

# 基于张量分解的神经影像分类模型

## 多元统计分析期末汇报

王一鑫

兰州大学萃英学院

2025 年 6 月 19 日



- ① 引言
- ② 建模方法
- ③ 模拟研究
- ④ 参考文献

- ① 引言
- ② 建模方法
- ③ 模拟研究
- ④ 参考文献

# 全文概要

- 张量又称多维数组，是矩阵的高维推广。与传统的矩阵（二维数据）不同，张量能够处理更高维度的数据，因此它是多维数据表示的理想工具。

# 全文概要

- 张量又称多维数组，是矩阵的高维推广。与传统的矩阵（二维数据）不同，张量能够处理更高维度的数据，因此它是多维数据表示的理想工具。
- 随着计算能力的突破性进展，张量方法在医学影像分析这类跨学科领域的应用日益扩展。



## ① 引言

## ② 建模方法

张量及其分解

基于不同损失函数的分类模型

先验估计

后验推断

贝叶斯张量模型

## ③ 模拟研究

## ④ 参考文献

## ① 引言

## ② 建模方法

张量及其分解

基于不同损失函数的分类模型

先验估计

后验推断

贝叶斯张量模型

## ③ 模拟研究

## ④ 参考文献



# 张量的概念

## 张量

张量是一个多维数组， $d$  阶张量是一个具有  $d$  个维度的数组，表示为：

$$\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$$

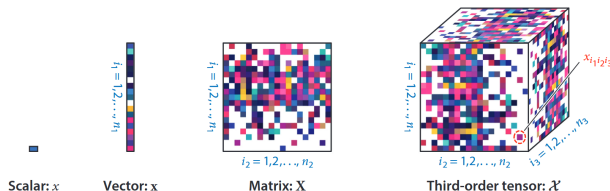


图 1: 标量，向量，矩阵和三阶张量<sup>[1]</sup>

# Tucker 分解

- 张量分解是一种将高维张量分解为若干低维因子组合的技巧，其中一种为 Tucker 分解，它是张量的高阶主成分分析 (PCA)，将张量分解为核心张量与各模式因子矩阵的乘积。

# Tucker 分解

- 张量分解是一种将高维张量分解为若干低维因子组合的技巧，其中一种为 Tucker 分解，它是张量的高阶主成分分析 (PCA)，将张量分解为核心张量与各模式因子矩阵的乘积。
- 具体地，给定一个张量  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ ，Tucker 分解表示为：

# Tucker 分解

- 张量分解是一种将高维张量分解为若干低维因子组合的技巧，其中一种为 Tucker 分解，它是张量的高阶主成分分析 (PCA)，将张量分解为核心张量与各模式因子矩阵的乘积。
- 具体地，给定一个张量  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ ，Tucker 分解表示为：

## Tucker 分解

$$\mathcal{X} \approx \mathcal{C} \times_1 Q^1 \times_2 Q^2 \cdots \times_d Q^d = \sum_{j_1=1}^{m_1} \sum_{j_2=1}^{m_2} \cdots \sum_{j_d=1}^{m_d} c_{j_1 j_2 \dots j_d} \mathbf{q}_{j_1}^1 \circ \mathbf{q}_{j_2}^2 \circ \cdots \circ \mathbf{q}_{j_d}^d \quad (1)$$

# Tucker 分解

- 张量分解是一种将高维张量分解为若干低维因子组合的技巧，其中一种为 Tucker 分解，它是张量的高阶主成分分析 (PCA)，将张量分解为核心张量与各模式因子矩阵的乘积。
- 具体地，给定一个张量  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ ，Tucker 分解表示为：

## Tucker 分解

$$\mathcal{X} \approx \mathcal{C} \times_1 Q^1 \times_2 Q^2 \cdots \times_d Q^d = \sum_{j_1=1}^{m_1} \sum_{j_2=1}^{m_2} \cdots \sum_{j_d=1}^{m_d} c_{j_1 j_2 \dots j_d} \mathbf{q}_{j_1}^1 \circ \mathbf{q}_{j_2}^2 \circ \cdots \circ \mathbf{q}_{j_d}^d \quad (1)$$

- 其中， $\mathcal{C} \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_d}$  是核心张量， $Q^k \in \mathbb{R}^{n_k \times m_k}$  是因子矩阵。

