Exact Expressions for the Generalization Error in Statistical Federated Learning

Yaiza Bermudez
joint work with Samir Perlaza (Inria), Iñaki Esnaola (University of Sheffield) and H. Vincent Poor (Princeton University)

FedMalin General Assembly
November 27, 2025, Montpellier, France

Outline

Representation of Algorithms

P_Z Unknown Data Generating Distribution

$\triangle(\mathcal{X} \times \mathcal{Y})$

Representation of Algorithms

$z = ((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n))$

Training Dataset

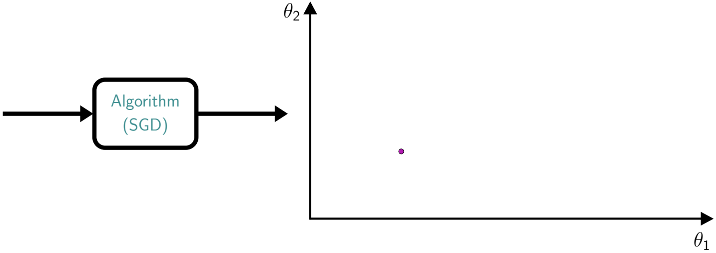
Algorithm
(SGD)

[1] W. Azizian, F. Lutzeler, J. Malick, and P. Mertikopoulos, "**What is the long-run distribution of stochastic gradient descent? A large deviations analysis**," in Proc. of the 41st International Conference on Machine Learning, July 2024, pp. 2168 – 2229.

Representation of Algorithms

$$z = ((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$$

Training Dataset

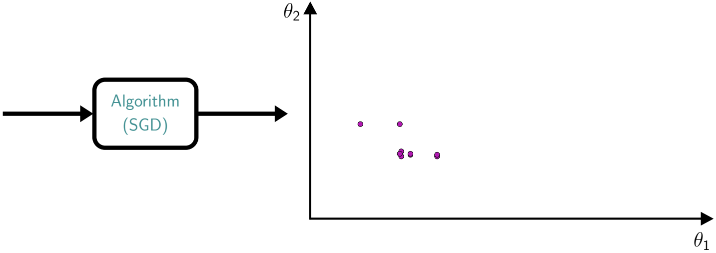


[1] W. Azizian, F. Lutzeler, J. Malick, and P. Mertikopoulos, "**What is the long-run distribution of stochastic gradient descent? A large deviations analysis**," in Proc. of the 41st International Conference on Machine Learning, July 2024, pp. 2168 – 2229.

Representation of Algorithms

$$z = ((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$$

Training Dataset

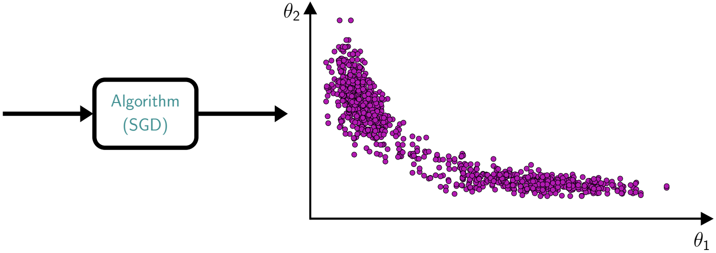


[1] W. Azizian, F. Lutzeler, J. Malick, and P. Mertikopoulos, "**What is the long-run distribution of stochastic gradient descent? A large deviations analysis**," in Proc. of the 41st International Conference on Machine Learning, July 2024, pp. 2168 – 2229.

Representation of Algorithms

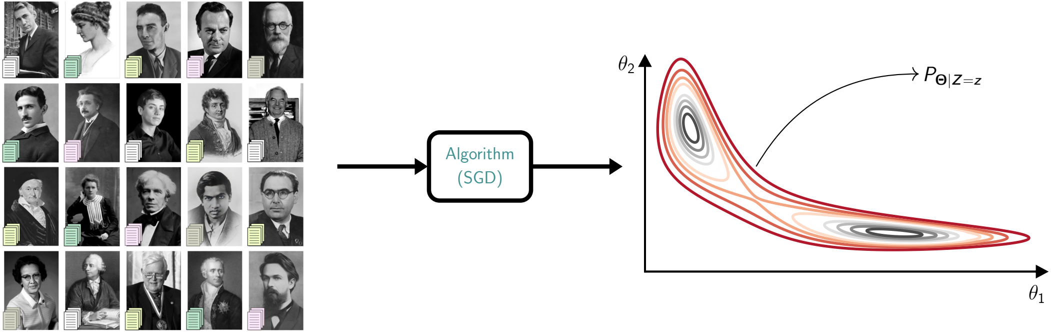
$z = ((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n))$

Training Dataset



[1] W. Azizian, F. Lutzeler, J. Malick, and P. Mertikopoulos, "**What is the long-run distribution of stochastic gradient descent? A large deviations analysis**," in Proc. of the 41st International Conference on Machine Learning, July 2024, pp. 2168 – 2229.

Representation of Algorithms



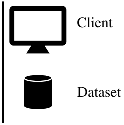
$z = ((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n))$

Training Dataset

[1] W. Azizian, F. Lutzeler, J. Malick, and P. Mertikopoulos, "**What is the long-run distribution of stochastic gradient descent? A large deviations analysis**," in Proc. of the 41st International Conference on Machine Learning, July 2024, pp. 2168 – 2229.

Table of Contents

Federated Learning: Overview



For all $k \in \{1, 2, \ldots, K\}$, the training dataset of client k, $\boldsymbol{z}_k \in \mathcal{Z}_k^{n_k} \triangleq (\mathcal{X}_k \times \mathcal{Y}_k)^{n_k}$:

$$\boldsymbol{z}_k \triangleq ((x_{k,1}, y_{k,1}), (x_{k,2}, y_{k,2}), \ldots, (x_{k,n_k}, y_{k,n_k})),$$

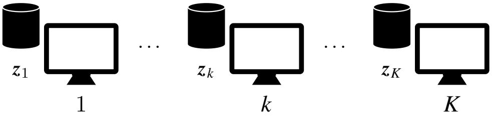
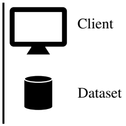


Figure: Federated learning scheme

Federated Learning: Overview



For all $k \in \{1, 2, \dots, K\}$, the training dataset of client k, $\boldsymbol{z}_k \in \mathcal{Z}_k^{n_k} \triangleq (\mathcal{X}_k \times \mathcal{Y}_k)^{n_k}$:

$$\boldsymbol{z}_k \triangleq \left((x_{k,1}, y_{k,1}), (x_{k,2}, y_{k,2}), \dots, (x_{k,n_k}, y_{k,n_k})\right),$$

