带有错误标签的张量数据的稳健多分类模型

张家瑞*,樊亚莉#

上海理工大学理学院, 上海

收稿日期: 2024年3月21日; 录用日期: 2024年6月10日; 发布日期: 2024年6月18日

摘要

传统机器学习方法大多都是基于正确标签的训练数据进行监督学习,但实际观测到的训练数据标签极可能受到污染,而错误标签的存在会导致传统模型产生有偏估计。现存的关于错误标签的稳健模型往往基于向量数据进行分类,面对存在错误标签的高阶张量数据时只能将其转化为低阶格式,由此产生过拟合问题且破坏张量结构。针对上述问题提出一种稳健的张量多分类模型(RMLTMLR),基于最小γ-散度估计、张量管道秩及相应的核范数来处理带有错误标签的低秩张量,在利用张量结构特点的同时使模型对污染标签具有稳健性,提高多分类准确率。进行的大量实验表明RMLTMLR模型在不同类别和污染程度的张量数据上有着优良的分类效果,与非稳健的模型相比,分类准确率显著提升。

关键词

图像多分类,错误标签,低秩张量,张量管道秩,机器学习

Robust Multiclass Models for Mislabeled Tensor Data

Jiarui Zhang*, Yali Fan#

College of Science, University of Shanghai for Science and Technology, Shanghai

Received: Mar. 21st, 2024; accepted: Jun. 10th, 2024; published: Jun. 18th, 2024

Abstract

Most of the traditional machine learning methods perform supervised learning based on training data with correct labels. However, the actual observed training data labels are likely to be contaminated, and the existence of wrong labels will lead to biased estimates of the traditional model. The existing robust models for mislabel classification are often based on vector data. When facing

文章引用: 张家瑞, 樊亚莉. 带有错误标签的张量数据的稳健多分类模型[J]. 运筹与模糊学, 2024, 14(3): 242-255. DOI: 10.12677/orf.2024.143262

^{*}第一作者。

[#]通讯作者。

high-order tensor data with mislabels, they have to transform it into low-order format, resulting in overfitting problem and damage to the tensor structure. Aiming at the above problems, a robust tensor multi-classification model (RMLTMLR) is proposed, which is based on minimum γ -divergence estimation, tensor tubal rank and the corresponding nuclear norm to deal with low-rank tensors with wrong labels. The model is robust to contaminated labels while taking advantage of the structural characteristics of tensors, and improves the accuracy of multi-classification. A large number of experiments show that the RMLTMLR model has excellent classification effects on tensor data with different categories and pollution levels, and the classification accuracy is significantly improved compared with the non-robust model.

Keywords

Image Multi-Classification, Mislabel, Low-Rank Tensor, Tensor Tubal Rank, Machine Learning

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

图像分类问题作为现代计算机视觉领域的基本问题之一,多年来被持续关注。在过去的几十年里,众多研究者已提出了许多算法和方法来解决图像分类的各种具体问题,其中包括经典的机器学习方法和最新的深度学习技术。深度学习方法,如 ResNet [1]、SENet [2]等,其分类精度高,但其训练所需数据量较大、参数学习耗时长且输入输出间不具备可解释性,并不适用于一些数据采集难度大、对分类器可解释性要求较高的场合。因此本文暂不考虑深度学习方法,而是将关注点仍旧放在传统机器学习方法的推广与改进上。

传统机器学习方法大多是监督学习,并且数据所对应的标签均是正确的。但在实际应用中并不总能得到正确的标签,由于人工标注错误、噪声扰动或采集数据不完整等各种原因,数据对应的标签往往不完全准确,这使得面向带有错误标签的稳健分类成为一个重要但具有挑战性的问题。标签噪声的存在具有许多潜在的负面影响,如模型分类准确度降低、推断模型的复杂性提升、必要的训练样本数量会增加等。其中,标签噪声会造成有偏估计从而导致模型的分类准确度降低是最为显著的负面影响。为克服这些问题,一些学者进行了不同的尝试。Teng [3] [4] [5]分别在 2000 年、2001 年和 2005 年提出了 3 种方法,这些方法在分类时会采用一些受标签噪声影响更小的算法,以达到预期的分类精度。然而,这些方法处理标签噪声只是为了避免过拟合,并没有真正考虑到标签噪声。Brodley 和 Friedl [6]尝试使用过滤方法来提高训练数据的质量,他们在训练开始之前识别并处理噪声标签,重新标记或直接删除被错误标记的实例。过滤方法成本低且易于实现,但对于某些样本,这种方法可能会删除大量数据,导致模型的训练效果受到影响。于是一些学者针对特定的算法进行改进,使其对标签噪声具有一定稳健性。

在众多可以用来解决图像分类问题的机器学习算法中,考虑到决策树模型面向高维数据时对微小变化较为敏感,支持向量机模型在面对样本重叠和噪声较多的数据表现不佳,因此逻辑回归模型(MLR)因其可解释性以及精确的分类结果成为解决错误标签问题的热门方法。

以二分类逻辑回归模型为例,设 $Y_0 = \{0,1\}$ 为真实标签,X为解释变量的 p 维随机向量,则 $P(Y_0 = 1|X = x)$ 满足条件标签概率模型

$$\pi(x;\omega,b) = \frac{\exp\{\omega^{\mathsf{T}}x + b\}}{1 + \exp\{\omega^{\mathsf{T}}x + b\}}.$$
 (1)

当样本标签被污染时,设Y为存在错误分类的标签,便存在错标概率

$$\eta_0(x) = P(Y = 1|Y_0 = 0, X = x), \eta_1(x) = P(Y = 0|Y_0 = 1, X = x).$$
 (2)

其中 η_0 表示负类数据中错误分类的概率, η_1 表示正类数据中错误分类的概率。而Y的概率不再遵循式(1),而是遵循

$$P(Y=1|X=x) = \eta_0(x)\{1-\pi(x;\omega,b)\} + \{1-\eta_1(x)\}\pi(x;\omega,b).$$
(3)

为了解决错误标记的问题,基于式(3)对 $\eta_j(x)$ 进行不同建模开发了一些稳健逻辑回归方法。Copas [7] 考虑相等和恒定的错标概率,即 $\eta_0(x) = \eta_1(x) = \eta$,提出恒定错标逻辑回归模型。Wainer 等人[8]和 Komori 等人[9]则分别在 2007 年和 2016 年提出了不对称的错标逻辑回归模型,即假设 $\eta_0(x) = \eta$, $\eta_1(x) = 0$ 。 Hayashi [10]扩展了 Takenouchi 和 Eguchi [11]提出的 η -boost 方法,提出了一种用于二元分类的稳健增强模型。上述所有方法的稳健性都来自底层的权重函数,并且假定错标概率具有一定的参数形式。由于最大似然对错标概率很敏感,因此对错标概率建模并不简单。但在实际应用中,我们关心的是真实的分类概率而不是错标概率这一干扰参数,花费过多精力为其建模似乎有些得不偿失。因此 Hung 等人[12]使用最小 γ -散度估计来取代最小 Kullback-Leibler (KL)散度估计,提出一种不需要对错误标签概率建模的 γ -逻辑回归方法,拟合错误标签的响应,通过降低可疑实例的权重来使模型对带错误标签的数据具有稳健性。

