

基于双层规划的代价敏感SVM超参数优化方法

王文琪*, 赵慧秀#

南京理工大学数学与统计学院, 江苏 南京

收稿日期: 2026年5月5日; 录用日期: 2026年5月27日; 发布日期: 2026年6月8日

摘要

针对类别不均衡且漏判代价更高的二分类问题, 提出一种基于双层规划的代价敏感SVM超参数优化方法。该方法在训练集上采用RN策略修正类别权重, 在验证集上以风险最小化学习惩罚系数, 并利用forward-mode超梯度递推完成求解。UCI Breast Cancer Wisconsin数据集实验表明, raw30特征空间下召回率由92.86%提高到95.24%, 总代价由32降至22; 在统一100次评估预算下, 相比anneal可更快达到更低验证损失。结果表明, 该方法能有效降低高代价漏判风险, 但RN增益受特征表达方式影响。

关键词

双层规划, 代价敏感支持向量机, 超参数优化, 相对数量, 预算公平比较

Hyperparameter Optimization of Cost-Sensitive SVM via Bilevel Optimization

Wenqi Wang*, Huixiu Zhao#

School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing Jiangsu

Received: May 5, 2026; accepted: May 27, 2026; published: June 8, 2026

Abstract

For binary classification tasks with class imbalance and higher false-negative cost, a cost-sensitive SVM hyperparameter optimization method based on bilevel optimization is proposed. The method applies the RN strategy on the training set to revise class weights, learns the penalty parameter by minimizing risk on the validation set, and solves the problem using forward-mode hypergradient recursion. Experiments on the UCI Breast Cancer Wisconsin dataset show that, in the raw30 feature

*第一作者。

#通讯作者。

space, recall improves from 92.86% to 95.24%, while the total cost decreases from 32 to 22. Under a unified budget of 100 evaluations, the proposed method reaches a lower validation loss faster than anneal. The results indicate that the method can effectively reduce high-cost false-negative risk, although the gain brought by RN depends on the feature representation.

Keywords

Bilevel Optimization, Cost-Sensitive Support Vector Machine, Hyperparameter Optimization, Relative Number, Budget-Fair Comparison

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在医学诊断、金融风控与入侵检测等二分类任务中,类别不均衡与误分类代价不对称常同时存在。与总体准确率相比,少数类样本被漏判所带来的风险往往更高,模型不仅需要获得较好的区分能力,还应在精度与召回之间体现符合应用需求的代价偏好。支持向量机(support vector machine, SVM)基于结构风险最小化思想,具有良好的泛化能力,因此在上述场景中得到了广泛应用[1]。然而,标准软间隔 SVM 默认对不同类别施加相同的误分类惩罚,在类别分布偏斜或漏判代价明显更高时,决策边界容易向多数类倾斜,从而导致假阴性偏多。围绕这一问题,已有研究通过为不同类别设置差异化惩罚[2],构建代价敏感支持向量机(cost-sensitive support vector machine, CS-SVM),以实现对错类型风险的再加权[3]。

现有 CS-SVM 的一个关键问题在于类别权重的设定。工程实践中常采用按类别样本数量反比设置惩罚比例的方式,以提高少数类样本的影响力。这类方法实现简单,在一定程度上能够缓解类别不均衡带来的边界偏置,但其隐含前提是类别难度主要由样本数量决定。事实上,类内分布的离散程度、边界噪声和类间重叠同样会影响分类器对不同类别的识别难度。若仅依据样本数量构造代价权重,可能低估或高估少数类的真实判别难度,进而影响模型对高代价漏判风险的调节能力。因此,有必要在权重设计中引入反映类内结构的信息,使类别惩罚更贴近数据本身的分布特征。

另一方面,SVM 及其代价敏感扩展的性能对惩罚参数选择高度敏感。传统超参数搜索多采用网格搜索、贝叶斯优化和模拟退火等黑盒方法。此类方法适用性较强,但在连续搜索空间中往往需要较多函数评估,且难以充分利用模型结构信息。当实验预算有限时,搜索效率与结果稳定性都会受到约束。双层规划为超参数学习提供了统一建模思路[4]-[6]:将“给定超参数后的模型训练”表示为下层问题,将“验证集上的泛化风险最小化”表示为上层问题,从而可以通过超梯度直接更新超参数,减少枚举式搜索带来的开销[7][8]。

基于上述考虑,本文提出一种基于双层规划的代价敏感 SVM 超参数优化方法。该方法在双层框架下统一建模类别权重与惩罚参数学习,并在传统数量加权基础上引入相对数量(Relative Number, RN)修正策略,以同时考虑类别样本量和类内结构差异。为保证求解过程可解析且可复现,本文采用可微逻辑损失构造经验风险,通过 forward-mode 超梯度递推与梯度下降实现惩罚参数的高效学习。实验结果表明,该方法在漏判代价更高的设定下能够有效降低假阴性风险,并在预算受限条件下表现出较好的寻优效率。

