

# **Hierarchical heterogeneous analysis of high-dimensional data based on interactions (Test Part)**

Yan Ren

2021103739

2022 年 10 月 5 日

# 目录

<b>1</b>	<b>Simple Set1-Linear</b>	<b>3</b>
1.1	Estimation . . . . .	3
1.1.1	Numerical Setting . . . . .	5
1.2	Results . . . . .	5
<b>2</b>	<b>Simple Set2-Two Part</b>	<b>7</b>
2.1	Estimation . . . . .	7
2.2	Numerical Setting . . . . .	8
2.3	Results . . . . .	9
<b>3</b>	<b>Simple Set3-Interaction</b>	<b>9</b>
3.1	Estimation . . . . .	9
3.2	Numerical Setting . . . . .	11
3.3	Package Results . . . . .	11
3.4	Iteration Results . . . . .	11
3.4.1	One Class . . . . .	12
3.4.2	Multiple Class-Fixed $q_{C_i}$ . . . . .	14
3.4.3	Multiple Class-Unknown $q_{C_i}$ . . . . .	17
<b>4</b>	<b>Test Summary</b>	<b>18</b>

# 1 Simple Set1-Linear

## 1.1 Estimation

尝试最简单的设定验证代码的准确性，模型简化为多分类的回归问题，在类别内部有

$$Y_i = X_i^T \beta_k, \text{ if } i \in \text{Class } k \quad (1.1)$$

目标函数现简化为 minimize 负对数似然函数

$$l = -\log P_Y(y) = -\sum_{i=1}^n \log \sum_{k=1}^K \pi_k f_k(y_i | X_i, \beta_k) \quad (1.2)$$

引入 hidden variable  $C_i$  用以指示第  $i$  个样本所属类别，其后验分布分布列为

$$q_{C_i}(c_i) = \frac{\prod_{k=1}^K (\pi_k f_k(y_i))^{I(C_i=k)}}{\sum_{k=1}^K \pi_k f_k(y_i)} \quad (1.3)$$

进而（根据[Gaussian mixture models and the EM algorithm]）

$$\begin{aligned} \log_Y(y) &= \log \sum_{k=1}^K p_{Y,C}(y, c) \\ &= \log E_{q_C} \left( \frac{p_{Y,C}(y, c)}{q_C(c)} \right) \\ &\geq E_{q_C} \left( \frac{p_{Y,C}(y, c)}{q_C(c)} \right) \\ &= E_{q_C} (\log p_{Y,C}(y, c)) - E_{q_C} (\log q_C(c)) \end{aligned} \quad (1.4)$$

其中  $E_{q_C} (\log q_C(c))$  与  $\beta_k$  取值无关，对参数进行估计时仅考虑第一项，此时问题转化为最大化  $E_{q_C} \log p_{Y,C}(y, c)$

$$\begin{aligned}
E_{q_C} \log p_{Y,C}(y, k) &= E_{q_C} \log (p_C(c) p_{Y|C}(y)) \\
&= E_{q_C} \log \left( \prod_{i=1}^n p_{C_i}(c_i) p_{Y_i|C_i}(y_i) \right) \\
&= E_{q_C} \log \left( \prod_{i=1}^n \prod_{k=1}^K \pi_{c_i} f_{Y_i|C_i}(y_i) \right)^{I(C_i=k)} \\
&= E_{q_C} \sum_{i=1}^n \sum_{k=1}^K I(C_i = k) (\log \pi_{c_i} + \log f_{Y_i|C_i}(y_i)) \\
&= \sum_{i=1}^n \sum_{k=1}^K E_{q_C} I(C_i = k) (\log \pi_{c_i} + \log f_{Y_i|C_i}(y_i)) \\
&= \sum_{i=1}^n \sum_{k=1}^K q_{C_i}(k) (\log \pi_{c_i} + \log f_{Y_i|C_i}(y_i)) \\
&= \sum_{i=1}^n \sum_{k=1}^K q_{C_i}(k) \left( \log \pi_{c_i} + \log \rho_k - \frac{1}{2} \log 2\pi - \frac{\rho_k^2}{2} (y_i - \mu_{ik})^2 \right)
\end{aligned} \tag{1.5}$$

其中  $C$  为与参数无关的项，对于类别  $k$ ，优化目标为

$$\max \quad \sum_{i=1}^n q_{C_i}(k) \left( C - \frac{\rho_k^2}{2} (y_i - X_i^T \beta_k)^2 \right) \tag{1.6}$$

$\Longleftrightarrow$

$$\begin{aligned}
\min \quad l_k &= \sum_{i=1}^n q_{C_i}(k) (y_i - X_i^T \beta_k)^2 \\
&= (y - X \beta_k)^T W_k (y - X \beta_k)
\end{aligned} \tag{1.7}$$

其中  $W_k = \text{diag}(q_{C_1}(k), \dots, q_{C_n}(k))$ ,  $y = (y_1, \dots, y_n)^T \in \mathbb{R}^{n \times 1}$ ,  $X = (X_1, \dots, X_n)^T \in \mathbb{R}^{n \times p}$ ,  $\beta_k \in \mathbb{R}^{p \times 1}$

转化为加权最小二乘问题（文献 [1] page 3），解为

$$\hat{\beta}_k = (X^T W_k X)^{-1} X^T W_k y \tag{1.8}$$

$\rho_k$  的选取会对分类效果产生影响。可以认为各类别共用一个  $\rho$ （即认为各个类别内部的随机误差离散程度相同），这种情况下的更新式为 1.9；还可以认为各类别  $\rho_k$  分别更新不共用相同值，此时更新式为 1.10，注意在这种情况下判断样本归类不仅取决于其距离哪个类别的中心更近，还受类别  $\rho_k$  当前取值的影响，当一个类别参数收敛情况较好时，对应  $\rho_k$  应很大（ $\sigma_k = 1/\rho_k$ ）。