上述方法均基于 X 是向量类型数据实现的。然而,对图像分类问题来说,大量的图像更多以彩色图像、视频等形式存在,这些图像和视频自然呈现高阶张量形式。若将高阶张量数据处理为向量,则会产生高维向量,因此增加计算复杂度,也会损失高维数据的结构信息[13] [14]。而文献[15] [16]表明,进一步探索张量结构信息,不仅可以使张量模型更具稳健性,还可以用更少的参数更加结构化地表示高维数据,因此,将传统地机器学习分类算法扩展到张量空间便成为了近年来学者们关注的热点。Tao 等人[17] 提出了一种监督张量学习(STL)方案,将支持向量机扩展为支持张量机(STMs)。基于 Tao 等人的研究,Kotsia 等人[18] [19]将 STM 中的秩一权重张量更改为 CP 格式和 Tucker 格式,进一步推广支持张量机。通过将权重张量建模为 CP 格式,Tan 等人[20]将逻辑回归模型推广到张量。而 Zhang 等人[21]则利用张量管道秩及其核范数构建了一种面向低秩张量的图像多分类逻辑回归模型。

本文受 Hung 等人[12]研究成果的启发,基于 Zhang 等人[21]提出的低秩张量多分类模型,提出一种新的面向错误标签张量数据的稳健图像多分类模型,称为稳健错标多分类模型。该模型的优点和贡献在于:

- 1)提出了一种新的面向错误标签的稳健张量多分类方法,既能够避免张量结构信息的损失,同时又能在面对标签污染数据时达到较好的分类结果,具有稳健性;
- 2) 本文方法不需要对误标概率进行建模,得到的加权估计方程不需要包含任何偏差校正项,可以自动进行偏差校正;
- 3) 在仿真数据和真实数据上进行大量实验,通过常用的多分类指标评价模型。结果表明在不同多分类情况下,本文模型面向标签被污染数据时的分类精度与非稳健的模型相比取得明显提升。

接下来,本文将在第一节介绍相关工作,第二节提出本文方法,并在第三节进行实验说明,最后在第四节进行总结。

2. 预备知识

为了更好地说明本文方法,本章将对文中使用的符号进行简单说明,回顾重点定义,并介绍相关方法。

2.1. 符号说明

在本文中,我们将张量认为是向量和矩阵推广到高阶的多维数组,用花体大写字母 $A \in \mathbb{R}^{I_i \times I_2 \times \dots \times I_d}$ 来表示 d 阶张量,用 $A(i_1,i_2,\dots,i_d)$ 表示张量的每个元素,其中 $1 \le i_k \le I_k$, $k = 1,\dots,d$;对于三阶张量 $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$,分别将其第 i 个前向切片、侧向切片和正向切片表示为 $\mathcal{B}(i,::)$ 、 $\mathcal{B}(:,i,:)$ 和 $\mathcal{B}(:,::)$,用 $\mathcal{B}(i,j,:)$ 表示管道。用大写字母 A,B,\dots 来表示矩阵;用加粗的小写字母 a,b,\dots 来表示向量;用小写字母 a,b,\dots 来表示标量。我们用 $\|A\|_a$, $\|A\|_a$, $\|A\|_a$ 和 $\|A\|_a$ 来表示 I_0 范数、 I_a 范数、核范数和 I_a 范数。

对于有 C 种可能标签的训练数据集 $\left\{\mathcal{X}_i \in \mathbb{R}^{I_1 \times I_2 \times ... \times I_d} \middle| i = 1, \cdots, N \right\}$,通常使用独热编码来表示标签 $\left\{ \mathbf{y}_i \in \mathbb{R}^C \middle| i = 1, \cdots, N \right\}$ 。具体来说,我们将 \mathcal{X}_i 的标签编码为 $\mathbf{y}_i = \begin{bmatrix} 0, \cdots, 1, \cdots, 0 \end{bmatrix}^T \in \mathbb{R}^C$,若 \mathcal{X}_i 属于第 j 类,那 么 \mathbf{y}_i 除了第 j 个元素为 1 外其余元素皆为 0。我们定义 $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N \end{bmatrix}^T \in \mathbb{R}^{N \times C}$ 被污染的标签矩阵, \mathbf{Y}_0 为真实的标签矩阵, $\hat{\mathbf{Y}}$ 为根据模型预测的标签矩阵。

2.2. 相关定义

定义1 张量模 k 乘积。

对于一个张量 $A \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$ 与一个矩阵 $M \in \mathbb{R}^{m \times I_k}$,二者的模k乘积记作 $\mathcal{C} = A \times_{\iota} M$,其定义如下:

$$C(i_{1}, \dots, i_{k-1}, j, i_{k+1}, \dots, i_{d}) = A \times_{k} M = \sum_{i_{k}=1}^{I_{k}} M(j, i_{k}) A(i_{1}, \dots, i_{k-1}, i_{k}, i_{k+1}, \dots, i_{d}).$$

$$(4)$$

定义2张量内积。

对于张量 $A \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$ 和 $B \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$,它们的内积 $\langle A, B \rangle$ 定义为

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_d=1}^{I_d} \mathcal{A}(i_1, i_2, \dots, i_d) \mathcal{B}(i_1, i_2, \dots, i_d).$$
 (5)

定义3 张量的 Frobenius 范数。

张量 $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$ 的 Frobenius 范数定义为

$$\|A\|_{F} = \sqrt{\langle A, A \rangle}.$$
 (6)

定义4 正交张量[22]。

对于三阶张量 $A \in \mathbb{R}^{I_1 \times I_2 \times I_3}$,若它满足 $A^* * A = A * A^* = \mathcal{I}$,则它为正交张量。

定义5 张量-张量积(T积)[22]。

对于三阶张量,设 $A \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, $\mathcal{B} \in \mathbb{R}^{I_2 \times m \times I_3}$,我们定义

$$\operatorname{unfold}(\mathcal{A}) = \begin{bmatrix} \mathcal{A}(:,:,1) \\ \mathcal{A}(:,:,2) \\ \vdots \\ \mathcal{A}(:,:,I_3) \end{bmatrix} = \begin{bmatrix} \mathcal{A}^{(1)} \\ \mathcal{A}^{(2)} \\ \vdots \\ \mathcal{A}^{(I_3)} \end{bmatrix}. \tag{7}$$

$$fold(unfold(A)) = A.$$
(8)

$$bcirc(A) = \begin{bmatrix} A^{(1)} & A^{(I_3)} & \cdots & A^{(2)} \\ A^{(2)} & A^{(1)} & \cdots & A^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(I_3)} & A^{(I_3-1)} & \cdots & A^{(1)} \end{bmatrix}.$$
(9)