# Tucker 分解示意图

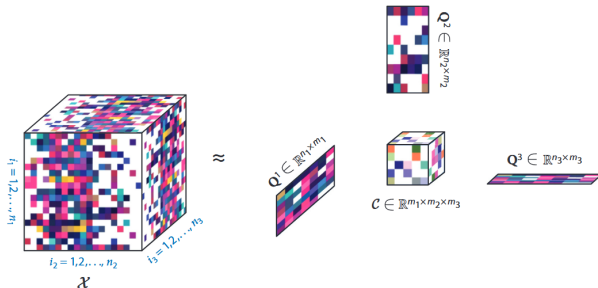


图 2: Tucker 分解示意图<sup>[1]</sup>

# PARAFAC 分解

- 特别地，当  $m_1 = m_2 = \cdots = m_d = R$ ，且核心张量限制为对角型时被称为 PARAFAC 分解. 它将张量近似为一组秩为 1 的张量的和。具体而言，给定一个张量  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ ，其分解形式为：

# PARAFAC 分解

- 特别地，当  $m_1 = m_2 = \cdots = m_d = R$ ，且核心张量限制为对角型时被称为 PARAFAC 分解. 它将张量近似为一组秩为 1 的张量的和。具体而言，给定一个张量  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ ，其分解形式为：

## PARAFAC 分解

$$\mathcal{X} \approx \sum_{r=1}^R \beta_1^{(r)} \circ \beta_2^{(r)} \circ \cdots \circ \beta_d^{(r)} \quad (2)$$



# PARAFAC 分解

- 特别地，当  $m_1 = m_2 = \cdots = m_d = R$ ，且核心张量限制为对角型时被称为 PARAFAC 分解. 它将张量近似为一组秩为 1 的张量的和。具体而言，给定一个张量  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ ，其分解形式为：

## PARAFAC 分解

$$\mathcal{X} \approx \sum_{r=1}^R \beta_1^{(r)} \circ \beta_2^{(r)} \circ \cdots \circ \beta_d^{(r)} \quad (2)$$

- 其中  $\beta_1, \cdots, \beta_d$  是长度为  $p_1, \cdots, p_d$  的向量. PARAFAC 分解将系数  $p_1 \times \cdots \times p_d$  降低至  $R(p_1 + p_2 + \cdots + p_d)$ ，提供了有效的降维.

# PARAFAC 分解图示

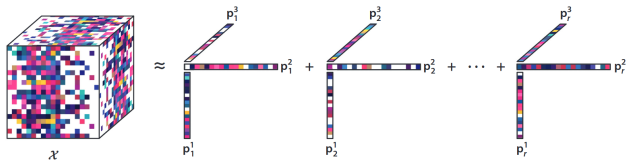


图 3: PARAFAC 分解示意图<sup>[1]</sup>

## ① 引言

## ② 建模方法

张量及其分解

基于不同损失函数的分类模型

先验估计

后验推断

贝叶斯张量模型

## ③ 模拟研究

## ④ 参考文献

# 损失函数

- 损失函数就像是我们用来衡量“错误”的工具。在机器学习中，我们通过这个工具来评判模型训练性能。

# 损失函数

- 损失函数就像是我们用来衡量“错误”的工具。在机器学习中，我们通过这个工具来评判模型训练性能。
- 贝叶斯方法中，损失函数转化为不同类型的似然(likelihood)。通过损失函数结合模型参数的先验假设和真实观测值来更新参数。通过这种方式，我们可以得到对模型参数的更好估计。

# 损失函数

- 损失函数就像是我们用来衡量“错误”的工具。在机器学习中，我们通过这个工具来评判模型训练性能。
- 贝叶斯方法中，损失函数转化为不同类型的似然 (likelihood)。通过损失函数结合模型参数的先验假设和真实观测值来更新参数。通过这种方式，我们可以得到对模型参数的更好估计。
- 我们使用了两种常用的损失函数，分别是以支持向量机为代表的铰链损失 (hinge loss) 和逻辑回归损失 (logistic regression loss)。

# Hinge Loss

## Hinge Loss

$$\mathcal{L}(y|\beta) = \frac{1}{\sigma^2} \max(1 - yf(\mathbf{x}; \beta), 0) + R \quad (3)$$

这里的  $y \in \{-1, 1\}$  是二元输出， $f(\cdot)$  是协变量  $\mathbf{x}$  的线性或非线性的函数， $\beta$  是需要从数据中估计的参数， $\sigma^2$  是调整参数。

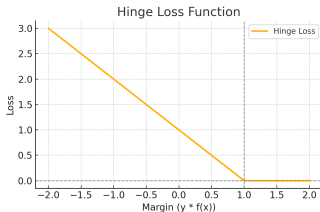


图 4: SVM 的铰链损失

# 伪似然 (pseudo-likelihood) 方法

- SVM 并没有明确的似然函数，不能直接在贝叶斯框架下建模。针对该问题，Polson 和 Scott 提出了一种伪似然 (pseudo-likelihood) 的方法<sup>[3]</sup>。



