最大分隔面算法中的对偶问题推导

DeepSeek

2026-03-13

1 原始问题 (Primal Problem)

支持向量机（SVM）的核心目标是寻找一个超平面 $w^T x + b = 0$，在保证正确分类训练样本的同时，使分类间隔（Margin）最大化。这一目标可以转化为如下的凸二次规划问题：

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

其中，w 是法向量，b 是偏置，$y_i \in \{+1, -1\}$ 是样本标签。该问题的目标是最小化 $\|w\|^2/2$，等价于最大化几何间隔 $2/\|w\|$。

2 拉格朗日函数的构建

为了将有约束优化问题转化为无约束优化问题，我们引入拉格朗日乘子 $\alpha_i \geq 0$。

2.1 函数定义

定义拉格朗日函数 $L(w, b, \alpha)$ 如下：

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i \left[y_i(w^T x_i + b) - 1 \right] \tag{2}$$

该函数由原始目标函数与约束项的加权组合构成。原始问题可以表示为如下的 **min-max** 形式：

$$\min_{w,b} \max_{\alpha_i \geq 0} L(w, b, \alpha) \tag{3}$$

2.2 符号逻辑解释：为何是减号？

在构造拉格朗日函数时，约束项前的负号并非随意设定的，它背后的逻辑在于通过“惩罚”机制确保约束条件的履行：

1. **约束条件的标准化**：SVM 的原始约束为 $y_i(w^T x_i + b) - 1 \geq 0$。我们将这一项记作 $g_i(w, b)$。
2. **违反约束的代价**：如果某个样本 i 违反了约束（即 $g_i(w, b) < 0$），那么在内部的极大化操作 $\max_{\alpha_i \geq 0} L$ 中，由于 g_i 是负数，前面带有减号的项 $-\alpha_i g_i$ 就变成了正数。此时，只要让 α_i 趋向于 $+\infty$，整个函数值 L 就会趋向于 $+\infty$。
3. **强制执行**：外部的最小化操作 $\min_{w,b}$ 绝不会选择一个使结果为无穷大的解。因此，这种结构强制要求 w 和 b 必须满足所有 $g_i(w, b) \geq 0$ 的条件。此时 $\max_{\alpha_i \geq 0}[-\alpha_i g_i]$ 的最大值只能在 $\alpha_i g_i = 0$ 时取得，即 0。

简而言之，**减号配合非负乘子** α_i 构成了一个强大的“惩罚项”，确保了优化过程在合法的可行域内进行。

3 对偶问题的转化 (max-min)

利用拉格朗日对偶性，通过交换最小化和最大化的顺序，我们得到对偶问题的 **max-min** 形式：

$$\max_{\alpha_i \geq 0} \min_{w,b} L(w, b, \alpha) = \max_{\alpha_i \geq 0} \min_{w,b} \left(\frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i [y_i(w^T x_i + b) - 1] \right) \tag{4}$$

3.1 求解步骤与梯度计算

1. **内部极小化**：首先固定 α，对 w 和 b 求极小值。根据多元函数求极值的必要条件，令 L 对 w 和 b 的偏导数（梯度）为 0：

   - **对 w 求梯度**：

     $$\nabla_w L(w, b, \alpha) = w - \sum_{i=1}^{N} \alpha_i y_i x_i = 0 \implies w = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{5}$$

     该式揭示了最优权重向量 w 本质上是样本点 x_i 的线性组合。

   - **对 b 求偏导**：

     $$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N} \alpha_i y_i = 0 \implies \sum_{i=1}^{N} \alpha_i y_i = 0 \tag{6}$$

     这是一个关于拉格朗日乘子的等式约束。

2. **代入消元过程**：将上述两个结论代回拉格朗日函数 L 以消去 w 和 b：

$$
\begin{aligned}
L(w,b,\alpha) &= \frac{1}{2}w^T w - \sum_{i=1}^{N}\alpha_i y_i w^T x_i - \sum_{i=1}^{N}\alpha_i y_i b + \sum_{i=1}^{N}\alpha_i \\
&= \frac{1}{2}w^T\left(\sum_{j=1}^{N}\alpha_j y_j x_j\right) - w^T\left(\sum_{i=1}^{N}\alpha_i y_i x_i\right) - b\underbrace{\sum_{i=1}^{N}\alpha_i y_i}_{=0} + \sum_{i=1}^{N}\alpha_i \\
&= \sum_{i=1}^{N}\alpha_i - \frac{1}{2}w^T w \\
&= \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\left(\sum_{i=1}^{N}\alpha_i y_i x_i\right)^T\left(\sum_{j=1}^{N}\alpha_j y_j x_j\right)
\end{aligned}
\tag{7}
$$

