第 18 章 对偶支持向量机

18.1 对偶 SVM 的动机

命题 18.1.1 (非线性硬间隔 SVM 的 QP 刻画)

给定训练集 $\{(x_n, y_n)\}_{n=1}^N$ 与特征映射 $\Phi : \mathbb{R}^d \to \mathbb{R}^D$，非线性硬间隔 SVM 等价于在 $D+1$ 维空间求解

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

需求：大间隔（参数不多）却能通过 Φ 产生复杂边界。

挑战：当 D 很大甚至无限时，$D+1$ 个变量与 N 条约束的 QP 难以直接求解。

目标：消除对维度 D 的显式依赖，引出后续核技巧。

命题 18.1.2 (拉格朗日乘子：约束转无约束)

令拉格朗日乘子 $\alpha_n \geq 0$，$n = 1, \dots, N$。对硬间隔 SVM 的原始问题

$$\min_{b,w} \tfrac{1}{2} w^\top w \quad \text{s.t. } y_n(w^\top z_n + b) \geq 1,$$

构造拉格朗日函数

$$\mathcal{L}(b, w, \alpha) = \tfrac{1}{2} w^\top w + \sum_{n=1}^N \alpha_n \big[1 - y_n(w^\top z_n + b)\big].$$

等价刻画

$$\min_{b,w} \max_{\alpha \geq 0} \mathcal{L}(b, w, \alpha) = \begin{cases} +\infty, & \text{若存在约束被违反，即存在 k 使得} y_k(w^\top z_k + b) \leq 1; \\ \tfrac{1}{2} w^\top w, & \text{若全部约束满足，即对任意 k 有} y_k(w^\top z_k + b) \geq 1. \end{cases}$$

结果　所有显式约束被隐藏于 $\max\limits_{\alpha \geq 0} \mathcal{L}(b, w, \alpha)$ 中，实现“约束优化 $\to$ 无约束鞍点问题”的转化。

例题 18.1 选择题：硬间隔 SVM 的拉格朗日函数

考虑两个变换后的样本 $(\mathbf{z}_1, +1)$ 和 $(\mathbf{z}_2, -1)$，其中 $\mathbf{z}_2 = -\mathbf{z}_1 = -\mathbf{z}$。硬间隔 SVM 的拉格朗日函数 $\mathcal{L}(b, \mathbf{w}, \boldsymbol{\alpha})$ 为：

1) $\frac{1}{2}\mathbf{w}^T\mathbf{w} + \alpha_1(1 + \mathbf{w}^T\mathbf{z} + b) + \alpha_2(1 + \mathbf{w}^T\mathbf{z} + b)$
2) $\frac{1}{2}\mathbf{w}^T\mathbf{w} + \alpha_1(1 - \mathbf{w}^T\mathbf{z} - b) + \alpha_2(1 - \mathbf{w}^T\mathbf{z} + b)$
3) $\frac{1}{2}\mathbf{w}^T\mathbf{w} + \alpha_1(1 + \mathbf{w}^T\mathbf{z} + b) + \alpha_2(1 + \mathbf{w}^T\mathbf{z} - b)$
4) $\frac{1}{2}\mathbf{w}^T\mathbf{w} + \alpha_1(1 - \mathbf{w}^T\mathbf{z} - b) + \alpha_2(1 - \mathbf{w}^T\mathbf{z} - b)$

解答　正确选项为 $\boxed{2}$。硬间隔 SVM 的拉格朗日函数为：

$$\mathcal{L} = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \alpha_1(1 - y_1(\mathbf{w}^T\mathbf{z}_1 + b)) + \alpha_2(1 - y_2(\mathbf{w}^T\mathbf{z}_2 + b))$$

代入 $y_1 = +1$，$y_2 = -1$，$\mathbf{z}_2 = -\mathbf{z}_1 = -\mathbf{z}$，化简得：

$$\mathcal{L} = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \alpha_1(1 - \mathbf{w}^T\mathbf{z} - b) + \alpha_2(1 - \mathbf{w}^T\mathbf{z} + b)$$

■

18.2 拉格朗日对偶 SVM

命题 18.2.1 (拉格朗日对偶问题)

记拉格朗日函数：

$$\mathcal{L}(b,w,\alpha)=\frac{1}{2}w^\top w+\sum_{n=1}^{N}\alpha_n\left[1-y_n(w^\top z_n+b)\right],\quad \alpha_n\geq 0$$

