哈爾濱工業大學

模式识别与机器学习实验报告

实验__二__

题　　目	逻辑回归
学　　院	未来技术学院
专　　业	人工智能
学　　号	2023112419
学　　生	陈铠
任课教师	刘扬

哈尔滨工业大学计算机科学与技术学院

2025 年秋季

一、 实验内容

（一）逻辑回归简要介绍

逻辑回归（Logistic Regression）是以二分类为主的判别式、监督学习的分类模型。

1 逻辑回归的核心任务

逻辑回归的核心任务是针对输入的特征向量$\boldsymbol{x} \in \mathbb{R}^d$学习映射关系：

$$f(\boldsymbol{x};\theta): \mathbb{R}^d \to [0,1]$$

通过学习数据特征，建立拟合函数$f(\boldsymbol{x};\theta)$来**估计后验概率**$P(y=1|\boldsymbol{x})$，再基于概率阈值判别最终类别y。

在本实验中，主要研究逻辑回归的**线性二分类**问题。而事实上，也可以将逻辑回归算法拓展更广泛的问题上，例如通过对特征进行非线性映射，可以学习到非线性决策边界；通过设计 One-vs-Rest 或 Softmax 回归，可以实现多分类问题。

2 逻辑回归的数学模型

逻辑回归选用的拟合函数为

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}},$$

其中$g(x)$为 sigmoid 函数。$\theta^T x = \sum_{i=0}^{n} \theta_i x_i\,, x_0 = 1$。

对于待分类样本，有$\hat{y} = \mathbb{I}[\theta^T \boldsymbol{x} \geq 0]$

3 逻辑回归的优化方法

3.1 符号定义

表 1 本实验使用符号

名称	符号	描述
特征矩阵	$X \in \mathbb{R}^{m\times n}$	每行$x^{(i)} \in \mathbb{R}^n$为第i个样本，含n个特征（常置第一个特征为 1）
输出向量	$y \in \mathbb{R}^m$	每行$y^{(i)}$为第i个样本的预测值（0 或 1）
参数向量	$\theta \in \mathbb{R}^n$	模型待优化变量
预测向量	$h \in \mathbb{R}^m$	每行$h_i = h_\theta(x^{(i)})$

记训练集为$\{(x^{(i)}, y^{(i)})\}_i^m$且样本独立同分布，使用极大似然估计 MLE 以及交叉熵损失，可以得到

$$L(\theta) = P(\mathrm{y} \mid X;\theta) = \prod_{i=1}^{m} p(y^{(i)} \mid x^{(i)};\theta) = \prod_{i=1}^{m} (h_\theta(x^{(i)}))^{y^{(i)}} (1-h_\theta(x^{(i)}))^{1-y^{(i)}}$$

$$\ell(\theta) = \log\, L(\theta) = \sum_{i=1}^{m} y^{(i)} \log\, h(x^{(i)}) + (1-y^{(i)}) \log\,(1-h(x^{(i)}))$$

于是损失函数为

$$J(\theta) = -\frac{1}{m}[y^T \log\, h + (1-y)^T \log\,(1-h)] + \frac{\lambda}{2m}\theta^T\theta$$

其中$\frac{\lambda}{2m}\theta^T\theta$为正则化项

模型的训练目标是

$$\theta^* = \underset{\theta}{argmin} J(\theta)$$

3.2 梯度下降法求解

梯度下降核心是求解损失函数梯度向量$\nabla J(\theta) \in \mathbb{R}^n$，再沿负梯度方向更新参数单个参数的偏导数为

$$\frac{\partial}{\partial\theta_j} J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)} + \frac{\lambda}{m}\theta_j$$

于是

$$\nabla J(\theta) == -\frac{1}{m}X^T(y-h) + \frac{\lambda}{m}\theta$$

更新公式为

$$\begin{aligned}\theta^{(t+1)} &= \theta^{(t)} - \alpha \cdot \nabla J\left(\theta^{(t)}\right)\theta^{(t+1)} \\ &= \theta^{(t)} - \alpha\left[-\frac{1}{m}X^T\left(y - h^{(t)}\right) + \frac{\lambda}{m}\theta^{(t)}\right] \\ &= \theta^{(t)} + \frac{\alpha}{m}X^T(y - h^{(t)}) - \frac{\alpha\lambda}{m}\theta^{(t)}\end{aligned}$$

