# 第 17 章 线性支持向量机

## 17.1 大间隔分离超平面

**命题 17.1.1 (线性分类的选择)**

若未来输入 $x$ 受（近似高斯）噪声扰动，则

- 噪声强度与 $x$ 到分离超平面的距离呈正相关；
- 距离越大，对噪声的容忍度越高；
- 因此，噪声容忍度 $\propto$ 超平面到最近样本点 $X_n$ 的距离。

结论 最右侧的超平面（即与最近样本点距离最大的超平面）具有最大的噪声容忍度，从而对过拟合更加鲁棒。

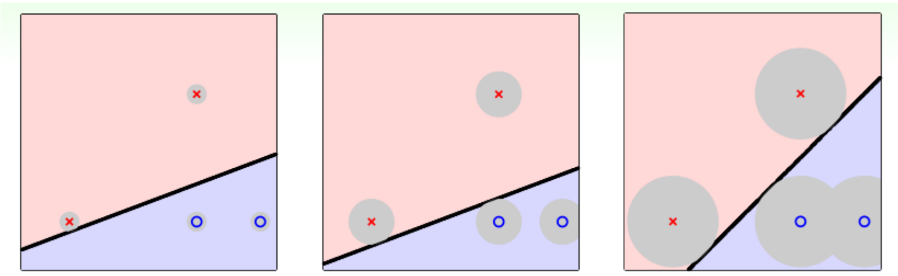

图 17.1.1: 线性分类示意图

**定义 17.1.1 (最大间隔分离超平面（Large-Margin Separating Hyperplane））**

给定训练集 $\mathcal{D} = \{(x_n, y_n)\}_{n=1}^{N}$，其中 $y_n \in \{-1, +1\}$。

优化目标

$$\max_{w} \quad \mathrm{margin}(w) \quad \text{s.t.} \quad y_n w^\top x_n > 0, \quad n = 1, \ldots, N$$

间隔定义

$$\mathrm{margin}(w) = \min_{n=1,\ldots,N} \mathrm{distance}(x_n, w),$$

即超平面到最近样本点的距离，亦称"胖度"。

结论：寻找能够正确分类所有样本且间隔最大的分离超平面。

## 17.2 标准大间隔问题

**命题 17.2.1 (点到超平面的距离：几何推导与最大间隔)**

设超平面 $\mathcal{H}$ 由仿射方程

$$w^\top x + b = 0, \qquad w \in \mathbb{R}^d,\ b \in \mathbb{R}$$

给出，其中 $w \neq 0$。对任意点 $x \in \mathbb{R}^d$，其到超平面 $\mathcal{H}$ 的欧几里得距离 $\mathrm{distance}(x, b, w)$ 的推导如

下：
几何推导
1. 取超平面 $\mathcal{H}$ 上任意一点 $x'$，满足 $w^\top x' + b = 0$。
2. 向量 $x - x'$ 指向点 $x$；将其投影到法向量 $w$ 上即得垂直距离：

$$\text{distance}(x, b, w) = \frac{|w^\top (x - x')|}{\|w\|} = \frac{|w^\top x + b|}{\|w\|},$$

最大间隔（**maximal margin**）
若 $\mathcal{H}$ 是分离超平面，即对所有训练样本 $(x_n, y_n)$ 有

$$y_n(w^\top x_n + b) > 0, \qquad n = 1, \dots, N,$$

则定义训练集到该平面的几何间隔（functional margin 已归一化）为

$$\text{margin}(b, w) \triangleq \min_{n=1,\dots,N} \frac{y_n(w^\top x_n + b)}{\|w\|}.$$

最大化 $\text{margin}(b, w)$ 的超平面 $(b^*, w^*)$ 称为最大间隔分离超平面，满足

$$(b^*, w^*) = \arg\max_{b,w} \min_{n} \frac{y_n(w^\top x_n + b)}{\|w\|}.$$

该平面对所有样本具有最大鲁棒裕度，从而具备最优的泛化保证。

**命题 17.2.2 (最大间隔超平面的归一化推导)**

设训练集 $\{(x_n, y_n)\}_{n=1}^N$ 线性可分。最大间隔超平面可由以下等价优化刻画：
**1.** 原始几何间隔
对任意候选超平面 $(b, w)$，定义

$$\text{margin}(b, w) = \min_n \frac{y_n(w^\top x_n + b)}{\|w\|},$$

其中约束为 $y_n(w^\top x_n + b) > 0$, $n = 1, \dots, N$。
**2.** 标度不变性
超平面方程 $w^\top x + b = 0$ 与 $\lambda w^\top x + \lambda b = 0$ $(\lambda > 0)$ 描述同一平面，故几何间隔仅取决于方向 $(w/\|w\|)$ 与偏移 $b/\|w\|$，整体缩放不影响几何距离。
**3.** 特殊归一化
不失一般性，固定函数间隔为 1：令