3. **外部极大化**：整理后得到仅包含对偶变量 α 的最终对偶目标函数：

$$
\begin{aligned}
\max_{\alpha} \quad & \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) \\
\text{s.t.} \quad & \sum_{i=1}^{N}\alpha_i y_i = 0 \\
& \alpha_i \geq 0, \quad i = 1,\ldots,N
\end{aligned}
\tag{8}
$$

4 强对偶性的直观理解

通常情况下，对偶问题的解只是原始问题解的一个下界（弱对偶）。但在 SVM 中，由于满足以下两个关键性质，两者的解完全相等，即满足 **强对偶性 (Strong Duality)**：

- **凸性 (Convexity)**：目标函数和约束空间都是凸的，这意味着问题不存在局部最优陷阱，形状如同一个规则的碗。
- **Slater 条件**：只要训练数据在数学上是线性可分的（即存在至少一个超平面能完美分开两类），则强对偶性成立。

得益于强对偶性，我们通过求解计算上更简便的对偶变量 α，即可获得原始问题的最优分类器。

5 强对偶性的数学定义与条件

通常情况下，对偶问题的最优值总是小于或等于原始问题的最优值（即弱对偶性）。但在 SVM 中，由于问题具备良好的数学性质，两者可以取到相同的值，即满足 **强对偶性 (Strong Duality)**。

为了严谨地说明这一点，我们需要给出以下两个核心条件的数学定义：

5.1 凸性 (Convexity)

在最优化理论中，一个问题被称为凸优化问题，需要满足目标函数是凸函数，且约束集合是凸集。

1. **凸函数定义**：函数 $f(w)$ 被称为凸函数，当且仅当对于定义域内的任意两点 w_1, w_2 及 $\lambda \in [0, 1]$，均满足：

$$f(\lambda w_1 + (1-\lambda)w_2) \le \lambda f(w_1) + (1-\lambda)f(w_2) \tag{9}$$

在 SVM 中，$f(w) = \frac{1}{2}\|w\|^2$ 是一个开口向上的二次函数，其二阶导数恒大于 0，因此是严格凸函数。

2. **凸集定义**：集合 C 被称为凸集，意味着集合内任意两点的连线仍在该集合内。在 SVM 中，线性约束 $y_i(w^T x_i + b) \ge 1$ 定义的是一系列半空间的交集，这在数学上保证了可行域是一个凸集。

5.2 Slater 条件 (Slater's Condition)

Slater 条件是确保强对偶性成立的一个充分条件。对于一个凸优化问题，如果存在一个点 (w, b) 使得所有的不等式约束都严格成立，则强对偶性成立。

严格数学表述：对于原始问题的约束 $g_i(w, b) = 1 - y_i(w^T x_i + b) \le 0$，如果存在至少一组参数 (w, b)，使得：

$$g_i(w, b) < 0, \quad \forall i = 1, \dots, N \tag{10}$$

则称该问题满足 Slater 条件。

直观物理意义：这意味着训练数据集必须是 **线性可分** 的。即存在一个超平面，不仅能把两类样本分开，还能让所有样本点都不落在最大分隔面的边缘上（即所有点距离超平面的函数间隔都严格大于 1）。

5.3 结论

由于 SVM 的目标函数 $\frac{1}{2}\|w\|^2$ 是凸的，且在数据线性可分时满足 Slater 条件，因此根据拉格朗日对偶理论，该问题满足：

$$d^* = \max_{\alpha \ge 0} \min_{w,b} L(w, b, \alpha) = \min_{w,b} \max_{\alpha \ge 0} L(w, b, \alpha) = p^* \tag{11}$$

其中 d^* 为对偶问题的最优值，p^* 为原始问题的最优值。这使得我们通过求解 α 获得的解与直接求解 w, b 获得的结果完全一致。

6 一份给《数据挖掘》课堂的特别提醒

亲爱的同学们，虽然刚才这段推导看起来逻辑丝滑、公式如画，但这主要是因为 LLM 特别擅长模仿数学专家的“优雅姿态”。请务必记住：LLM 可以是你的超级助教，但它偶尔也会在微小的符号或逻辑推导上“一本正经地胡说八道”。

在《数据挖掘》的学习旅程中，LLM 生成的代码和文档只是起点而非终点。请像检查代码 Bug 一样去复核它的每一个 $\sum$ 和 α。最硬核的智力成果依然凝结在人类编写的教材中——那才是不会因为断网或提示词干扰而动摇的真理。总之，**要做 LLM 的驾驭者，不要做它的复读机**。毕竟，期末考试时，坐在考场里的可不是那个会跳 Latex 代码的对话框！