其中$\frac{\alpha\lambda}{m}\theta^{(t)}$为正则项。

3.3 牛顿法求解

牛顿法使用二阶导数矩阵 Hessian 矩阵$H \in \mathbb{R}^{n\times n}$加速收敛

$$H = \frac{1}{m}X^T DX + \frac{\lambda}{m}I$$

其中$D \in \mathbb{R}^{m\times m}$是对角矩阵，对角元素$D_{ii} = h_i(1-h_i)$，$I \in \mathbb{R}^{n\times n}$是单位矩阵
更新公式为

$$\theta^{(t+1)} = \theta^{(t)} - H^{-1}\nabla J\left(\theta^{(t)}\right)$$

$$= \theta^{(t)} - \left(\frac{1}{m}X^T D^{(t)} X + \frac{\lambda}{m}I\right)^{-1} - [\frac{1}{m}X^T(y - h^{(t)}) + \frac{\lambda}{m}\theta^{(t)}]$$

其中$\frac{\lambda}{m}I$，$\frac{\lambda}{m}\theta^{(t)}$为正则项。

（二）实验研究内容

1 研究样本数量、样本分布条件的影响

调整样本数量为 100，200，类朴素贝叶斯分布与不满足类朴素贝叶斯分布。

2 研究两种求解方法和正则化的影响，

分别为：**梯度下降+无正则化**、**梯度下降+正则化**、**牛顿方法+无正则化**、**牛顿方法+正则化**。

3 研究模型在真实数据上的分类表现

使用 UCI 数据集（https://archive.ics.uci.edu/dataset/267/banknote+authentication）。共 1372 个样本，特征维度为 4，类别为 2