则二者的 T 积 A*B 大小为 $I_1 \times m \times I_3$, 其定义为

$$A * B = \text{fold}(\text{bcirc}(A) \cdot \text{unfold}(B)). \tag{10}$$

定义 6 张量的奇异值分解(T-SVD) [23]。

设 $A \in \mathbb{R}^{I_1 \times I_2 \times I_3}$,那么它可以分解为

$$\mathcal{A} = \mathcal{U} * \mathcal{G} * \mathcal{V}^*. \tag{11}$$

其中 $U \in \mathbb{R}^{I_1 \times I_1 \times I_3}$, $V \in \mathbb{R}^{I_2 \times I_2 \times I_3}$ 是正交的, $G \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ 为每个正向切片都是对角矩阵的对角张量(图 1)。

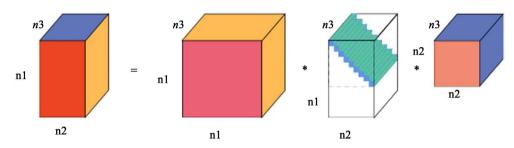


Figure 1. The t-SVD of third order tensors **图 1.** 三阶张量的张量奇异值分解

定义7 张量管道秩[24] [25]。

对于张量 $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$,其张量管道秩记为 $rank_t(\mathcal{A})$,定义为 \mathcal{G} 的非零的奇异管道的数量,其中 \mathcal{G} 来源于 \mathcal{A} 的 T-SVD,即 $\mathcal{A} = \mathcal{U} * \mathcal{G} * \mathcal{V}^*$,于是有

$$rank_{t}(\mathcal{A}) = \#\{i, \mathcal{G}(i, i, :) \neq 0\}. \tag{12}$$

根据矩阵奇异值的递减性质和逆快速傅里叶变换,文献[23]表明了 \mathcal{G} 的第一个正向切片 $\mathcal{G}(:,:,1)$ 决定张量管道秩,即

$$rank_{t}(A) = \#\{i, \mathcal{G}(i, i, 1) \neq 0\}. \tag{13}$$

定义8张量核范数[23]。

设 $A = \mathcal{U} * \mathcal{G} * \mathcal{V}^*$ 是 $A \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ 的T-SVD,则张量A的核范数定义为

$$\|\mathcal{A}\|_{*} := \langle \mathcal{G}, \mathcal{I} \rangle = \sum_{i=1}^{r} \mathcal{G}(i, i, 1). \tag{14}$$

其中 $r = rank_{\cdot}(A)$, \mathcal{I} 为单位张量,即第一个正向切片是单位矩阵其他正向切片都是零的张量。

2.3. 相关工作

2.3.1. 最小 γ-散度估计

设 g 为数据生成分布, f_{θ} 为参数 θ 的模型分布,设 θ_{0} 为感兴趣的真实参数值,则 g 和 f_{θ} 之间的 γ -散度定义为

$$D_{\gamma}(g, f_{\theta}) = \frac{1}{\gamma(\gamma + 1)} \left\{ \|g\|_{\gamma + 1} - \int \left(\frac{f_{\theta}}{\|f_{\theta}\|_{\gamma + 1}} \right)^{\gamma} g \right\}. \tag{15}$$

其中 $\|f_{\theta}\|_{\gamma+1} = \left(\int f_{\theta}^{\gamma+1}\right)^{\frac{1}{\gamma+1}}$ 。 当 $\gamma \to 0$ 时, $\lim_{\gamma \to 0} D_{\gamma}(g, f_{\theta}) = \int \ln\left(\frac{g}{f_{\theta}}\right)g$,即为 KL 散度。

用最小 γ -散度估计 θ 。如下式所示

$$\underset{\theta}{\arg\min} D_{\gamma}(g, f_{\theta}) = \underset{\theta}{\arg\max} \int \left(\frac{f_{\theta}}{\|f_{\theta}\|_{\gamma+1}}\right)^{\gamma} g. \tag{16}$$

当 g 属于参数类 $\{f_{\theta}:\theta\in\Theta\}$ 且参数值为 θ_0 时,上式在 $\theta=\theta_0$ 处取得最优解。这保证了最小 γ -散度估计的一致性[26]。而在有污染的情况下, $g=cf_{\theta_0}+(1-c)h$,其中 h 为污染分布, 1-c 为污染比例。通过计算得到

$$D_{\gamma}(g, f_{\theta}) = \left\{ cD_{\gamma}(f_{\theta}, f_{\theta}) + \frac{B_{\gamma}(c, h; \theta)}{\gamma(\gamma + 1)} \right\} + \frac{\left\| cf_{\theta_{0}} + (1 - c)h \right\|_{\gamma + 1} - c \left\| f_{\theta_{0}} \right\|_{\gamma + 1}}{\gamma(\gamma + 1)}.$$
(17)

其中 $B_{\gamma}(c,h;\theta) = (1-c)\int \left(\frac{f_{\theta}}{\|f_{\theta}\|_{\gamma+1}}\right)^{\gamma}h$ 。 忽略不涉及 θ 的项,最小化上式等价于最小化

$$cD_{\gamma}(f_{\theta}, f_{\theta}) + \frac{B_{\gamma}(c, h; \theta)}{\gamma(\gamma + 1)} \approx cD_{\gamma}(f_{\theta}, f_{\theta}). \tag{18}$$

若对某些 γ , 当 θ 在 θ_0 的领域时偏差 $B_{\gamma}(c,h;\theta)$ 小的可以忽略不计,则上式的约等于成立。约等号的右边在 $\theta=\theta_0$ 处最小。

Fujisawa 等人[27]和 Kanamori 等人[28]的研究表示,最小化过程受比例 c 和污染 h 的影响较小,因此我们可以以可忽略的偏差很好的估计 θ 。

2.3.2. 基于向量的稳健 γ-逻辑回归

当模型实际由(3)生成时,稳健的 γ-散度可以用来推断模型(1)。其原因在 Hung [12]的研究成果中由 如下引理揭示:

引理 1 式(3)中污染的 Y的分布可以表示为目标分布 $P(Y_0 = y | X = x)$ 和错标诱导分布 h(y | x) 的混合分布,

$$P(Y = y | X = x) = c(x)P(Y_0 = y | X = x) + \{1 - c(x)\}h(y|x).$$
(19)