在测试实验中，会同时尝试两种策略，共用  $\rho$  会减轻样本无法归类到距离其最近的类别中心的问题，使用不同  $\rho_k$  可以方便观测到各个类别的收敛情况。

$$\hat{\rho}^{(0)} = n / \left( \sum_{i=1}^n \sum_{k=1}^K (y_i - \mu_{ik})^2 \right) \quad (1.9)$$

$$\hat{\rho}_k^{(0)} = \left( \sum_{i=1}^n q_{C_i}(k) \right) / \left( \sum_{i=1}^n \sum_{k=1}^K (y_i - \mu_{ik})^2 \right), k = 1, \dots, K \quad (1.10)$$

### 1.1.1 Numerical Setting

四个类别真实参数设置见表 1

	Comp.1	Comp.2	Comp.3	Comp.4
$\beta_{k1}$	-2	-1	1	2
$\beta_{k2}$	-2	-1	1	2
$\beta_{k3}$	-2	-1	1	2
$\beta_{k4}$	-2	-1	1	2

## 1.2 Results

使用 *flexmix* 包和 *fmrs* 包进行对比。

*fmrs*: 随机初始化 ( $N(0, 1)$ ) 后得到正确结果。

*flexmix*: 无需给定初始值，得到正确结果，同时，如果两个类别距离很近，会自动合并为一类。

1 为 EM 算法计算流程，以不同类别共用相同  $\rho$  为例，不同类别分别更新  $\rho_k$  的流程也类似，只需要替换  $\rho_k$  的更新公式即可。

1. 收敛效果：使用与调用 *fmrs* 包时相同初值直接进行计算，即相同一组  $N(0, 4)$  的随机数初始化，迭代约13步得到正确结果，改变方差结果类似，若全 0 初始化各个类别的参数估计始终一致，无法收敛。
2. 收敛条件（停止迭代）：两次迭代得到的参数估计值  $\|\hat{\beta}_k^{(iter)} - \hat{\beta}_k^{(iter-1)}\|_2^2$  小于临界值，满足该条件未必收敛到真值，但继续运算不会再有变化。

---

**Algorithm 1:** 最简设定下的算法（各类别共用  $\rho$ ）

---

**Input** : 观测到的  $y$ , 最大类别个数  $K$ , 参数维度  $p$

**Output:** 各样本的分类结果,  $\beta_k, k = 1, \dots, K$  的估计值

```
1 初始化  $\hat{\beta}_k^{(0)}$ （正态分布随机数）；
2 初始化  $\pi_k^{(0)} = 1/K, k = 1, \dots, K$ ；
3 iter = 1.
4 repeat
5   计算基于  $\hat{\beta}_k^{(iter-1)}$  得到的后验概率  $q_C$ 
6   if  $\sum_{k'=1}^K \pi_{k'} f_{k'}(y_i) \neq 0$  and iter > 1 then
7      $q_{C_i}^{(iter)}(k) = \frac{\pi_k^{(iter-1)} f_k(y_i; \hat{\beta}_k^{(iter-1)})}{\sum_{k'=1}^K \pi_{k'}^{(iter-1)} f_{k'}(y_i; \hat{\beta}_{k'}^{(iter-1)})}, k = 1, \dots, K$ 
8   else
9      $q_{C_i}^{(iter)}(k) = \frac{\pi_k^{(iter-1)} / \|y_i - X_i^T \hat{\beta}_k^{(iter-1)}\|_2^2}{\sum_{k'=1}^K \pi_{k'}^{(iter-1)} / \|y_i - X_i^T \hat{\beta}_{k'}^{(iter-1)}\|_2^2}, k = 1, \dots, K$ 
10  end if
11 更新  $\pi_k^{(iter)} = \frac{1}{n} \sum_{i=1}^n q_{C_i}^{(iter)}(k)$ 
12 for  $k$  in  $1 : K$  do
13    $W_k = \text{diag}(q_{C_1}(k), \dots, q_{C_n}(k))$ ;
14   更新  $\hat{\beta}_k^{(iter)} = (X^T W_k X)^{-1} X^T W_k y$ ;
15 end for
16 更新  $\hat{\rho}^{2(iter)} = n / \left( \sum_{i=1}^n \sum_{k=1}^K q_{C_i}^{(iter)}(k) (y_i - X_i^T \hat{\beta}_k^{(iter)})^2 \right)$ ;
17 iter = iter + 1
18 until 收敛（参数差的二范数小于临界值）；
```

