

双层规划在支持向量机超参数选取问题的应用

陈 骞, 徐梦薇*

河北工业大学理学院, 天津

收稿日期: 2024年9月18日; 录用日期: 2024年10月11日; 发布日期: 2024年10月23日

摘 要

支持向量机(SVM)作为一种高效的分类模型, 其性能在很大程度上取决于超参数的选择。本文将SVM的超参数选择问题重新构建为一个双层规划问题, 结合正向模式法和梯度下降法来解决这一问题, 从而获得优化后的SVM模型。为了应对高维数据的挑战, 本文采用了主成分分析法(PCA)对原始数据进行降维处理, 从而提升了SVM模型在高维小样本数据上的表现。通过与当前流行的三种方法: 网格搜索、贝叶斯优化和模拟退火算法进行比较, 结果表明, 采用双层规划方法得到的SVM模型准确率为98.2%, 召回率为100%, 训练时间为0.768 s, 分别优于其他三种方法, 说明本文提出的方法得到的模型具有更好的预测效果。

关键词

双层规划, 支持向量机, 超参数选择, 主成分分析

The Application of Bilevel Optimization in Support Vector Machine Hyperparameter Selection Problem

Qian Chen, Mengwei Xu*

College of Science, Hebei University of Technology, Tianjin

Received: Sep. 18th, 2024; accepted: Oct. 11th, 2024; published: Oct. 23rd, 2024

Abstract

As an efficient classification model, the performance of Support Vector Machine (SVM) depends largely on the hyperparameter selection. In this paper, the hyperparameter selection problem of

*通讯作者。

SVM is reconstructed into a bilevel optimization problem, which is combined with the forward mode method and Gradient descent method to solve this problem, resulting in an optimized SVM model. To tackle the challenges posed by high-dimensional data, this paper employs Principal Component Analysis (PCA) for dimensionality reduction on the original data, thereby enhancing the performance of the SVM model on high-dimensional, small-sample datasets. Comparing the results with three currently popular methods—grid search, Bayesian optimization, and simulated annealing—shows that the SVM model obtained through the proposed bilevel optimization method achieves an accuracy of 98.2%, a recall of 100%, and a training time of 0.768 seconds, outperforming the other three methods. This indicates that the model obtained through our proposed approach has better predictive effectiveness.

Keywords

Bilevel Optimization, Support Vector Machine, Hyperparameter Selection, Principal Component Analysis

Copyright © 2024 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

双层规划(Bilevel Optimization, BLO)是一种起源于 Stackelberg 博弈的数学规划方法,它在多个科学和工程领域中已被证明是一种强大的问题解决框架。在机器学习领域, BLO 为支持向量机等模型的超参数选取问题提供了一种高效的解决方案。与传统的网格搜索等方法相比, BLO 利用其内在的优化层次结构,能够更精确地定位最优超参数,在提高模型性能的同时,显著提升搜索效率。例如 Okuno 等人[1]针对具有非光滑特性的正则化问题,提出了双层规划模型用于选取超参数,与传统的网格搜索方法相比,该方法能够更有效地找到最优超参数。Kunisch 和 Pock [2]将变分图像去噪模型的超参数选取问题表述为一个双层规划问题,提出了半光滑牛顿法求解,和传统方法相比,该算法收敛速度快,计算量小,模型在测试集上泛化效果好。李庆娜等人[3]将支持向量分类的超参数选取问题重新表述为一个双层规划问题,提出了一种基于 Sholtes 收敛的全局松弛法交叉验证算法求解,并证明该算法有更快的收敛速度,得到的支持向量机模型性能更好。Moore 等人[4]将支持向量回归问题中的超参数选取问题表述为一个非光滑的双层规划模型,并提出了近似算法来解决这个双层模型。Kristin P. Bennett 等人[5]将支持向量回归的超参数选取问题表述为双层规划问题求解,证明了该方法相比网格搜索具有更高的灵活性和处理更多超参数的能力。

本文首先采用主成分分析方法对影响乳腺癌的多种因素进行降维处理,结合正向模式法(Forward mode)和梯度下降法(Gradient Descent Method)得到支持向量机的超参数,同时训练得到了 SVM 分类模型。该模型可以用于高维小样本数据的分类,具有较高的泛化能力和较快的训练速度。

