 $\boxed{3}$。根据二次核函数公式 $K_{\Phi_2}(\mathbf{x}, \mathbf{x}') = 1 + \mathbf{x}^T\mathbf{x}' + (\mathbf{x}^T\mathbf{x}')^2$，代入 $\mathbf{x}^T\mathbf{x}' = 10$ 得：

$$
K_{\Phi_2}(\mathbf{x}, \mathbf{x}') = 1 + 10 + 10^2 = 111
$$

■

19.2 多项式核

命题 19.2.1 (通用二阶多项式核的三种等价形式)

对任意输入 $x, x' \in \mathbb{R}^d$，二阶多项式核可写成三种等价但几何意义不同的形式：

形式一：带常数项与交叉项

$$\Phi_2(x) = \big(1,\, x_1, \ldots, x_d,\, x_1^2, \ldots, x_d^2\big), \qquad K_2(x, x') = 1 + x^\top x' + (x^\top x')^2.$$

形式二：带缩放系数

$$\Phi_2(x) = \big(1,\, \sqrt{2}x_1, \ldots, \sqrt{2}x_d,\, x_1^2, \ldots, x_d^2\big), \qquad K_2(x, x') = 1 + 2x^\top x' + (x^\top x')^2.$$

形式三：带额外缩放与平方项

$$\Phi_2(x) = \big(1,\, \sqrt{2\gamma}x_1, \ldots, \sqrt{2\gamma}x_d,\, \gamma x_1^2, \ldots, \gamma x_d^2\big), \qquad K_2(x, x') = 1 + 2\gamma x^\top x' + \gamma^2 (x^\top x')^2.$$

最常用紧凑形式

$$K_2(x, x') = (1 + x^\top x')^2, \quad \text{其中隐含系数}\gamma = 1.$$

几何差异　三种 Φ_2 的幂次相同，但内积缩放不同，导致特征空间中的距离与角度各异；而在 SVM 中，常用

$$K_2(x, x') = (1 + x^\top x')^2$$

以获得简洁表达与数值稳定。

命题 19.2.2 (通用多项式核与线性核)

1. 通用多项式核　对阶次 $Q \geq 1$、参数 $\gamma > 0, \zeta \geq 0$，定义

$$K_Q(x, x') = (\zeta + \gamma\, x^\top x')^Q.$$

该核隐式地将样本嵌入到度为 Q 的多项式特征空间 Φ_Q，使 SVM 可在高维空间实现大间隔多项式分类，而计算复杂度与原始维度 d 无关。

2. 线性核（特殊情形）　取 $Q = 1, \gamma = 1, \zeta = 0$ 得

$$K_{\text{linear}}(x, x') = x^\top x',$$

即为标准内积核，对应原始线性 SVM，可直接在原始空间高效求解。

19.3 高斯核

命题 19.3.1 (无限维特征变换的核函数)

若存在高效可计算的核函数 $K(x, x')$，则即使特征映射 $\Phi(x)$ 为无限维，仍可实际应用。

1. 高斯核的无限维展开对一维输入 $x \in \mathbb{R}$，令

$$K(x, x') = \exp\!\big(-(x - x')^2\big) = \exp(-x^2)\exp(-x'^2)\exp(2xx').$$

利用泰勒级数

$$\exp(2xx') = \sum_{i=0}^{\infty} \frac{(2xx')^i}{i!},$$

可得

$$K(x,x') = \exp(-x^2)\exp(-x'^2)\sum_{i=0}^{\infty}\frac{2^i}{i!}x^i x'^i = \Phi(x)^\top\Phi(x'),$$

其中

$$\Phi(x) = \exp(-x^2)\Big(1, \sqrt{2}\,x, \sqrt{\tfrac{2^2}{2!}}\,x^2, \dots, \sqrt{\tfrac{2^i}{i!}}\,x^i, \dots\Big)^\top$$

为无限维向量。

2. 一般高斯核（**RBF** 核）对任意维度 d，定义

$$K(x,x') = \exp\big(-\gamma\|x-x'\|^2\big), \quad \gamma > 0.$$

该核对应无限维特征空间，但仅需 $O(d)$ 时间即可计算核值，实现“无限维变换 + 内积”一步完成。

命题 19.3.2 (高斯核 SVM 的假设与机制)

