

核方法 (Kernel Method)

1. 背景 (Background)

核方法 (Kernel Method) 的引入可以从两个角度来理解：模型角度和优化角度。这也是核方法的核心思想。

1. 从模型角度：非线性带来的高维转换

在之前的章节中，我们讨论了不同的线性模型：

数据特性 (Data Characteristic)	适用算法 (Algorithm)	备注 (Note)
线性可分 (Linearly Separable)	PLA (Perceptron Learning Algorithm)	仅适用于无噪声的完美线性可分数据
有一点错误 (A little noise)	Pocket Algorithm	容忍少量的分类错误
严格非线性 (Strictly Non-Linear)	Kernel Method (Feature Transformation)	需要映射到高维空间才能线性可分

对于严格非线性的数据 X ，我们在原始的输入空间 (Input Space) 无法找到一个线性的超平面将正负样本分开。

解决办法是：将数据从低维空间 \mathcal{X} 映射到高维空间 \mathcal{Z} 。

$$\mathcal{X} \xrightarrow{\phi} \mathcal{Z} \quad (\text{Feature Space}) \quad (1)$$

$$x \rightarrow \phi(x) \quad (2)$$

根据 **Cover's Theorem**，在高维空间中，非线性映射后的模式比在低维空间更容易线性可分。

映射后，我们可以应用标准的线性算法（如 PLA, Hard-Margin SVM），即：

Linear Model in $\mathcal{Z} \iff$ Non-Linear Model in \mathcal{X}

2. 从优化角度：对偶表示带来的内积运算

既然映射到高维空间 \mathcal{Z} 就可以用线性算法解决非线性问题，那为什么这会成为一个难题？因为高维空间的计算非常昂贵（维度灾难）。

然而，如果我们观察 SVM 的优化问题（以 Hard-Margin SVM 为例）：

- **Primal Problem (原问题):**

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2} w^T w \\ \text{s.t.} \quad & y_i (w^T \phi(x_i) + b) \geq 1, \quad i = 1, \dots, N \end{aligned} \quad (3)$$

直接求解原问题需要显式计算 $\phi(x_i)$ ，这在高维空间通常是不可行的。

- **Dual Problem (对偶问题):**

通过拉格朗日乘子法，我们将问题转化为对偶形式：

$$\begin{aligned}
\min_{\lambda} \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j y_i y_j \underbrace{\phi(x_i)^T \phi(x_j)}_{\text{Inner Product}} - \sum_{i=1}^N \lambda_i \\
\text{s.t.} \quad & \lambda_i \geq 0, \quad \sum_{i=1}^N \lambda_i y_i = 0
\end{aligned} \tag{4}$$

我们发现，在对偶问题中，数据点仅以内积 (Inner Product) 的形式出现： $\langle \phi(x_i), \phi(x_j) \rangle$ 。这意味着我们不需要显式地知道 $\phi(x)$ 是什么，只要我们能直接计算出映射后向量的内积即可。

3. 核技巧 (Kernel Trick)

这就引入了核技巧 (Kernel Trick)：

$$K(x, x') = \langle \phi(x), \phi(x') \rangle \tag{5}$$

如果存在这样一个核函数 $K(x, x')$ ，能够直接计算出 $\phi(x)$ 和 $\phi(x')$ 的内积，我们就可以完全避开显式的高维映射，直接在低维空间完成高维空间的计算。

- **Kernel Method (思想)**: 非线性带来高维转换 ($\mathcal{X} \rightarrow \mathcal{Z}$)。
- **Kernel Trick (计算)**: 对偶表示带来内积运算，通过核函数避免显式映射。

2. 正定核 (Positive Definite Kernel)

我们已经知道，核方法的核心在于找到一个函数 $K(x, z)$ ，使得它等于在高维空间中特征向量的内积： $K(x, z) = \langle \phi(x), \phi(z) \rangle$

那么，什么样的函数 $K(x, z)$ 才能作为核函数呢？或者说，给定义一个函数 K ，我们如何判断它是否存在对应的映射 ϕ ？

我们通常所说的核函数，严格来说是 **正定核 (Positive Definite Kernel)**。我们可以从两个角度来定义它。

定义 I：基于映射 (Mapping)