---

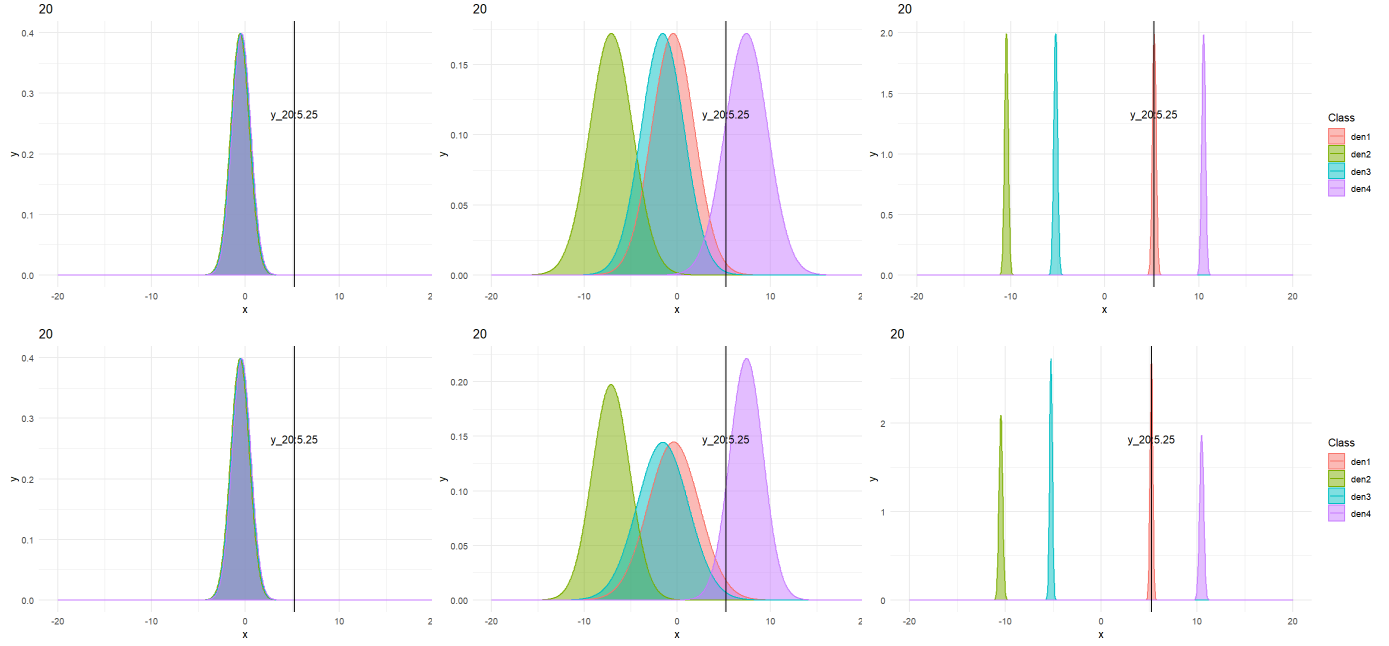


图 1: 简单设定下各类别分布随迭代轮数变化图 (样本 20 为例, 第一行为相同类别共用  $\rho$  的设定, 第二行为各类别分别更新各自  $\rho_k$  的设定)

在实际运算中发现, 当参数估计近乎正确时  $\hat{\rho}_k$  会非常大, 即  $\sigma_k$  会很小, 导致样本在不同高斯分布密度函数的取值全为 0, 此时使用真值距离各类别中心的欧氏距离计算概率, 保证代码正常运行的同时降低本轮  $\hat{\rho}$ , 以更好收敛到真值。

3. 是否收敛到真值:  $\|\hat{y}^{(iter)} - \hat{\beta}_k^{(iter-1)}\|_2^2$  距离小于一定值认为迭代收敛到真值。模拟数据知道真实参数可以用这个条件进行判断, 真实计算中没有该判断条件。

## 2 Simple Set2-Two Part

### 2.1 Estimation

该部分本质上依旧是不包含交互项的多分类回归问题, 在类别内部有

$$Y_i = X_i^T \beta_k + Z_i^T \alpha_k, \text{ if } i \in \text{Class } k \quad (2.1)$$

其中  $\beta_k \in \mathbb{R}^p$ ,  $\alpha_k \in \mathbb{R}^q$ , 若记  $\tilde{X}_i^T = (X_i^T, Z_i^T)$ ,  $\theta_k^T = (\beta_k^T, \alpha_k^T)$  可以将模型改写为  $Y_i = \tilde{X}_i^T \theta_k$ , 此时完全退化为最简单线性设定, 本部分的重点在于验证两部分依次各自更新迭代公式的正确性。

对于类别  $k$ ，可以对不同参数写出优化目标，计算时依次迭代即可。

**Estimate  $\beta_k$**

$$\begin{aligned} \min \quad l_k &= \sum_{i=1}^n q_{C_i}(k) (y_i - X_i^T \beta_k - Z_i^T \alpha_k) \\ &= (y^{\beta_k} - X \beta_k)^T W_k (y^{\beta_k} - X \beta_k) \end{aligned} \quad (2.2)$$

其中  $W_k = \text{diag}(q_{C_1}(k), \dots, q_{C_n}(k))$ ，则  $\beta_k$  的估计值为

$$\hat{\beta}_k = (X^T W_k X)^{-1} X^T W_k y^{\beta_k} \quad (2.3)$$

**Estimate  $\alpha_k$**

$$\begin{aligned} \min \quad l_k &= \sum_{i=1}^n q_{C_i}(k) (y_i - X_i^T \beta_k - Z_i^T \alpha_k) \\ &= (y^{\alpha_k} - Z \alpha_k)^T W_k (y^{\alpha_k} - Z \alpha_k) \end{aligned} \quad (2.4)$$

$\alpha_k$  的估计值为

$$\hat{\alpha}_k = (Z^T W_k Z)^{-1} Z^T W_k y^{\alpha_k} \quad (2.5)$$

值得注意的是，将两部分参数当做一体的方法更新公式 2.6 与分别更新  $\beta_k, \alpha_k$  的公式 2.7 并不等价。公式 2.6 中  $\beta_k$  的更新不依赖于当前  $\alpha_k$  的取值，反之亦反；而公式 2.7 中  $y^{\beta_k} = y - Z \alpha_k, y^{\alpha_k} = y - X \beta_k$ ， $\beta_k$  与  $\alpha_k$  的更新相互依赖。直观理解上来看，二者计算目标不同，推导方法参照分块矩阵求逆。

$$\begin{pmatrix} \hat{\beta}_k \\ \hat{\alpha}_k \end{pmatrix} = ((X \ Z)^T W_k (X \ Z))^{-1} (X \ Z)^T W_k y \quad (2.6)$$

$$\begin{cases} \hat{\beta}_k = (X^T W_k X)^{-1} X^T W_k y^{\beta_k} \\ \hat{\alpha}_k = (Z^T W_k Z)^{-1} Z^T W_k y^{\alpha_k} \end{cases} \quad (2.7)$$

## 2.2 Numerical Setting

令  $p = 4, q = 3$  四个类别真实参数设置见表 2



表 2: 两部分参数设定

	Comp.1	Comp.2	Comp.3	Comp.4
$\beta_{k1}$	-2	-1	1	2
$\beta_{k2}$	-2	-1	1	2
$\beta_{k3}$	0	0	0	0
$\beta_{k4}$	0	0	0	0
$\alpha_{k1}$	-2	-1	1	2
$\alpha_{k2}$	-2	-1	1	2
$\alpha_{k3}$	0	0	0	0