2. 基于双层规划的代价敏感 SVM 超参数优化

本文面向不均衡二分类场景,构建代价敏感支持向量机的双层超参数学习模型。方法由两部分组成:

一是在训练集上利用相对数量(Relative Number, RN)修正类别权重, 以表征类别不均衡与类内结构差异; 二是以验证集风险最小化为目标, 对惩罚系数 C 进行梯度型优化。为保证上层超梯度可计算, 本文在经验风险项中采用可微的逻辑损失, 并以 forward-mode 递推实现双层求解。为构造可微的双层优化模型, 本文在软间隔 SVM 的代价敏感建模框架下, 采用 logistic 损失对经验风险进行平滑近似, 从而支持超梯度求解。整体建模思路可概括为“下层训练 - 上层验证 - RN 修正 - forward-mode 更新”的统一框架。

2.1. 代价敏感 SVM 与权重修正策略

设样本集为

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N, x_i \in \mathbb{R}^d, y_i \in \{+1, -1\} \quad (1)$$

并将其划分为训练集 \mathcal{T} 、验证集 \mathcal{V} 、测试集 \mathcal{S} 。定义线性判别函数为

$$f_{\theta}(x) = w^{\top}x + b, \theta = (w, b) \quad (2)$$

其中, $w \in \mathbb{R}^d$ 为权重向量, $b \in \mathbb{R}$ 为偏置项。为获得可微的风险表达, 采用逻辑损失

$$\ell(\theta; x_i, y_i) = \log\left(1 + \exp\left(-y_i(w^{\top}x_i + b)\right)\right) \quad (3)$$

为刻画代价不对称与类别不均衡, 引入按类别常数定义的样本权重。记训练集中正类、负类样本索引集合分别为

$$\begin{aligned} \mathcal{T}_+ &= \{i \in \mathcal{T} : y_i = +1\} \\ \mathcal{T}_- &= \{i \in \mathcal{T} : y_i = -1\} \end{aligned} \quad (4)$$

对应样本数分别为

$$m = |\mathcal{T}_+|, n = |\mathcal{T}_-| \quad (5)$$

传统代价敏感 SVM 常按数量反比设置惩罚比例, 其本质是仅依据类别样本数调整少数类与多数类的重要性。该方式实现简单, 但难以反映类内结构差异。为此, 本文在样本数量之外进一步引入类内离散程度, 对类别权重进行 RN 修正。

分别定义两类样本的类中心为

$$\begin{aligned} \mu_+ &= \frac{1}{m} \sum_{i \in \mathcal{T}_+} x_i \\ \mu_- &= \frac{1}{n} \sum_{i \in \mathcal{T}_-} x_i \end{aligned} \quad (6)$$

样本到所属类中心的欧氏距离为

$$\begin{aligned} d_i^+ &= \|x_i - \mu_+\|_2, i \in \mathcal{T}_+ \\ d_i^- &= \|x_i - \mu_-\|_2, i \in \mathcal{T}_- \end{aligned} \quad (7)$$

两类距离均值分别为

$$\begin{aligned} \bar{d}_+ &= \frac{1}{m} \sum_{i \in \mathcal{T}_+} d_i^+ \\ \bar{d}_- &= \frac{1}{n} \sum_{i \in \mathcal{T}_-} d_i^- \end{aligned} \quad (8)$$

对应标准差定义为

$$\begin{aligned} s_+ &= \sqrt{\frac{1}{m} \sum_{i \in T_+} (d_i^+ - \bar{d}_+)^2} \\ s_- &= \sqrt{\frac{1}{n} \sum_{i \in T_-} (d_i^- - \bar{d}_-)^2} \end{aligned} \tag{9}$$

据此, 构造两类的离散系数

$$\begin{aligned} cv_+ &= \frac{s_+}{\bar{d}_+ + \varepsilon} \\ cv_- &= \frac{s_-}{\bar{d}_- + \varepsilon} \end{aligned} \tag{10}$$

其中, $\varepsilon > 0$ 为防止分母为零的极小常数。进一步定义相对数量为

$$\begin{aligned} rn_+ &= \frac{m}{cv_+} \\ rn_- &= \frac{n}{cv_-} \end{aligned} \tag{11}$$