其中 $h(y|x) = \left\{\frac{\eta_0(x)}{\eta_0(x) + \eta_1(x)}\right\}^y \left\{\frac{\eta_1(x)}{\eta_0(x) + \eta_1(x)}\right\}^{1-y}$,且 $1-c(x) = \eta_0(x) + \eta_1(x)$ 为给定X = x时的条件污染比例。

引理 1 和式(18)表明从污染数据中推断真实概率 $P(Y_0 = y|X)$ 的可能性,因为最小 y-散度估计可以忽略 h(y|x)的影响。由此可以得到

$$g(y|x) = c(x) f(y|x; \omega_0) + \{1 - c(x)\} h(y|x).$$
 (20)

其中c(x)和h(y|x)的定义与引理1一致,而

$$f(y|x;\omega) = \left\{\pi(x;\omega)\right\}^{y} \left\{1 - \pi(x;\omega)\right\}^{1-y}.$$
 (21)

从而得到基于 γ-逻辑回归的稳健估计模型,其目标函数为

$$\underset{\omega}{\arg\min} E_{X} \left[D_{Y} \left\{ g\left(\bullet \middle| X \right), f_{\theta}\left(\bullet \middle| X ; \omega \right) \right\} \right] = \underset{\omega}{\arg\max} E_{X,Y} \left\{ \left(\frac{f\left(Y \middle| X ; \omega \right)}{\left\| f\left(\bullet \middle| X ; \omega \right) \right\|_{Y+1}} \right)^{Y} \right\}. \tag{22}$$

其中 $\|f(\bullet|X;\omega)\|_{\gamma+1} = \left[\left\{\pi(X;\omega)\right\}^{\gamma+1} + \left\{1-\pi(X;\omega)\right\}^{\gamma+1}\right]^{\frac{1}{\gamma+1}}$, E_X 和 $E_{X,Y}$ 分别表示对X 和(X,Y) 的期望。 $\hat{\omega}_{\gamma}$ 表示对模型参数 ω_{γ} 的估计值:

$$\omega_{\gamma} = \arg\min \frac{1}{N} \sum_{i=0}^{N} \left(\frac{f(Y_i | X_i; \omega)}{\left\| f(\bullet | X_i; \omega) \right\|_{\gamma+1}} \right)^{\gamma} = \arg\min \frac{1}{N} \sum_{i=0}^{N} \left(\frac{\exp\{Y_i (\gamma + 1)\omega^{\mathsf{T}} X_i\}}{1 + \exp\{(\gamma + 1)\omega^{\mathsf{T}} X_i\}} \right)^{\frac{\gamma}{\gamma+1}}.$$
 (23)

求偏导且令 $S_{\nu}(\hat{o}_{\nu})=0$ 得

$$S_{\gamma}(\omega) = \frac{1}{N} \sum_{i=0}^{N} \psi_{\gamma,i}(\omega) \left\{ Y_i - \pi \left(X_i; (\gamma + 1)\omega \right) \right\} X_i. \tag{24}$$

其中, 权函数

$$\psi_{\gamma,i}(\omega) = \left(\frac{\exp\{Y_i(\gamma+1)\omega^{\mathrm{T}}X_i\}}{1+\exp\{(\gamma+1)\omega^{\mathrm{T}}X_i\}}\right)^{\frac{\gamma}{\gamma+1}}.$$
(25)

 $\psi_{\gamma,i}(\omega)$ 对不匹配 $\left(Y_i,\omega^{\mathsf{T}}X_i\right)$ 的实例进行降权,因此上式的稳健性显而易见,且稳健性由 γ 的值控制。 当 $\gamma=0$ 时, $S_{\gamma}(\omega)=\frac{1}{N}\sum_{i=0}^{N}\left\{Y_i-\pi\left(x_i;(\gamma+1)\omega\right)\right\}X_i$,未调整权重,即为非稳健的估计方程。越大的 γ 意味着模型越稳健,但模型的效率也将降低。因此需要在实验中选择合适的 γ 来确定模型的效率和稳健性。

但该模型只适用于向量格式的数据,若直接将张量数据转化为向量将会损失结构信息,降低分类准确性。因此我们考虑面向张量的多分类方法。

2.3.3. 非稳健错标张量多分类模型

Zhang 等人[21]提出的非稳健错标张量多分类模型(NRMLTMLR, 其中 NR 表示 non-robust),通过对低秩权重张量的秩加以约束的方式充分考虑了张量结构,最大程度避免损失张量结构信息。

给定 C 类张量训练数据集 $\{(\mathcal{X}_i, y_i), i=1, \dots, N\}$, NRMLTMLR 的目标函数为

$$\min_{\mathcal{W}^{r},b^{r}} - \frac{1}{N} \sum_{i=1}^{N} \sum_{r=1}^{C} Y_{ir} \log \left(\frac{e^{\left\langle \mathcal{W}^{r}, \mathcal{X}_{i} \right\rangle + b^{r}}}{\sum_{i=1}^{C} e^{\left\langle \mathcal{W}^{j}, \mathcal{X}_{i} \right\rangle + b^{j}}} \right) + \lambda_{1} \sum_{r=1}^{C} \left\| \mathcal{W}^{r} \right\|_{*}. \tag{26}$$

其中 $\mathcal{X}_i \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$, $\mathcal{W}' \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$, $r = 1, \cdots, C$, \mathcal{W}' 为属于第 r 类的张量所对应的权重张量, $\left\| \mathcal{W}' \right\|_*$ 作为对应的张量核范数代表着权重张量的结构信息, $Y_{ir} \in \mathbb{R}^{N \times C}$ 为标签矩阵。

但该方法并未考虑标签受到污染的情况,不具有稳健性。因此受 y-逻辑回归的稳健估计模型的启发,我们基于 NRMLTMLR 模型提出了一种基于低秩张量的稳健多分类模型,可以在最大限度利用张量结构信息的同时,对带有错误标签的数据具有稳健性。

3. 稳健错标张量多分类模型(RMLTMLR)

受到式(23)和式(25)的启发,我们提出了一种面对带有错误标签数据具有稳健性的张量多分类模型。该模型在多分类逻辑回归模型的基础上进行改进,利用了逻辑回归的优势,使模型在错误标签和张量数据情况下仍然具有优良的性质。

给定 C 类张量训练数据集 $\{(\mathcal{X}_i, y_i), i=1,\cdots,N\}$,得到 RMLTMLR 的目标函数为

$$\min_{\mathcal{W}^{r},b^{r}} -\frac{1}{N} \sum_{i=1}^{N} \sum_{r=1}^{C} Y_{ir} \log \left(\frac{e^{(\gamma+1)(\langle \mathcal{W}^{r}, \mathcal{X}_{i} \rangle + b^{r})}}{\sum_{j=1}^{C} e^{(\gamma+1)(\langle \mathcal{W}^{j}, \mathcal{X}_{i} \rangle + b^{j})}} \right) + \lambda \sum_{r=1}^{C} \| \mathcal{W}^{r} \|_{*}.$$

$$(27)$$

其中 $\mathcal{X}_i \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$, $\mathcal{W}^r \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d}$, $r = 1, \dots, C$, \mathcal{W}^r 为属于第r类的张量所对应的权重张量。 γ 是反应稳健性的权重,当 $\gamma = 0$ 时退化为非稳健的模型。 $Y_{ir} \in \mathbb{R}^{N \times C}$ 为标签矩阵。