$$\min_n y_n(w^\top x_n + b) = 1.$$

在此标度下，几何间隔简化为

$$\text{margin}(b, w) = \frac{1}{\|w\|}.$$

**4.** 最终优化形式
最大化间隔等价于最小化权重范数：

$$\min_{b,w} \quad \frac{1}{2}\|w\|^2 \quad \text{s.t.} \quad y_n(w^\top x_n + b) \geq 1, \quad n = 1, \dots, N.$$

该凸二次规划的唯一解 $(b^*, w^*)$ 即为最大间隔分离超平面。

## 17.3 支持向量机

**定义 17.3.1 (支持向量机（Support Vector Machine，SVM）)**

支持向量机是一种监督式学习方法，用于分类与回归分析。其基本思想是在特征空间中构造一个（或一组）能够最大化类别之间间隔的超平面，使得不同类别的样本点被正确划分，并且离最近样本的距离尽可能远，从而提高模型的泛化能力。

更具体地，给定训练集

$$\{(x_i, y_i)\}_{i=1}^N \subset \mathbb{R}^d \times \{-1, +1\},$$

SVM 求解如下凸二次规划问题：

$$\begin{aligned} \min_{w,b,\xi} \quad & \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^N \xi_i, \\ \text{s.t.} \quad & y_i(w^\top x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, N, \end{aligned}$$

其中 $w \in \mathbb{R}^d$ 为法向量，$b \in \mathbb{R}$ 为偏置，$\xi_i$ 为松弛变量，$C \geq 0$ 为惩罚参数。

满足 $y_i(w^\top x_i + b) = 1$ 的训练样本称为支持向量，它们决定了最优超平面的位置与方向。

当数据线性不可分时，SVM 通过核函数（Kernel trick）将样本映射到高维（甚至无限维）特征空间，使得在高维空间中线性可分，并使用相同的二次规划框架求解。

因此，支持向量机可视为：

- 特征空间上的最大间隔线性分类器；
- 可借助核技巧推广为非线性分类器；
- 最终模型仅由支持向量决定，具有稀疏性和良好泛化能力。

**命题 17.3.1 (SVM 的二次规划刻画)**

对线性可分训练集 $\{(x_n, y_n)\}_{n=1}^N$，支持向量机等价于求解凸二次规划：

$$\min_{b,w} \frac{1}{2}\|w\|^2 \quad \text{s.t. } y_n(w^\top x_n + b) \geq 1,\ n = 1, \dots, N$$

令 $u = \begin{pmatrix} b \\ w \end{pmatrix} \in \mathbb{R}^{d+1}$，可进一步写成标准 QP 形式：

$$\min_u \frac{1}{2}u^\top Q u + p^\top u \quad \text{s.t. } Au \geq c$$

其中：

$$Q = \begin{pmatrix} 0 & 0_d^\top \\ 0_d & I_d \end{pmatrix}, \quad p = 0_{d+1}$$

$$A = \begin{bmatrix} y_1(1, x_1^\top) \\ y_2(1, x_2^\top) \\ \vdots \\ y_N(1, x_N^\top) \end{bmatrix}, \quad c = \mathbf{1}_N$$

该凸问题存在唯一全局最优解 $(b^*, w^*)$，其对应的超平面即为最大间隔分离超平面。

**算法 17.3.1:** 线性硬间隔 SVM（QP 求解器实现）

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

**输出:** 超平面参数 $(b, w)$ 与决策函数 $g_{\text{SVM}}$

**1. 构造 QP 参数**

$p \leftarrow 0_{d+1} \in \mathbb{R}^{d+1}$

对每个 $n = 1, \ldots, N$：

$$a_n \leftarrow y_n \begin{bmatrix} 1 \\ x_n \end{bmatrix} \in \mathbb{R}^{d+1}, \quad c_n \leftarrow 1$$

$$Q \leftarrow \begin{pmatrix} 0 & 0_d^\top \\ 0_d & I_d \end{pmatrix} \in \mathbb{R}^{(d+1)\times(d+1)}$$