The loss function of client k:

$$\ell_k : \mathcal{M}_k \times \mathcal{Z}_k \to [0, +\infty),$$

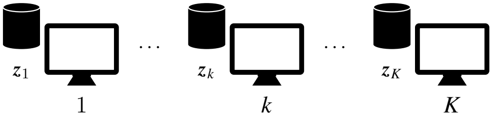


Figure: Federated learning scheme

Federated Learning: Overview

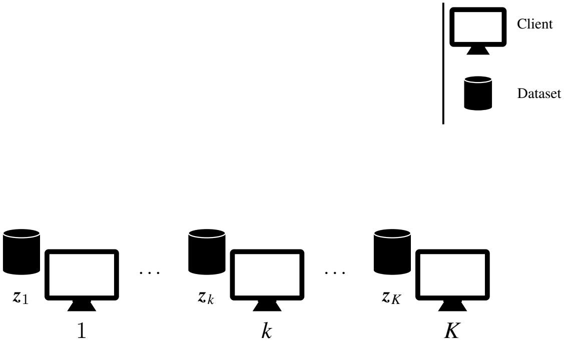


Figure: Federated learning scheme

For all $k \in \{1, 2, \ldots, K\}$, the training dataset of client k, $\boldsymbol{z}_k \in \mathcal{Z}_k^{n_k} \triangleq (\mathcal{X}_k \times \mathcal{Y}_k)^{n_k}$:

$$\boldsymbol{z}_k \triangleq \left((x_{k,1}, y_{k,1}), (x_{k,2}, y_{k,2}), \ldots, (x_{k,n_k}, y_{k,n_k})\right),$$

The loss function of client k:

$$\ell_k : \mathcal{M}_k \times \mathcal{Z}_k \to [0, +\infty),$$

The *empirical risk* induced by $\boldsymbol{\theta} \in \mathcal{M}_k$, with respect to the dataset $\boldsymbol{z}_k$:

$$\mathsf{L}_k : \begin{cases} \mathcal{Z}_k^{n_k} \times \mathcal{M}_k \longrightarrow [0, +\infty) \\ (\boldsymbol{z}_k, \boldsymbol{\theta}) \longmapsto \frac{1}{n_k} \sum_{i=1}^{n_k} \ell_k\left(\boldsymbol{\theta}, x_{k,i}, y_{k,i}\right), \end{cases}$$

[2] Samir M Perlaza, Gaetan Bisson, Iñaki Esnaola, Alain Jean-Marie, Stefano Rini. **Empirical Risk Minimization with Relative Entropy Regularization.** IEEE Transactions on Information Theory, 2024, 70 (7), pp.5122-5161.

Federated Learning: Overview

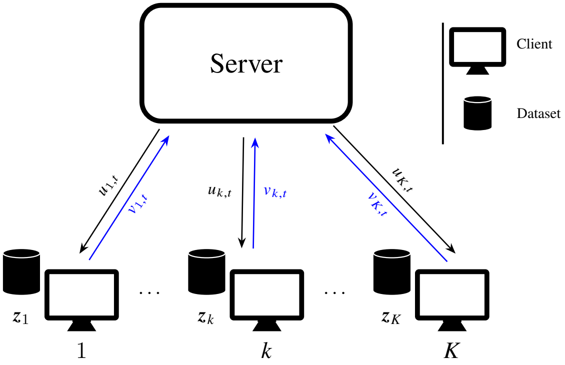


Figure: Federated learning scheme

For all $k \in \{1, 2, \dots, K\}$ and for $t \in \{0, \dots, m\}$,

- $\boldsymbol{z}_k \in \mathcal{Z}_k^{n_k}$;
- $u_{k,t} \in \mathcal{U}_k$;
- $v_{k,t} \in \mathcal{V}_k$;
- $\boldsymbol{\theta}_k \in \mathcal{M}_k$.

Federated Learning: Overview

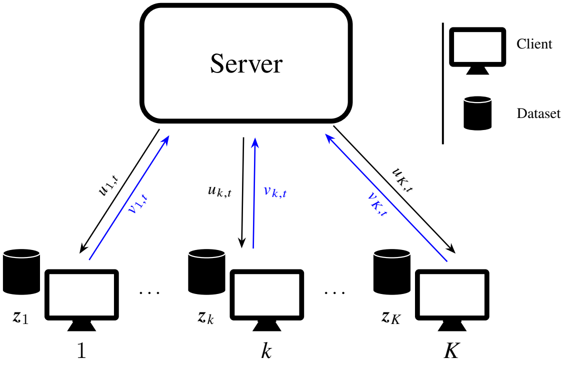


Figure: Federated learning scheme

For all $k \in \{1, 2, \ldots, K\}$ and for $t \in \{0, \ldots, m\}$,

- $\boldsymbol{z}_k \in \mathcal{Z}_k^{n_k}$;
- $u_{k,t} \in \mathcal{U}_k$;
- $v_{k,t} \in \mathcal{V}_k$;
- $\boldsymbol{\theta}_k \in \mathcal{M}_k$.

Example (FedAvg [3])

Let $\mathcal{U}_k = \mathcal{V}_k = \mathcal{M}_k$ and $v_{k,0} = \varnothing$, $u_{k,0} = \boldsymbol{\theta}_0$ then, for $t \in \{1, \ldots, m\}$,

$$v_{k,t} = \nabla \mathrm{L}_k \left(\boldsymbol{z}_k, \, u_{k,t-1} \right),$$

$$u_{k,t} = u_{k,t-1} \; - \; \eta \sum_{i=1}^{K} \frac{n_i}{n_0} \, v_{i,t}.$$

[3] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "**Communication-efficient learning of deep networks from decentralized data**", Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 1273–1282, 2017

Federated Learning: Overview

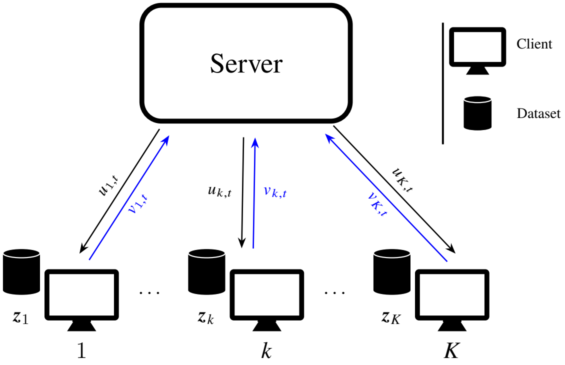


Figure: Federated learning scheme

For all $k \in \{1, 2, \ldots, K\}$ and for $t \in \{0, \ldots, m\}$,

- $\boldsymbol{z}_k \in \mathcal{Z}_k^{n_k}$;
- $u_{k,t} \in \mathcal{U}_k$;
- $v_{k,t} \in \mathcal{V}_k$;
- $\boldsymbol{\theta}_k \in \mathcal{M}_k$.

