arXiv:2601.05441v1 [stat.ML] 9 Jan 2026

简短说明

全局视角学习问题

GETACHEW K. BEFEKADU*

摘要。本简短说明探讨了动态优化委托代理设定下的学习问题。在该设定中，允许智能体对学习过程持有全局视角，即根据某些由主方共享的聚合信息，按照事物的相对重要性或其真实关系来观察事物的能力。而主方在聚合过程中对智能体的学习过程施加影响，其主要任务是解决一个高层优化问题：在条件矩约束模型下，将其表述为经验似然估计量；该模型还纳入了有关智能体样本外预测性能的信息，以及一组仅供主方使用的私有数据集。具体而言，我们提出了一个连贯的数学论证，这是刻画这一抽象委托代理学习框架背后学习过程所必需的，尽管我们承认，仍有一些概念和理论问题有待解决。

关键词。聚合算法、经验似然估计量、专家建议、梯度系统、学习问题、非线性函数建模、矩约束模型、委托代理、点估计量。

1. 引言。在这篇简短的说明中，我们探讨动态优化委托代理情境下的学习问题。在这种情境中，智能体能够从全局视角看待学习过程，即基于主方共享的某些聚合信息，按照事物的相对重要性或真实关系来审视问题。而主方在聚合过程中对智能体的学习过程施加影响，其主要任务是解决一个高层优化问题，该问题被表述为条件矩约束模型下的经验似然估计量，同时还考虑了智能体的样本外预测性能信息，以及仅主方可用的一组私有数据集（例如，参见 [1], [2], [3], [4] 和 [5]，了解有关矩约束下经验似然方法的进一步讨论）。在此，我们给出一个连贯的数学论证，以刻画这一抽象的动态优化委托代理学习框架背后的学习过程。值得注意的是，由于智能体之间固有的反馈行为，尽管主方和智能体都不一定了解样本分布或彼此数据集的质量，所提出的学习框架仍在稳定性和一致性方面具有显著优势。最后值得指出的是，这种学习框架能够为利用委托代理情境、具有全局视角的协作学习问题带来新的见解（例如，参见 [6], [7], [8] 或[9] 中的相关讨论），不过我们也承认，诸如小样本性质等若干概念和理论问题仍有待解决。

本简短说明其余部分安排如下。第2节中，我们给出形式化问题陈述，并介绍一种利用动态优化委托代理设定、具有全局视角的学习框架。第3节中，我们提出一套连贯的数学论证，这是刻画这一抽象学习框架背后学习过程所必需的。本节还介绍一种通用算法，为实现这类具有全局视角的学习问题提供可行方法。

*GETACHEW K. BEFEKADU 任职于美国天主教大学工程、物理与计算学院电气与计算机工程系，地址：美国华盛顿特区 20064。电子邮箱：BEFEKADU@CUA.EDU