2. 基本原理

2.1. 支持向量机

支持向量机(Support Vector Machine, SVM)是由 Vapnik 等人[6]基于统计学习理论中的结构风险最小化和 VC 维理论[7]提出的一种机器学习方法,近年来, SVM 已经被广泛应用于多个领域,如故障检测、

医疗诊断和图像分类等。设数据集 $(x_i, y_i), i=1, 2, \dots, N, x_i \in R^n, y_i \in \{+1, -1\}$, y_i 为标签值, N 为样本量。SVM 的核心目标是找到一个能够最大化分类间隔的决策边界, 即超平面 $\omega^T x + b$ 。当数据是线性可分时, SVM 可以找到决策边界, 然而在现实世界中, 数据往往不是完全线性可分的, 为了解决这一问题, SVM 引入了软间隔的概念, 即允许一些数据点违反间隔规则, 以此让模型拥有更好的泛化能力[8]。SVM 通过决策函数 $f(x) := \text{sgn}(\omega^T x + b)$ 进行分类, 其中 ω, b 分别是超平面的系数和偏置项。根据 SVM 的目标可以得到如下含有不等式约束的凸二次规划问题:

$$\begin{aligned} \min_{\omega, b} & 1/2 \|\omega\|_2^2 + C \sum_{i=1}^m \xi_i \\ \text{s.t.} & y_i (\omega^T x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, 2, \dots, n. \end{aligned} \quad (1)$$

C 叫做正则化参数, 控制对超出边界样本的惩罚程度, 在选择了适合的核函数 $k(x_i, x_j)$ 、适合的精度以及正则化参数后, 构造求解在约束条件下的凸二次规划优化问题:

$$\begin{aligned} \min_{\alpha} & 1/2 \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(x_i, x_j) - \sum_{i=1}^n \alpha_i \\ \text{s.t.} & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, \dots, n. \end{aligned} \quad (2)$$

2.2. 双层规划

双层规划问题(Bilevel Optimization, BLO)最早来源于, 1955 年 Stackelberg 在其研究市场经济的著作中, 提出了双层规划问题的近似模型。由于强大的建模能力, BLO 已被公认为各种机器学习和计算机视觉应用的重要工具。使用 BLO 选择超参数已经被广泛用于影像诊断, 非凸非光滑问题以及支持向量机等问题中, 双层规划强大的编程理论和算法的可用性使机器学习模型和方法的新研究成为可能。

从数学上讲, BLO 可表述为:

$$\begin{aligned} \min_{x \in X, y \in Y} & F(x, y) \\ \text{s.t.} & y \in S(x) := \arg \min_{y \in Y} f(x, y) \end{aligned} \quad (3)$$

其中 $F, f: R^n \times R^m \rightarrow R$ 分别是上下层目标函数, x, y 则各自对应上下层的决策变量, $S(x)$ 是下层目标函数给定 x 下的最优解集。双层规划将一个问题嵌入到另一个问题中, 下层问题的解作为上层问题的约束。

2.3. 主成分分析

在对高维数据进行分类时, 变量个数太多会增加模型训练的复杂性, 影响分类的效果。主成分分析作为一种广泛使用的多元统计方法, 利用坐标变换将存在相关性的原始变量转换为主成分, 可以显著减少变量的个数, 同时保证降维后的特征尽可能反映原有变量的内部结构信息。使用时将数据划分为训练集和测试集, 并根据训练集的方差累计贡献率确定主成分个数。在测试阶段, 将测试集数据与训练集的投影矩阵相乘并获得测试集特征。其中方差贡献率越大说明该主成分包含的有效信息越多。

3. 双层规划求解超参数

我的研究目标是利用双层规划方法选择支持向量机的超参数, 得到支持向量机分类模型, 提高分类模型精度, 同时简化模型并提供合理解释。通过研读相关方向的文献, 我建立了如下的双层模型。

$$\begin{aligned} & \min_{C \in R} \sum_{i=1}^{m_1} \log(1 + e^{-\bar{y}_i(\bar{x}_i^T \omega)}) \\ \omega \in \arg \min_{\omega \in R^n} & 1/2 \|\omega\|_2^2 + C \sum_{i=1}^{m_2} J_{\omega(\hat{x}_i, \hat{y}_i)} \\ & C \geq 0, \end{aligned} \tag{4}$$