Communication constraints:

- **uplink** information transmission rate: $\overline{R}_k$;
- **downlink** information transmission rate: $\underline{R}_k$.

$$\boxed{|\mathcal{U}_k| \leqslant 2^{m\underline{R}_k} \quad \text{and} \quad |\mathcal{V}_k| \leqslant 2^{m\overline{R}_k}.}$$

[3] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "**Communication-efficient learning of deep networks from decentralized data**", Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 1273–1282, 2017

Table of Contents

Federated Learning: Communication

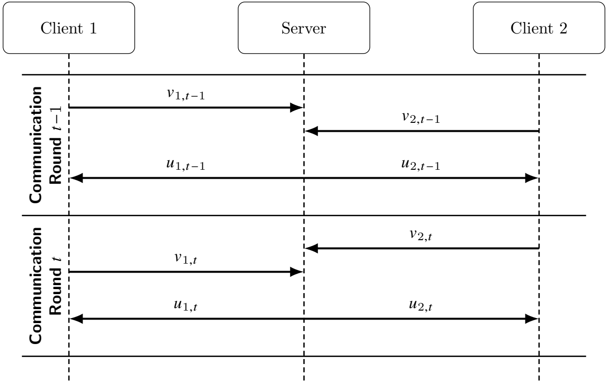


Figure: Two consecutive communications rounds, namely $t-1$ and t, with $t \in \{2, 3, \ldots, m\}$, in the case in which $K = 2$.

Federated Learning: Communication

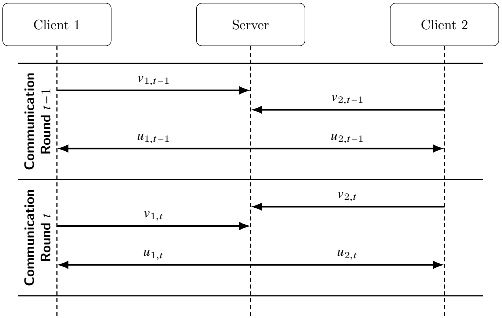


Figure: Two consecutive communications rounds, namely $t-1$ and t, with $t \in \{2, 3, \ldots, m\}$, in the case in which $K = 2$.

For all $k \in \{1, 2, \ldots, K\}$ and for $t \in \{0, \ldots, m\}$,

- all messages sent to the server during round t:

$$\boldsymbol{v}^{(t)} = \left(v_{1,t}, v_{2,t}, \ldots, v_{K,t}\right)^{\top} \in \mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_K;$$

- all messages sent by client k during all previous rounds, including round t:

$$\boldsymbol{v}_k^{(t)} = \left(v_{k,1}, v_{k,2}, \ldots, v_{k,t}\right)^{\top} \in \mathcal{V}_k^t.$$

Federated Learning: Communication

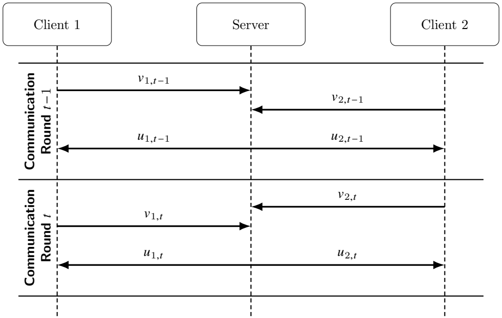


Figure: Two consecutive communications rounds, namely $t-1$ and t, with $t \in \{2, 3, \ldots, m\}$, in the case in which $K = 2$.

For all $k \in \{1, 2, \ldots, K\}$ and for $t \in \{0, \ldots, m\}$,

- all messages sent to the server during round t:

$$\boldsymbol{v}^{(t)} = (v_{1,t}, v_{2,t}, \ldots, v_{K,t})^{\top} \in \mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_K;$$

- all messages sent by client k during all previous rounds, including round t:

$$\boldsymbol{v}_k^{(t)} = (v_{k,1}, v_{k,2}, \ldots, v_{k,t})^{\top} \in \mathcal{V}_k^t.$$

Let $\underline{\boldsymbol{v}}^{(t)} \in (\mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_K)^t$ be such that,

$$\underline{\boldsymbol{v}}^{(t)} \triangleq \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,t} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,t} \\ \vdots & \vdots & \ddots & \vdots \\ v_{K,1} & v_{K,2} & \cdots & v_{K,t} \end{bmatrix}$$
$$= \left(\boldsymbol{v}_1^{(t)} \; \boldsymbol{v}_2^{(t)} \; \cdots \; \boldsymbol{v}_K^{(t)} \right)^{\top}$$
$$= \left(\boldsymbol{v}^{(1)} \; \boldsymbol{v}^{(2)} \; \cdots \; \boldsymbol{v}^{(t)} \right).$$

Federated Learning: Communication

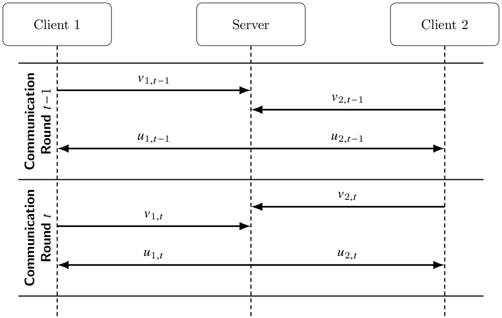


Figure: Two consecutive communications rounds, namely $t-1$ and t, with $t \in \{2, 3, \ldots, m\}$, in the case in which $K = 2$.

For all $k \in \{1, 2, \ldots, K\}$ and for $t \in \{0, \ldots, m\}$,

- all messages sent by the server during round t:

$$\boldsymbol{u}^{(t)} = (u_{1,t}, u_{2,t}, \ldots, u_{K,t})^{\top} \in \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_K;$$

- all messages sent by the server to client k during all previous rounds, including round t:

$$\boldsymbol{u}_k^{(t)} = (u_{k,1}, u_{k,2}, \ldots, u_{k,t})^{\top} \in \mathcal{U}_k^t.$$

Let $\underline{\boldsymbol{u}}^{(t)} \in (\mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_K)^t$ be such that,

$$\underline{\boldsymbol{u}}^{(t)} \triangleq \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,t} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,t} \\ \vdots & \vdots & \ddots & \vdots \\ u_{K,1} & u_{K,2} & \cdots & u_{K,t} \end{bmatrix} = \left(\boldsymbol{u}_1^{(t)} \ \boldsymbol{u}_2^{(t)} \ \cdots \ \boldsymbol{u}_K^{(t)} \right)^{\top} = \left(\boldsymbol{u}^{(1)} \ \boldsymbol{u}^{(2)} \ \cdots \ \boldsymbol{u}^{(t)} \right).$$

Table of Contents

Federated Learning: Statistical Modelling

$P_{\underline{\mathbf{\Theta}},\, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)}, \boldsymbol{Z}_0} \in \triangle\left(\mathcal{M}_0 \times \mathcal{U}_0^m \times \mathcal{V}_0^m \times \mathcal{Z}_0\right)$, with

$$
\begin{aligned}
\mathcal{M}_0 &\triangleq \mathcal{M}_1 \times \mathcal{M}_2 \times \ldots \times \mathcal{M}_K \subset \mathbb{R}^{d \times K}; \\
\mathcal{U}_0 &\triangleq \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_K; \\
\mathcal{V}_0 &\triangleq \mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_K; \\
\mathcal{Z}_0 &\triangleq \mathcal{Z}_1^{n_1} \times \mathcal{Z}_2^{n_2} \times \ldots \times \mathcal{Z}_K^{n_K}.
\end{aligned}
$$