则 RN 给出的类别比率为

$$\rho = \frac{rn_-}{rn_+} \tag{12}$$

式中, ρ 同时综合了样本数量与类内离散程度。当正类样本更少且内部更集中时, ρ 更大, 对应更强的正类惩罚; 当正类离散程度较大时, ρ 会相应减小, 以避免仅凭样本数过度放大惩罚比例。为保证不同权重模式下风险项尺度可比, 本文采用归一化映射, 将未归一化类别权重 $\tilde{\alpha}_+ = \rho$ 、 $\tilde{\alpha}_- = 1$ 转换为

$$\begin{aligned} \alpha_+ &= \frac{(m+n)\tilde{\alpha}_+}{m\tilde{\alpha}_+ + n\tilde{\alpha}_-} \\ \alpha_- &= \frac{(m+n)\tilde{\alpha}_-}{m\tilde{\alpha}_+ + n\tilde{\alpha}_-} \end{aligned} \tag{13}$$

于是样本权重可写为

$$\alpha_i = \begin{cases} \alpha_+, & y_i = +1 \\ \alpha_-, & y_i = -1 \end{cases} \tag{14}$$

该设定保留了代价敏感 SVM “提高少数类误分惩罚” 的基本思想, 同时将类内结构信息纳入类别权重之中。

2.2. 双层规划模型

在给定权重 α_i 后, 下层问题负责在训练集上学习分类参数。本文将代价敏感训练目标定义为

$$g(\theta; C) = \frac{1}{2} \|w\|_2^2 + \frac{C}{|T|} \sum_{i \in T} \alpha_i \ell(\theta; x_i, y_i), \quad C \geq 0 \tag{15}$$

其中, 第一项为 L_2 正则项, 用于控制模型复杂度; 第二项为加权经验风险, 用于体现类别不均衡与误分类代价差异; C 为惩罚系数, 用于平衡两部分的相对作用。

与传统网格搜索、贝叶斯优化或模拟退火等黑盒方法不同, 本文不将 C 视为独立于模型训练的外部搜索变量, 而是通过双层规划将其直接嵌入学习过程。定义验证集上的平均损失为

$$F(\theta) = \frac{1}{|\mathcal{V}|} \sum_{j \in \mathcal{V}} \ell(\theta; x_j, y_j) \quad (16)$$

则本文的双层规划模型可写为

$$\min_{C \in [C_{\min}, C_{\max}]} F(\theta^*(C)) \quad (17)$$

$$\text{s.t. } \theta^*(C) = \arg \min_{\theta} g(\theta; C) \quad (18)$$

式(17)~(18)表明: 上层问题以验证集泛化风险最小化为目标学习最优惩罚系数, 下层问题在给定惩罚系数下完成代价敏感分类器训练。这样, 类别权重由训练数据中的 RN 结构决定, 而惩罚系数则由验证风险驱动自适应更新, 二者共同构成本文方法的核心。在当前线性判别设置下, 本文仅对惩罚系数 C 进行双层学习, 类别权重由既定权重模式生成, 不再作为连续超参数参与外层优化。

2.3. 超梯度求解与代价定义

由于式(18)的解析解难以直接获得, 本文采用迭代优化近似求解下层问题。设内层学习率为 $\eta_{\theta} > 0$, 则第 t 次迭代写为

$$\theta_{t+1} = \theta_t - \eta_{\theta} \nabla_{\theta} g(\theta_t; C), \quad t = 0, 1, \dots, T-1 \quad (19)$$

为了计算上层目标关于惩罚系数 C 的梯度, 引入 forward-mode 灵敏度变量

$$Z_t = \frac{d\theta_t}{dC} \quad (20)$$

若初始点 θ_0 与 C 无关, 则有 $Z_0 = 0$ 。对式(19)关于 C 求导, 可得递推关系

$$Z_{t+1} = Z_t - \eta_{\theta} \left(\nabla_{\theta\theta}^2 g(\theta_t; C) Z_t + \nabla_{\theta C}^2 g(\theta_t; C) \right) \quad (21)$$

在获得内层终点 θ_T 与对应灵敏度 Z_T 后, 由链式法则可得上层超梯度

$$\nabla_C F(\theta_T) = \nabla_{\theta} F(\theta_T)^{\top} Z_T \quad (22)$$

进一步采用外层梯度下降更新惩罚系数

$$C^{(k+1)} = \Pi_{[C_{\min}, C_{\max}]} \left(C^{(k)} - \eta_C \nabla_C F(\theta_T) \right) \quad (23)$$

其中, $\eta_C > 0$ 为外层学习率, $\Pi_{[C_{\min}, C_{\max}]}$ 表示对区间 $[C_{\min}, C_{\max}]$ 的投影算子。