式(26)的第一项与第二项中同时出现了变量 \mathcal{W}' ,为降低求解难度,我们增加约束条件进行解耦,对应目标函数如下

$$\min_{\mathcal{W}^{r}, \mathcal{P}, b^{r}} \quad -\frac{1}{N} \sum_{i=1}^{N} \sum_{r=1}^{C} Y_{ir} \log \left(\frac{e^{(\gamma+1)(\langle \mathcal{W}^{r}, \mathcal{X}_{i} \rangle + b^{r})}}{\sum_{j=1}^{C} e^{(\gamma+1)(\langle \mathcal{W}^{j}, \mathcal{X}_{i} \rangle + b^{j})}} \right) + \lambda \sum_{r=1}^{C} \left\| \mathcal{P}^{r} \right\|_{*},$$

$$s.t \qquad \qquad \mathcal{W}^{r} = \mathcal{P}^{r}.$$
(28)

根据文献[29]的方法, 我们采用 ADMM 算法解决上述优化问题, 得到增广拉格朗日函数为

$$\mathcal{L}(\mathcal{X}, \mathcal{W}, b, \mathcal{P}, \mathcal{Q}, \mu) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{r=1}^{C} Y_{ir} \log \left(\frac{e^{(\gamma+1)(\langle \mathcal{W}^r, \mathcal{X}_i \rangle + b^r)}}{\sum_{j=1}^{C} e^{(\gamma+1)(\langle \mathcal{W}^j, \mathcal{X}_i \rangle + b^j)}} \right) + \lambda \sum_{r=1}^{C} \left\| \mathcal{P}^r \right\|_* + \sum_{r=1}^{C} \left\langle \mathcal{Q}^r, \mathcal{P}^r - \mathcal{W}^r \right\rangle + \frac{\mu}{2} \sum_{r=1}^{C} \left\| \mathcal{P}^r - \mathcal{W}^r \right\|_F^2.$$

$$(29)$$

为了方便表示,我们将张量堆叠成高一维度的张量,记作 $\mathcal{W} = \mathcal{W}_{stock}^r \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d \times C}$, $\mathcal{P} = \mathcal{P}_{stock}^r \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d \times C}$, $\mathcal{Q} = \mathcal{Q}_{stock}^r \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_d \times C}$ 为拉格朗日乘子张量, μ 和 λ 为惩罚参数。为了便于计算,本文采用了与 Yin 等人[30]相同的方法,将惩罚参数取相同的初始值,并在迭代的每一步中更新。根据经验,我们选择了 10^{-3} 作为参数的初始值。

该优化问题可以拆分为 5 个子问题,分别更新 W 、b 、 $\left\{\mathcal{P}^r\right\}_{r=1}^C$ 、Q 和 μ ,下面重点介绍其中 2 个子问题。

1) 更新 W。

W的子问题可以写作

$$\underset{\mathcal{W}}{\operatorname{arg\,min}} \quad J(\mathcal{X}, \mathcal{W}, b) + \frac{\mu}{2} \left\| \mathcal{P} - \mathcal{W} + \frac{\mathcal{Q}}{\mu} \right\|_{F}^{2}$$
(30)

其中

$$J(\mathcal{X}, \mathcal{W}, b) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{r=1}^{C} Y_{ir} \log \left(\frac{e^{(\gamma+1)(\langle \mathcal{W}^r, \mathcal{X}_i \rangle + b^r)}}{\sum_{j=1}^{C} e^{(\gamma+1)(\langle \mathcal{W}^j, \mathcal{X}_i \rangle + b^j)}} \right).$$
(31)

根据上式求W梯度为

$$\nabla_{\mathcal{W}} \mathcal{J}_{\mathcal{W}} = \frac{1}{N} \mathcal{X} \times_{(d+1)} \left[\left(\hat{Y} - Y \right) \cdot \psi_{\gamma} \left(\mathcal{W} \right) \right] - \mu \left(\mathcal{P} - \mathcal{W} + \frac{\mathcal{Q}}{\mu} \right). \tag{32}$$

其中Y为给出的训练标签, \hat{Y} 为预测标签, $\psi_{y}(W)$ 为权重函数,其表达式如下

$$\psi_{\gamma}(\mathcal{W}) = \left(\frac{e^{\hat{Y}(\gamma+1)[\langle \mathcal{W}^r, \mathcal{X} \rangle + b^r]}}{\sum_{j=1}^{C} e^{(\gamma+1)[\langle \mathcal{W}^j, \mathcal{X} \rangle + b^j]}}\right)^{\frac{\gamma}{\gamma+1}}.$$
(33)

 $(\hat{Y}-Y)$ 和 $\psi_{\gamma}(W)$ 之间作矩阵乘法,而 \mathcal{X} 和 $(\hat{Y}-Y)\cdot\psi_{\gamma}(W)$ 按照定义 1 作张量的模 d+1乘法。 代入梯度下降算法,在每次迭代期间通过 $\mathcal{W}^{t+1}=\mathcal{W}^t-\alpha*\nabla_{\mathcal{W}}\mathcal{J}_{\mathcal{W}}$ 最终可以最小化目标函数,其中 t为迭代次数, α 为步长。

2) 更新 $\left\{\mathcal{P}^r\right\}_{r=1}^C$

可以通过

$$\underset{\mathcal{P}}{\operatorname{arg\,min}} \quad \frac{\lambda}{\mu} \|\mathcal{P}\|_{*} + \frac{1}{2} \|\mathcal{P} - \mathcal{W} + \frac{Q}{\mu}\|_{F}^{2}. \tag{34}$$

来更新 $\left\{\mathcal{P}'\right\}_{r=1}^{C}$ 。该子问题可通过文献[28]中提出的张量奇异值收缩算法(TSVT)来解决,其算法流程如算法 1 所示。

算法 1. 张量奇异值收缩算法(TSVT)

输入: $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}, \tau > 0.$

输出: $\tilde{\mathcal{X}} = \mathcal{U} * \mathcal{G}_{\tau} * \mathcal{V}^*$.