Federated Learning: Statistical Modelling

$P_{\underline{\mathbf{\Theta}},\, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)}, \boldsymbol{Z}_0} \in \triangle \left(\mathcal{M}_0 \times \mathcal{U}_0^m \times \mathcal{V}_0^m \times \mathcal{Z}_0 \right)$, with

$$\begin{aligned} \mathcal{M}_0 &\triangleq \mathcal{M}_1 \times \mathcal{M}_2 \times \ldots \times \mathcal{M}_K \subset \mathbb{R}^{d \times K}; \\ \mathcal{U}_0 &\triangleq \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_K; \\ \mathcal{V}_0 &\triangleq \mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_K; \\ \mathcal{Z}_0 &\triangleq \mathcal{Z}_1^{n_1} \times \mathcal{Z}_2^{n_2} \times \ldots \times \mathcal{Z}_K^{n_K}. \end{aligned}$$

$$\boxed{P_{\underline{\mathbf{\Theta}},\, \underline{\boldsymbol{U}}^{(m)},\, \underline{\boldsymbol{V}}^{(m)},\, \boldsymbol{Z}_0} = P_{\underline{\mathbf{\Theta}},\, \underline{\boldsymbol{U}}^{(m)},\, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0} P_{\boldsymbol{Z}_0}}$$

with,

$$\begin{aligned} P_{\underline{\mathbf{\Theta}},\, \underline{\boldsymbol{U}}^{(m)},\, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0} &\in \triangle \left(\mathcal{M}_0 \times \mathcal{U}_0^m \times \mathcal{V}_0^m \mid \mathcal{Z}_0 \right); \text{ and} \\ P_{\boldsymbol{Z}_0} &\in \triangle \left(\mathcal{Z}_0 \right). \end{aligned}$$

Federated Learning: Statistical Modelling

$P_{\underline{\boldsymbol{\Theta}},\, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)}, \boldsymbol{Z}_0} \in \triangle\left(\mathcal{M}_0 \times \mathcal{U}_0^m \times \mathcal{V}_0^m \times \mathcal{Z}_0\right)$, with

$$
\begin{aligned}
\mathcal{M}_0 &\triangleq \mathcal{M}_1 \times \mathcal{M}_2 \times \ldots \times \mathcal{M}_K \subset \mathbb{R}^{d \times K};\\
\mathcal{U}_0 &\triangleq \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_K;\\
\mathcal{V}_0 &\triangleq \mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_K;\\
\mathcal{Z}_0 &\triangleq \mathcal{Z}_1^{n_1} \times \mathcal{Z}_2^{n_2} \times \ldots \times \mathcal{Z}_K^{n_K}.
\end{aligned}
$$

$$
\boxed{P_{\underline{\boldsymbol{\Theta}},\, \underline{\boldsymbol{U}}^{(m)},\, \underline{\boldsymbol{V}}^{(m)},\, \boldsymbol{Z}_0} = \underbrace{P_{\underline{\boldsymbol{\Theta}},\, \underline{\boldsymbol{U}}^{(m)},\, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0}}_{\text{Federated System}} P_{\boldsymbol{Z}_0}}
$$

with,

$$
\begin{aligned}
P_{\underline{\boldsymbol{\Theta}},\, \underline{\boldsymbol{U}}^{(m)},\, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0} &\in \triangle\left(\mathcal{M}_0 \times \mathcal{U}_0^m \times \mathcal{V}_0^m \mid \mathcal{Z}_0\right); \text{ and}\\
P_{\boldsymbol{Z}_0} &\in \triangle\left(\mathcal{Z}_0\right).
\end{aligned}
$$

Federated Learning: Statistical Modelling

$$P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)}, \boldsymbol{Z}_0} \in \triangle\left(\mathcal{M}_0 \times \mathcal{U}_0^m \times \mathcal{V}_0^m \times \mathcal{Z}_0\right), \text{ with}$$

$$\begin{aligned}
\mathcal{M}_0 &\triangleq \mathcal{M}_1 \times \mathcal{M}_2 \times \ldots \times \mathcal{M}_K \subset \mathbb{R}^{d \times K};\\
\mathcal{U}_0 &\triangleq \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_K;\\
\mathcal{V}_0 &\triangleq \mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_K;\\
\mathcal{Z}_0 &\triangleq \mathcal{Z}_1^{n_1} \times \mathcal{Z}_2^{n_2} \times \ldots \times \mathcal{Z}_K^{n_K}.
\end{aligned}$$

$$\boxed{P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)}, \boldsymbol{Z}_0} = \underbrace{P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0}}_{\text{Federated System}} \underbrace{P_{\boldsymbol{Z}_0}}_{\text{Datasource}}}$$

with,

$$\begin{aligned}
P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0} &\in \triangle\left(\mathcal{M}_0 \times \mathcal{U}_0^m \times \mathcal{V}_0^m \mid \mathcal{Z}_0\right); \text{ and}\\
P_{\boldsymbol{Z}_0} &\in \triangle\left(\mathcal{Z}_0\right).
\end{aligned}$$

The Machine Learning Algorithm

Assumption

The model chosen by client k, with $k \in \{1, 2, \ldots, K\}$, is obtained by sampling a probability measure exclusively conditioned on

- *The dataset $\boldsymbol{z}_k$;*
- *The m messages received from the server, $\boldsymbol{u}_k^{(m)}$; and*
- *The m messages sent to the server, $\boldsymbol{v}_k^{(m)}$.*

The Machine Learning Algorithm

Assumption

The model chosen by client k, with $k \in \{1, 2, \ldots, K\}$, is obtained by sampling a probability measure exclusively conditioned on

- *The dataset $\boldsymbol{z}_k$;*
- *The m messages received from the server, $\boldsymbol{u}_k^{(m)}$; and*
- *The m messages sent to the server, $\boldsymbol{v}_k^{(m)}$.*

The *machine learning algorithm* used by client k:

$$\boxed{P_{\Theta_k | \boldsymbol{U}_k^{(m)}, \boldsymbol{V}_k^{(m)}, \boldsymbol{Z}_k} \in \triangle\left(\mathcal{M}_k | \mathcal{U}_k^m \times \mathcal{V}_k^m \times \mathcal{Z}_k^{n_k}\right).}$$

The Messenger of Client k

Assumption

The message sent by client k to the server at communication round t, with $k \in \{1, 2, \ldots, K\}$ and $t \in \{1, 2, \ldots, m\}$, is obtained by sampling a probability measure exclusively conditioned on

- *The dataset $\boldsymbol{z}_k$;*
- *The $t-1$ messages previously received from the server, $\boldsymbol{u}_k^{(t-1)}$; and*
- *The $t-1$ messages previously sent to the server, $\boldsymbol{v}_k^{(t-1)}$.*

The Messenger of Client k

Assumption

The message sent by client k to the server at communication round t, with $k \in \{1, 2, \ldots, K\}$ and $t \in \{1, 2, \ldots, m\}$, is obtained by sampling a probability measure exclusively conditioned on