2. 问题设定。 本节给出形式化问题陈述，以概述我们具有全局视角的抽象学习框架。具体而言，所提出的学习框架利用动态优化委托代理设定，包含以下核心理念：(i) 数据集： 共有 k 个数据集，即 $D^{(i)} = \{(x_{j_i}, y_{j_i})\}_{j_i=1}^{n_i}$，其中 $i = 1$、2、.....、 k，每个数据集的大小为 n_i，且 $\sum_{i=1}^{k} n_i = n$。这些数据集可以通过有放回或无放回自助抽样，从给定的原始数据集中生成，例如用于处理大型数据集的分而治之范式；也可以仅代表与高维非线性函数建模中的点估计问题相关的分布式／去中心化观测。(ii)智能体： 共有 k个智能体，编号为智能体-1、智能体-2、.....、智能体-k。它们的主要任务是寻找一组参数估计 $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$，即从有限维参数空间中选取参数，使函数 $h(x;\theta) \in \mathcal{H}$（即来自给定的假设函数空间类别 $\mathcal{H}$）能够最佳地描述相应的训练数据集 $D^{(i)}$，其中 $i = 1$、2、.....、k。此外，我们假设每个智能体都在某种全局视角下，利用梯度动力系统的离散时间近似来更新其参数估计，但更新过程受其对应训练数据集的引导。(iii) 主方： 主方对聚合中智能体的学习过程施加影响，其任务是求解一个高层优化问题，该问题被设定为条件矩约束模型下的经验似然估计量，既考虑智能体在样本外的预测性能信息，也考虑仅主方可用的一组私有数据集。此外，我们假设主方在每个迭代时刻使用一份单独的数据 (x_{n+1}, y_{n+1})，评估智能体在样本外的预测性能。条件矩约束则针对一组私有数据集 $\{s_1, s_2, \ldots, s_k\}$进行，其支撑集为 $\mathbb{R}^{d_s}$；这组数据集仅主方可用，但能为智能体提供有关学习过程的全局视角，即根据事物的相对重要性或其真实关系来看待问题。图2.1展示了所提出的具有全局视角的抽象学习框架的函数结构。在动态优化委托代理设定下，该框架会利用各种信息，例如仅智能体可用、作为其学习过程或任务一部分的一组局部数据集；用于评估智能体样本外预测性能的单独数据集；以及仅主方可用、用于对学习过程施加影响的一组私有数据集。

从数学构造的角度来看，寻找一组最优参数估计 $\theta^* \in \Theta$，可以通过聚合一组稳态解来实现。这些稳态解对应于以下一族梯度动力系统，其时间演化 由相应的训练数据集 $D^{(i)}, i = 1, 2, \ldots, k$引导，即：

$$\dot{\theta}^{(i)}(t) = -\nabla_\theta J^{(i)}\big(\theta^{(i)}(t), D^{(i)}\big), \quad \theta^{(i)}(0) = \theta_0, \quad \left(\text{with } \dot{\theta}(t) \triangleq \frac{d}{dt}\theta(t)\right), \tag{2.1}$$

其中，项 $J^{(i)}(\theta^{(i)}, D^{(i)}) = \frac{1}{n_i}\sum_{j_i=1}^{n_i} \ell\big(y_{j_i}, h\big(x_{j_i};\, \theta^{(i)}\big)\big)$，并采用适当的损失函数 ℓ，以量化所研究的假设非线性泛函模型 $h(\cdot\,;\theta) \in \mathcal{H}$与数据集 $D^{(i)}$之间的拟合程度，其中 $i = 1, 2, \ldots, k$。需要注意的是，这种方法存在一系列缺陷，因为解可能会陷入局部最小值，而非全局最小值解（例如，见 [10]）。解决这一困难的一种方法，是将上述梯度动力系统族嵌入一个通用的抽象学习框架中；该框架能够促进具有全局视角的学习过程，并支持具有内在反馈行为的聚合以及智能体之间的协作。这正是本短文关注的重点。

一般假设： 在本文中，我们假设对于每个 h

$i \in \{1, 2, \ldots, k\}$，是一致利普希茨连续的，并且进一步满足以下增长条件

$$\left|\nabla_\theta J^{(i)}(\theta, D^{(i)})\right|^2 \leq L_{\text{Lip}}\left(1 + |\theta|^2\right), \quad \forall\, \theta \in \Theta, \tag{2.2}$$