经过若干轮外层迭代后, 可得到最优惩罚系数 C^* , 并据此训练最终分类器。最终模型对测试样本 x 的预测为

$$\hat{y} = \text{sign}(w^{\top} x + b) \quad (24)$$

在代价不对称场景下, 仅以准确率评价分类器并不充分。为更直接反映高代价错误的实际风险, 本文定义总代价指标为

$$\text{Cost} = c_{FN} \cdot FN + c_{FP} \cdot FP \quad (25)$$

其中, FN 和 FP 分别表示假阴性与假阳性样本数, c_{FN} 和 c_{FP} 分别表示对应的单位代价系数。由于本文关注“漏判代价更高”的应用设定, 式(25)将作为后续实验中评价 RN 修正与双层寻优有效性的核心指标。

3. 实验设置与结果分析

3.1. 数据集与实验设置

本文采用 UCI Breast Cancer Wisconsin (Diagnostic)数据集进行实验验证。该数据集共包含 569 个样本和 30 维实值特征, 其中恶性样本记为正类, 良性样本记为负类。数据按 6:2:2 的比例划分为训练集、验证集和测试集, 分别用于下层模型训练、上层超参数学习以及最终性能评估。为保证不同方法之间的可比性, 所有对比实验均在相同的数据划分和随机种子条件下完成。

为考察特征表示对分类性能的影响, 实验分别在原始 30 维特征空间和经标准化后主成分分析降维得到的 6 维特征空间中进行, 分别记为 raw30 和 pca6。对比方法包括双层规划(bilevel)、网格搜索(grid)、贝叶斯优化(bayes)以及模拟退火(anneal)。主结果比较统一在不加权模式(none)下进行, 以考察不同超参数搜索策略本身的差异; 在此基础上, 再在双层规划框架内比较不加权(none)、数量比(count)与相对比(rn)三种权重模式, 从而分析 RN 修正的实际作用。

需要说明的是, 标准化所用的均值和标准差、PCA 投影矩阵, 以及 RN 中类中心、距离均值和离散系数等统计量均仅在训练集上估计, 再分别应用于验证集和测试集。上层超参数学习仅使用验证集信息, 测试集仅在模型与超参数确定后用于一次性最终评估。

从计算开销看, RN 策略只在每次数据切分确定后计算一次类别中心、样本到类中心的距离及其均值和离散系数。设训练集样本数 N_r , 特征维数为 d , 则 RN 统计的主要时间复杂度为 $O(N_r d)$, 属于一次性预处理开销。相比之下, 单次外层超梯度更新需要完成 T 步内层参数迭代及相应的 forward-mode 灵敏度递推, 其主要计算量与训练集上的梯度计算同阶, 可近似记为 $O(TN_r d)$ 。在当前仅优化单个惩罚系数 C 的设定下, 灵敏度变量与模型参数同维, 因此单次超梯度更新的额外开销是可控的。由于本文更关注不同方法在统一评估预算下的总体运行效率, 实际运行时间的比较放在 2.3 节结合预算公平实验一并给出。

评价指标采用准确率、精确率、召回率以及总代价 Cost。考虑到本文关注漏判代价更高的应用设定, 代价系数取 $c_{FN} = 10$ 、 $c_{FP} = 1$, 即

$$Cost = 10 \cdot FN + 1 \cdot FP$$

其中, FN 和 FP 分别表示假阴性和假阳性样本数。该指标能够更直接地反映模型在代价敏感场景中的实际风险。

3.2. 基线条件下的分类结果

Table 1. Comparison of test set performance of different methods in unweighted mode