其中 m_1 为验证集的样本点个数, m_2 为训练集的样本点个数, C 是超参数, $J_w(\cdot)$ 为 logistic 损失函数, $J_{\omega}(\hat{x}_i, \hat{y}_i) = \frac{1}{2}(1 + \hat{y}_i) \log(1 + e^{-\hat{x}_i^T \omega}) + \frac{1}{2}(1 - \hat{y}_i) \log(1 + e^{\hat{x}_i^T \omega})$. 上层问题是最小化验证集的损失, 下层问题是最小化训练集的损失。由于 Logistic Loss 的凸性以及连续可微性, 使得相应的下层优化问题成为强凸且是可微的, 对于每一个 $C \in R$, 下层都有唯一的 $\omega(C)$, 根据 Franceschi 等人的研究[9], 当下层问题是强凸且可微时正向模式算法是收敛的, 因此上述方法可以应用于本文。

设(4)上下层目标函数分别为 $F(C, \omega), f(C, \omega)$,

$$F(C, \omega) = \sum_{i=1}^{m_1} \log(1 + e^{-\bar{y}_i(\bar{x}_i^T \omega)}), f(C, \omega) = 1/2 \|\omega\|_2^2 + C \sum_{i=1}^{m_2} J_{\omega(\hat{x}_i, \hat{y}_i)} \tag{5}$$

根据链式法则有:

$$\nabla_C F(C, \omega) = \nabla_C F(C, \omega) + \nabla_{\omega} F(C, \omega) \frac{d\omega^*(C)}{dC} = \sum_{i=1}^{m_1} \frac{-\exp\{-\bar{y}_i(\bar{x}_i^T \omega)\} \bar{y}_i \bar{x}_i^T}{1 + \exp\{-\bar{y}_i(\bar{x}_i^T \omega)\}} \frac{d\omega^*(C)}{dC} \tag{6}$$

下层目标函数对 ω 求导得:

$$\nabla_{\omega} f(C, \omega) = \omega + \frac{1}{2} C \sum_{i=1}^{m_2} \left[(1 + \hat{y}_i) \frac{-\hat{x}_i \exp\{-\hat{x}_i^T \omega\}}{1 + \exp\{-\hat{x}_i^T \omega\}} + (1 - \hat{y}_i) \frac{\hat{x}_i \exp\{\hat{x}_i^T \omega\}}{1 + \exp\{\hat{x}_i^T \omega\}} \right] \tag{7}$$

下层使用梯度下降法则有:

$$\omega_t = \omega_{t-1} - \eta (\nabla_{\omega_{t-1}} f(C, \omega_{t-1})), t = 1, 2, \dots, T. \tag{8}$$

根据链式法则有:

$$\begin{aligned} \frac{d\omega_t}{dC} &= \frac{\partial \omega_t}{\partial \omega_{t-1}} \frac{d\omega_{t-1}}{dC} + \frac{\partial \omega_t}{\partial C} \\ \frac{\partial \omega_t}{\partial \omega_{t-1}} &= 1 - \eta \left[1 + \frac{C}{2} \sum_{i=1}^{m_2} \left(\frac{(1 + \hat{y}_i) \hat{x}_i \hat{x}_i^T \exp\{-\hat{x}_i^T \omega\}}{(1 + \exp\{-\hat{x}_i^T \omega\})^2} + \frac{(1 - \hat{y}_i) \hat{x}_i \hat{x}_i^T \exp\{\hat{x}_i^T \omega\}}{(1 + \exp\{\hat{x}_i^T \omega\})^2} \right) \right] \\ \frac{\partial \omega_t}{\partial C} &= -\frac{\eta}{2} \left[\sum_{i=1}^{m_2} \left(\frac{(1 + \hat{y}_i) - \hat{x}_i \exp\{-\hat{x}_i^T \omega\}}{(1 + \exp\{-\hat{x}_i^T \omega\})^2} + \frac{(1 - \hat{y}_i) \hat{x}_i \exp\{\hat{x}_i^T \omega\}}{(1 + \exp\{\hat{x}_i^T \omega\})^2} \right) \right] \end{aligned} \tag{9}$$

定义矩阵:

$$Z_t = \frac{d\omega_t}{dC}, A_t = \frac{\partial \omega_t}{\partial \omega_{t-1}}, B_t = \frac{\partial \omega_t}{\partial C} \tag{10}$$