# 伪似然 (pseudo-likelihood) 方法

- SVM 并没有明确的似然函数，不能直接在贝叶斯框架下建模。针对该问题，Polson 和 Scott 提出了一种伪似然 (pseudo-likelihood) 的方法<sup>[3]</sup>。
- 具体而言，伪似然可以被表示为一种带有潜在变量  $\rho$  的位置-尺度混合正态分布 (location-scale mixture of normals)，这种表示法可以为后验推断提供高效的吉布斯采样器。

# 伪似然 (pseudo-likelihood) 方法

- SVM 并没有明确的似然函数，不能直接在贝叶斯框架下建模。针对该问题，Polson 和 Scott 提出了一种伪似然 (pseudo-likelihood) 的方法<sup>[3]</sup>。
- 具体而言，伪似然可以被表示为一种带有潜在变量  $\rho$  的位置-尺度混合正态分布 (location-scale mixture of normals)，这种表示法可以为后验推断提供高效的吉布斯采样器。

## 伪似然方法

$$\begin{aligned} L &= \prod_{i=1}^n L_i(y_i | \mathbf{x}_i, \beta, \sigma^2) = \prod_{i=1}^n \left\{ \frac{1}{\sigma^2} \exp\left\{-\frac{2}{\sigma^2} \max(1 - y_i f(\mathbf{x}; \beta), 0)\right\} \right\} \\ &= \int_0^\infty \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi} \rho_i} \exp\left(-\frac{(1 + \rho_i - y_i f(\mathbf{x}; \beta))^2}{2\rho_i \sigma^2}\right) d\rho_i \end{aligned} \quad (4)$$

# Logistic Loss

## Logistic Loss

$$\mathcal{L}(y = 1|\beta) = \frac{\exp\{f(\mathbf{x}; \beta)\}}{1 + \exp\{f(\mathbf{x}; \beta)\}} \quad (5)$$

其中， $f(\mathbf{x}; \beta)$  代表协变量对逻辑损失的贡献， $\beta$  为待估参数。

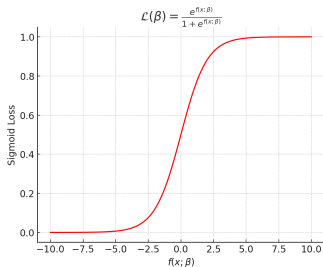


图 5: 逻辑回归损失函数

# Polya-Gamma 潜变量

- 在贝叶斯框架中，逻辑回归的似然函数在分析上并不方便处理，这使得直接从后验分布中采样变得困难。针对此问题，我们通常使用 Polya-Gamma 潜变量来实现<sup>[4]</sup>。

# Polya-Gamma 潜变量

- 在贝叶斯框架中，逻辑回归的似然函数在分析上并不方便处理，这使得直接从后验分布中采样变得困难。针对此问题，我们通常使用 Polya-Gamma 潜变量来实现<sup>[4]</sup>。

## Polya-Gamma 分布

若随机变量  $X$  有如下形式

$$X \stackrel{D}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - 1/2)^2 + c^2/(4\pi^2)} \quad (6)$$

则称  $X$  服从参数为  $b > 0, c \in \mathbb{R}$  的 Polya-Gamma 分布，记作  $X \sim PG(b, c)$ 。其中  $g_k$  服从 Gamma 分布  $Ga(b, 1)$ ，相互独立。 $\stackrel{D}{=}$  表示在分布意义下相等。

# Polya-Gamma 潜变量

- 通过引入 Polya-Gamma 潜在变量可以将对数优势比参数化的二项式似然表示为关于 Polya-Gamma 分布的高斯混合<sup>[4]</sup>。逻辑损失函数可以通过对潜在的 Polya-Gamma 变量进行边缘化处理而得到，其关系如下所示：

# Polya-Gamma 潜变量

- 通过引入 Polya-Gamma 潜在变量可以将对数优势比参数化的二项式似然表示为关于 Polya-Gamma 分布的高斯混合<sup>[4]</sup>。逻辑损失函数可以通过对潜在的 Polya-Gamma 变量进行边缘化处理而得到，其关系如下所示：

## 损失函数

$$\frac{(e^{f(\cdot)})^y}{(1 + e^{f(\cdot)})^b} = 2^{-b} e^{\kappa \psi} \int_0^\infty e^{-\omega \psi^2 / 2} p(\omega) d\omega, \quad b > 0, \quad \kappa = y - \frac{b}{2} \quad (7)$$

# Polya-Gamma 潜变量

- 通过引入 Polya-Gamma 潜在变量可以将对数优势比参数化的二项式似然表示为关于 Polya-Gamma 分布的高斯混合<sup>[4]</sup>。逻辑损失函数可以通过对潜在的 Polya-Gamma 变量进行边缘化处理而得到，其关系如下所示：

## 损失函数

$$\frac{(e^{f(\cdot)})^y}{(1 + e^{f(\cdot)})^b} = 2^{-b} e^{\kappa \psi} \int_0^\infty e^{-\omega \psi^2 / 2} p(\omega) d\omega, \quad b > 0, \quad \kappa = y - \frac{b}{2} \quad (7)$$