## 2.3 Results

如果根据  $Y_i = \tilde{X}_i^T \theta_k$  使用最简情况方法进行求解，可收敛，但初始值的设置需要比较小心。当初始值为一组来自  $N(0, 0.1)$  的随机数时候依旧很快（27轮左右）收敛到真值，但若抽样正态分布方差较大，会出现参数估计不再变化但未收敛到真值的情况。

## 3 Simple Set3-Interaction

### 3.1 Estimation

以上为有分组的简单回归情况，现考虑原模型的协变量构成，即： $X, Z, W$ （对应回归参数为  $\beta, \alpha, \eta$ ，其中  $\eta$  又由  $\beta, \gamma$  计算得到，真实需要更新的参数为  $\beta, \alpha, \gamma$ ），其中  $W$  为交互项， $Z$  为环境相关自变量， $X$  为基因相关自变量。考虑加入交互项，目标函数  $\mu_{ik}$  的成分发生变化，影响和  $\beta, \gamma$  相关的求解。不影响  $\alpha$  的求解。

与式 1.5 形式类似有

$$\begin{aligned}
 l &= \sum_{i=1}^n \sum_{k=1}^K q_{C_i}(k) \left( C - \frac{\rho^2}{2} (y_i - \mu_{ik})^2 \right) \\
 &= \sum_{i=1}^n \sum_{k=1}^K q_{C_i}(k) \left( C - \frac{\rho^2}{2} (y_i - X_i^T \beta_k - Z_i^T \alpha_k - W_i^T \eta_k)^2 \right)
 \end{aligned} \tag{3.1}$$

为了方便从交互项中按照不同需要拆分出系数，验证以下几种表达形式的等价（数值验证）

$$\begin{aligned}
W_i^T \eta_i &= W_i^T (I_q \otimes \beta_k \odot \gamma_k) \\
&= \sum_{s=1}^q W_i^{(s)T} \beta_k \odot \gamma_{ks} \\
&= \left( \sum_{s=1}^q W_i^{(s)} \odot \gamma_{ks} \right)^T \beta_k \\
&= (W_i \odot (I_q \otimes \beta_k))^T \gamma_k
\end{aligned} \tag{3.2}$$

对于类别  $k$ ，可以对不同参数写出优化目标

**Estimate  $\alpha_k$**

$$\begin{aligned}
\min \quad l_k &= \sum_{i=1}^n q_{C_i}(k) (y_i - X_i^T \beta_k - Z_i^T \alpha_k - W_i^T \eta_k) \\
&= (y^{\alpha_k} - Z \alpha_k)^T W_k (y^{\alpha_k} - Z \alpha_k)
\end{aligned} \tag{3.3}$$

其中  $y_i^{\alpha_k} = y_i - X_i^T \beta_k - W_i^T \eta_k$ ,  $y^{\alpha_k} = (y_1^{\alpha_k}, \dots, y_n^{\alpha_k})^T$ ,  $W_k = \text{diag}(q_{C_1}(k), \dots, q_{C_n}(k))$ , 则  $\alpha_k$  的估计值为

$$\hat{\alpha}_k = (Z^T W_k Z)^{-1} Z^T W_k y^{\alpha_k} \tag{3.4}$$

**Estimate  $\beta_k$**

$$\begin{aligned}
\min \quad l_k &= \sum_{i=1}^n q_{C_i}(k) (y_i - X_i^T \beta_k - Z_i^T \alpha_k - W_i^T \eta_k) \\
&= (y^{\beta_k} - X^{\beta_k} \beta_k)^T W (y^{\beta_k} - X^{\beta_k} \beta_k)
\end{aligned} \tag{3.5}$$

其中  $X^{\beta_k} = (X_1^{\beta_k}, \dots, X_n^{\beta_k})^T \in \mathbb{R}^{n \times p}$ ,  $X_i^{\beta_k} = X_i + (\sum_{s=1}^q W_i^{(s)} \odot \gamma_{ks}) \in \mathbb{R}^{p \times 1}$ ,  $y_i^{\beta_k} = y_i - Z_i^T \alpha_k$ ,  $y_i^{\beta_k} = (y_1^{\beta_k}, \dots, y_n^{\beta_k})^T$ ,  $W_k = \text{diag}(q_{C_1}(k), \dots, q_{C_n}(k))$ , 则  $\beta_k$  的估计值为

$$\hat{\beta}_k = (X^{\beta_k T} W_k X^{\beta_k})^{-1} X^{\beta_k T} W_k y^{\beta_k} \tag{3.6}$$

**Estimate  $\gamma_k$**

$$\begin{aligned}
\min \quad l_k &= \sum_{i=1}^n q_{C_i}(k) (y_i - X_i^T \beta_k - Z_i^T \alpha_k - W_i^T \eta_k) \\
&= (y^{\gamma_k} - X^{\gamma_k} \gamma_k)^T W (y^{\gamma_k} - X^{\gamma_k} \gamma_k)
\end{aligned} \tag{3.7}$$

其中  $X^{\gamma_k} = (X_1^{\gamma_k}, \dots, X_n^{\gamma_k})^T \in \mathbb{R}^{n \times pq}$ ,  $X_i^{\gamma_k} = W_i \odot (I_q \otimes \beta_k) \in \mathbb{R}^{pq \times 1}$ ,  $y_i^{\gamma_k} = y_i - X_i^T \beta_k - Z_i^T \alpha_k$ ,  $y_i^{\gamma_k} = (y_1^{\gamma_k}, \dots, y_n^{\gamma_k})^T$ ,  $W_k = \text{diag}(q_{C_1}(k), \dots, q_{C_n}(k))$ , 则  $\gamma_k$  的估计值为

$$\hat{\gamma}_k = (X^{\gamma_k T} W_k X^{\gamma_k})^{-1} X^{\gamma_k T} W_k y^{\gamma_k} \quad (3.8)$$

