支持向量机 (SVM) 原理详解

1. 问题定义

给定训练数据集 $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$，其中：

- $\mathbf{x}_i \in \mathbb{R}^d$：特征向量
- $y_i \in \{+1, -1\}$：类别标签

目标是学习一个分离超平面：

$$\mathbf{w}^T\mathbf{x} + b = 0$$

决策函数为：

$$f(\mathbf{x}) = \mathrm{sign}(\mathbf{w}^T\mathbf{x} + b)$$

2. 函数间隔与几何间隔

- **函数间隔**：$\hat{\gamma}_i = y_i(\mathbf{w}^T\mathbf{x}_i + b)$
- **几何间隔**：$\gamma_i = \dfrac{y_i(\mathbf{w}^T\mathbf{x}_i + b)}{\|\mathbf{w}\|} = \dfrac{\hat{\gamma}_i}{\|\mathbf{w}\|}$

3. 硬间隔SVM（线性可分）

优化问题：

$$\begin{aligned} &\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{s.t.} \quad y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, \quad i = 1, \ldots, n \end{aligned}$$

4. 软间隔SVM（线性不可分）

引入松弛变量 $\xi_i \geq 0$ 和惩罚参数 $C > 0$：

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

5. 对偶问题与核技巧

拉格朗日函数：

$$L(\mathbf{w}, b, \boldsymbol{\alpha}, \boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i[y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 + \xi_i] - \sum_{i=1}^{n}\mu_i\xi_i$$

对偶问题：

$$\begin{aligned} &\max_{\boldsymbol{\alpha}} \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ &\text{s.t.} \quad 0 \le \alpha_i \le C, \quad \sum_{i=1}^{n}\alpha_i y_i = 0, \quad i = 1, \ldots, n \end{aligned}$$

决策函数：

$$f(\mathbf{x}) = \operatorname{sign}\left(\sum_{i=1}^{n}\alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right)$$

6. 常用核函数

- 线性核：$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T\mathbf{x}_j$
- 多项式核：$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma\mathbf{x}_i^T\mathbf{x}_j + r)^d$
- RBF核：$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$
- Sigmoid核：$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma\mathbf{x}_i^T\mathbf{x}_j + r)$

7. 支持向量的重要性

- 支持向量是满足 $\alpha_i > 0$ 的样本点
- 决策边界仅由支持向量决定
- 具有稀疏性的优点