- 其中  $\omega \sim \text{PG}(b, 0)$ ,  $p(\omega)$  表示 Polya-Gamma 分布的密度函数。



# Polya-Gamma 潜变量

在该等式基础上，完整的数据增强似然函数为

$$L = \prod_{i=1}^n \frac{(e^{f_i})^{y_i}}{1 + e^{f_i}} = \prod_{i=1}^n 2^{-1} e^{\kappa_i \psi_i} \int_0^\infty e^{-\omega_i f_i^2 / 2} p(\omega_i) d\omega_i \quad (8)$$

其中  $\kappa_i = y_i - \frac{1}{2}$ ,  $b = 1$ ,  $\omega_i \sim \text{PG}(1, 0)$ 。

## ① 引言

## ② 建模方法

张量及其分解

基于不同损失函数的分类模型

先验估计

后验推断

贝叶斯张量模型

## ③ 模拟研究

## ④ 参考文献

# 先验估计

## 线性预测模型

$$f_i = \langle \mathbf{X}_i, \mathbf{B} \rangle + \mathbf{z}_i' \boldsymbol{\gamma} \quad (9)$$

- 这里的  $\mathbf{X}_i$  和  $\mathbf{z}_i$  分别表示第  $i$  个样本的影像预测变量与其他特征。符号  $\langle \cdot, \cdot \rangle$  表示内积算子。张量系数矩阵  $\mathbf{B}$  用于量化图像在分类模型中的作用， $\boldsymbol{\gamma}$  是一个维度为  $p_z + 1$  的向量，用以捕捉补充协变量的影响。
- 对张量  $\mathbf{B}$  进行 PARAFAC 分解，这里的  $\mathbf{B} \in \bigotimes_{j=1}^d \mathbb{R}^{p_j}$ ，有如下分解形式

$$\mathbf{B} \approx \sum_{r=1}^R \beta_1^{(r)} \circ \beta_2^{(r)} \circ \dots \circ \beta_d^{(r)} \quad (10)$$

# multiway Dirichlet generalized double Pareto , M-DGDP

- 在贝叶斯框架下需要有先验假设, 对这里  $\beta_j^{(r)}$  的先验选择, 我们可以采用多向 Dirichlet 广义双帕累托 (multiway Dirichlet generalized double Pareto , M-DGDP) 分布<sup>[5]</sup>。
- Guhanियogi 等人证明了<sup>[5]</sup> 使用该方法作为先验可以在贝叶斯张量回归中实现自动稀疏性控制、低秩建模与不失大信号的精确建模, 同时具备对称性和后验一致性的理论保证。
- 具体地, 该先验在各组分之间以可交换方式诱导收缩效应, 其中全局尺度参数为  $\tau \sim \text{Ga}(a_\tau, b_\tau)$ , 并在每个组分中进行调整为  $\tau_r = \phi_r \tau$ ,  $r = 1, \dots, R$ , 其中  $\Phi = (\phi_1, \dots, \phi_R) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_R)$ , 其作用是鼓励在假设的 PARAFAC 分解中向低秩方向收缩。
- 此外, 令  $\mathbf{W}_{jr} = \text{diag}(w_{jr,1}, \dots, w_{jr,p_j})$ ,  $j = 1, \dots, d$  且  $r = 1, \dots, R$  表示边缘特异的尺度参数。

## 23 / 47

## ① 引言

## ② 建模方法

张量及其分解

基于不同损失函数的分类模型

先验估计

后验推断

贝叶斯张量模型

## ③ 模拟研究

## ④ 参考文献

# MCMC 算法

- 蒙特卡罗法 (Monte Carlo method) 是通过从概率模型的随机抽样进行近似数值计算的方法。马尔可夫链蒙特卡罗法 (Markov Chain Monte Carlo, MCMC), 则是以马尔可夫链 (Markov chain) 为概率模型的蒙特卡罗法。马尔可夫链蒙特卡罗法构建一个马尔可夫链, 使其平稳分布就是要进行抽样的分布, 首先基于该马尔可夫链进行随机游走, 产生样本的序列, 之后使用该平稳分布的样本进行近似数值计算。

# MCMC 算法

- 蒙特卡罗法 (Monte Carlo method) 是通过从概率模型的随机抽样进行近似数值计算的方法。马尔可夫链蒙特卡罗法 (Markov Chain Monte Carlo, MCMC), 则是以马尔可夫链 (Markov chain) 为概率模型的蒙特卡罗法。马尔可夫链蒙特卡罗法构建一个马尔可夫链, 使其平稳分布就是要进行抽样的分布, 首先基于该马尔可夫链进行随机游走, 产生样本的序列, 之后使用该平稳分布的样本进行近似数值计算。
- 吉布斯采样和 Metropolis-Hastings 算法是常用的 MCMC 方法。