$X^{\beta_k}$  依赖于当前  $\gamma_k$  估计值,  $X^{\gamma_k}$  也依赖于当前  $\beta_k$  的估计值。有交互项的情况不再可以直接化为最简单的线性回归情况, 即不可以一次性算出一轮参数的更新。综上, 更新总结如式 3.9

$$\begin{cases} \hat{\beta}_k = (X^{\beta_k T} W_k X)^{-1} X^{\beta_k T} W_k y^{\beta_k} \\ \hat{\alpha}_k = (Z^T W_k Z)^{-1} Z^T W_k y^{\alpha_k} \\ \hat{\gamma}_k = (X^{\gamma_k T} W_k X)^{-1} X^{\gamma_k T} W_k y^{\gamma_k} \end{cases} \quad (3.9)$$

### 3.2 Numerical Setting

1. 主要实验设置:  $p = 4, q = 3, n = 200, K = 4$ , 真实参数设置如表 3.3
2. 辅助实验设置:  $p = 2, q = 2, n = 200, K = 4$ , 真实参数设置如表 3.3.

### 3.3 Package Results

注意在有交互项的模型中, 包得到的结果和上述更新不同。形式

$$\begin{aligned} y_i &= \sum_{k=1}^K q_{C_i}(k) (X_i^T \beta_k + Z_i^T \alpha_k + W_i^T \eta_k) \\ &= \sum_{k=1}^K q_{C_i}(k) (X_i^T \beta_k + Z_i^T \alpha_k + (W_i \odot (I_q \otimes \beta_k))^T \gamma_k) \end{aligned} \quad (3.10)$$

加权最小二乘法得到的是  $\beta_k, \alpha_k, \gamma_k$  的更新,  $\eta_k$  是根据其他参数计算出来的。而直接用包得到的参数为  $\beta_k, \alpha_k, \eta_k$ ,  $\gamma_k$  需要再计算。当然也可以重新定义  $\gamma_k$  相关的自变量为  $W_i \odot (I_q \otimes \beta_k)$ , 但该表达式包含  $\beta_k$ 。

*fmrs*:  $p = 2, q = 2$  效果不好

*flexmix*:  $p = 2, q = 2$  计算正确

直接计算流程类似于算法 1, 固定其他参数依次更新  $\beta_k, \alpha_k, \gamma_k$  即可。

### 3.4 Iteration Results

总结实验经验结果如下:

	Comp.1	Comp.2	Comp.3	Comp.4
$\beta_{k1}$	2	2	-2	-2
$\beta_{k2}$	2	2	-2	-2
$\alpha_{k1}$	2	2	-2	-2
$\alpha_{k2}$	2	2	-2	-2
$\gamma_{k1}$	3	1	-1	-3
$\gamma_{k2}$	3	1	-1	-3
$\gamma_{k3}$	3	1	-1	-3
$\gamma_{k4}$	3	1	-1	-3

1. 迭代过程中需要实时使用最新结果，而不是一轮迭代中统一使用上一轮结果，这样可以避免参数距离真值的距离来回震荡，形成两三条曲线；
2. 矩阵求逆出现特征值过小情况应加上一个小数乘以对角阵，小数取 0.1 已经足够大，否则会影响参数估计；
3. 更新单个参数或更新一对参数检验时，定位问题所在是  $\beta_k$  和  $\gamma_k$  的交互项；
4. 原始问题非凸，依赖初始值，固定初始化方法相比随机初始化较稳定。

### 3.4.1 One Class

模型不收敛有两个的可能：（1）参数迭代估计过程中出错；（2）后验概率相关。为逐一排除错误，首先去除后验概率的影响，先对单类别进行实验。取不收敛异常样例（seed 11, K4）进行分析，选取固定初始化方法和 7 组随机初始化，得到收敛曲线及使用对应参数进行回归估计得到残差分布图与 MSE 如图 2。

随机设置 200 组（199 组  $N(0, 0.01)$  初始值+1组固定初始值），得到的参数距离真值的距离总结如图 3。

当前做法为，在迭代开始前得到 200 组初始值，依次代入循环，若某一次使用其结果进行回归得到的 MSE 小于临界值（暂时设置为 1），则停止尝试其他初值，认为已经得到比较好的结果；如果试遍所有 200 个初始值依旧没有得到足够小的 MSE，则选取相对较好的结果，此时结果依旧不理想，应回查，现有实验中尚未遇到这种情况，最多 72 次初始值尝试后就得到理想 MSE。