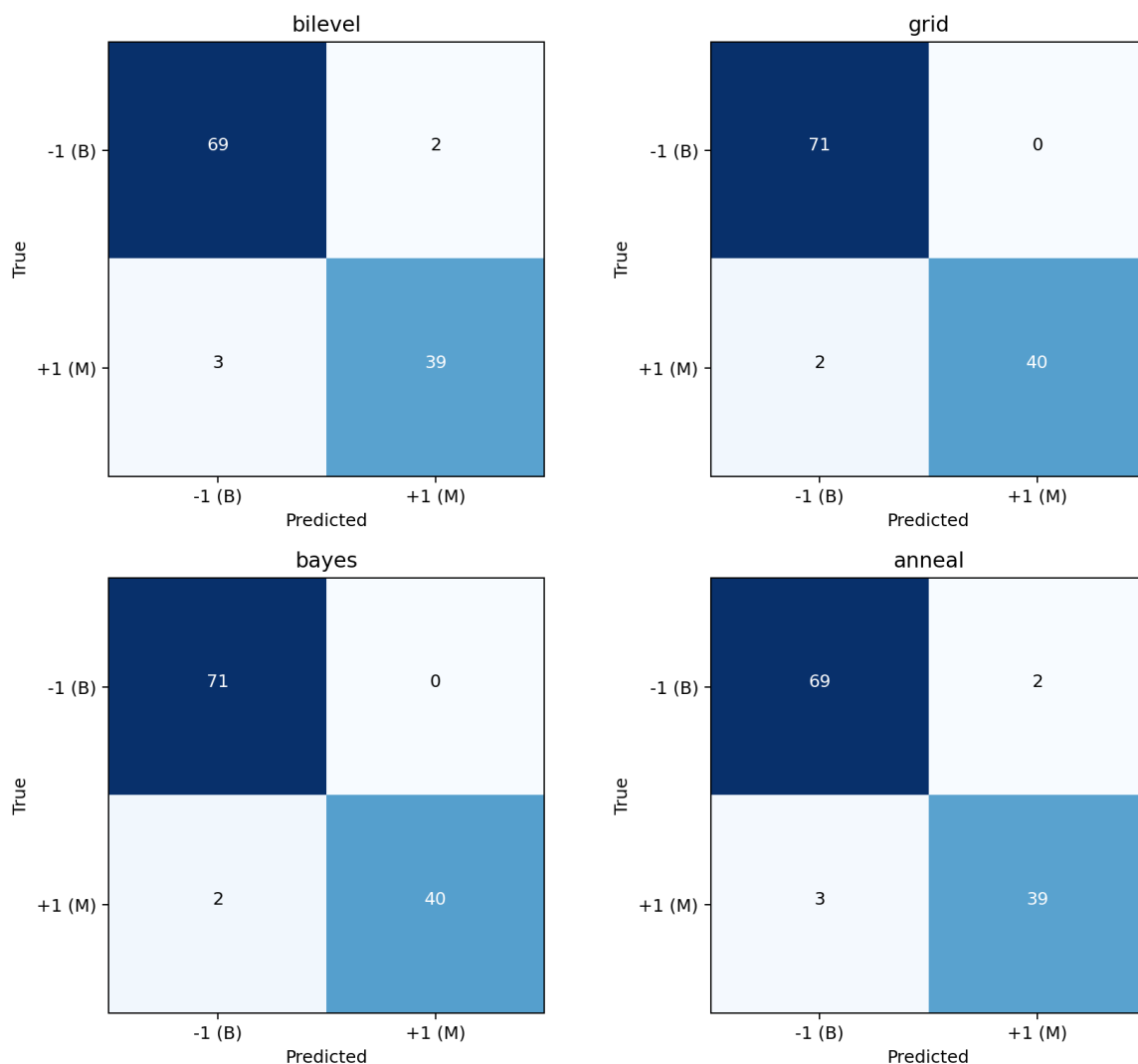
表 1. 不加权模式下不同方法的测试集性能比较

特征空间	方法对比	Acc%	Pre%	Rec%	FP	FN	Cost
raw30	bilevel	95.58	95.12	92.86	2	3	32
raw30	grid	98.23	100.00	95.24	0	2	20
raw30	bayes	98.23	100.00	95.24	0	2	20
raw30	anneal	95.58	95.12	92.86	2	3	32
pca6	bilevel	96.46	95.24	95.24	2	2	22
pca6	grid	96.46	95.24	95.24	2	2	22
pca6	bayes	96.46	95.24	95.24	2	2	22
pca6	anneal	95.58	93.02	95.24	3	2	23

表 1 给出了不加权模式下不同方法在测试集上的分类结果。从 raw30 的结果看, grid 与 bayes 的分类性能更优, 二者的准确率均为 98.23%, 精确率均为 100.00%, 召回率均为 95.24%, 且假阴性均为 2; 相比之下, bilevel 与 anneal 的结果一致, 准确率为 95.58%, 召回率为 92.86%, 假阴性均为 3。说明在原始特征空间中, 若不引入额外权重修正, bilevel 在主性能指标上并不占优, 其主要差距体现在漏判数量略多。

在 pca6 条件下, 不同方法之间的差异明显缩小。bilevel、grid 与 bayes 的准确率均为 96.46%, 召回率均为 95.24%, 对应 $FP=2$ 、 $FN=2$; anneal 略低, 准确率为 95.58%, 精确率为 93.02%, 但召回率保持不变。说明降维后方法之间的边界差异被压缩, 不同搜索方法在测试集上的错误分布趋于接近。

由图 1 可见, 在 raw30 条件下, grid 与 bayes 的优势主要来自假阴性更少且假阳性为 0; 而 bilevel 与 anneal 的错误分布完全一致。相比之下, 在 pca6 条件下, bilevel、grid 与 bayes 的混淆矩阵趋于一致, 说明降维后不同方法的分类边界差异被明显缩小。因此, 图 1 与表 1 共同表明: 在不加权基线条件下, bilevel 具备可用的分类能力, 但其主性能优势并不稳定, 尤其在 raw30 条件下并不优于 grid 和 bayes。