- 1) 对 \mathcal{X} 的每个正向切片做快速傅里叶变换, 计算 $\overline{\mathcal{X}} = \mathrm{fft}(\mathcal{X},[],3)$;
- 2) 在 $\overline{\chi}$ 的每个正向切片上做矩阵 SVT:

for
$$i = 1, \dots, \left[(I_3 + 1)/2 \right]$$

 $\left[\mathcal{U}, \mathcal{G}, \mathcal{V} \right] = SVD(\overline{Y}^i);$
 $\overline{\mathcal{Z}}^{(i)} = \mathcal{U} \cdot (\mathcal{G} - \tau) \cdot \mathcal{V}^*.$

end

for
$$\mathbf{i} = \left[\left(I_3 + 1 \right) / 2 \right] + 1, \dots, I_3$$

$$\overline{\mathcal{Z}}^{(i)} = \operatorname{conj} \left(\overline{\mathcal{Z}}^{I_3 - i - 3} \right);$$

end.

- 3) 对 \overline{z} 的每个正向切片做逆快速傅里叶变换,计算 $\tilde{\mathcal{X}}=\mathrm{ifft}\left(\overline{\mathcal{Z}},[],3\right)$;
- 4) 输出 $\tilde{\mathcal{X}} = \mathcal{U} * \mathcal{G}_{\tau} * \mathcal{V}^*$.

RMLTMLR 算法的更新过程则以伪代码形式表示为算法 2。

算法 2. 稳健错标张量多分类算法(RMLTMLR)

输入: $\{(\mathcal{X}, y_i), i=1,\dots,N\}, \lambda, \alpha, \gamma.$

输出: W,b.

1) 初始化:

 $\mathcal{W} = \operatorname{zero}\left(\left[I_{1}, I_{2}, I_{3}, C\right]\right), \mathcal{P} = \operatorname{zero}\left(\left[I_{1}, I_{2}, I_{3}, C\right]\right), b = \operatorname{zero}\left(\left[C, 1\right]\right), \mu = 1e^{-3}, \rho = 1.01;$

- 2) 计算 ψ_{γ} , \hat{Y} : 按照式(32)计算 ψ_{γ} ; 按照式(32)计算 \hat{Y} ;
- 3) 更新 W,b:

续表

for
$$i = 1, \dots, 25$$

$$\mathcal{W} = \mathcal{W} - \alpha * \nabla_{\mathcal{W}} \mathcal{F}_{\mathcal{W}};$$

$$b = b - \alpha * \nabla_{b} \mathcal{F}_{b};$$
end.
4) 用算法 1 更新 \mathcal{P} ;
5) 更新 \mathcal{Q}, μ :
$$\mathcal{Q} = \mathcal{Q} + \mu * (\mathcal{P} - \mathcal{W}), \mu = \rho * \mu;$$
6) 输出 \mathcal{W}, b .

4. 实验

本章选择彩色图像作为真实数据进行一系列实验以测试模型的分类效果,由此证实本文提出模型对于包含错误标签的张量图像分类具有稳定性,并以不同评价指标来全面评估本文模型的效果。本章使用的数据集为加州理工大学的 101 类彩色图像识别数据集,该数据集共包含 102 类 9145 副彩色图像,取该数据集包含不同数据量的不同子集,分别进行二分类、五分类和七分类的三个实验。

为了方便实验,我们对带标签的张量数据集 $\{(\mathcal{X}_i, \mathcal{Y}_i), i=1,\cdots, N\}$ 做以下处理:对图像数据 \mathcal{X}_i ,由于数据集中不同图像的格式并不一致,我们统一将图像格式重塑为 $32 \times 32 \times 3$,即令 $\mathcal{X}_i \in \mathbb{R}^{32 \times 32 \times 3}$;对标签数据 \mathcal{Y}_i ,按照不同比重选择部分标签进行错误标注。为了更好地对比实验结果,我们对实验数据做如下处理。我们选用的彩色图像识别数据集所带的标签是干净标签,我们将使用干净标签的数据分类的实验称为 Naive 方法,作为实验的对照组来反映该数据集一般情况下分类的效果;接着我们对标签按照不同比例进行错误标注,二分类实验为例,随机选择部分标签,将其中正类标签标为负类,负类标签标为正类,即完成对该部分标签的错误标注。我们使用 NRMLTMLR 模型和 RMLTMLR 模型对进行过标签错误标注的张量图像数据集进行分类,通过比对分类效果来体现本文提出的 RMLTMLR 模型在错误标注的张量数据分类问题上的优越性。理论上讲 RMLTMLR 模型的分类效果应该好于 NRMLTMLR 的分类效果,且不好于 Naive 方法的分类效果。

我们的实验使用交叉验证方法,在参数选择上采用网格搜索法来选用使模型效果最佳的参数。根据 经验选择 λ 的取值范围为 [le – 5,le3], α 的取值范围为 [0.1,1.5], γ 则按照张齐航[31]的实验结果选择最 优值 γ = 1。

本章共使用 4 种常见的多分类评价指标,分别为准确率、MacroF1、MicroF1 和 Hamming 距离。其中,准确率指正类和负类中预测正确的数量占总量的比例,MacroF1 和 MicroF1 是 F1-Score 的多分类扩展,Hamming 距离衡量预测标签与真实标签之间的距离。准确率、MacroF1 和 MicroF1 越大,Hamming 距离越小,表示模型的效果越好。

本章所有实验均在搭载酷睿 11 代处理器 64 位操作系统的计算机上使用 Matlab2023b 软件上实现。

4.1. 二分类

本节选用手风琴和船锚 2 个类别共计 86 张彩色图像作为二分类数据集,分别按照错误标签(mislabel) 0.10、0.15 和 0.20 的比重进行实验。考虑到每次实验划分的训练集和测试集不同,Naive 方法的分类结果 会有差别,因此在不同 mislabel 比重下,三种方法均使用相同的 \mathcal{X}_i ,区别在于 Naive 方法使用不带噪声的标签数据 y_i ,其余两种方法使用带噪声的标签数据。最终得到二分类实验结果如下表 1:

Table 1. Results of two classification experiments 表 1. 二分类实验结果

Mislabel	0.10			0.15			0.20		
Method	Naive	NRMLTMLR	RMLTMLR	Naive	NRMLTMLR	RMLTMLR	Naive	NRMLTMLR	RMLTMLR
Acc	0.9231	0.8077	0.8462	0.8846	0.7692	0.8077	0.8846	0.6923	0.7308
macroF1	0.9091	0.7826	0.8182	0.8696	0.7000	0.7619	0.8889	0.7143	0.7586
microF1	0.9091	0.7826	0.8182	0.8696	0.7000	0.7619	0.8889	0.7143	0.7586
Hamming	0.0769	0.1923	0.1538	0.1154	0.2308	0.1923	0.1154	0.3077	0.2692

如上表 1 所示,在不同的错误标签比重下,本文提出方法的评价结果均好于直接使用 NRMLTMLR 模型分类的结果,在评价结果上更接近 Naive 方法的分类结果,因此证实了本文方法在二分类实验下对于含错误标签的张量图像的分类稳健性。