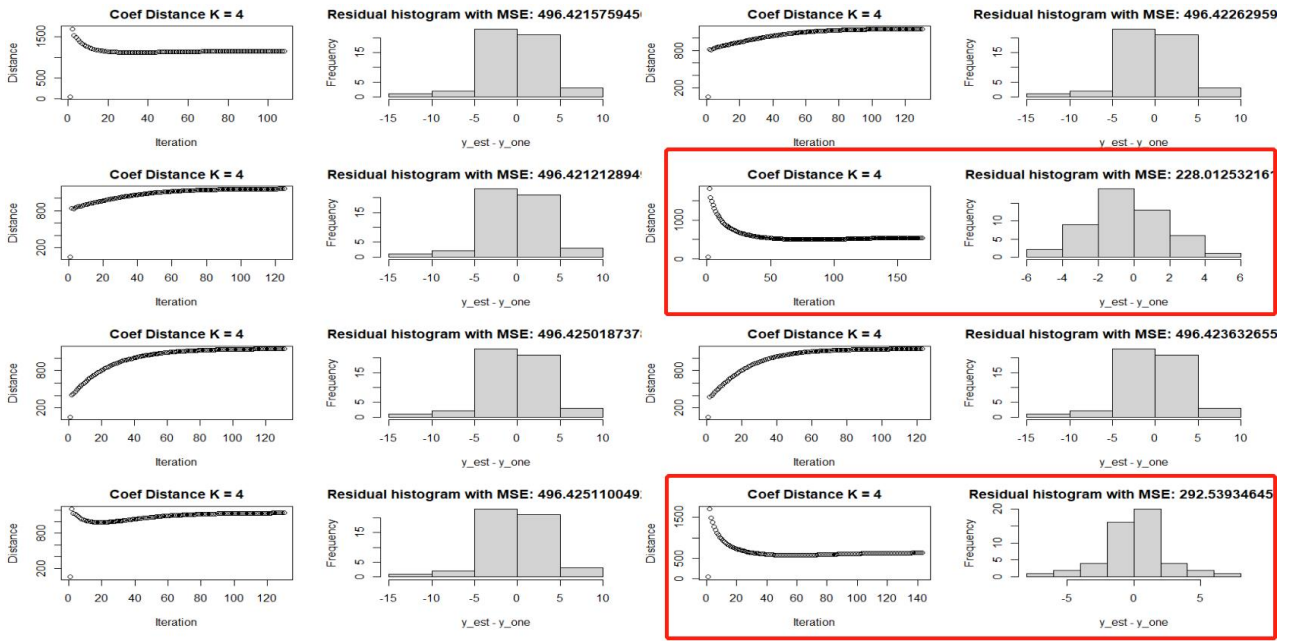


图 2: 多组初始化下, 参数到真值距离随迭代轮数的变化 (seed 11,K4)

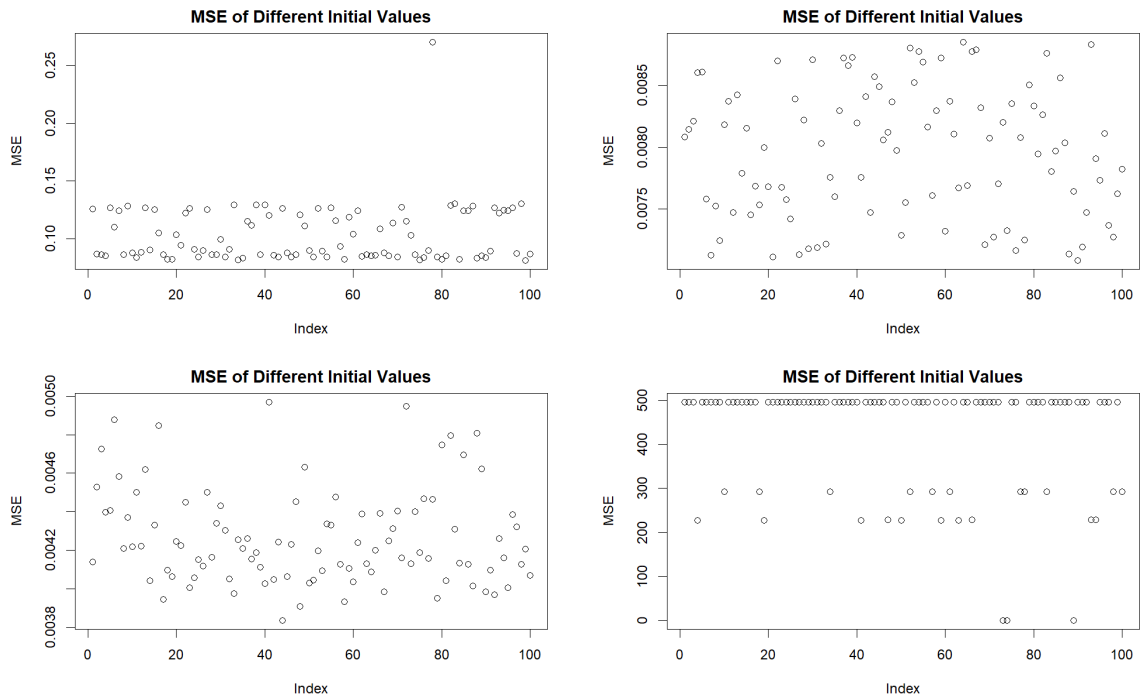


图 3: 200 组初始化下最终得到的 MSE (seed 11,K 1,2,3,4, 注意纵坐标尺度差异)

	Comp.1	Comp.2	Comp.3	Comp.4
$\beta_{k1}$	2	2	-2	-2
$\beta_{k2}$	2	2	-2	-2
$\beta_{k3}$	0	0	0	0
$\beta_{k4}$	0	0	0	0
$\alpha_{k1}$	2	2	-2	-2
$\alpha_{k2}$	2	2	-2	-2
$\alpha_{k3}$	0	0	0	0
$\gamma_{k1}$	3	1	-1	-3
$\gamma_{k2}$	3	1	-1	-3
$\gamma_{k3}$	3	1	-1	-3
$\gamma_{k4}$	3	1	-1	-3
$\gamma_{k5}$	0	0	0	0
$\gamma_{k6}$	0	0	0	0
$\gamma_{k7}$	0	0	0	0
$\gamma_{k8}$	0	0	0	0
$\gamma_{k9}$	0	0	0	0
$\gamma_{k10}$	0	0	0	0
$\gamma_{k11}$	0	0	0	0
$\gamma_{k12}$	0	0	0	0

### 3.4.2 Multiple Class-Fixed $q_{C_i}$

多组初始值的设置行之有效，将单类实验嵌入多类别框架，即恢复样本属于不同类别的设定，但已知其真值分组，此时多分类实验等价于同时进行多组单类实验，各个类别的参数互不影响，但只有当每个类别都正常收敛时才会认为一次多类实验成功。结果展示如图 4,5,7。

此处实验样例均为之前单分类至少有一个分类异常的数据，seed 11 生成的数据较难估计参数，9 组初始值不足以估计出真实结果，尝试 100 组初始值，使用对应参数估计得到的回归模型 MSE 如图 7。