对于某个常数 $L_{\text{Lip}} > 0$。

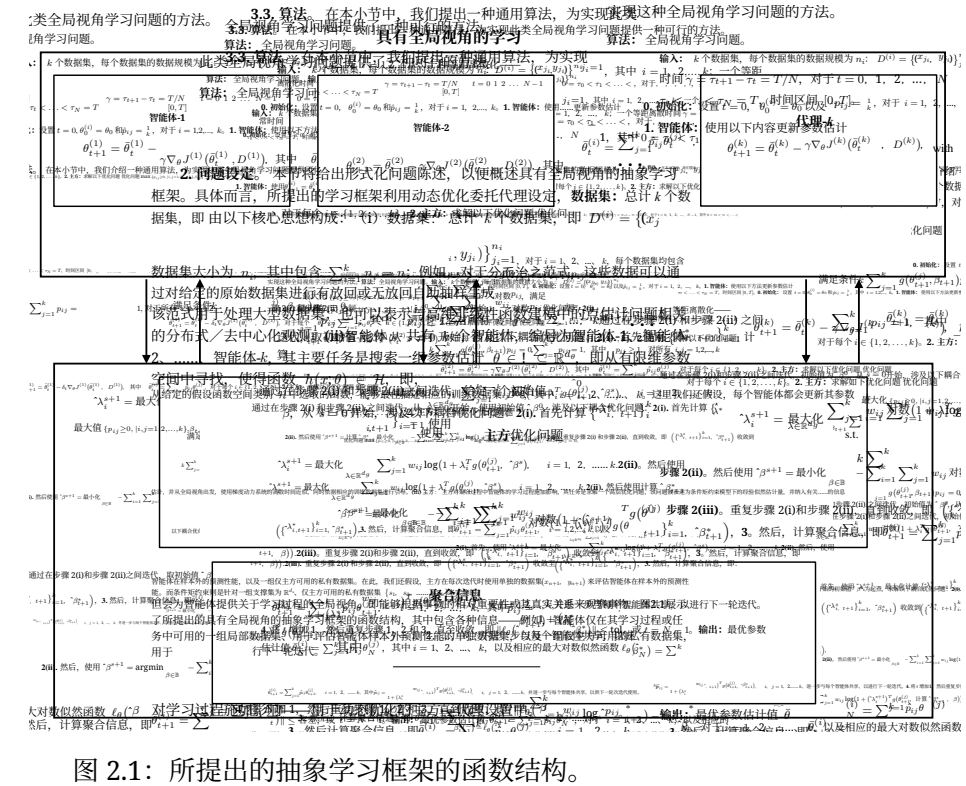

图 2.1：所提出的抽象学习框架的函数结构。

3. 主要结果。在本节中，我们介绍主要结果，并提出连贯的数学论证，以刻画这一动态优化委托代理设定下抽象学习框架背后的学习过程。

3.1. 智能体的参数更新模型。下文中，我们允许每个智能体使用方程 (2.1) 中梯度系统的离散时间近似来更新其参数估计，该系统的时间演化由其各自的训练数据集 $D^{(i)}$ 引导，其中 $i \in \{1, 2, \ldots, k\}$，即：

$$\theta_{t+1}^{(i)} = \bar{\theta}_t^{(i)} - \gamma_t^{(i)} \nabla_\theta J^{(i)}\big(\bar{\theta}_t^{(i)}, D^{(i)}\big), \quad \text{with} \quad \bar{\theta}_t^{(i)} = \sum\nolimits_{j=1}^{k} p_{ij} \theta_t^{(j)}, \tag{3.1}$$

其中 – $\sum_{j=1}^{k} p_{ij} = 1$，其中 $p_{ij} \geq 0$，对于 $i,\ j \in \{1, 2, \ldots, k\}$ 是概率质量信息（用于汇总智能体的参数估计），由主方在每个迭代时刻与智能体共享，–

$\bar{\theta}_t^{(i)} = \sum_{j=1}^{k} \theta_t^{(j)}$，对于 $i = 1$、2、.....、k，是智能体在其参数更新模型中使用的聚合信息，并且

– $\gamma_t^{(i)}$ 是离散化时间常数，即，对于 $t=0$、1、2、.....、$N-1$以及 $i=1$、2、.....、k，其中 $0=\tau_0<\tau_1<\ldots<\tau_t<\ldots<\tau_N=T$，时间区间为 $[0,T]$。