**2. 调用 QP 求解器**

$(b, w) \leftarrow \text{QP}(Q,\, p,\, [a_n^\top]_{n=1}^N,\, [c_n]_{n=1}^N)$

**3. 返回模型**

$g_{\text{SVM}}(x) \leftarrow \text{sign}(w^\top x + b)$

**性质**：硬间隔保证所有样本位于“胖边界”之外，无违反。

## 17.4 大间隔超平面的理由

**命题 17.4.1 (大间隔限制二分法)**

设“大间隔算法” $\mathcal{A}_\rho$ 满足：

- 若存在间隔 margin(g) $\geq \rho$ 的假设，则返回该假设；
- 否则返回 $\varnothing$。

对比示例

- $\mathcal{A}_0$（等价于 PLA）可打散“一般”三个输入；
- $\mathcal{A}_{1.126}$（比 SVM 更严格）可能无法打散任何三个输入。

结论　增大间隔 $\rho$ 减少了可实现二分法（dichotomies）的数量，从而

$$d_{\text{VC}}(\mathcal{A}_\rho)\downarrow \quad \Rightarrow \quad \text{泛化性能提升。}$$

**命题 17.4.2 (大间隔算法的 VC 维上界)**

设大间隔算法 $\mathcal{A}_\rho$ 在半径为 $R$ 的超球 $\mathcal{B}_R \subset \mathbb{R}^d$ 上运行，则其 VC 维满足：

$$d_{\text{VC}}(\mathcal{A}_\rho) \leq \min\left\{\left\lceil \frac{R^2}{\rho^2} \right\rceil + 1,\ d+1\right\}$$

特例：二维单位圆当 $\mathcal{X}$ 为单位圆 $S^1 \subset \mathbb{R}^2$ 时：

- $\rho = 0$（退化为感知器）：$d_{\text{VC}} = 3$，可打散任意三点；
- $\rho > \frac{\sqrt{3}}{2}$：因任意三点中必有两点弧距 $\leq \sqrt{3}$，故无法被打散，$d_{\text{VC}} < 3$。

结论　间隔 $\rho$ 越大，$d_{\text{VC}}(\mathcal{A}_\rho)$ 越小，模型复杂度越低，泛化保证越强。

**命题 17.4.3 (大间隔超平面的双重收益)**

令 $\mathcal{H}_\rho$ 表示间隔至少为 $\rho$ 的超平面集合。

**1.** 线性场景

$\mathcal{H}_\rho$ 本身可实现边界有限，VC 维 $d_{\mathrm{VC}}(\mathcal{H}_\rho) \le d+1$；保持模型简单，对泛化有利。

**2.** 引入特征变换 $\Phi$

通过映射 $\Phi : \mathbb{R}^d \to \mathbb{R}^D$ 后，$\mathcal{H}_\rho$ 在高维空间可构造复杂决策边界，同时

$$d_{\mathrm{VC}}(\mathcal{H}_\rho \circ \Phi) \le \min\left\{\frac{R^2}{\rho^2}+1,\, D+1\right\}$$

仍受控；既降低 $E_{\mathrm{in}}$，又维持较低的模型复杂度。

结论　大间隔超平面 + 任意特征变换 $\Phi$ 产生“非线性 SVM”：边界可复杂，参数数量可庞大，但有效 VC 维依旧受限，在降低训练误差的同时保持优异泛化性能。

**例题 17.1 选择题：大间隔算法的 VC 维上界**

考虑在 $\mathcal{Z}$-空间中运行大间隔算法 $\mathcal{A}_\rho$，其中 $\rho = \frac{1}{4}$，特征向量 $\mathbf{z} = \phi(\mathbf{x})$ 的维度为 1126（不包括 $z_0$），且 $\|\mathbf{z}\| \le 1$。使用公式 $\min\left(\frac{R^2}{\rho^2}, d\right) + 1$ 计算 $d_{\mathrm{VC}}(\mathcal{A}_\rho)$ 的上界：

1) 5
2) 17
3) 1126
4) 1127

**解答**　正确选项为 $\boxed{2}$。根据公式 $d_{\mathrm{VC}}(\mathcal{A}_\rho) \le \min\left(\frac{R^2}{\rho^2}, d\right) + 1$：
$R = 1$，$\rho = \frac{1}{4}$，故 $\frac{R^2}{\rho^2} = 16$。$d = 1126$，因此 $\min(16, 1126) = 16$。最终上界为 $16 + 1 = 17$。■

## 17.5 总结

笔记 [线性支持向量机]

- 大间隔分离超平面：直观上对噪声更鲁棒。
- 标准大间隔问题：在特定分离尺度下最小化 $\|w\|$。
- 支持向量机：通过二次规划“轻松”求解。
- 大间隔超平面的理由：二分法数量更少，泛化性能更优。