- *The dataset $\boldsymbol{z}_k$;*
- *The $t-1$ messages previously received from the server, $\boldsymbol{u}_k^{(t-1)}$; and*
- *The $t-1$ messages previously sent to the server, $\boldsymbol{v}_k^{(t-1)}$.*

The *messenger* of client k during communication round t:

$$\boxed{P_{V_{k,t}|\boldsymbol{U}_k^{(t-1)}, \boldsymbol{V}_k^{(t-1)}, \boldsymbol{Z}_k} \in \triangle\left(\mathcal{V}_k | \mathcal{U}_k^{t-1} \times \mathcal{V}_k^{t-1} \times \mathcal{Z}_k^{n_k}\right).}$$

The Server

Assumption

The messages sent by the server at communication round t to all clients, $\boldsymbol{u}^{(t)}$, with $t \in \{1, 2, \ldots, m\}$, are obtained by sampling a probability measure exclusively conditioned on

- *The t messages received from each client, $\underline{\boldsymbol{v}}^{(t)}$; and*
- *The $t - 1$ messages previously sent to the clients, $\underline{\boldsymbol{u}}^{(t-1)}$.*

The Server

Assumption

The messages sent by the server at communication round t to all clients, $\boldsymbol{u}^{(t)}$, with $t \in \{1, 2, \ldots, m\}$, are obtained by sampling a probability measure exclusively conditioned on

- *The t messages received from each client, $\underline{\boldsymbol{v}}^{(t)}$; and*
- *The $t-1$ messages previously sent to the clients, $\underline{\boldsymbol{u}}^{(t-1)}$.*

The *server* during communication round t:

$$\boxed{P_{\boldsymbol{U}^{(t)}|\underline{\boldsymbol{U}}^{(t-1)},\underline{\boldsymbol{V}}^{(t)}} \in \triangle\left(\mathcal{U}_0|\mathcal{U}_0^{t-1} \times \mathcal{V}_0^t\right).}$$

Statistical Federated Learning

$$P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0} = \left(\prod_{k=1}^{K} P_{\boldsymbol{\Theta}_k | \boldsymbol{U}_k^{(m)}, \boldsymbol{V}_k^{(m)}, \boldsymbol{Z}_k} \right) \prod_{t=1}^{m} \left(P_{\boldsymbol{U}^{(t)} | \underline{\boldsymbol{U}}^{(t-1)}, \underline{\boldsymbol{V}}^{(t-1)}} \prod_{j=1}^{K} P_{V_{j,t} | \boldsymbol{U}_j^{(t-1)}, \boldsymbol{V}_j^{(t-1)}, \boldsymbol{Z}_j} \right).$$

Learning Algorithm

Statistical Federated Learning

$$P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0} = \left(\prod_{k=1}^{K} P_{\boldsymbol{\Theta}_k | \boldsymbol{U}_k^{(m)}, \boldsymbol{V}_k^{(m)}, \boldsymbol{Z}_k} \right) \prod_{t=1}^{m} \left(P_{\boldsymbol{U}^{(t)} | \underline{\boldsymbol{U}}^{(t-1)}, \underline{\boldsymbol{V}}^{(t-1)}} \prod_{j=1}^{K} P_{V_{j,t} | \boldsymbol{U}_j^{(t-1)}, \boldsymbol{V}_j^{(t-1)}, \boldsymbol{Z}_j} \right).$$

Learning Algorithm

Server

Statistical Federated Learning

$$P_{\underline{\boldsymbol{\Theta}},\underline{\boldsymbol{U}}^{(m)},\underline{\boldsymbol{V}}^{(m)}|\boldsymbol{Z}_0} = \left(\prod_{k=1}^{K} P_{\boldsymbol{\Theta}_k|\boldsymbol{U}_k^{(m)},\boldsymbol{V}_k^{(m)},\boldsymbol{Z}_k}\right) \prod_{t=1}^{m} \left(P_{\boldsymbol{U}^{(t)}|\underline{\boldsymbol{U}}^{(t-1)},\underline{\boldsymbol{V}}^{(t-1)}} \prod_{j=1}^{K} P_{V_{j,t}|\boldsymbol{U}_j^{(t-1)},\boldsymbol{V}_j^{(t-1)},\boldsymbol{Z}_j}\right).$$

Learning Algorithm

Server

Messenger

Example

Example (FedAvg)

$$P_{U_{k,t}|U_{k,t-1}=u_{k,t-1},\, \boldsymbol{V}^{(t)}=\boldsymbol{v}^{(t)}}\left(\{u_{k,t}\}\right) = \mathbb{1}_{\left\{u_{k,t}=u_{k,t-1}-\eta\sum_{i=1}^{K}\frac{n_i}{n_0}v_{i,t}\right\}} \in \triangle\left(\mathcal{U}_k\right),$$

$$P_{V_{k,t}|U_{k,t-1}=u_{k,t-1},\, \boldsymbol{Z}_k=\boldsymbol{z}_k}\left(\{v_{k,t}\}\right) = \mathbb{1}_{\left\{v_{k,t}=\nabla \mathsf{L}_k\left(\boldsymbol{z}_k,\, u_{k,t-1}\right)\right\}} \in \triangle\left(\mathcal{V}_k\right),$$

$$P_{\boldsymbol{\Theta}_k|U_{k,t}=u_{k,t},\, \boldsymbol{Z}_k=\boldsymbol{z}_k}\left(\{\boldsymbol{\theta}_k\}\right) = \mathbb{1}_{\left\{\boldsymbol{\theta}_k=u_{k,t}\right\}} \in \triangle\left(\mathcal{M}_k\right).$$

Example

Example (FedAvg)

$$P_{U_{k,t}|U_{k,t-1}=u_{k,t-1},\,\boldsymbol{V}^{(t)}=\boldsymbol{v}^{(t)}}\left(\{u_{k,t}\}\right) = \mathbb{1}_{\left\{u_{k,t}=u_{k,t-1}-\eta\sum_{i=1}^{K}\frac{n_i}{n_0}v_{i,t}\right\}} \in \triangle\left(\mathcal{U}_k\right),$$

$$P_{V_{k,t}|U_{k,t-1}=u_{k,t-1},\,\boldsymbol{Z}_k=\boldsymbol{z}_k}\left(\{v_{k,t}\}\right) = \mathbb{1}_{\left\{v_{k,t}=\nabla\mathsf{L}_k\left(\boldsymbol{z}_k,\,u_{k,t-1}\right)\right\}} \in \triangle\left(\mathcal{V}_k\right),$$

$$P_{\boldsymbol{\Theta}_k|U_{k,t}=u_{k,t},\,\boldsymbol{Z}_k=\boldsymbol{z}_k}\left(\{\boldsymbol{\theta}_k\}\right) = \mathbb{1}_{\left\{\boldsymbol{\theta}_k=u_{k,t}\right\}} \in \triangle\left(\mathcal{M}_k\right).$$