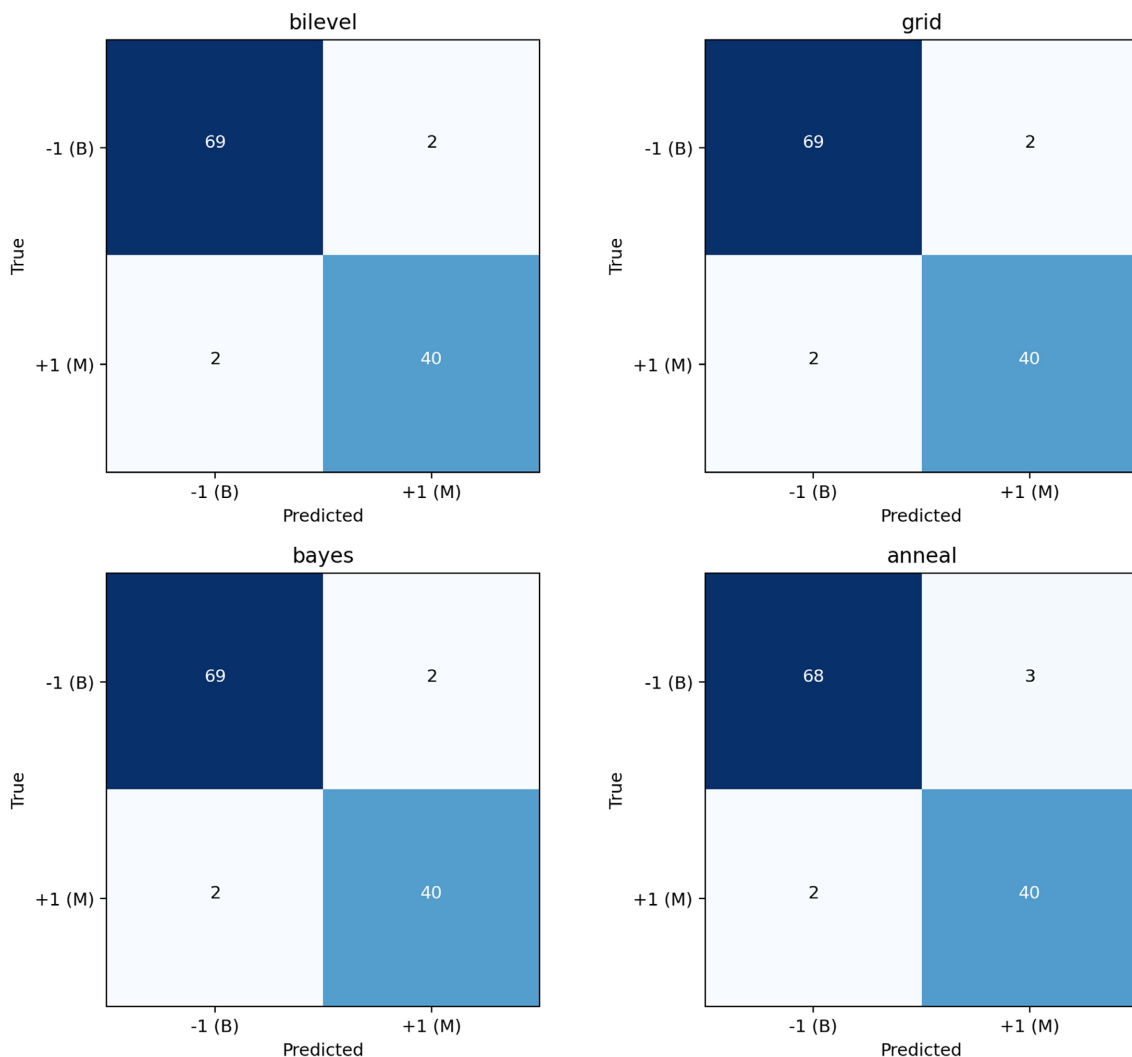


Figure 1. Confusion matrices of different methods under raw30 and pca6 conditions

图 1. 不同方法在 raw30 与 pca6 条件下的混淆矩阵

需要指出的是, 表 1 中部分方法出现相同结果, 并不必然意味着数据泄露或实现错误。在固定数据切分下, 本文仅优化单个惩罚系数 C , 而测试集指标由离散预测标签决定, 因此多个不同的 C 可能对应同一组预测结果, 从而产生相同的混淆矩阵和分类指标。

3.3. 统一预算下的寻优结果

除了最终分类性能外, 本文进一步比较不同方法在有限预算下的寻优效率。为避免不同方法预算上限不一致带来的比较偏差, 本文将 bilevel、grid、bayes 和 anneal 的验证评估预算统一设为 100 次, 并据此重新绘制预算比较曲线。图 2 中的横轴为验证集评估次数 eval_count, 纵轴为当前最优验证损失 best_val_loss, 累计耗时作为辅助审计信息记录。该设置使不同方法能够在相同评估预算下进行直接比较, 从而更完整地反映其收敛过程与最终表现。