那么, 上式可以写为:

$$Z_t = A_t Z_{t-1} + B_t, t = 1, 2, \dots, T. \tag{11}$$

利用正向模式法(Forward mode)方法, 根据下层的最后一次迭代 Z_T 优化 C :

$$Z_T = (A_T Z_{T-1} + B_T) = (A_T A_{T-1} Z_{T-2} + A_T B_{T-1} + B_T) = \dots = \sum_{t=1}^T \left(\prod_{s=t+1}^T A_s \right) B_t. \quad (12)$$

则有:

$$\nabla_C F(C, \omega) = \frac{-\exp(-\bar{y}_i (\bar{x}_i^T \omega)) \bar{y}_i \bar{x}_i^T}{1 + \exp(-\bar{y}_i (\bar{x}_i^T \omega))} \left(\sum_{t=1}^T \left(\prod_{s=t+1}^T A_s \right) B_t \right) \quad (13)$$

然后再利用梯度下降法更新 C :

$$C_k = C_{k-1} - \eta \nabla_C F(C, \omega), k = 1, 2, \dots, K.$$

当满足一定的迭代条件时得到最优的超参数, 然后通过训练得到 SVM 模型。

4. 实证分析

4.1. 数据介绍

本文采用公开的“UCI Breast Cancer Dataset”数据集进行验证。此数据集共 569 个实例, 涉及 30 个特征指标(10 个平均值, 10 个标准差, 10 个最值): radius-细胞半径(单位: 像素); texture-纹理; perimeter-细胞核周长; area-细胞核面积; smoothness-平滑度; compactness-紧凑度; concavity-凹度; concave points-凹点; symmetry-对称性; fractal dimension-分形维数; 数据分为良性和恶性肿瘤两类(target: 患者是否患乳腺癌(1 = 是, -1 = 否)), 其中正例样本为恶性, 共计 212 个, 负例样本为良性, 共计 357 个, 部分原始数据如表 1。

Table 1. Part of original data

表 1. 部分原始数据

| 是否患癌 | 半径 | 纹理 | 周长 | 区域 | 平滑度 | 紧凑度 | 分形尺寸 | 凹面点 | 对称 |
|------|--------|--------|-------|-------|----------|---------|----------|---------|---------|
| 1 | 1.095 | 0.9053 | 8.589 | 153.4 | 0.006399 | 0.04904 | 0.006193 | 0.01587 | 0.03003 |
| 1 | 0.5435 | 0.7339 | 3.398 | 74.08 | 0.005225 | 0.01308 | 0.003532 | 0.0134 | 0.01389 |
| 1 | 0.7456 | 0.7869 | 4.585 | 94.03 | 0.00615 | 0.04006 | 0.004571 | 0.02058 | 0.0225 |
| -1 | 0.4956 | 1.156 | 3.445 | 27.23 | 0.00911 | 0.07458 | 0.009208 | 0.01867 | 0.05963 |
| 1 | 0.7572 | 0.7813 | 5.438 | 94.44 | 0.01149 | 0.02461 | 0.005115 | 0.01885 | 0.01756 |
| -1 | 0.3345 | 0.8902 | 2.217 | 27.19 | 0.00751 | 0.03345 | 0.005082 | 0.01137 | 0.02165 |
| 1 | 0.4467 | 0.7732 | 3.18 | 53.91 | 0.004314 | 0.01382 | 0.002179 | 0.01039 | 0.01369 |

主成分分析法常用于减少数据集的维数, 本文利用 Pycharm 软件对 30 个因素进行主成分分析, 从主成分分析结果可以得出, 仅前六项的方差累计就达到了 0.897, 即保留前 6 个主成分就可以概括原始数据的绝大部分信息, 然后根据主成分得分系数矩阵, 计算出 6 个主成分的值整理之后数据见表 2。

4.2. 实验过程及结果

首先, 分别比较双层规划、网格搜索、贝叶斯算法、退火算法得到的支持向量机模型的在原始数据集上的训练速度和经过 PCA 处理后的数据集上的训练速度, 实验结果如表 3 所示。实验结果表明, 无论是哪种算法, 在经过 PCA 处理的数据集上, 训练速度都得到了提升, 这证明了 PCA 在提高数据处理效率方面的显著优势; 并且, PCA 处理的前后双层规划得到的 SVM 都保持了最快的训练速度。