4.2. 五分类

为了进一步测试RMLTMLR方法在带噪声的张量图像多分类问题中的表现,本节选用手风琴、飞机、船锚、蚂蚁和木桶 5 个标签 960 张彩色图像作为五分类数据集,分别按照错误标签(mislabel) 0.25、0.30 和 0.35 的比重进行实验,得到的实验结果如下表 2 所示:

Table 2. Results of five classification experiments 表 2. 五分类实验结果

Mislabel	0.25			0.30			0.35		
Method	Naive	NRMLTMLR	RMLTMLR	Naive	NRMLTMLR	RMLTMLR	Naive	NRMLTMLR	RMLTMLR
Acc	0.8576	0.7986	0.8021	0.8958	0.6667	0.8368	0.8785	0.6319	0.8299
macroF1	0.9872	0.9696	0.9617	0.9910	0.8302	0.9650	0.9874	0.8193	0.9749
microF1	0.9872	0.9696	0.9617	0.9910	0.8302	0.9650	0.9874	0.8193	0.9749
Hamming	0.1424	0.2014	0.1979	0.1042	0.3333	0.1632	0.1215	0.3681	0.1701

如上表 2 所示,在不同的错误标签比重下,本文提出方法几乎在所有评价指标下的结果好于直接使用 NRMLTMLR 模型分类的结果。随着错误标签比重增大,可以看到 NRMLTMLR 的分类准确率逐渐减小,而 RMLTMLR 的分类准确率维持在一个稳定的状态下且始终高于 80%。该结果证实了本文方法在五分类实验下对于含错误标签的张量图像的分类稳健性。

4.3. 七分类

本节中我们选用手风琴、飞机、船锚、蚂蚁、木桶、鱼和河狸 7 个标签 1055 张彩色图像作为七分类数据集,分别按照错误标签(mislabel) 0.20、0.25 和 0.30 的比重进行实验,得到的实验结果如下表 3 所示。

如下表 3 所示,本文提出方法在七分类的数据集上评价表现依旧好于直接使用 NRMLTMLR 模型分类,在评价结果上也更接近 Naive 方法的分类结果。当 NRMLTMLR 方法在不同的错误标签比重下分类准确率变化过大的情况下,RMLTMLR 方法不仅保持了准确率均在 72%以上,并且保持了分类结果稳定,进一步证明了本文方法能够做到对带有错误标签的张量图像的稳健分类。

Table 3. Results of seven classification experiments 表 3. 七分类实验结果

mislabel	0.20			0.25			0.30		
method	Naive NRMLTMLR RMLTMLR		Naive	Naive NRMLTMLR RMLTMLR		Naive	NRMLTMLR	RMLTMLR	
acc	0.8170	0.7066	0.7666	0.7886	0.0505	0.7224	0.8013	0.5741	0.7350
macroF1	0.9935	0.9033	0.9525	0.9917	0.0066	0.9674	0.9869	0.8189	0.9459
microF1	0.9935	0.9033	0.9525	0.9917	0.0066	0.9674	0.9869	0.8189	0.9459
hamming	0.1830	0.2934	0.2334	0.2114	0.9495	0.2776	0.1987	0.4259	0.2650

4.4. 耗费时间

为了比对模型的复杂度,我们分别记录了每次实验在 MATLAB R2021b 上运行所消耗的时间,具体结果如下表 4 所示:

Table 4. Model time 表 4. 模型所耗时间

class	二分类				五分类		七分类		
mislabel	0.10	0.15	0.20	0.25	0.30	0.35	0.20	0.25	0.30
NRMLTMLR	8.87	8.59	11.45	89.28	94.44	94.53	106.63	112.34	108.30
RMLTMLR	27.62	37.23	51.11	420.56	434.69	449.40	409.01	465.66	417.26

通过上表 4 可以看出,由于增加了权重参数的缘故,模型的复杂度增加,RMLTMLR 方法在各项实验中消耗的时间约为 NRMLTMLR 方法的五倍,该实验结果也符合理论预期。鉴于 RMLTMLR 模型在面向包含错误标签的张量图像数据集时,表现出的分类效果和分类稳定性远优于 NRMLTMLR 模型,因此我们认为多出的消耗时间是值得的。

5. 总结

本文面向带有错误标签的张量图像数据,基于多分类逻辑回归模型,提出了一种能够直接处理张量格式数据并对带噪声标签具有稳健性的图像多分类模型。该模型利用 γ -散度、张量管道秩及其核范数,通过增加权重函数的方式降低预测标签与训练标签不匹配的实例的权重,从而提升了稳健性,使其在实验中具有优良的表现,为含错误标签的张量图像数据分类提供了新思路。

本文对比了非稳健错标张量多分类模型(NRMLTMLR)和本文方法稳健错标张量多分类模型(RMLTMLR),在加州理工大学的 101 类彩色图像识别数据集的二分类、五分类和七分类子集上进行实验。在实验过程中取不同的错标比重以对比两个方法在不同错标比重不同类别数据集下的分类情况。结果表明,在二分类数据集上,当错标比重分别为 0.10, 0.15 和 0.20 时,本文提出的 RMLTMLR 较 NRMLTMLR 的分类准确率均提升了 3.85 个百分点;在五分类数据集上,当错标比重分别为 0.25、0.30 和 0.35 时,RMLTMLR 较 NRMLTMLR 的分类准确率则分别提升了 0.35、17.01 和 19.8 个百分点;在七分类数据集上,当错标比重为 0.20、0.25 和 0.30 时,RMLTMLR 较 NRMLTMLR 的分类准确率分别提升了 6.00、67.19 和 16.09 个百分点。因此,本文提出的 RMLTMLR 方法在含有错误标签的多分类张量数据集上具有较为显著的优势,也在实验方面证明了本文创新性地融合 y-散度方法增加了权重参数以提升模型对标签污染数据集分类的有效性。

考虑到张量管道秩只适用于三阶张量,面对更高阶的张量,后续需要对其余不同张量秩进行研究。 另外,除了使用 γ-散度方法,还可以参考背景建模的思路对错误标签进行建模,从该角度进一步提升模型的稳健性与适用性,以此优化模型。