二、 实验环境

操作系统：Windows 11
实验平台：pycharm
解释器版本：Python 3.12
工具包：

表 2 本实验使用工具包

工具包名称	版本
pip	25.2
matplotlib	3.10.6
numpy	2.3.3
scikit-learn	1.7.2

三、 实验结果及分析

（一）样本数量、样本分布条件的影响

在该项中选用**梯度下降+无正则化**的求解方法

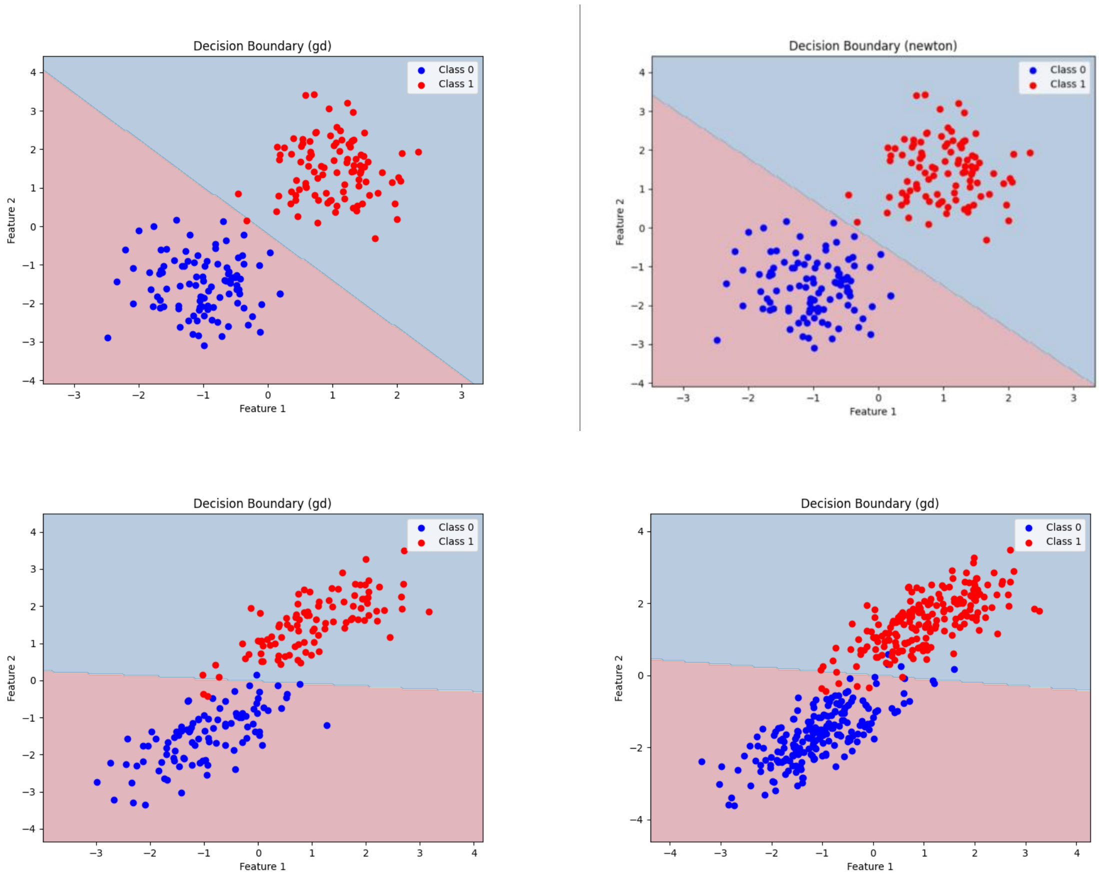

图 1-4 样本数量、样本分布条件的影响训练结果

表 3 样本数量、样本分布条件的影响

样本量/类朴素贝叶斯分布	训练轮次	训练集 acc	验证集 acc	训练集 loss
100/是	200	0.994	1.000	0.0181
200/是	200	0.991	1.000	0.0187
100/否	200	0.981	1.000	0.0664
200/否	200	0.978	0.975	0.0751

样本分布的影响：当样本符合类朴素贝叶斯分布时，训练集 acc≥0.99、loss≤0.0187、验证集 acc=1.0，说明逻辑回归在数据满足线性可分或近似线性分布（如类朴素贝叶斯分布）时，模型拟合效率高、泛化稳定性强，且参数少。反之，逻辑回归性能略有下降，证明逻辑回归对数据分布敏感，依赖“特征与标签线性相关”假设