Table 2. Part of PCA data**表 2.** 部分主成分分析数据

| 主成分 1 | 主成分 2 | 主成分 3 | 主成分 4 | 主成分 5 | 主成分 6 |
|-------------|--------------|--------------|--------------|--------------|--------------|
| 9.18475521 | 1.94687003 | -1.122178766 | -3.630536408 | 1.194059478 | -1.410183639 |
| 2.385702629 | -3.764859063 | -0.528827374 | -1.117280773 | 1.117280773 | -0.028631162 |
| 5.728855491 | -1.074228589 | 0.55126254 | -0.91128084 | 0.176930218 | -0.540976145 |
| 7.11669126 | 10.26655564 | -3.229947535 | -0.152412923 | 2.958275431 | -3.050737497 |
| 3.931842467 | -1.946358977 | 1.388544953 | -2.938054169 | -0.546266745 | 1.225416405 |

Table 3. Time comparison for models**表 3.** 不同算法下模型训练耗时对比

| 算法 | PCA 前训练耗时/s | PCA 后训练耗时/s |
|--------|-------------|-------------|
| 双层规划算法 | 1.321 | 0.768 |
| 网格搜索 | 52.140 | 4.551 |
| 贝叶斯算法 | 46.136 | 7.452 |
| 退火算法 | 3.734 | 2.658 |

其次, 在相同的数据集上我们对 HSGD 算法得到的模型和其他算法得到的模型进行性能对比, 使用精确度(Pre)、准确率(Acc)、召回率(Rec)作为实验结果的评判标准, 表达式为:

$$\begin{aligned} \text{Pre} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ \text{Acc} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\ \text{Rec} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \end{aligned} \quad (14)$$

如表 4 所示, 双层规划得到的 SVM 模型在乳腺癌测试集上的精确度优于网格搜索得到的 SVM、贝叶斯算法得到的 SVM, 分别高了 5.286%、2.300%; 准确率优于网格搜索得到的 SVM、贝叶斯算法得到的 SVM、退火算法得到的 SVM, 分别高了 5.218%、0.800%、4.014%; 召回率为 100%, 优于贝叶斯算法得到的 SVM 和退火算法得到的 SVM, 分别高了 5.263%、12.2%。

Table 4. Comparison of experimental results**表 4.** 实验结果对比

| 模型 | 精确度% | 准确率% | 召回率% |
|-----------|--------|--------|--------|
| 双层规划-SVM | 95.000 | 98.200 | 100 |
| 网格搜索-SVM | 89.714 | 92.982 | 94.737 |
| 贝叶斯算法-SVM | 92.700 | 97.400 | 100 |
| 退火算法 SVM | 96.296 | 94.186 | 87.800 |

然后, 给出由四种方法得到的 SVM 在乳腺癌数据集上的学习曲线, 学习曲线是评估机器学习模型性

能的重要工具, 它们显示了模型在不同训练样本大小下的表现, 红色曲线代表训练分数(Training score)表示模型在训练集上的性能, 绿色曲线代表交叉验证分数(Cross-validation score)表示模型在交叉验证集上的性能, 体现模型在未知数据集上的表现。图 1 表明, 双层规划得到的 SV (Bi-SVM)表现尤为突出, 随着训练样本数量的增加, 其训练分数和交叉验证分数之间的差距减小, 几乎重叠, 说明双层规划得到的 SVM 具有显著的泛化能力, 网格搜索得到的 SVM (Grid search-SVM)、退火算法得到的 SVM (ANN-SVM)和贝叶斯算法得到的 SVM (Bayes-SVM)随着训练样本数量增加, 训练分数和交叉验证分数之间的差距虽然也在逐渐减小, 但相较于双层规划得到的 SVM, 它们在训练分数和交叉验证分数之间的差距减小的速度略慢, 说明其泛化能力略低。总体来看, 双层规划得到的 SVM 在模型的泛化能力上展现出了明显的优势。

