Federated Stochastic Approximation of the EM Algorithm

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

To be completed

1 SAEM algorithm for Federated Learning

- SA of the expectations
- Monte Carlo approximations
- Exponential Family
- Statistics stored on devices and parameters computed on device
- FedAvg on central server $--$ $>$ boils down to the Averaging of parameters that we may do in standard central settings (Polyak)
- Applications to PK-PD modeling

2 Notations and Algorithm

We minimize the negated log incomplete data likelihood

$$\min_{\theta \in \Theta} \overline{L}(\theta) := L(\theta) + r(\theta) \quad \text{with } L(\theta) = \frac{1}{n} \sum_{i=1}^{n} L_i(\theta) := \frac{1}{n} \sum_{i=1}^{n} \big\{ -\log g(y_i; \theta) \big\} \,, \tag{1}$$

Consider a curved exponential family

$$f(z_i, y_i; \theta) = h(z_i, y_i) \text{exp}(\langle S(z_i, y_i) \,|\, \phi(\theta) \rangle - \psi(\theta)) \,, \tag{2}$$

Then EM reads

$$\overline{s}_i(\theta) := \int_{\mathsf{Z}} S(z_i, y_i) p(z_i | y_i; \theta) \mu(\mathrm{d} z_i) \,, \tag{3}$$

and the *M-step* is given by

$$\overline{\theta}(\overline{s}(\theta)) := \underset{\vartheta \in \theta}{\arg\min} \, \big\{ \mathrm{R}(\vartheta) + \psi(\vartheta) - \langle \overline{s}(\theta) \,|\, \phi(\vartheta) \rangle \big\} \,. \tag{4}$$

In the case where the expectations are intractable, then (3) becomes:

$$\tilde{S}^{(k+1)} := \frac{1}{n} \sum_{i=1}^{n} \tilde{S}_i^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{M_k} \sum_{m=1}^{M_k} S(z_{i,m}^{(k)}, y_i) \,, \tag{5}$$

2.1 Periodic averaging of the local models

Algorithm 1 FL-SAEM with parameter averaging

Input: .
Init: $\theta_0 \in \Theta \subseteq \mathbb{R}^d$, as the global model and $\bar{\theta}_0 = \frac{1}{n}\sum_{i=1}^{n} \theta_0$.
for $r = 1$ to R **do**
  for parallel for device $i \in D^r$ **do**
    Set $\hat{\theta}_i^{(0,k)} = \hat{\theta}^{(k)}$.
    for $t = 1$ to T **do**
      Draw M samples $\{z_{i,m}^{(t,k)}\}_{m=1}^{M}$ under model $\hat{\theta}_i^{(t,k)}$
      Compute the surrogate sufficient statistics $\tilde{S}_i^{(t,k+1)}$
      Update local model:

$$\hat{\theta}_i^{(t,k+1)} = \overline{\theta}(\tilde{S}_i^{(t,k+1)})$$

    end for
    Devices send $\hat{\theta}_i^{(T,k+1)}$ to server.
  end for
  Server computes **the average of the local models**:

$$\hat{\theta}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_i^{(T,k+1)}$$

  and send global model back to the devices.
end for

2.2 Periodic averaging of the local statistics

Algorithm 2 FL-SAEM with statistics averaging

Input: .
Init: $\theta_0 \in \Theta \subseteq \mathbb{R}^d$, as the global model and $\bar{\theta}_0 = \frac{1}{n}\sum_{i=1}^{n} \theta_0$.
for $r = 1$ to R **do**
  for parallel for device $i \in D^r$ **do**
    Set $\hat{\theta}_i^{(0,k)} = \hat{\theta}^{(k)}$.
    for $t = 1$ to T **do**
      Here one local iteration, $T = 1$
      Draw M samples $z_{i,m}^{(k)}$ under model $\hat{\theta}_i^{(t,k)}$
      Compute the surrogate sufficient statistics $\tilde{S}_i^{(t,k+1)}$
    end for
    Devices send local statistics $\tilde{S}_i^{(t,k+1)}$ to server.
  end for
  Server computes **global model using the aggregated statistics**:

$$\hat{\theta}^{(k+1)} = \overline{\theta}(\tilde{S}^{(t,k+1)})$$

  where $\tilde{S}^{(t,k+1)} = (\tilde{S}_i^{(t,k+1)}, i \in D_r)$ and send global model back to the devices.
end for