# Gibbs Sampling

---

## Algorithm 1: Gibbs Sampling (吉布斯抽样)

---

**Input:** 目标密度函数  $p(x)$ , 函数  $f(x)$ ; 收敛步数  $m$ , 迭代步数  $n$

**Output:** 随机样本  $\{x_{m+1}, \dots, x_n\}$  及函数样本均值  $f_{mn}$

- 1 初始化: 设初始样本  $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_k^{(0)})^\top$ .
  - 2 **for**  $i = 1$  **to**  $n$  **do**
  - 3     设上一步样本为  $x^{(i-1)} = (x_1^{(i-1)}, x_2^{(i-1)}, \dots, x_k^{(i-1)})^\top$
  - 4     **for**  $j = 1$  **to**  $k$  **do**
  - 5         从条件分布  $p(x_j | x_1^{(i)}, \dots, x_{j-1}^{(i)}, x_{j+1}^{(i-1)}, \dots, x_k^{(i-1)})$  采样  $x_j^{(i)}$ .
  - 6     **end**
  - 7     得到当前样本  $x^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_k^{(i)})^\top$ .
  - 8 **end**
  - 9 抽取后  $n - m$  个样本组成样本集合  $\{x^{(m+1)}, \dots, x^{(n)}\}$
  - 10 计算函数样本均值:  $f_{mn} = \frac{1}{n-m} \sum_{i=m+1}^n f(x^{(i)})$
-

## ① 引言

## ② 建模方法

张量及其分解

基于不同损失函数的分类模型

先验估计

后验推断

贝叶斯张量模型

## ③ 模拟研究

## ④ 参考文献

# 贝叶斯张量模型

- 令  $y \in \mathbb{R}$  表示一个响应值,  $\mathbf{z} \in \mathbb{R}^p$ ,  $\mathbf{X} \in \bigotimes_{j=1}^d \mathbb{R}^{p_j}$ , 我们有如下的张量回归模型:

$$y|\gamma, \mathbf{B}, \sigma \sim \mathcal{N}(\mathbf{z}'\gamma + \langle \mathbf{X}, \mathbf{B} \rangle, \sigma^2)$$

$$\mathbf{B} = \sum_{r=1}^R \mathbf{B}_r, \quad \mathbf{B}_r = \beta_1^{(r)} \circ \dots \circ \beta_d^{(r)}$$

$$\gamma \sim \pi_\gamma, \quad \beta_j^{(r)} \sim \pi_{\beta} \quad (12)$$

- 先前提出的多向先验(11) 可为张量回归模型 (12) 的大多数参数提供吉布斯采样方案。我们依赖于边缘化与分块策略来减少以下参数的自相关性

$$\left\{ \left( \beta_j^{(r)}, w_{jr}; 1 \leq j \leq d, 1 \leq r \leq R \right), (\Phi, \tau), \gamma \right\}$$

从  $[\alpha, \Phi, \tau | \mathbf{B}, \mathbf{W}]$ ,  $[\mathbf{B}, \mathbf{W} | \Phi, \tau, \gamma, \mathbf{y}]$ , 以及  $[\gamma | \mathbf{B}, \mathbf{y}]$  依次进行采样。具体算法 BT-SVM 和 BT-LR 太过复杂, 并未列出。

## ① 引言

## ② 建模方法

## ③ 模拟研究

模拟数据生成

模型评估

MRI 脑肿瘤分类

## ④ 参考文献

## ① 引言

## ② 建模方法

## ③ 模拟研究

模拟数据生成

模型评估

MRI 脑肿瘤分类

## ④ 参考文献

# 模拟数据生成

- 我们通过几种模拟设置来阐述方法的性能，并使用其他已有的方法进行比较。考虑四种不同类型的张量系数  $\mathbf{B}$  来生成二元结果，设置如下：

# 模拟数据生成

- 我们通过几种模拟设置来阐述方法的性能，并使用其他已有的方法进行比较。考虑四种不同类型的张量系数  $\mathbf{B}$  来生成二元结果，设置如下：
- 场景 1 张量  $\mathbf{B}$  由秩  $R_0 = 3$  和维度  $p = c(48, 48)$  的秩- $R$  PARAFAC 分解构建。每个  $\beta$  边缘  $\beta_j^{(r)}$  都是从独立的二项分布  $\text{Binomial}(2, 0.2)$  生成的。在构建张量之后，我们将张量  $\mathbf{B}$  单元格的最大值设置为 1。