基金项目

国家自然科学基金资助项目(12371308)。

参考文献

- [1] He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90
- [2] Hu, J., Shen, L. and Sun, G. (2018). Squeeze-and-Excitation Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, 18-23 June 2018, 7132-7141. https://doi.org/10.1109/cvpr.2018.00745
- [3] Teng, C.M. (2000) Evaluating Noise Correction. In: Mizoguchi, R. and Slaney, J., Eds., PRICAI 2000 Topics in Artificial Intelligence, Springer, Berlin, 188-198. https://doi.org/10.1007/3-540-44533-1 22
- [4] Teng, C.M. (2001) A Comparison of Noise Handling Techniques. *Proceedings of the Fourteenth International Florida Artificial Intelligence Research Society Conference*, Key West, 21-23 May 2001, 269-273.
- [5] Teng, C.M. (2005) Dealing with Data Corruption in Remote Sensing. In: Famili, A.F., Kok, J.N., Peña, J.M., Siebes, A. and Feelders, A., Eds., Advances in Intelligent Data Analysis VI, Springer, Berlin, 452-463. https://doi.org/10.1007/11552253 41
- [6] Brodley, C.E. and Friedl, M.A. (19996) Identifying and Eliminating Mislabeled Training Instances. Proceedings of the Thirteenth National Conference on Artificial Intelligence, Portland, 4-8 August 1996, 799-805.
- [7] Copas, J.B. (1988) Binary Regression Models for Contaminated Data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **50**, 225-253. https://doi.org/10.1111/j.2517-6161.1988.tb01723.x
- [8] Wainer, H., Bradlow, E.T. and Wang, X. (2007). Testlet Response Theory and Its Applications. Cambridge University Press, Cambridge. https://doi.org/10.1017/cbo9780511618765
- Komori, O., Eguchi, S., Ikeda, S., Okamura, H., Ichinokawa, M. and Nakayama, S. (2015) An Asymmetric Logistic Regression Model for Ecological Data. *Methods in Ecology and Evolution*, 7, 249-260. https://doi.org/10.1111/2041-210x.12473
- [10] Hayashi, K. (2011) A Boosting Method with Asymmetric Mislabeling Probabilities Which Depend on Covariates. Computational Statistics, 27, 203-218. https://doi.org/10.1007/s00180-011-0250-8
- [11] Takenouchi, T. and Eguchi, S. (2004) Robustifying Adaboost by Adding the Naive Error Rate. *Neural Computation*, **16**, 767-787. https://doi.org/10.1162/089976604322860695
- [12] Hung, H., Jou, Z. and Huang, S. (2017) Robust Mislabel Logistic Regression without Modeling Mislabel Probabilities. *Biometrics*, **74**, 145-154. https://doi.org/10.1111/biom.12726
- [13] Song, K., Nie, F., Han, J. and Li, X. (2017) Parameter Free Large Margin Nearest Neighbor for Distance Metric Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 31, 2555-2561. https://doi.org/10.1609/aaai.v31i1.10861
- [14] Cai, D., He, X., Hu, Y., Han, J. and Huang, T. (2007). Learning a Spatially Smooth Subspace for Face Recognition. 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, 17-22 June 2007, 1-7. https://doi.org/10.1109/cvpr.2007.383054
- [15] Liu, J., Zhu, C., Long, Z., Huang, H. and Liu, Y. (2021) Low-Rank Tensor Ring Learning for Multi-linear Regression. Pattern Recognition, 113, Article ID: 107753. https://doi.org/10.1016/j.patcog.2020.107753
- [16] Koniusz, P., Wang, L. and Cherian, A. (2022) Tensor Representations for Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 648-665. https://doi.org/10.1109/tpami.2021.3107160
- [17] Tao, D., Li, X.L., Hu, W.M., et al. (2005) Supervised Tensor Learning. Fifth IEEE International Conference on Data Mining (ICDM'05), Houston, 27-30 November 2005, 8.
- [18] Kotsia, I., Guo, W. and Patras, I. (2012) Higher Rank Support Tensor Machines for Visual Recognition. *Pattern Recognition*, **45**, 4192-4203. https://doi.org/10.1016/j.patcog.2012.04.033
- [19] Irene, K. and Ioannis, P. (2011) Support Tucker Machines. *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, 20-25 June 2011, 633-640.

- [20] Tan, X., Zhang, Y., Tang, S.L., et al. (2012) Logistic Tensor Regression for Classification. Proceedings of the Third Sino-Foreign-Interchange Conference on Intelligent Science and Intelligent Data Engineering, Nanjing, 15-17 October 2012, 589-597.
- [21] 张家瑞,胡毓榆,唐开煜,樊亚莉.基于张量低管道秩的图像多分类模型[J]. 建模与仿真, 2024, 13(3): 3980-3997.
- [22] Kilmer, M.E. and Martin, C.D. (2011) Factorization Strategies for Third-Order Tensors. Linear Algebra and Its Applications, 435, 641-658. https://doi.org/10.1016/j.laa.2010.09.020
- [23] Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z. and Yan, S. (2020) Tensor Robust Principal Component Analysis with a New Tensor Nuclear Norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42, 925-938. https://doi.org/10.1109/tpami.2019.2891760
- [24] Kilmer, M.E., Braman, K., Hao, N. and Hoover, R.C. (2013) Third-Order Tensors as Operators on Matrices: A Theoretical and Computational Framework with Applications in Imaging. SIAM Journal on Matrix Analysis and Applications, 34, 148-172. https://doi.org/10.1137/110837711
- [25] Zhang, Z., Ely, G., Aeron, S., Hao, N. and Kilmer, M. (2014). Novel Methods for Multilinear Data Completion and De-Noising Based on Tensor-SVD. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 23-28 June 2014, 3842-3849. https://doi.org/10.1109/cvpr.2014.485
- [26] Jones, M.C. (2001) A Comparison of Related Density-based Minimum Divergence Estimators. *Biometrika*, 88, 865-873. https://doi.org/10.1093/biomet/88.3.865
- [27] Fujisawa, H. and Eguchi, S. (2008) Robust Parameter Estimation with a Small Bias against Heavy Contamination. *Journal of Multivariate Analysis*, **99**, 2053-2081. https://doi.org/10.1016/j.jmva.2008.02.004
- [28] Kanamori, T. and Fujisawa, H. (2015) Robust Estimation Under Heavy Contamination Using Unnormalized Models. *Biometrika*, **102**, 559-572. https://doi.org/10.1093/biomet/asv014
- [29] Hu, Y., Fan, Y., Song, Y. and Li, M. (2023) A General Robust Low-Rank Multinomial Logistic Regression for Corrupted Matrix Data Classification. Applied Intelligence, 53, 18564-18580. https://doi.org/10.1007/s10489-022-04424-0
- [30] Yin, M., Zeng, D., Gao, J., Wu, Z. and Xie, S. (2018) Robust Multinomial Logistic Regression Based on RPCA. IEEE Journal of Selected Topics in Signal Processing, 12, 1144-1154. https://doi.org/10.1109/jstsp.2018.2872460
- [31] 张齐航. 信用评分中拒绝推断问题的研究[D]: [硕士学位论文]. 厦门: 厦门大学, 2021. https://doi.org/10.27424/d.cnki.gxmdu.2021.002140