值得注意的是，将生成的所有初始值都尝试一遍后再选择最优结果效率低但比较稳妥；

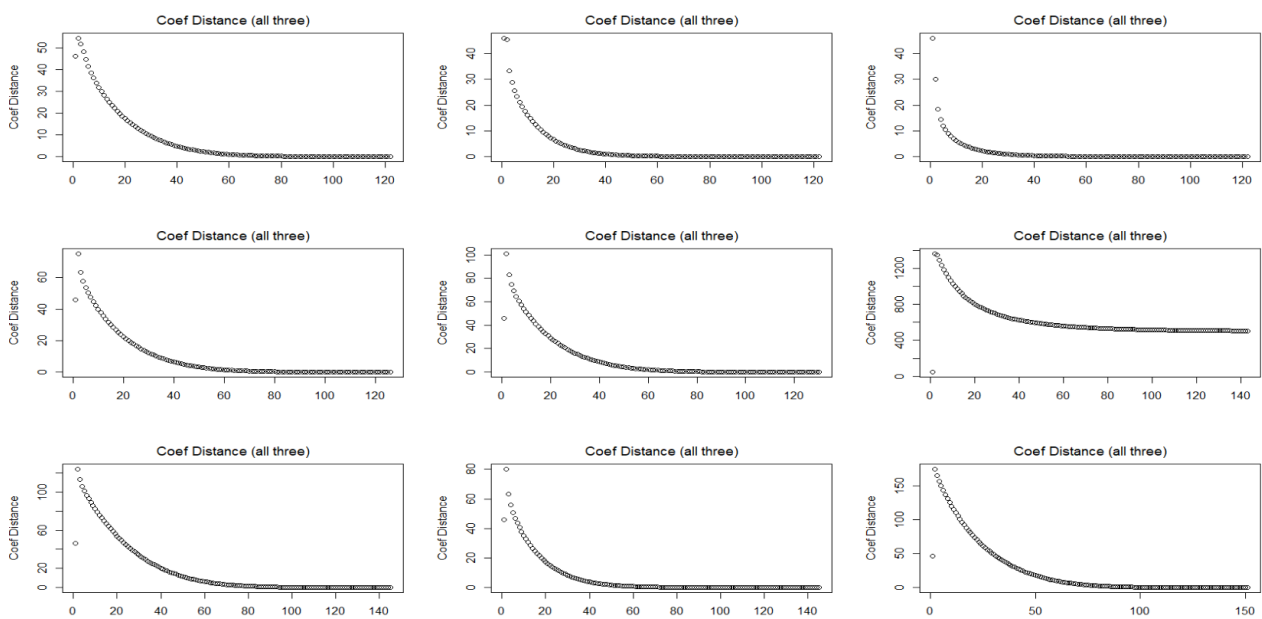


图 4: 多组初始化下, 参数到真值距离随迭代轮数的变化 (seed 111)

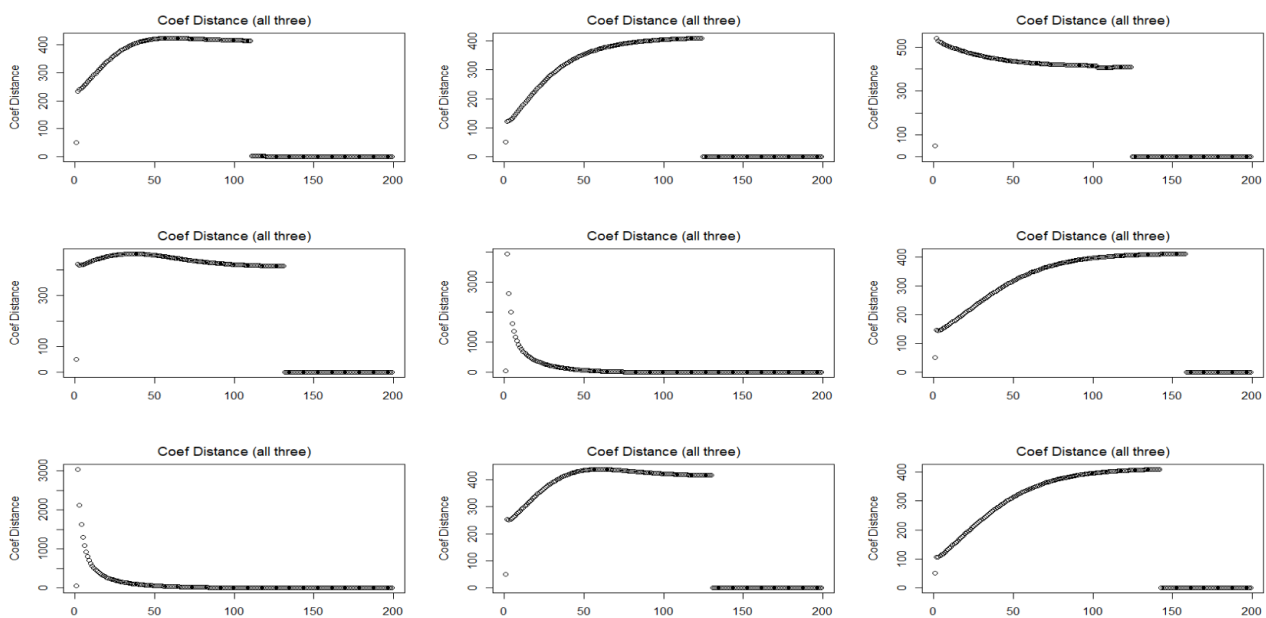


图 5: 多组初始化下, 参数到真值距离随迭代轮数的变化 (seed 14)