Table of Contents

Generalization Error: Meta Federated Learning Algorithm

Lemma

The marginal probability measure in $\triangle(\mathcal{M}_k \mid \mathcal{Z}_0)$ *of the probability measure* $P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0}$, *denoted* $P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0} \in \triangle(\mathcal{M}_k \mid \mathcal{Z}_0)$, *is such that for all measurable sets* $\mathcal{A} \subseteq \mathcal{M}_k$ *and for all* $\boldsymbol{z}_0 = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_K) \in \mathcal{Z}_0$,

$$P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0 = \boldsymbol{z}_0}(\mathcal{A}) = \sum_{\left(\underline{\boldsymbol{u}}^{(m)}, \underline{\boldsymbol{v}}^{(m)}\right) \in \mathcal{U}_0^m \times \mathcal{V}_0^m} a_{\underline{\boldsymbol{u}}^{(m)}, \underline{\boldsymbol{v}}^{(m)}, \boldsymbol{z}_0} P_{\boldsymbol{\Theta}_k | \boldsymbol{U}_k^{(m)} = \boldsymbol{u}_k^{(m)},\, \boldsymbol{V}_k^{(m)} = \boldsymbol{v}_k^{(m)},\, \boldsymbol{Z}_k = \boldsymbol{z}_k}(\mathcal{A}).$$

Moreover, $a_{\underline{\boldsymbol{u}}^{(m)}, \underline{\boldsymbol{v}}^{(m)}, \boldsymbol{z}_0} \geqslant 0$ *and* $\sum_{\left(\underline{\boldsymbol{u}}^{(m)}, \underline{\boldsymbol{v}}^{(m)}\right) \in \mathcal{U}_0^m \times \mathcal{V}_0^m} a_{\underline{\boldsymbol{u}}^{(m)}, \underline{\boldsymbol{v}}^{(m)}, \boldsymbol{z}_0} = 1.$

$$a_{\underline{\boldsymbol{u}}^{(m)}, \underline{\boldsymbol{v}}^{(m)}, \boldsymbol{z}_0} \triangleq \prod_{t=1}^{m} P_{\boldsymbol{U}^{(t)} | \underline{\boldsymbol{U}}^{(t-1)} = \underline{\boldsymbol{u}}^{(t-1)}, \underline{\boldsymbol{V}}^{(t)} = \underline{\boldsymbol{v}}^{(t)}}\left(\left\{\boldsymbol{u}^{(t)}\right\}\right) \prod_{j=1}^{K} P_{V_{j,t} | \boldsymbol{U}_j^{(t-1)} = \boldsymbol{u}_j^{(t-1)}, \boldsymbol{V}_j^{(t-1)} = \boldsymbol{v}_j^{(t-1)}, \boldsymbol{Z}_j = \boldsymbol{z}_j}\left(\left\{v_{j,t}\right\}\right).$$

Table of Contents

Generalization Error: Definition

For some fixed dataset $\boldsymbol{z}_k \in \mathcal{Z}_k^{n_k}$,

$$\mathsf{R}_{k,\boldsymbol{z}_k} : \begin{cases} \triangle\left(\mathcal{M}_k\right) \longrightarrow [0, +\infty) \\ \qquad P \longmapsto \int \mathsf{L}_k\left(\boldsymbol{z}_k, \boldsymbol{\theta}\right) \mathrm{d}P\left(\boldsymbol{\theta}\right) \end{cases}.$$

Generalization Error: Definition

For some fixed dataset $\boldsymbol{z}_k \in \mathcal{Z}_k^{n_k}$,

$$\mathsf{R}_{k,\boldsymbol{z}_k} : \begin{cases} \triangle\left(\mathcal{M}_k\right) \longrightarrow [0,+\infty) \\ \qquad\quad P \longmapsto \int \mathsf{L}_k\left(\boldsymbol{z}_k, \boldsymbol{\theta}\right) \mathrm{d}P\left(\boldsymbol{\theta}\right) \end{cases}.$$

Definition (Generalization Error)

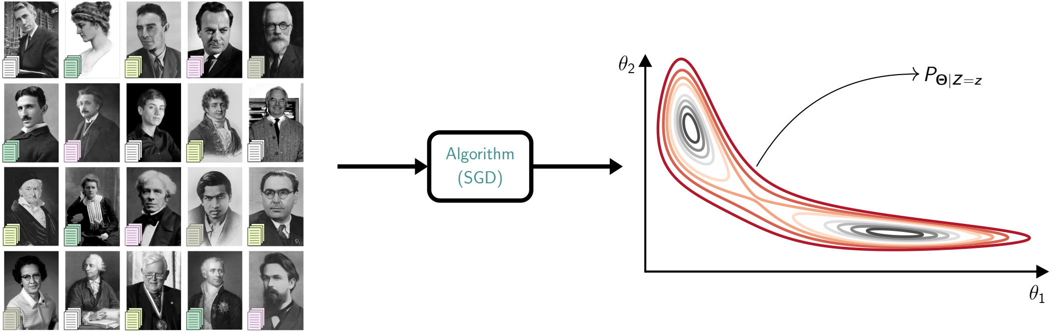
In the federated learning system described by the measure $P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0}$, under the assumption that datasets $\boldsymbol{z}_0 = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_K) \in \mathcal{Z}_0$ are obtained by sampling a probability measure $P_{\boldsymbol{Z}_0} \in \triangle\left(\mathcal{Z}_0\right)$, the generalization error at client k is

$$\mathbf{G}_k\left(P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) \triangleq \iint \left(\mathsf{R}_{k,\hat{\boldsymbol{z}}_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0 = \boldsymbol{z}_0}\right) - \mathsf{R}_{k,\boldsymbol{z}_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0 = \boldsymbol{z}_0}\right) \right) \mathrm{d}P_{\boldsymbol{Z}_k}(\hat{\boldsymbol{z}}_k) \mathrm{d}P_{\boldsymbol{Z}_0}(\boldsymbol{z}_0).$$

[4] Samir M. Perlaza and Xinying Zou. "**The Generalization Error of Machine Learning Algorithms**". November, 2024.

Table of Contents

Representation of Algorithms



$z = ((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n))$

Training Dataset

[1] W. Azizian, F. Lutzeler, J. Malick, and P. Mertikopoulos, "**What is the long-run distribution of stochastic gradient descent? A large deviations analysis**," in Proc. of the 41st International Conference on Machine Learning, July 2024, pp. 2168 – 2229.

Optimization problems

Let $\lambda_k > 0$,

$$\boxed{\min_{P \in \triangle_{Q_{\mathbf{\Theta}_k}}(\mathcal{M}_k)} \mathsf{R}_{k,\boldsymbol{z}_k}(P) + \lambda_k \,\mathsf{D}\left(P \| Q_{\mathbf{\Theta}_k}\right)}$$

Let $\lambda_k < 0$,

$$\boxed{\max_{P \in \triangle_{Q_{\mathbf{\Theta}_k}}(\mathcal{M}_k)} \mathsf{R}_{k,\boldsymbol{z}_k}(P) + \lambda_k \,\mathsf{D}\left(P \| Q_{\mathbf{\Theta}_k}\right)}$$

[5] Samir M. Perlaza and Gaetan Bisson. "**Variations on the expectation due to changes in the probability measure**". Entropy, 27(8:865):1–20, Aug. 2025.