请注意，支撑集位于参数空间上的概率质量信息，以及一组提供给主方的私有数据集，都是主方所负责的高层优化问题的解。

3.2. 委托方的优化问题。本小节假设主方和智能体遵循双向信息共享协议，即规定双方如何共享各自信息的流程。假设在每个迭代时刻结束时，所有智能体同时向主方报告当前更新后的参数估计值。那么，主方将求解一个高层优化问题，该问题被表述为条件矩约束模型下的经验似然估计量，同时还考虑智能体在样本外的预测性能信息，以及仅主方掌握的一组私有信息。具体而言，假设我们有 k一组独立参数估计值 $(\theta^{(1)},\ \theta^{(2)},\ .....,\ \theta^{(k)})$，其支撑集为$\Theta\subseteq\mathbb{R}^{d_\theta}$，接着我们通过条件矩约束模型来确定/估计一组未知参数$\beta=(\mu,\ \sigma^2)\in\mathbb{B}\subseteq\mathbb{R}\times\mathbb{R}_+$

$$E\left[g\big((x_{n+1},y_{n+1}\big),\theta,\beta\big)\,\middle|\,s\right]=\begin{bmatrix}E\left[\big(y_{n+1}-h\big(x_{n+1};\theta\big)-\mu\big)\,\middle|\,s\right]\\ E\left[\big(\big(y_{n+1}-h\big(x_{n+1};\theta\big)-\mu\big)^2-\sigma^2\big)\,\middle|\,s\right]\end{bmatrix}=0,\quad(3.2)$$

其中，可测函数 g 定义为

$$g\big(\big(x_{n+1},y_{n+1}\big),\theta,\beta\big)=\begin{bmatrix}\big(y_{n+1}-h\big(x_{n+1};\theta\big)-\mu\big)\\ \big(\big(y_{n+1}-h\big(x_{n+1};\theta\big)-\mu\big)^2-\sigma^2\big)\end{bmatrix}\in\mathbb{R}^{d_g},$$

而量 s（其支撑集为 $\mathbb{R}^{d_s}$）假定包含能够为智能体提供全局视角的信息。注意，可测函数 g: $(x,y)\times\Theta\times\mathbb{B}\to\mathbb{R}^{d_g}$(其中$\mathbb{R}^{d_g}\ \equiv\ \mathbb{R}^2$)提供了关于智能体参数估计值 θ的预测误差信息。

接下来，对于每个迭代时刻 $t\ =\ 0,\ 1,\ .....,\ \ N-1,$ 令 p_{ij}对每个 $i,j\ \in\ \{1,2,\ldots,k\}$是由一个支撑集为 $\{\theta_{t+1}^{(1)},\theta_{t+1}^{(2)},\ldots,\theta_{t+1}^{(k)}\}\times\{s_1,\ s_2,\ldots,s_k\}$的离散概率分布在 $\big(\theta_{t+1}^{(i)},s_j\big)$ 处赋予的概率质量。此外，令 w_{ij} 为以下权重:

$$w_{ij}=\frac{\psi\big(\frac{s_i-s_j}{h_{bw}}\big)}{\sum_{j=1}^{k}\psi\big(\frac{s_i-s_j}{h_{bw}}\big)},\quad i,j=1,2,\ldots,k,\qquad(3.3)$$

对于适当选择的核函数 ψ和带宽 h_{bw}。此外，如果我们定义如下平滑对数似然函数 $\ell_\theta\big(\beta_{t+1}\big)$,

$$\ell_\theta\big(\beta_{t+1}\big)=\sum\nolimits_{i=1}^{k}\sum\nolimits_{j=1}^{k}w_{ij}\log p_{ij}.\qquad(3.4)$$

那么，我们得到如下优化问题

$$\max_{\{p_{ij}\geq 0,\,|i,j=1,2,\ldots,k\},\beta_{t+1}} \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}\log p_{ij} \tag{3.5}$$

s.t.

$$k\sum_{j=1}^{k} g\big(\theta_{t+1}^{(j)},\beta_{t+1}\big)p_{ij}=0,$$

$$\sum_{j=1}^{k} p_{ij}=1,\quad \text{with}\quad p_{ij}\geq 1,\ \text{for all } i,j=1,2,\ldots,k,$$