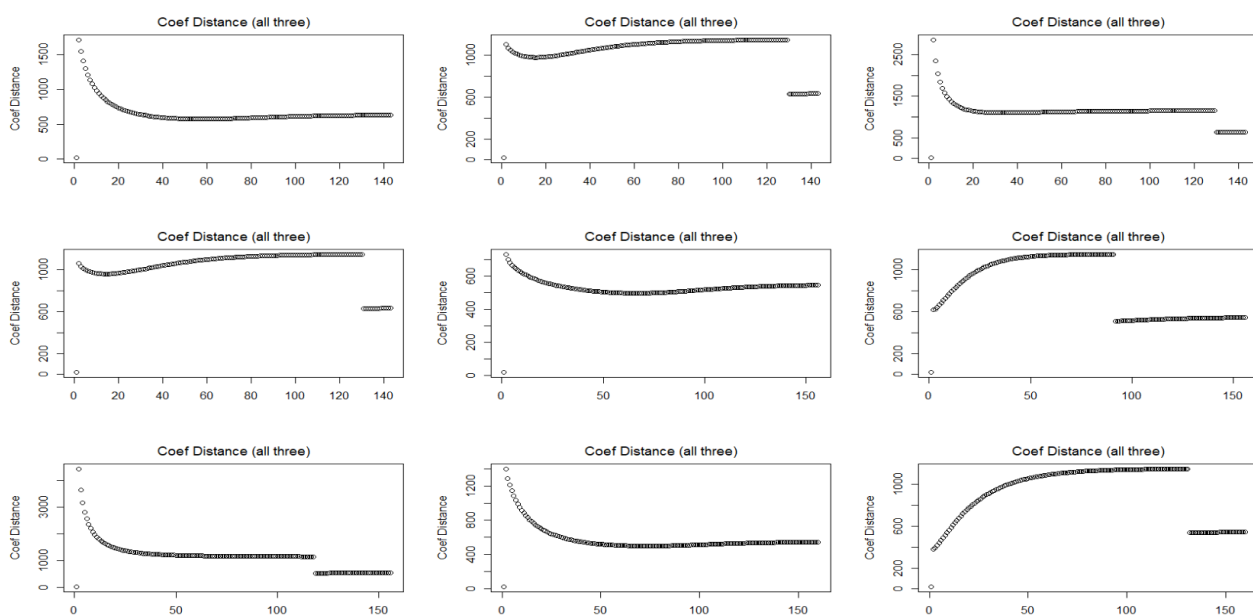


图 6: 多组初始化下, 参数到真值距离随迭代轮数的变化 (seed 11)

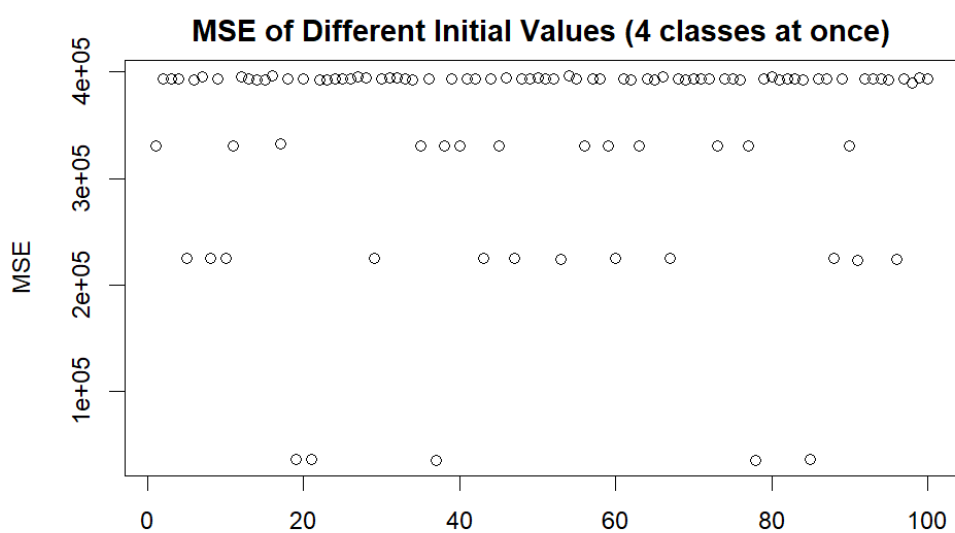


图 7: 100 组初始化下最终得到的 MSE (seed 11, 所有分类)



MSE 小于临界值的方法不容易找到合适的临界值，比如在图 7 的例子中，最小的 MSE 是 35883.49，对应的参数估计结果为表 3（真值见表 3.3） MSE 数值虽大但收敛到几乎为真值的地方，临界值和数据的尺度有关。

表 3: seed 11 数据在 100 组初始值中的最佳收敛结果

	Class 1	Class 2	Class 3	Class 4
$\beta_{k1}$	2.051	1.997	-1.999	-2.010
$\beta_{k2}$	1.980	1.998	-1.996	-2.015
$\alpha_{k1}$	1.985	1.996	-1.996	-1.998
$\alpha_{k2}$	1.994	1.994	-1.994	-1.999
$\gamma_{k1}$	2.911	1.002	-1.000	-2.983
$\gamma_{k2}$	3.033	1.000	-0.999	-2.976
$\gamma_{k3}$	2.931	1.002	-1.000	-2.985
$\gamma_{k4}$	3.026	0.999	-1.000	-2.978

### 3.4.3 Multiple Class-Unknown $q_{C_i}$

进一步，取消已知  $q_{C_i}$  的限制，将各个样本的后验概率估计也加入迭代过程

## 4 Test Summary

1. 初始化参数而非后验概率，EM 算法顺序不可倒置否则无法与原包结果进行对比
2.  $\rho$  的更新方式，按照类别分和不同类别取平均作每个类别  $1/\sigma$  效果没有明显差别，为了防止某类异常影响整体效果当前各类别有自己的方差。考虑到及时在最终应用场景中，目标类别也不会太多，这样的设定应该不会增加很大计算复杂度，若有需要再进行调整

## References

- [1] Bettina Grün and Friedrich Leisch. Fitting finite mixtures of linear regression models with varying & fixed effects in r. 01 2006.
- [2] Mengyun Wu, Qingzhao Zhang, and Shuangge Ma. Structured gene-environment interaction analysis. *Biometrics*, 76(1):23–35, 2020.