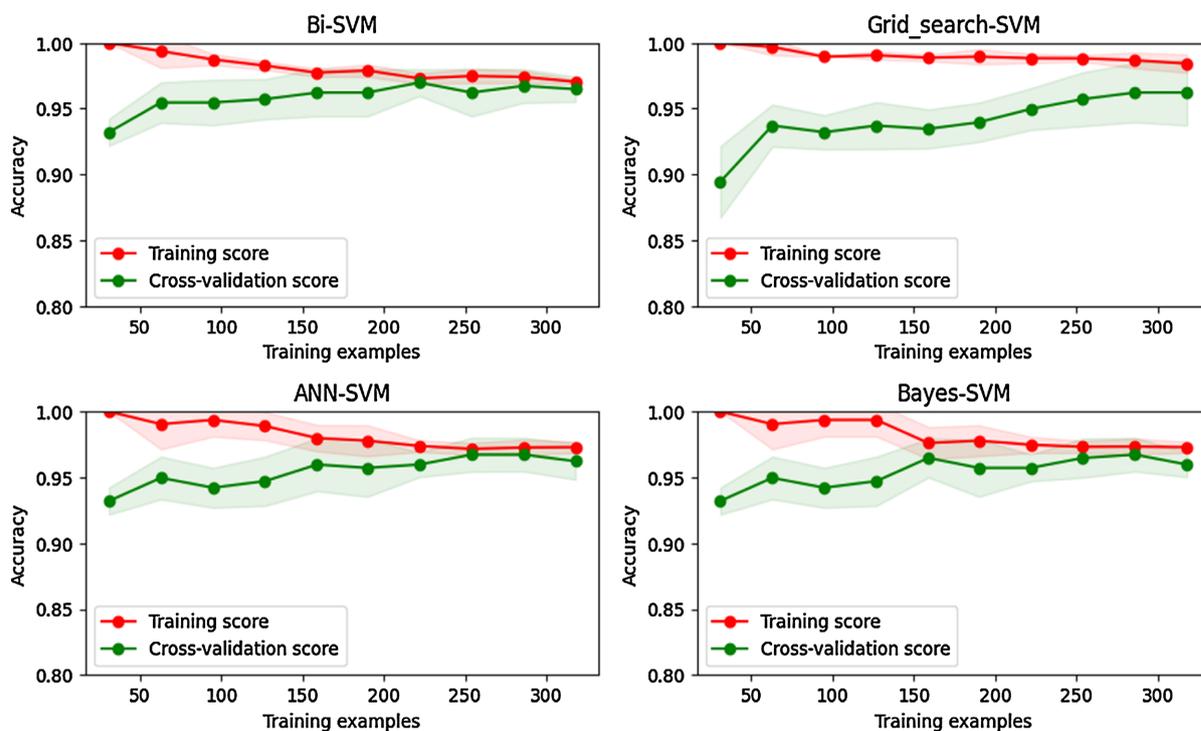


Figure 1. Learning curve

图 1. 学习曲线

此外在测试集上我们对双层规划得到的模型和其他三个模型进行分类结果进行可视化, 得到了如图 2 所示的混淆矩阵热图。混淆矩阵热图是用来描述分类模型的性能的图像, 它展示了每个类别的真实值与模型预测值之间的关系, 每一行代表一个真实类别, 每一列代表一个预测类别, 对角线上的值(从左上角到右下角)表示正确预测的样本数, 即真正例(TP)和真负例(TN)。由图 2 可知, 通过双层规划得到的 SVM 模型在精确度上略低于其他三个模型, 但其准确率是最高的, 并且在召回率方面, 双层规划得到的 SVM 模型和由贝叶斯优化得到的 SVM 模型表现均为最佳。

5. 结论

本文提出了 SVM 超参数选择的双层规划模型, 使用基于梯度的算法结合 Forward mode 方法求解该双层模型, 并对高维数据集使用主成分分析降维处理, 经过实验验证, 本文由双层规划得到的 SVM 模型的在乳腺癌测试集上的精确度优于网格搜索得到的 SVM、贝叶斯算法得到的 SVM, 分别高了 5.28%、