需要主方在每个迭代时刻 $t=0$、1、.....、$N-1$ 求解。请注意，方程 3.5 中的上述优化问题可利用拉格朗日乘子（即基于 KKT 最优性条件（例如，参见 [11]））方便地求解如下

$$\mathscr{L}\big(\beta_{t+1}\big)=\sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}\log p_{ij}-\sum_{i=1}^{k}\eta_i\left(\sum_{j=1}^{k}p_{ij}-1\right)$$
$$-k\sum_{i=1}^{k}\lambda_i^T\sum_{j=1}^{k} g\big(\theta_{t+1}^{(j)},\beta_{t+1}\big)p_{ij}, \tag{3.6}$$

其中，对于 $i=1$、2、.....、k，$\lambda_i\in\mathbb{R}^{d_g}$ 和 $\eta_i\in\mathbb{R}$ 是与方程 3.5 中约束条件对应的拉格朗日乘子。于是，容易看出解为

$$\hat{p}_{ij}=\frac{w_{ij}}{1+\lambda_i^T g\big(\theta_{t+1}^{(j)},\beta_{t+1}\big)},\quad i,j=1,2,\ldots,k, \tag{3.7}$$

而对于每个 $\beta\in\mathbb{B}$，拉格朗日乘子 λ_i（$i=1$、2、.....、k）满足

$$\sum_{j=1}^{k}\frac{w_{ij}g\big(\theta_{t+1}^{(j)},\beta_{t+1}\big)}{1+\lambda_i^T g\big(\theta_{t+1}^{(j)},\beta_{t+1}\big)}=0. \tag{3.8}$$

请注意，对于初始值 $\hat{}\,\beta^0$，从 $s=0$ 开始，并对每个迭代时刻 $t=0$、1、.....、$N-1$，以下通用算法在以下两个相互耦合的优化步骤之间迭代

$$\hat{\lambda}_i^{s+1}=\arg\max_{\lambda\in\mathbb{R}^{d_g}}\sum_{j=1}^{k} w_{ij}\log\big(1+\lambda_i^T g\big(\theta_{t+1}^{(j)},\hat{\beta}^s\big),\quad i=1,2,\ldots,k \tag{3.9}$$

and

$$\hat{\beta}^{s+1}=\arg\min_{\beta\in\mathbb{B}}-\sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}\log\big(1+\big(\hat{\lambda}_i^{s+1}\big)^T g\big(\theta_{t+1}^{(j)},\beta\big)\big) \tag{3.10}$$

$\hat{}$ 可用于确定一组最优解 $\lambda_{i,\ t+1}^*$，适用于 $i=1$、2、.....、k 和 $\hat{}\,\beta_{t+1}^*$。请注意，在下一个迭代周期中，主方基于 $\big(\{\hat{}\,\lambda_{i,\ t+1}^*\}_{i=1}^{k},\ \hat{}\,\beta_{t+1}^*\big)$、$\{\theta_{t+1}^{(i)}\}_{i=1}^{k}$和 $\{w_{i,j}\}_{i}^{k},\ _{j=1}$，随后以某种聚合信息的形式与每个智能体共享信息，即:

$$\bar{\theta}_{t+1}^{(i)}=\sum_{j=1}^{k}\hat{p}_{ij}\theta_{t+1}^{(j)},\quad i=1,2,\ldots,k, \tag{3.11}$$

with

$$\hat{p}_{ij}=\frac{w_{ij}}{1+\big(\hat{\lambda}_{i,t+1}^*\big)^T g\big(\theta_{t+1}^{(j)},\hat{\beta}_{t+1}^*\big)},\quad i,j=1,2,\ldots,k.$$