给定核函数

$$K(x,x') = \exp\big(-\gamma\|x-x'\|^2\big), \quad \gamma > 0,$$

高斯核（又称 RBF 核）SVM 的最终假设为

$$g_{\mathrm{SVM}}(x) = \mathrm{sign}\Big(\sum_{n\in\mathcal{I}_{\mathrm{SV}}}\alpha_n y_n \exp\big(-\gamma\|x-x_n\|^2\big) + b\Big).$$

该式可视作以各支撑向量 x_n 为中心、系数为 $\alpha_n y_n$ 的高斯径向基函数的线性组合。

支撑向量机制

- 通过核技巧，将高阶甚至无限维特征变换 $\Phi(x)$ 简化为高效核值 $K(x,x')$。
- 无需显式存储 w，仅需保留少量支撑向量及其 α_n。
- 高斯 SVM 实现“无限维线性分类”，泛化性能由大间隔理论保证。

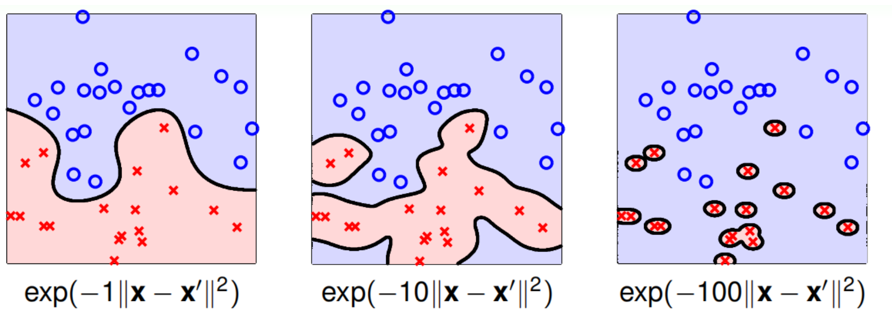

图 19.3.1: 硬间隔高斯核函数拟合图像

命题 19.3.3 (高斯核 SVM 的实际表现与过拟合风险)

高斯核

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

随参数 γ 增大呈如下行为：

- γ 较小：高斯函数平坦，决策边界平滑，欠拟合风险；

- γ 中等：边界与数据分布匹配良好；
- γ 较大：高斯函数尖锐，决策边界剧烈波动，存在过拟合风险。

重要提醒　即使采用大间隔框架，高斯 SVM 仍可能因 γ 过大而过拟合。因此需通过交叉验证或验证集对 γ（以及软间隔参数 C）进行谨慎选择。

例题 19.2 选择题：高斯核函数的极限行为

考虑高斯核函数 $K(\mathbf{x},\mathbf{x}') = \exp\left(-\gamma\|\mathbf{x}-\mathbf{x}'\|^2\right)$。当 $\gamma \to \infty$ 时，该核函数收敛到：

1) $K_{\text{lim}}(\mathbf{x},\mathbf{x}') = 0$
2) $K_{\text{lim}}(\mathbf{x},\mathbf{x}') = [\![\mathbf{x}=\mathbf{x}']\!]$
3) $K_{\text{lim}}(\mathbf{x},\mathbf{x}') = [\![\mathbf{x}\neq\mathbf{x}']\!]$
4) $K_{\text{lim}}(\mathbf{x},\mathbf{x}') = 1$

解答　正确选项为 $\boxed{2}$。当 $\gamma \to \infty$ 时：

- 若 $\mathbf{x}=\mathbf{x}'$，则 $K(\mathbf{x},\mathbf{x}') \to 1$；
- 若 $\mathbf{x}\neq\mathbf{x}'$，则 $K(\mathbf{x},\mathbf{x}') \to 0$。

这等价于指示函数 $[\![\mathbf{x}=\mathbf{x}']\!]$。■

19.4 核函数比较

命题 19.4.1 (线性核的优缺点)

核函数

$$K(x,x') = x^\top x'$$

缺点

- 表达能力受限——并非所有数据都能被线性分离。

优点

- 计算高效：可利用专门的原始 QP 求解器快速求解。
- 可解释性强：权重 w 与支撑向量均具有直观含义。
- 基本原则：先尝试线性模型，必要时再引入非线性核。