样本量的影响：类朴素贝叶斯分布下，逻辑回归在较小样本量的情况下均能保持极高准确率和泛化率，但当样本量较大时其训练时间、性能消耗增加，效果可能下降。

（二）两种求解方法和正则化的影响

在该项中选择样本数量为 200，分布为类朴素贝叶斯分布。

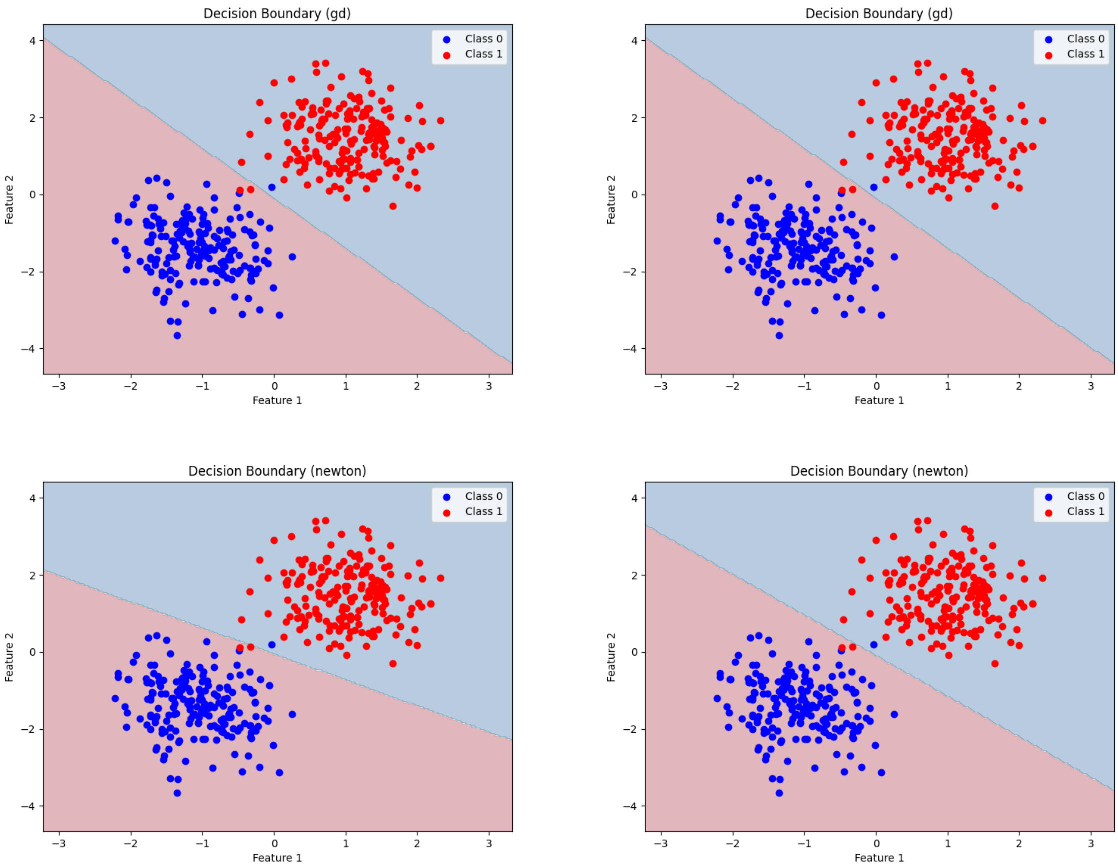

图 5-8 求解方法和正则化的影响训练结果

表 4 求解方法和正则化的影响

求解方法/正则化	训练轮次	训练集 acc	验证集 acc	训练集 loss
梯度下降/否	200	0.991	1.000	0.0187
梯度下降/是	200	0.991	1.000	0.0235
牛顿法/否	14	0.991	1.000	0.0132
牛顿法/是	10	0.991	1.000	0.0215

求解方法的影响：牛顿法显著降低了逻辑回归的训练轮次，说明其在解决该类问题上的高效性，但考虑到其求解方法需要计算 Hessian 矩阵，当样本特征映射至高维后，求解复杂度将极大增加，这是其局限性。

正则化的影响：逻辑回归加入正则化后 acc 不变、loss 略升，说明其过拟合控制能力有限，正则化仅通过惩罚参数抑制复杂度，无法从模型结构上避免过拟合。

（三）模型在真实数据上的分类表现

表 5 真实数据集上的表现

求解方法/正则化	训练轮次	训练集 acc	验证集 acc	训练集 loss
梯度下降/否	200	0.985	0.971	0.0418
梯度下降/是	200	0.985	0.971	0.0461
牛顿法/否	17	0.991	0.985	0.0183
牛顿法/是	10	0.993	0.985	0.0355

可以发现，模型在真实数据集上仍有较好的表现，且与自建数据集上的表现相近，验证了模型的正确性。

四、 结论

样本符合类朴素贝叶斯分布时，模型训练精度高（acc≥0.991）、损失低（loss≤0.0187）、泛化能力强（验证集 acc≥1.0）；而非该分布下，模型性能显著下降，且样本量增大会进一步加剧拟合难度与泛化能力下滑。

牛顿法求解效率与性能均优于梯度下降，无论在类朴素贝叶斯分布数据还是真实数据中，牛顿法的收敛轮次仅为梯度下降的 1/10-1/20，且能实现更高的训练精度与更低的训练损失，是逻辑回归模型更高效的求解方法。

正则化在实验条件下作用有限：在样本符合类朴素贝叶斯分布（无过拟合）或真实数据（轻微过拟合风险）中，正则化仅轻微提高训练损失，对验证集泛化精度无影响；仅在牛顿法+真实数据的组合中，正则化小幅提升训练精度，但整体未改变模型核心性能。

五、 参考文献

[1]（美）SHELDON AXLER 著；杜现坤，刘大艳，马晶译. 线性代数应该这样学 第 3 版[M]. 北京：人民邮电出版社, 2016.10.
[2]周志华著. 机器学习[M]. 北京：清华大学出版社, 2016.01.
[3]谢文睿，秦州编著. 机器学习公式详解[M]. 北京：人民邮电出版社, 2021.03.
[4]李航著. 统计学习方法 第 2 版[M]. 北京：清华大学出版社, 2019.05.