# 模拟数据生成

- 我们通过几种模拟设置来阐述方法的性能，并使用其他已有的方法进行比较。考虑四种不同类型的张量系数  $\mathbf{B}$  来生成二元结果，设置如下：
- 场景 1 张量  $\mathbf{B}$  由秩  $R_0 = 3$  和维度  $p = c(48, 48)$  的秩- $R$  PARAFAC 分解构建。每个  $\beta$  边缘  $\beta_j^{(r)}$  都是从独立的二项分布  $\text{Binomial}(2, 0.2)$  生成的。在构建张量之后，我们将张量  $\mathbf{B}$  单元格的最大值设置为 1。
- 场景 2 张量图像由秩  $R_0 = 3$  的秩- $R$  PARAFAC 分解模拟。这里，我们没有从已知分布生成张量边缘，而是手动设置了  $\beta_j^{(r)}$  的每个值。



# 模拟数据生成

- 我们通过几种模拟设置来阐述方法的性能，并使用其他已有的方法进行比较。考虑四种不同类型的张量系数  $\mathbf{B}$  来生成二元结果，设置如下：
- 场景 1 张量  $\mathbf{B}$  由秩  $R_0 = 3$  和维度  $p = c(48, 48)$  的秩- $R$  PARAFAC 分解构建。每个  $\beta$  边缘  $\beta_j^{(r)}$  都是从独立的二项分布  $\text{Binomial}(2, 0.2)$  生成的。在构建张量之后，我们将张量  $\mathbf{B}$  单元格的最大值设置为 1。
- 场景 2 张量图像由秩  $R_0 = 3$  的秩- $R$  PARAFAC 分解模拟。这里，我们没有从已知分布生成张量边缘，而是手动设置了  $\beta_j^{(r)}$  的每个值。
- 场景 3 与从 PARAFAC 分解生成 2D 张量图像不同，张量系数  $\mathbf{B}$  对于矩形区域设置为 1，否则为 0. 非零元素约占总面积的 30%。

# 模拟数据生成

- 我们通过几种模拟设置来阐述方法的性能，并使用其他已有的方法进行比较。考虑四种不同类型的张量系数  $\mathbf{B}$  来生成二元结果，设置如下：
- 场景 1 张量  $\mathbf{B}$  由秩  $R_0 = 3$  和维度  $p = c(48, 48)$  的秩- $R$  PARAFAC 分解构建。每个  $\beta$  边缘  $\beta_j^{(r)}$  都是从独立的二项分布  $\text{Binomial}(2, 0.2)$  生成的。在构建张量之后，我们将张量  $\mathbf{B}$  单元格的最大值设置为 1。
- 场景 2 张量图像由秩  $R_0 = 3$  的秩- $R$  PARAFAC 分解模拟。这里，我们没有从已知分布生成张量边缘，而是手动设置了  $\beta_j^{(r)}$  的每个值。
- 场景 3 与从 PARAFAC 分解生成 2D 张量图像不同，张量系数  $\mathbf{B}$  对于矩形区域设置为 1，否则为 0. 非零元素约占总面积的 30%。
- 场景 4 张量系数  $\mathbf{B}$  对于圆形区域设置为 1，否则为 0. 非零元素约占总面积的 10%。

# 模拟数据图示

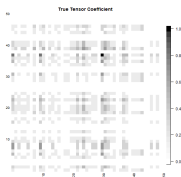


图 6: Scene 1

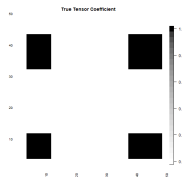


图 7: Scene 2

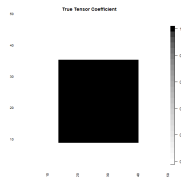


图 8: Scene 3

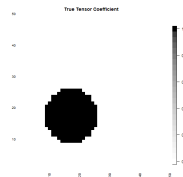


图 9: Scene 4

## ① 引言

## ② 建模方法

## ③ 模拟研究

模拟数据生成

模型评估

MRI 脑肿瘤分类

## ④ 参考文献

# 模型评估

- 我们使用均方根误差（Root Mean Squared Error, RMSE）与相关系数（correlation coefficient）来评估单元层级张量系数的点估计精度。同时，为了衡量分类准确性，我们计算误分类率（misclassification error）与 F1 分数（F1 score）。

# 模型评估

- 我们使用均方根误差（Root Mean Squared Error, RMSE）与相关系数（correlation coefficient）来评估单元层级张量系数的点估计精度。同时，为了衡量分类准确性，我们计算误分类率（misclassification error）与 F1 分数（F1 score）。
- 实验中，我们将数据按 70:30 的比例划分为训练集与测试集。用于系数估计与特征选择性能评估的指标在训练集上计算，而分类性能指标在测试集上进行评估。