结论　线性核是 SVM 工具箱中不可或缺的基础利器。

命题 19.4.2 (多项式核的优缺点)

核函数

$$K(x,x') = (\zeta + \gamma x^\top x')^Q,\quad \gamma > 0,\ \zeta \geq 0,\ Q \in \mathbb{N}.$$

缺点

- 当阶次 Q 较大时易出现数值不稳定。
- 若 $|\zeta + \gamma x^\top x'| < 1$，则核值趋近 0；若大于 1，则迅速放大，导致数值尺度失衡。
- 含 γ, ζ, Q 三个超参数，调参复杂。

优点

- 表达能力优于线性核，可处理非线性可分数据。

- 明确控制多项式阶次 Q，便于解释模型复杂度。
- 对低阶 Q（如 $Q = 2, 3$）往往足够有效。

建议　多项式核通常仅在 Q 较小时实用；高阶时可考虑先在显式高维特征 $\Phi(x)$ 上做线性 SVM，以规避数值与调参难题。

命题 19.4.3 (高斯核的优缺点)

核函数

$$K(x, x') = \exp\big(-\gamma\|x - x'\|^2\big), \quad \gamma > 0.$$

缺点

- 模型"黑箱"化：无限维特征空间导致 w 不可解释。
- 计算速度慢于线性核。
- 表达能力过强，易过拟合。

优点

- 比线性/多项式核更强大，能刻画复杂决策边界。
- 数值稳定，不会出现多项式核的极端放大或缩小。
- 仅一个超参数 γ，调参相对简单。

结论　高斯核是实践中使用最广的核之一，但需通过交叉验证谨慎选择 γ 与软间隔参数 C，以防过拟合。

命题 19.4.4 (其他有效核函数与 Mercer 条件)

核函数 $K(x, x')$ 体现特定相似度 $\Phi(x)^\top\Phi(x')$；"任意相似度 $\Rightarrow$ 有效核"并不成立。

Mercer 充要条件　设 $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ 为候选核，记

$$K_{ij} = K(x_i, x_j), \quad \mathbf{K} = [K_{ij}] \in \mathbb{R}^{N\times N}.$$

则 K 为有效核当且仅当

- 对称性：$K(x, x') = K(x', x)$；
- 半正定性：对任意有限点集 $\{x_n\}_{n=1}^N$，Gram 矩阵

$$\mathbf{K} = \begin{bmatrix} \Phi(x_1)^\top\Phi(x_1) & \cdots & \Phi(x_1)^\top\Phi(x_N) \\ \vdots & \ddots & \vdots \\ \Phi(x_N)^\top\Phi(x_1) & \cdots & \Phi(x_N)^\top\Phi(x_N) \end{bmatrix}$$

  半正定，即 $\mathbf{K} \succeq 0$。

自定义核　任意满足上述 Mercer 条件的对称半正定函数均可作为新核；然而构造与验证通常较为困难。

例题 19.3 选择题：无效核函数的识别

下列哪个不是有效的核函数?（提示：考虑两个 1 维向量 $\mathbf{x}_1 = (1)$ 和 $\mathbf{x}_2 = (-1)$，验证 Mercer 条件。）

1) $K(\mathbf{x}, \mathbf{x}') = (-1 + \mathbf{x}^T\mathbf{x}')^2$
2) $K(\mathbf{x}, \mathbf{x}') = (0 + \mathbf{x}^T\mathbf{x}')^2$
3) $K(\mathbf{x}, \mathbf{x}') = (1 + \mathbf{x}^T\mathbf{x}')^2$

4) $K(\mathbf{x}, \mathbf{x}') = (-1 - \mathbf{x}^T \mathbf{x}')^2$

解答 正确选项为 $\boxed{1}$。构造核矩阵并验证半正定性：

- 选项 1 的核矩阵为 $\begin{pmatrix} 0 & 4 \\ 4 & 0 \end{pmatrix}$，行列式为 $-16 < 0$，不满足 Mercer 条件。
- 其他选项的核矩阵均半正定。

■

19.5 总结

笔记 [核支持向量机]

- 核技巧：核函数充当“变换 + 内积”的快捷方式。
- 多项式核：嵌入经过特殊缩放的多项式特征变换。
- 高斯核：隐式嵌入无限维特征变换。
- 核函数比较：线性核追求效率，高斯核追求表达能力。