Generalization Error: Gibbs Measure

Definition

Given the function L_k in (1), with $k \in \{1, 2, \ldots, K\}$; a σ-finite measure $Q_{\Theta_k} \in \triangle(\mathcal{M}_k)$; and a $\lambda_k \in \mathbb{R}\backslash\{0\}$, the probability measure $P^{(Q_{\Theta_k}, \lambda_k)}_{\Theta_k | \boldsymbol{Z}_k} \in \triangle(\mathcal{M}_k | \mathcal{Z}_k^{n_k})$ is said to be an $(\mathsf{L}_k, Q_{\Theta_k}, \lambda_k)$-Gibbs conditional probability measure if for all $\boldsymbol{z}_k \in \mathcal{S}_k$, $\mathsf{K}_{k, Q_{\Theta_k}, \boldsymbol{z}_k}\left(-\frac{1}{\lambda_k}\right) < +\infty$; for some set $\mathcal{S}_k \subseteq \mathcal{Z}_k^{n_k}$; and for all $(\boldsymbol{z}_k, \boldsymbol{\theta}_k) \in \mathcal{S}_k \times \operatorname{supp} Q_{\Theta_k}$,

$$\frac{\mathrm{d}P^{(Q_{\Theta_k}, \lambda_k)}_{\Theta_k | \boldsymbol{Z}_k = \boldsymbol{z}_k}}{\mathrm{d}Q_{\Theta_k}}(\boldsymbol{\theta}_k) = \exp\left(-\frac{1}{\lambda_k}\mathsf{L}_k(\boldsymbol{z}_k, \boldsymbol{\theta}_k) - \mathsf{K}_{k, Q_{\Theta_k}, \boldsymbol{z}_k}\left(-\frac{1}{\lambda_k}\right)\right),$$

where the function $\mathsf{K}_{k, Q_{\Theta_k}, \boldsymbol{z}_k}$ is defined as follows

$$\mathsf{K}_{k, Q_{\Theta_k}, \boldsymbol{z}_k} : \begin{cases} \mathbb{R} \longrightarrow \mathbb{R} \\ t \longmapsto \log\left(\int \exp\left(t\, \mathsf{L}_k(\boldsymbol{z}_k, \boldsymbol{\theta}_k)\right) \mathrm{d}Q_{\Theta_k}(\boldsymbol{\theta}_k)\right). \end{cases}$$

Optimization problems

Let $\lambda_k > 0$,

$$\boxed{\min_{P \in \triangle_{Q_{\boldsymbol{\Theta}_k}}(\mathcal{M}_k)} \mathsf{R}_{k,\boldsymbol{z}_k}(P) + \lambda_k \, \mathsf{D}\left(P \| Q_{\boldsymbol{\Theta}_k}\right).}$$

Let $\lambda_k < 0$,

$$\boxed{\max_{P \in \triangle_{Q_{\boldsymbol{\Theta}_k}}(\mathcal{M}_k)} \mathsf{R}_{k,\boldsymbol{z}_k}(P) + \lambda_k \, \mathsf{D}\left(P \| Q_{\boldsymbol{\Theta}_k}\right).}$$

Optimization problems

Let $\lambda_k > 0$,

$$\boxed{\min_{P \in \triangle_{Q_{\boldsymbol{\Theta}_k}}(\mathcal{M}_k)} \mathrm{R}_{k,\boldsymbol{z}_k}(P) + \lambda_k \,\mathrm{D}\left(P \| Q_{\boldsymbol{\Theta}_k}\right).}$$

Let $\gamma \; = \; \mathrm{D}\left(P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_k = \boldsymbol{z}_k} \| Q_{\boldsymbol{\Theta}_k}\right)$,

$$\boxed{\begin{aligned} \min_{P \in \triangle_{Q_{\boldsymbol{\Theta}_k}}(\mathcal{M}_k)} & \mathrm{R}_{k,\boldsymbol{z}_k}(P) \\ \text{s.t. } & \mathrm{D}\left(P \| Q_{\boldsymbol{\Theta}_k}\right) \leq \gamma. \end{aligned}}$$

Let $\lambda_k < 0$,

$$\boxed{\max_{P \in \triangle_{Q_{\boldsymbol{\Theta}_k}}(\mathcal{M}_k)} \mathrm{R}_{k,\boldsymbol{z}_k}(P) + \lambda_k \,\mathrm{D}\left(P \| Q_{\boldsymbol{\Theta}_k}\right).}$$

Let $\gamma \; = \; \mathrm{D}\left(P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_k = \boldsymbol{z}_k} \| Q_{\boldsymbol{\Theta}_k}\right)$,

$$\boxed{\begin{aligned} \max_{P \in \triangle_{Q_{\boldsymbol{\Theta}_k}}(\mathcal{M}_k)} & \mathrm{R}_{k,\boldsymbol{z}_k}(P) \\ \text{s.t. } & \mathrm{D}\left(P \| Q_{\boldsymbol{\Theta}_k}\right) \leq \gamma. \end{aligned}}$$

[5] Samir M. Perlaza and Gaetan Bisson. "**Variations on the expectation due to changes in the probability measure**". Entropy, 27(8:865):1–20, Aug. 2025.

Optimization problems

Let $\lambda_k > 0$,

$$\boxed{\min_{P \in \triangle_{Q_{\boldsymbol{\Theta}_k}}(\mathcal{M}_k)} \mathsf{R}_{k,\boldsymbol{z}_k}(P) + \lambda_k \,\mathsf{D}\left(P \| Q_{\boldsymbol{\Theta}_k}\right).}$$

Let $\gamma \;=\; \mathsf{D}\left(P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k} \| Q_{\boldsymbol{\Theta}_k}\right)$,

$$\boxed{\begin{array}{rl}\displaystyle\min_{P \in \triangle_{Q_{\boldsymbol{\Theta}_k}}(\mathcal{M}_k)} & \mathsf{R}_{k,\boldsymbol{z}_k}(P) \\ \text{s.t.} & \mathsf{D}\left(P \| Q_{\boldsymbol{\Theta}_k}\right) \leq \gamma.\end{array}}$$

The Best !

Let $\lambda_k < 0$,

$$\boxed{\max_{P \in \triangle_{Q_{\boldsymbol{\Theta}_k}}(\mathcal{M}_k)} \mathsf{R}_{k,\boldsymbol{z}_k}(P) + \lambda_k \,\mathsf{D}\left(P \| Q_{\boldsymbol{\Theta}_k}\right).}$$

Let $\gamma \;=\; \mathsf{D}\left(P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k} \| Q_{\boldsymbol{\Theta}_k}\right)$,

$$\boxed{\begin{array}{rl}\displaystyle\max_{P \in \triangle_{Q_{\boldsymbol{\Theta}_k}}(\mathcal{M}_k)} & \mathsf{R}_{k,\boldsymbol{z}_k}(P) \\ \text{s.t.} & \mathsf{D}\left(P \| Q_{\boldsymbol{\Theta}_k}\right) \leq \gamma.\end{array}}$$

The Worse !