由图 2 可见, 在统一 100 次评估预算下, bilevel 在 raw30 和 pca6 条件下均能够较快进入较低的验证损失区域, 说明其在有限预算内具有较好的搜索效率。在 raw30 条件下, bilevel 在 100 次评估后的最优验证损失为 0.25, 累计耗时为 9.37 s, 折算单次评估平均耗时约为 0.09 s; anneal 在相同预算下的最优验

证损失为 0.30, 累计耗时为 10.91 s, 折算单次评估平均耗时约为 0.11 s。作为对照, grid 和 bayes 在 100 次评估预算内的最优验证损失分别为 0.05 和 0.05, 累计耗时分别为 10.70 s 和 10.96 s。

统一100次评估预算下不同方法的预算公平比较结果

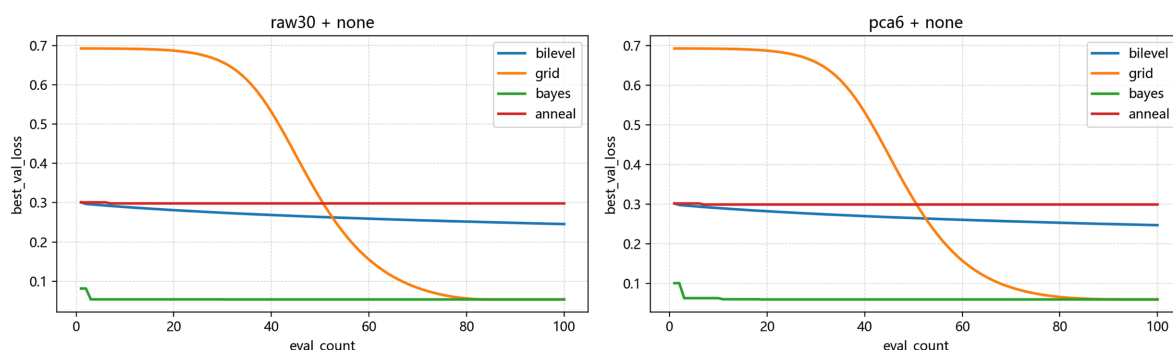


Figure 2. Budget fairness comparison results of different methods under a unified 100 evaluation budget

图 2. 统一 100 次评估预算下不同方法的预算公平比较结果

在 *pca6* 条件下, *bilevel* 在 100 次评估后的最优验证损失为 0.25, 累计耗时为 0.93 s, 折算单次评估平均耗时约为 0.01 s; *anneal* 的最优验证损失为 0.30, 累计耗时为 1.01 s, 折算单次评估平均耗时约为 0.01 s。grid 和 bayes 在相同预算下的最优验证损失分别为 0.06 和 0.06, 累计耗时分别为 0.96 s 和 1.00 s。

结合 2.1 节的复杂度分析可以看出, RN 仅带来一次性的统计预处理开销, 而本文“高效”的主要依据仍然来自统一预算下的总体寻优过程。就当前实验结果而言, *bilevel* 相比 *anneal* 在 *raw30* 与 *pca6* 两种特征空间下均表现出更快达到较低验证损失的趋势, 且单次评估平均耗时也略低于 *anneal*, 说明 *forward-mode* 驱动的双层更新在当前单超参数设置下具有较好的时间效率。与此同时, grid 与 bayes 在统一预算下的最终验证损失仍明显低于 *bilevel*, 这表明不同方法的差异主要体现在搜索路径、计算代价与达到较优解的速度上, 而不完全体现为单一方法在所有指标上的全面占优。

因此, 在统一预算口径下, 本文关于效率的结论可进一步表述为: *bilevel* 相比 *anneal* 仍表现出更快达到较低验证损失的趋势, 而对于 grid 和 bayes, 本文更关注其在相同预算约束下的收敛轨迹与最终结果, 而不再依赖不同上限预算下的局部区间比较。

3.4. RN 修正作用分析

在完成不同搜索方法的基线比较后, 进一步在 *bilevel* 框架内分析 RN 权重修正的作用。

Table 2. Ablation results of bilevel under different weighting modes

表 2. *Abilevel* 在不同权重模式下的消融结果

特征	权重	Acc/%	Pre/%	Rec/%	FP	FN	Cost
raw30	<i>none</i>	95.58	95.12	92.86	2	3	32
raw30	<i>count</i>	96.46	95.24	95.24	2	2	22
raw30	<i>rn</i>	96.46	95.24	95.24	2	2	22
<i>pca6</i>	<i>none</i>	96.46	95.24	95.24	2	2	22
<i>pca6</i>	<i>rn</i>	95.58	93.02	95.24	3	2	23