对任意固定的 $\alpha'\geq 0$，(由于 $max\geq any$)，因此有：

$$\min_{b,w}\left(\max_{\alpha\geq 0}\mathcal{L}(b,w,\alpha)\right)\geq\min_{b,w}\mathcal{L}(b,w,\alpha')$$

取右侧对 α' 取最大，可得：

$$\min_{b,w}\left(\max_{\alpha\geq 0}\mathcal{L}(b,w,\alpha)\right)\geq\max_{\alpha'\geq 0}\min_{b,w}\mathcal{L}(b,w,\alpha')$$

这是由于右侧是左侧的一个下界，而最大化这个下界可以得到一个更紧的下界。

其中右侧即为拉格朗日对偶问题：

$$\max_{\alpha'\geq 0}\min_{b,w}\mathcal{L}(b,w,\alpha')$$

这表示在外层对 α 进行最大化，等价于对原始问题下界的最大化。

强对偶性的条件：对于二次规划问题，当满足以下条件时，弱对偶性的不等式可以变为等式：

- 原问题是凸的
- 原问题可行（如果数据 Φ 是可分的）
- 约束是线性的

这些条件被称为约束品性（constraint qualification）。

当强对偶性成立时，存在原问题和对偶问题的最优解 (b,w,α) 使得两者的最优值相等。♠

命题 18.2.2 (拉格朗日对偶的化简步骤)

记拉格朗日函数：

$$\mathcal{L}(b,w,\alpha)=\frac{1}{2}w^\top w+\sum_{n=1}^{N}\alpha_n\left[1-y_n(w^\top z_n+b)\right],\quad \alpha_n\geq 0$$

1. 消去变量 b

对固定 α，考虑内层无约束最小化问题：

$$\min_{b,w}\mathcal{L}(b,w,\alpha)$$

在最优解处满足偏导数为零的条件：

$$\frac{\partial\mathcal{L}(b,w,\alpha)}{\partial b}=0\implies\sum_{n=1}^{N}\alpha_n y_n=0$$

这一条件使得 b 被消去，目标函数仅依赖于 w。

2. 消去变量 w

同样地，对 w 求偏导并令其为零：

$$\frac{\partial\mathcal{L}(b,w,\alpha)}{\partial w}=0\implies w=\sum_{n=1}^{N}\alpha_n y_n z_n$$

将 w 代入拉格朗日函数，得到仅含 α 的目标函数：

$$\mathcal{L}(\alpha) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2}\left\|\sum_{n=1}^{N} \alpha_n y_n z_n\right\|^2$$

最终对偶问题　通过上述消元过程，原问题转化为仅含 α 的二次规划：

$$\max_{\alpha}\left\{\sum_{n=1}^{N} \alpha_n - \frac{1}{2}\left\|\sum_{n=1}^{N} \alpha_n y_n z_n\right\|^2\right\} \quad \text{s.t.} \quad \sum_{n=1}^{N} \alpha_n y_n = 0,\ \alpha_n \geq 0.$$

命题 18.2.3 (KKT 最优性条件)

设原始问题为

$$\min_{b,w} \frac{1}{2} w^\top w \quad \text{s.t. } y_n(w^\top z_n + b) \geq 1$$

化简后的对偶问题为

$$\max_{\alpha}\left\{\sum_{n=1}^{N} \alpha_n - \frac{1}{2}\left\|\sum_{n=1}^{N} \alpha_n y_n z_n\right\|^2\right\} \quad \text{s.t.} \quad \sum_{n=1}^{N} \alpha_n y_n = 0,\ \alpha_n \geq 0$$

设 (b, w, α) 为原始-对偶最优解，则下列 Karush–Kuhn–Tucker (KKT) 条件成立，且在凸 QP 中既是必要也是充分条件：

1. 原始可行性

$$y_n(w^\top z_n + b) \geq 1, \qquad n = 1, \ldots, N.$$

2. 对偶可行性

$$\alpha_n \geq 0, \qquad n = 1, \ldots, N.$$

3. 对偶-内点最优性

$$\sum_{n=1}^{N} \alpha_n y_n = 0, \qquad w = \sum_{n=1}^{N} \alpha_n y_n z_n.$$

4. 原始-内点最优性（互补松弛）

$$\alpha_n\left[1 - y_n(w^\top z_n + b)\right] = 0, \qquad n = 1, \ldots, N.$$

用途　利用上述 KKT 条件，可从最优 α 直接反演得到 (b, w)，从而完成 SVM 求解。

例题 18.2 选择题：KKT 条件的验证

对于单变量 w，考虑最小化 $\frac{1}{2}w^2$，约束为 $w \geq 1$ 和 $w \leq 3$。拉格朗日函数为 $\mathcal{L}(w, \alpha) = \frac{1}{2}w^2 + \alpha_1(1 - w) + \alpha_2(w - 3)$。下列哪些方程属于该优化问题的 KKT 条件？