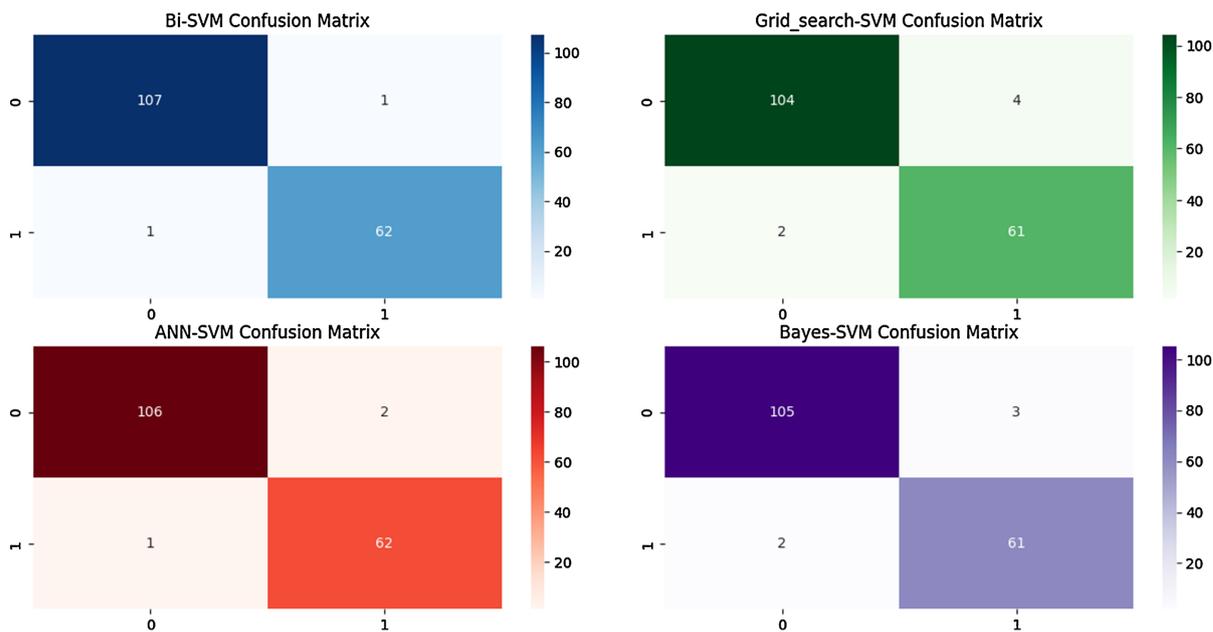


Figure 2. Confusion matrix heatmap

图 2. 混淆矩阵热图

2.300%；准确率优于网格搜索得到的 SVM、贝叶斯算法得到的 SVM、退火算法得到的 SVM，分别高了 5.218%、0.800%、4.014%；召回率为 100%，优于贝叶斯和退火算法得到的 SVM，分别高了 5.263%、12.2%。并且四种模型经过主成分分析后训练耗时都有所缩短，双层规划算法前后都保持了最快的训练速度。说明本文提出的方法具有实际意义与效果，在对 SVM 的超参数的选择中，能够发挥更好的性能和更高的效率。

基金项目

国家自然科学基金(NSFC), 项目编号为 12071342。河北省自然科学基金, 项目编号为 A2020202030。福州大学离散数学及其应用教育部重点实验室开放课题项目, 项目编号为 J20230701。

参考文献

- [1] Okuno, T., Takeda, A., Kawana, A., et al. (2021) On lp-Hyperparameter Learning via Bilevel Nonsmooth Optimization. *Journal of Machine Learning Research*, **22**, 1-47.
- [2] 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>
- [3] Qi, H. (2013) A Semismooth Newton Method for the Nearest Euclidean Distance Matrix Problem. *SIAM Journal on Matrix Analysis and Applications*, **34**, 67-93. <https://doi.org/10.1137/110849523>
- [4] 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. <https://doi.org/10.1109/icdmw.2009.74>
- [5] 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*, 16-21 July 2006. <https://doi.org/10.1109/ijcnn.2006.246935>
- [6] Vapnik, V.N. and Chervonenkis, A.Y. (1971) On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. *Theory of Probability & Its Applications*, **16**, 264-280. <https://doi.org/10.1137/1116025>
- [7] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1007/bf00994018>

- [8] 贺加贝. 基于改进 SFLA 算法对 SVM 算法超参数的优化[J]. 科技与创新, 2024(6): 39-41.
- [9] Franceschi, L., Donini, M., Frasconi, P. and Pontil M. (2017) Forward and Reverse Gradient-Based Hyperparameter Optimization. *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, 6-11 August 2017, 1165-1173.