# 模型评估

- 我们使用均方根误差（Root Mean Squared Error, RMSE）与相关系数（correlation coefficient）来评估单元层级张量系数的点估计精度。同时，为了衡量分类准确性，我们计算误分类率（misclassification error）与 F1 分数（F1 score）。
- 实验中，我们将数据按 70:30 的比例划分为训练集与测试集。用于系数估计与特征选择性能评估的指标在训练集上计算，而分类性能指标在测试集上进行评估。
- 为检验新模型的性能，我们选取两种现有的先进分类方法作为对比模型，分别是带有 Lasso 惩罚项的逻辑回归模型（Lasso Logistic Regression）和 L1 范数支持向量机模型（L1-norm Support Vector Machine, SVM）。

# 张量数据预测

- 图 10 展示了使用 BT-SVM 和 BT-LR 估算张量系数的情况。从图中可以看出，我们提出的方法能够广泛地恢复二维张量  $\mathbf{B}$  的形状，而不受其形状的影响，也不取决于张量信号是否通过 PARAFAC 分解构建。



# 张量数据预测

- 图 10 展示了使用 BT-SVM 和 BT-LR 估算张量系数的情况。从图中可以看出，我们提出的方法能够广泛地恢复二维张量  $\mathbf{B}$  的形状，而不受其形状的影响，也不取决于张量信号是否通过 PARAFAC 分解构建。
- 为了展示预测系数和真实系数之间的相关程度，我们通过图 11 展示了不同模型（LR 和 SVM）在不同情景下的表现。

# 张量数据预测

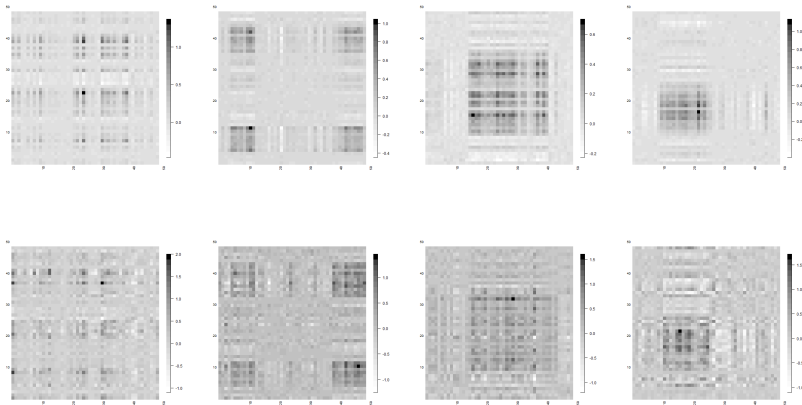


图 10: 估计的张量系数 (第一排由 BT-SVM 模型估计第二排由 BT-LR 模型估计)

## 残差图

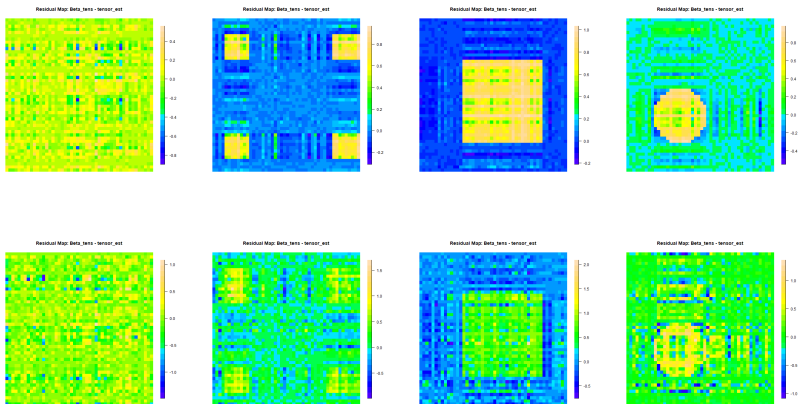


图 11: 残差图

# 分类性能

- 表 1 和表 3 中展示了评估模型性能的相关结果，分别对应场景 1-4。具体而言，表 1 展示了在二元结果  $Y$  由 SVM Loss 生成时的四个场景的结果，而表 3 则展示了当  $Y$  来自逻辑损失时的结果。这些结果表明，提出的两种方法（BT-SVM 和 BT-LR）在系数估计和分类性能上始终优于其他竞争的惩罚方法。

# 分类性能

- 表 1 和表 3 中展示了评估模型性能的相关结果，分别对应场景 1-4。具体而言，表 1 展示了在二元结果  $Y$  由 SVM Loss 生成时的四个场景的结果，而表 3 则展示了当  $Y$  来自逻辑损失时的结果。这些结果表明，提出的两种方法（BT-SVM 和 BT-LR）在系数估计和分类性能上始终优于其他竞争的惩罚方法。
- 当二元结果数据来自 SVM 损失时，BT-SVM 方法具有优越的系数估计（如表1 中较低的 RMSE 和较高的相关系数所示）和改进的分类精度（如表 1 中较低的误分类率和较高的 F1 分数所示）。即便在数据由逻辑损失生成时，依然有这一特点。

# 分类性能

表 1: Performance Comparison of Different Methods Across Scenarios; Y generated from SVM Loss.