从表 2 可见, 在 raw30 条件下, RN 修正能够有效降低漏判风险。具体而言, bilevel + none 的假阴性为 3, 总代价为 32; 引入 rn 后, 假阴性降为 2, 总代价降为 22, 降幅为 31.25%。同时, count 与 rn 在 raw30 下得到相同结果, 说明在当前原始特征空间中, RN 至少能够达到与传统数量加权一致的代价调节效果, 并显著改善 bilevel 基线模型的漏判表现。

相比之下, 在 pca6 条件下, RN 未继续带来收益。此时 none 模式下 $FP=2$ 、 $FN=2$ 、 $Cost=22$, 而 rn 使假阳性增加到 3, 假阴性保持不变, 导致 $Cost$ 上升至 23, 对应准确率和精确率均有所下降。说明 RN 的收益并非无条件成立, 而是与特征表示方式有关。对于本文当前实验而言, 最稳定、最有说服力的增益出现在 raw30 的主分析口径下。

因此, 本文更稳妥的结论是: RN 修正能够在 bilevel 框架内有效降低 raw30 条件下的假阴性数量并显著降低总代价, 但其收益具有条件性, 在降维特征空间中不一定保留, 也不足以据此证明 RN-bilevel 优于所有传统搜索方法。

4. 结束语

本文针对类别不均衡且误分类代价不对称的二分类问题, 提出了一种基于双层规划的代价敏感 SVM 超参数优化方法。该方法在训练阶段引入 RN 权重修正策略, 以类别样本数量和类内离散程度共同刻画类别权重, 并在此基础上构建“下层训练 - 上层验证”的双层优化模型, 利用 forward-mode 超梯度递推实现惩罚系数 C 的梯度型学习。由此, 类别权重设定与超参数寻优被统一到同一框架之中, 避免了传统黑盒搜索中训练与验证相分离的问题。

实验结果表明, 在 raw30 特征空间下, RN 修正能够有效降低假阴性数量, 使 bilevel 框架下的召回率由 92.86% 提高到 95.24%, 总代价由 32 降至 22, 说明该方法在漏判代价更高的设定下能够更好地体现代价敏感目标; 在 pca6 特征空间下, RN 未继续带来收益, 表明其改进效果与特征表达方式有关。同时, 在统一 100 次评估预算下, bilevel 相比 anneal 仍表现出更快达到较低验证损失的趋势, 说明该方法在当前单超参数学习设定下兼具一定的风险控制能力与寻优效率。

综合来看, 本文方法在 raw30 主分析口径下实现了较好的漏判控制与寻优效率提升。进一步结合复杂度分析可知, RN 统计仅带来一次性的预处理开销, 而单次超梯度更新的计算量在当前单超参数设置下是可控的。当前实验结论仍主要建立在单数据集、固定切分和单超参数学习设置之上, 后续仍需在更广泛的数据场景和重复实验条件下进一步验证其稳定性。

参考文献

- [1] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1023/a:1022627411411>
- [2] Veropoulos, K., Campbell, C. and Cristianini, N. (1999) Controlling the Sensitivity of Support Vector Machines. *Proceedings of the International Joint Conference on AI*, Stockholm, 31 July-6 August 1999, 55-60.
- [3] 郑恩辉, 李平, 宋执环. 代价敏感支持向量机[J]. 控制与决策, 2006, 21(4): 473-476.
- [4] Sinha, A., Malo, P. and Deb, K. (2018) A Review on Bilevel Optimization: From Classical to Evolutionary Approaches and Applications. *IEEE Transactions on Evolutionary Computation*, **22**, 276-295. <https://doi.org/10.1109/tevc.2017.2712906>
- [5] Kunisch, K. and Pock, T. (2013) A Bilevel Optimization Approach for Parameter Learning in Variational Models. *SIAM Journal on Imaging Sciences*, **6**, 938-983. <https://doi.org/10.1137/120882706>
- [6] Bennett, K.P., Hu, J., Ji, X.Y., Kunapuli, G. and Pang, J.S. (2006) Model Selection via Bilevel Optimization. *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, Vancouver, 16-21 July 2006, 1922-1929. <https://doi.org/10.1109/ijcnn.2006.246935>
- [7] Moore, G.M., Bergeron, C. and Bennett, K.P. (2009) Nonsmooth Bilevel Programming for Hyperparameter Selection.

2009 *IEEE International Conference on Data Mining Workshops*, Miami, 6 December 2009, 374-381.
<https://doi.org/10.1109/icdmw.2009.74>

- [8] Okuno, T., Takeda, A., Kawana, A., *et al.* (2021) On ℓ_p -Hyperparameter Learning via Bilevel Nonsmooth Optimization. *Journal of Machine Learning Research*, **22**, 1-47.