如果存在一个映射 $\phi : \mathcal{X} \rightarrow \mathcal{H}$ (其中 \mathcal{H} 是希尔伯特空间 Hilbert Space)，使得对任意 $x, z \in \mathcal{X}$ ，都有：

$$K(x, z) = \langle \phi(x), \phi(z) \rangle_{\mathcal{H}} \tag{6}$$

那么称 $K(x, z)$ 为正定核函数。

Hilbert Space (\mathcal{H}): 简而言之，是一个完备的 (Complete)、可能是无限维 (Infinite Dimensional) 的、被赋予内积 (Inner Product) 的线性空间。

- 完备性：对极限是封闭的 ($\lim_{n \rightarrow \infty} K_n = K \in \mathcal{H}$)。
- 线性空间：对加法和数乘封闭。
- 内积空间：定义了 $\langle f, g \rangle$ ，满足对称性、正定性、线性性。

定义 II：基于 Gram Matrix

通常我们很难直接找到映射 ϕ ，因此我们需要一个更可操作的定义。

如果函数 $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ 满足以下两个性质：

1. 对称性 (Symmetry):

$$K(x, z) = K(z, x)$$

2. 正定性 (Positive Definiteness):

对任意 N 个数据点 $x_1, x_2, \dots, x_N \in \mathcal{X}$, 其对应的 **Gram Matrix** $K = [K(x_i, x_j)]_{N \times N}$ 是半正定矩阵 (**Positive Semi-Definite, PSD**)。

即对任意非零向量 $\alpha \in \mathbb{R}^N$, 有:

$$\alpha^T K \alpha \geq 0$$

那么称 $K(x, z)$ 为正定核函数。

等价性 (Mercer's Theorem 相关)

这两个定义是等价的:

$$\exists \phi, \text{ s.t. } K(x, z) = \langle \phi(x), \phi(z) \rangle \iff \text{Gram Matrix is PSD} \quad (7)$$

这是一个非常重要的结论 (也就是 Mercer 定理的离散形式)。它告诉我们: 只要保证 K 是对称的, 且其 **Gram Matrix** 是半正定的, 我们就一定能找到对应的特征空间 \mathcal{H} 和映射 ϕ , 而无需知道 ϕ 的具体形式。

这为我们设计核函数提供了理论保证。

必要性证明 (Proof of Necessity: \Rightarrow)

这里我们证明: 如果 $K(x, z)$ 可以表示为特征空间中的内积 $K(x, z) = \langle \phi(x), \phi(z) \rangle$, 那么 K 一定是半正定的 (Positive Semi-Definite) 且对称的。

1. 对称性 (Symmetry)

已知 $K(x, z) = \langle \phi(x), \phi(z) \rangle$ 。

由于内积具有对称性 $\langle a, b \rangle = \langle b, a \rangle$,

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

所以 $K(x, z)$ 满足对称性。

2. 正定性 (Positive Definiteness)

我们需要证明 Gram Matrix K 是半正定的。

即证明: 对于任意 N 个点 x_1, \dots, x_N 和任意向量 $\alpha \in \mathbb{R}^N$, 都有 $\alpha^T K \alpha \geq 0$ 。

$$\begin{aligned} \alpha^T K \alpha &= \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j K_{ij} \\ &= \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j K(x_i, x_j) \\ &= \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \langle \phi(x_i), \phi(x_j) \rangle \\ &= \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \phi(x_i)^T \phi(x_j) \\ &= \left(\sum_{i=1}^N \alpha_i \phi(x_i) \right)^T \left(\sum_{j=1}^N \alpha_j \phi(x_j) \right) \\ &= \left\| \sum_{i=1}^N \alpha_i \phi(x_i) \right\|^2 \end{aligned} \quad (8)$$

由于范数的平方一定是非负的 (≥ 0), 所以:

$$\alpha^T K \alpha \geq 0$$

证毕。这说明 K 是半正定的。