1) $\alpha_1 \geq 0$ 且 $\alpha_2 \geq 0$
2) $w = \alpha_1 - \alpha_2$
3) $\alpha_1(1 - w) = 0$ 且 $\alpha_2(w - 3) = 0$
4) 以上所有

解答　正确选项为 $\boxed{4}$。KKT 条件包括：

- 拉格朗日乘子非负性：$\alpha_1 \geq 0$，$\alpha_2 \geq 0$；
- 梯度条件：$w = \alpha_1 - \alpha_2$；
- 互补松弛条件：$\alpha_1(1 - w) = 0$，$\alpha_2(w - 3) = 0$。

所有选项均满足 KKT 条件。■

18.3 求解对偶 SVM

命题 18.3.1 (支持向量机的对偶形式与 QP 形式)

硬间隔支持向量机（SVM）的对偶问题可以表示为以下凸二次规划（Quadratic Programming, QP）问题：

$$\begin{aligned}
\min_{\boldsymbol{\alpha}} \quad & \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \left(z_n^\top z_m\right) - \sum_{n=1}^{N}\alpha_n \\
\text{s.t.} \quad & \sum_{n=1}^{N}\alpha_n y_n = 0, \quad \alpha_n \geq 0, \quad n = 1, \ldots, N.
\end{aligned}$$

这是一个包含 N 个变量与 $N+1$ 条线性约束的凸二次规划问题，目标函数关于 $\boldsymbol{\alpha}$ 严格凸，故存在唯一的全局最优解。

为方便使用标准 QP 求解器，上述问题可转写为通用 QP 形式：

$$\begin{aligned}
\min_{\boldsymbol{\alpha}} \quad & \frac{1}{2}\boldsymbol{\alpha}^\top Q\boldsymbol{\alpha} + \boldsymbol{p}^\top\boldsymbol{\alpha} \\
\text{s.t.} \quad & A\boldsymbol{\alpha} \geq \boldsymbol{c},
\end{aligned}$$

其中各项定义如下：

$$Q_{nm} = y_n y_m\, z_n^\top z_m, \quad \boldsymbol{p} = -\mathbf{1}_N,$$

$$A = \begin{bmatrix} \boldsymbol{y}^\top \\ -\boldsymbol{y}^\top \\ I_N \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} 0 \\ 0 \\ \mathbf{0}_N \end{bmatrix}.$$

注 其中第一行是等式约束 $\sum\alpha_n y_n = 0$，写作两条不等式 $\boldsymbol{y}^\top\alpha \geq 0$ 与 $-\boldsymbol{y}^\top\alpha \geq 0$；第三部分 I_N 实现 $\alpha_n \geq 0$ 的约束。许多数值求解器（如 CVXOPT、quadprog）对等式与边界条件会做特殊处理以保证数值稳定性。

命题 18.3.2 (由最优 α 恢复最优 (b, w) 的 KKT 公式)

设 (b, w, α) 为原始-对偶最优解，则下列 KKT 条件成立：

1. 原始可行性

$$y_n(w^\top z_n + b) \geq 1, \qquad n = 1, \ldots, N.$$

2. 对偶可行性

$$\alpha_n \geq 0, \qquad n = 1, \ldots, N.$$

3. 对偶-内点最优性

$$\sum_{n=1}^{N}\alpha_n y_n = 0, \qquad w = \sum_{n=1}^{N}\alpha_n y_n z_n.$$

4. 互补松弛性

$$\alpha_n\big[1-y_n(w^\top z_n+b)\big]=0,\qquad n=1,\dots,N.$$

恢复公式

1. 最优 w：由对偶-内点最优性直接给出

$$w=\sum_{n=1}^{N}\alpha_n y_n z_n.$$

2. 最优 b：任取一个 $\alpha_k>0$（支撑向量），利用互补松弛得

$$1-y_k(w^\top z_k+b)=0\Longrightarrow b=y_k-w^\top z_k.$$

3. 最终的分类决策函数为：

$$g_{\text{SVM}}(x)=\operatorname{sign}\left(\sum_{n=1}^{N}\alpha_n y_n\left(z_n^\top z\right)+b\right).$$