最后，值得指出的是，当 $t = N-1$ 时，对数似然函数 $\ell_\theta\big(\hat{}\beta_N^*\big)$ 满足以下上界

$$\begin{aligned}\ell_\theta\big(\beta_N^*\big) &= \sum_{i=1}^k \sum_{j=1}^k w_{ij} \log \hat{p}_{ij} \\ &= \sum_{i=1}^k \sum_{j=1}^k \log\left(\frac{w_{ij}}{1+\big(\hat{\lambda}_{i,N}^*\big)^T g\big(\theta_N^{(j)}, \hat{\beta}_N^*\big)}\right) \\ &= \sum_{i=1}^k \sum_{j=1}^k \log w_{ij} - \sum_{i=1}^k \sum_{j=1}^k \log\Big(1+\big(\hat{\lambda}_{i,N}^*\big)^T g\big(\theta_N^{(j)}, \beta_N^*\big)\Big) \\ &\le -\sum_{i=1}^k \sum_{j=1}^k \log\Big(1+\big(\hat{\lambda}_{i,N}^*\big)^T g\big(\theta_N^{(j)}, \beta_N^*\big)\Big), \end{aligned} \tag{3.12}$$

由于 $w_{ij} \in (0,1]$，因为 $\sum_{i=1}^k \sum_{j=1}^k \log\ w_{ij} \le 0$。

3.3. 算法。在本小节中，我们提出一种通用算法，为实现此类全局视角学习问题提供可行的方法。

算法： 全局视角学习问题。

输入：(i) 智能体的数据集： k 个数据集，每个数据集的数据规模为 n_i, $D^{(i)} = \{(x_{j_i}, y_{j_i})\}_{j_i=1}^{n_i}$，其中 $i = 1,2,..,k$; **(ii) 等距离散时间：** $\gamma = \tau_{t+1} - \tau_t = T/N$，其中 $t = 0,1,2,\ldots,N-1$，且 $0 = \tau_0 < \tau_1 < \ldots < \tau_t < \ldots < \tau_N = T$，属于时间区间 $[0,T]$; 以及 **(iii) 主方的数据集：** **</style>一组私有数据集 和一项额外数据，**y_{n+1}。

0. 初始化： 设置 $t = 0$, $\theta_0^{(i)} = \theta_0$ 和 $\hat{}p_{ij} = 1/k$，其中 $i = 1,2,\ldots, k$。使用适当的核函数和带宽 h_{bw}，计算权重，其中$w_{ij} = \psi\big(\frac{s_i - s_j}{h_{bw}}\big) \Big/ \sum_{j=1}^k \psi\big(\frac{s_i - s_j}{h_{bw}}\big)$，对于 $i,\ j = 1,2,\ldots,k$。

1. 智能体： 使用以下方法更新参数估计

$$\theta_{t+1}^{(i)} = \bar{\theta}_t^{(i)} - \gamma \nabla_\theta J^{(i)}\big(\bar{\theta}_t^{(i)}, D^{(i)}\big), \quad \text{with} \quad \bar{\theta}_t^{(i)} = \sum_{j=1}^k \hat{p}_{ij} \theta_t^{(j)},$$

对于每个 $i \in \{1,2,\ldots,k\}$。

2. 主方： 求解以下优化问题

$$\begin{aligned} \max_{\{p_{ij}\ge 0,\, |i,j=1,2,\ldots,k\}, \beta_{t+1}} & \sum_{i=1}^k \sum_{j=1}^k w_{ij} \log p_{ij} \\ \text{s.t.} & \\ & k \sum_{j=1}^k g\big(\theta_{t+1}^{(j)}, \beta_{t+1}\big) p_{ij} = 0, \\ & \sum_{j=1}^k p_{ij} = 1, \quad \text{with} \quad p_{ij} \ge 1,\ \forall\, i,j \in \{1,2,\ldots,k\} \end{aligned}$$

通过在步骤 2(i) 和步骤 2 (ii) 之间迭代，初始值为 $\hat{}\beta^0$，从 $s = 0$ 开始，涉及以下耦合优化问题：

2(i). 首先使用以下方法计算 $\{\hat{}\lambda_{i,\ t+1}^*\}_{i=1}^k$

$$\hat{\lambda}_i^{s+1} = \arg\max_{\lambda \in \mathbb{R}^{d_g}} \sum_{j=1}^k w_{ij} \log\big(1 + \lambda_i^T g\big(\theta_{t+1}^{(j)}, \hat{\beta}^s\big)\big), \quad i = 1,2,\ldots,k.$$