Scenarios	Methods	RMSE	Corr.Coeff.	Mis. Class.	F1-score
Scenario 1	LR w/ lasso	0.186	0.118	0.500	0.576
	L1norm-SVM	0.129	0.233	0.413	0.523
	BT-SVM	0.127	0.591	0.340	0.653
	BT-LR	0.197	0.449	0.360	0.640
Scenario 2	LR w/ Lasso	0.395	0.057	0.520	0.487
	L1norm-SVM	0.393	0.168	0.433	0.591
	BT-SVM	0.285	0.820	0.193	0.803
	BT-LR	0.330	0.505	0.347	0.653

# 分类性能

表 2: Performance Comparison of Different Methods Across Scenarios; Y generated from SVM Loss.

Scenarios	Methods	RMSE	Corr.Coeff.	Mis. Class.	F1-score
Scenario 3	LR w/ Lasso	0.541	0.064	0.453	0.585
	L1norm-SVM	0.539	0.166	0.407	0.639
	BT-SVM	0.430	0.768	0.227	0.788
	BT-LR	0.454	0.570	0.320	0.684
Scenario 4	LR w/ Lasso	0.315	0.178	0.407	0.639
	L1norm-SVM	0.315	0.213	0.440	0.554
	BT-SVM	0.221	0.769	0.207	0.805
	BT-LR	0.276	0.529	0.347	0.679

# 分类性能

表 3: Performance Comparison of Different Methods Across Scenarios; Y generated from Lasso Loss.

Scenarios	Methods	RMSE	Corr.Coeff.	Mis. Class.	F1-score
Scenario 1	LR w/ lasso	0.175	0.078	0.493	0.580
	L1norm-SVM	0.192	0.208	0.527	0.448
	BT-SVM	0.120	0.701	0.220	0.793
	BT-LR	0.230	0.456	0.340	0.698
Scenario 2	LR w/ Lasso	0.395	0.128	0.547	0.474
	L1norm-SVM	0.393	0.215	0.453	0.534
	BT-SVM	0.233	0.880	0.140	0.844
	BT-LR	0.284	0.670	0.233	0.788



# 分类性能

表 4: Performance Comparison of Different Methods Across Scenarios; Y generated from Lasso Loss.

Scenarios	Methods	RMSE	Corr.Coeff.	Mis. Class.	F1-score
Scenario 3	LR w/ Lasso	0.542	0.050	0.427	0.467
	L1norm-SVM	0.539	0.149	0.413	0.544
	BT-SVM	0.426	0.811	0.187	0.823
	BT-LR	0.414	0.555	0.313	0.647
Scenario 4	LR w/ Lasso	0.316	0.127	0.447	0.518
	L1norm-SVM	0.315	0.193	0.473	0.530
	BT-SVM	0.221	0.739	0.213	0.790
	BT-LR	0.317	0.479	0.360	0.635

## ① 引言

## ② 建模方法

## ③ 模拟研究

模拟数据生成

模型评估

MRI 脑肿瘤分类

## ④ 参考文献

# 脑肿瘤 MRI 图像分类

- 数据集包含 155 张脑部肿瘤切片 MRI 图像和 98 张正常 MRI 图像。我们对比了 BT-SVM，BT-LR，带 Lasso 惩罚的逻辑回归以及 L1norm-SVM 对该数据集的分类性能。测试结果见表 5。

表 5: 不同模型的分类性能

Methods	Mis. Class.	F1-score
LR w/ lasso	0.333	0.702
L1norm-SVM	0.451	0.709
BT-SVM	0.211	0.855
BT-LR	0.303	0.763

- ① 引言
- ② 建模方法
- ③ 模拟研究
- ④ 参考文献

- [1] Bi X, Tang X, Yuan Y, et al. Tensors in statistics. *Annual Review of Statistics and Its Application*, 2021, 8(1): 345-368.
- [2] Tucker L R. Some mathematical notes on three-mode factor analysis[J]. *Psychometrika*, 1966, 31(3): 279-311.
- [3] Polson N G, Scott S L. Data augmentation for support vector machines[J]. 2011.
- [4] Polson N G, Scott J G, Windle J. Bayesian inference for logistic models using Pólya-Gamma latent variables. *Journal of the American Statistical Association*, 2013, 108(504): 1339-1349.
- [5] Guhaniyogi R, Qamar S, Dunson D B. Bayesian tensor regression[J]. *Journal of Machine Learning Research*, 2017, 18(79): 1-31.

*Thanks!*