结论　$\alpha_n>0$ 对应的样本即为支撑向量，位于“胖边界”上。

例题 18.3 选择题：硬间隔 SVM 的最优偏置 b

考虑两个变换后的样本 $(\mathbf{z}_1,+1)$ 和 $(\mathbf{z}_2,-1)$，其中 $\mathbf{z}_2=-\mathbf{z}_1=-\mathbf{z}$。假设硬间隔 SVM 的对偶问题最优解满足 $\alpha_1,\alpha_2>0$，则最优偏置 b 为：

1) -1
2) 0
3) 1
4) 无法确定

解答　正确选项为 $\boxed{2}$。根据 KKT 条件，最优解满足：

$$b=1-\mathbf{w}^T\mathbf{z}\quad \text{和}\quad b=-1+\mathbf{w}^T\mathbf{z}$$

联立得 $b=0$。■

18.4 对偶 SVM 的启示

命题 18.4.1 (“最胖”超平面的数据表示)

设训练集 $\{(x_n,y_n)\}_{n=1}^{N}$，经特征映射 $z_n=\Phi(x_n)$，则

支持向量机

$$w_{\text{SVM}}=\sum_{n=1}^{N}\alpha_n y_n z_n,\qquad \alpha_n \text{ 来自对偶最优解，且仅支撑向量对应的} \alpha_n>0.$$

PLA（感知器）

$$w_{\text{PLA}}=\sum_{n=1}^{N}\beta_n y_n z_n,\qquad \beta_n \text{ 为样本} x_n \text{ 被误分类并用于修正的次数.}$$

共性结论　所有基于梯度下降 / 随机梯度下降的 LogReg、LinReg（初值 $w_0=0$）同样满足

$$w=\sum_{n=1}^{N}\gamma_n y_n z_n,\quad \gamma_n\in\mathbb{R}.$$

术语　称 w 由数据 $\{z_n\}$ “线性表示”；SVM 仅用支撑向量（SVs）即可表示 w，实现稀疏数据依赖。

命题 18.4.2 (硬间隔 SVM 的两种等价形式)

1. 原始硬间隔 **SVM**

$$\min_{b,w} \quad \tfrac{1}{2}w^\top w \quad \text{s.t.} \quad y_n(w^\top z_n + b) \geq 1, \quad n = 1, \ldots, N.$$

2. 对偶硬间隔 **SVM**

$$\begin{aligned} \min_{\alpha} \quad & \tfrac{1}{2}\alpha^\top Q\alpha - \mathbf{1}^\top \alpha \\ \text{s.t.} \quad & y^\top \alpha = 0, \quad \alpha_n \geq 0, \quad n = 1, \ldots, N, \end{aligned}$$

其中 $Q_{n,m} = y_n y_m\, z_n^\top z_m$。

规模与适用场景

- 原始形式：$d+1$ 个变量，N 条约束；适合 $d \ll N$。
- 对偶形式：N 个变量，$N+1$ 条简单约束；适合 $N \ll d$。

物理含义

- 原始：直接寻找经特殊尺度化后的 (b, w)。
- 对偶：识别支撑向量 (z_n, y_n) 及其系数 α_n。

最终输出　两种形式均给出同一最优 (b, w)，对应最胖超平面；决策函数为

$$g_{\text{SVM}}(x) = \text{sign}\big(w^\top \Phi(x) + b\big).$$

例题 18.4 选择题：支持向量候选的数量

考虑在 5566 个样本上应用对偶硬间隔 SVM，得到 1126 个支持向量（SVs）。下列哪个可能是“胖边界”样本（SV 候选）的数量？

1) 0
2) 1024
3) 1234
4) 9999

解答　正确选项为 $\boxed{3}$。支持向量候选数量需满足：

- 大于或等于实际支持向量数量（1126）；
- 小于或等于总样本数（5566）。

选项 3（1234）符合条件。■

18.5 总结

笔记 [对偶支持向量机]

- 对偶 SVM 的动机：去除对维度 d 的依赖。
- 拉格朗日对偶 SVM：KKT 条件连接原始问题与对偶问题。
- 求解对偶 SVM：仍是二次规划，但可用专用求解器更高效地计算。
- 对偶 SVM 的启示：支持向量（SVs）刻画了“最胖”的超平面。