ˆ2 (ii)。然后使用 β_{t+1}^* 进行计算

$$\hat{\beta}^{s+1} = \arg\min_{\beta \in \mathbb{B}} -\sum_{i=1}^k \sum_{j=1}^k w_{ij} \log\big(1 + \big(\hat{\lambda}_i^{s+1}\big)^T g\big(\theta_{t+1}^{(j)}, \beta\big)\big).$$

2 (iii). 重复步骤 2(i)和步骤 2(ii)，直到收敛，即 $\Big(\{\hat{}\lambda_{i,t+1}^s\}_{i=1,}^k \hat{}\beta_{t+1}^s\Big)$ 收敛至

$$\Big(\{\hat{\lambda}_{i,t+1}^*\}_{i=1}^k, \hat{\beta}_{t+1}^*\Big),$$

3. 然后，计算聚合信息，即

$$\bar{\theta}_{t+1}^{(i)} = \sum\nolimits_{j=1}^{k} \hat{p}_{ij}\theta_{t+1}^{(j)}, \quad i = 1, 2, \ldots, k,$$

with

$$\hat{p}_{ij} = \frac{w_{ij}}{1 + \left(\hat{\lambda}_{i,t+1}^{*}\right)^{T} g\left(\theta_{t+1}^{(j)}, \hat{\beta}_{t+1}^{*}\right)}, \quad i, j = 1, 2, \ldots, k,$$

并在下一轮迭代中进一步与每个智能体共享。

4. 将 t 加 1，然后重复步骤 1、2 和 3，直至收敛，即 $\|\ell_\theta\left(\hat{\beta}_{t+1}^{*}\right) - \ell_\theta\left(\hat{\beta}_{t}^{*}\right)\| \leq \text{tol}$，或 $t = N - 1$。

输出：(i) 最优参数估计：$\bar{\theta}_{N}^{(i)} = \sum_{j=1}^{k} \hat{p}_{ij}\theta_{N}^{(j)}$，对于 $i = 1$、2、…、k，以及 **(ii)** 最大对数似然：$\ell_\theta\left(\hat{\beta}_{N}^{*}\right) = \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij} \log \hat{p}_{ij}$。

参考文献

[1] A.B. Owen。《经验似然》。Chapman & Hall/CRC，2001年。[2] P.J. Bickel、C.A.J. Klaassen、Y. Ritov 和 J.A. Wellner。《半参数模型的高效自适应估计》。Springer-Verlag，纽约，1998年。[3] G.W. Imbens、R.H. Spady 和 P. Johnson。矩条件模型推断的信息论方法。*Econometrica*，第 66 卷，333–57，1998年。[4] L.P. Hansen。广义矩估计量的大样本性质。*Econometrica*，第 50 卷，1029 54，1982–。[5] Y. Kitamura。计量经济学中的经验似然方法：理论与实践。载于 R. Blundell、W.K. Newey 和 T. Persson（编），经济学与计量经济学进展，理论与应用，第九届世界大会，第 *3* 卷，第 *174–237* 页。剑桥大学出版社，2007年。[6] G.K. Befekadu。用于提升一类学习问题泛化能力的朴素聚合算法。载于：第 *64* 届 *IEEE* 决策与控制会议论文集（*CDC*），2025年12月10–12日。巴西里约热内卢，第 6134–6137 页。[7] G.K. Befekadu。委托人—代理人协作学习问题的决策理论模型。*arXiv:2409. 16068 [*统计学（*M*）*]*, 2024。链接：https://doi.org/10.48550/arXiv.2409. 16068。[8] V.G. Vovk。基于专家建议的预测博弈。*J. Comput. Syst. Sci.,* 56(2)，153–173，1998年。[9] N. Littlestone 和 M. Warmuth。加权多数算法。*Inf. Comput.,* 108(2)，212–261，1994年。[10] M.W. Hirsch 和 S. Smale。《微分方程、动力系统与线性代数》。Academic Press，伦敦，1974年。[11] R.T. Rockafellar。《凸分析》。普林斯顿大学出版社，1970年。