[5] Samir M. Perlaza and Gaetan Bisson. "**Variations on the expectation due to changes in the probability measure**". Entropy, 27(8:865):1–20, Aug. 2025.

Table of Contents

Generalization Error: The First Closed Form Expression

Theorem

Consider the generalization error $\mathbf{G}_k\left(P_{\underline{\boldsymbol{\Theta}},\underline{\boldsymbol{U}}^{(m)},\underline{\boldsymbol{V}}^{(m)}|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right)$ *and assume that for all* $\boldsymbol{z}_0 \in \mathcal{Z}_0$, $P^{(m)}_{\boldsymbol{\Theta}_k} \ll Q_{\boldsymbol{\Theta}_k} \ll P^{(m)}_{\boldsymbol{\Theta}_k} \ll P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0} \ll Q_{\boldsymbol{\Theta}_k}$. *Then,*

$$\mathbf{G}_k\left(P_{\underline{\boldsymbol{\Theta}},\underline{\boldsymbol{U}}^{(m)},\underline{\boldsymbol{V}}^{(m)}|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) = \lambda_k\left(I\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) + L\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) + J_{\mathsf{L}_k,Q_{\boldsymbol{\Theta}_k},\lambda_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right)\right).$$

Generalization Error: The First Closed Form Expression

Theorem

Consider the generalization error $\mathbf{G}_k\left(P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)}|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right)$ *and assume that for all* $\boldsymbol{z}_0 \in \mathcal{Z}_0$, $P^{(m)}_{\boldsymbol{\Theta}_k} \ll Q_{\boldsymbol{\Theta}_k} \ll P^{(m)}_{\boldsymbol{\Theta}_k} \ll P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0} \ll Q_{\boldsymbol{\Theta}_k}$. *Then,*

$$\mathbf{G}_k\left(P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)}|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) = \lambda_k\left(I\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) + L\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) + J_{\mathsf{L}_k, Q_{\boldsymbol{\Theta}_k}, \lambda_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right)\right).$$

$$J_{\mathsf{L}_k, Q_{\boldsymbol{\Theta}_k}, \lambda_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right)$$
$$= \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k, \boldsymbol{z}_0) - \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k, \boldsymbol{z}_0),$$

where the probability measure $P^{(m)}_{\boldsymbol{\Theta}_k}$ is such that for all measurable sets $\mathcal{A} \subseteq \mathcal{M}_k$

$$P^{(m)}_{\boldsymbol{\Theta}_k}(\mathcal{A}) = \int P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}(\mathcal{A})\,\mathrm{d}P_{\boldsymbol{Z}_0}(\boldsymbol{z}_0);$$

$P^{(m)}_{\boldsymbol{\Theta}_k}P_{\boldsymbol{Z}_0} \in \triangle(\mathcal{M}_k \times \mathcal{Z}_0)$ is a product measure formed by $P^{(m)}_{\boldsymbol{\Theta}_k}$ and $P_{\boldsymbol{Z}_0}$; and
$P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0} \in \triangle(\mathcal{M}_k \times \}setZ_0)$ is the joint probability measure induced by $P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}$.

Generalization Error: The First Closed Form Expression

Theorem

Consider the generalization error $\mathbf{G}_k\left(P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)}|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right)$ *and assume that for all* $\boldsymbol{z}_0 \in \mathcal{Z}_0$, $P^{(m)}_{\boldsymbol{\Theta}_k} \ll Q_{\boldsymbol{\Theta}_k} \ll P^{(m)}_{\boldsymbol{\Theta}_k} \ll P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0} \ll Q_{\boldsymbol{\Theta}_k}$. *Then,*

$$\mathbf{G}_k\left(P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)}|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) = \lambda_k\left(I\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) + L\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) + J_{\mathsf{L}_k, Q_{\boldsymbol{\Theta}_k}, \lambda_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right)\right).$$

Warnings :

- λ_k does not influence the value of $\mathbf{G}_k\left(P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)}|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right)$;
- $I\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) + L\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) \geq 0$;
- $J_{\mathsf{L}_k, Q_{\boldsymbol{\Theta}_k}, \lambda_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) \in \mathbb{R}$.

Independence Hypothesis Test

$$
\begin{aligned}
& I\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) + L\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) \\
&= \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k, \boldsymbol{z}_0) - \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k, \boldsymbol{z}_0)
\end{aligned}
$$

Independence Hypothesis Test

$$
\begin{aligned}
&I\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) + L\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) \\
&= \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0) - \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0)
\end{aligned}
$$

Independence Hypothesis Test

$$I\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) + L\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right)$$
$$= \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0) - \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0)$$

Hypothesis:

$$H_0 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}$$

$$H_1 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(m)}_{\boldsymbol{\Theta}_k}P_{\boldsymbol{Z}_0}$$

Contribution of the Federation

$$
\begin{aligned}
& J_{\mathsf{L}_k, Q_{\boldsymbol{\Theta}_k}, \lambda_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) \\
= & \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0) - \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0),
\end{aligned}
$$

Contribution of the Federation

$$
\begin{aligned}
& J_{\mathsf{L}_k, Q_{\boldsymbol{\Theta}_k}, \lambda_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) \\
&= \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0) - \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0),
\end{aligned}
$$

Mismatch Hypothesis testing:

$$H_0 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0} P_{\boldsymbol{Z}_0}$$

$$H_1 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k} P_{\boldsymbol{Z}_0}$$

Contribution of the Federation

$$
\begin{aligned}
&J_{\mathsf{L}_k, Q_{\boldsymbol{\Theta}_k}, \lambda_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) \\
&= \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0) - \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0),
\end{aligned}
$$

Mismatch Hypothesis testing:

$$H_0 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}$$

$$H_1 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k}P_{\boldsymbol{Z}_0}$$

Contribution of the Federation

$$J_{\mathsf{L}_k, Q_{\boldsymbol{\Theta}_k}, \lambda_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right)$$
$$= \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k} P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0) - \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0} P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0),$$

Mismatch Hypothesis testing:

$$H_0 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0} P_{\boldsymbol{Z}_0}$$

$$H_1 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k} P_{\boldsymbol{Z}_0}$$

Contribution of the Federation

$$
\begin{aligned}
&J_{\mathsf{L}_k, Q_{\boldsymbol{\Theta}_k}, \lambda_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) \\
&= \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0) - \int \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right)\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}(\boldsymbol{\theta}_k,\boldsymbol{z}_0),
\end{aligned}
$$

Consider $\gamma > 0$,

$$
\mathcal{A}_\gamma = \left\{(\boldsymbol{\theta}_k, \boldsymbol{z}_0) \in \mathcal{M}_k \times \mathcal{Z}_0 : \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right) \geqslant \gamma\right\}
$$

The set $\mathcal{A}_\gamma$ is the acceptance region of $H_0 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}$.

The set $\mathcal{A}^{\mathsf{c}}_\gamma$ is the acceptance region of $H_1 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k}P_{\boldsymbol{Z}_0}$.

Consider $\gamma > 0$,

$$\mathcal{A}_\gamma = \left\{ (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \in \mathcal{M}_k \times \mathcal{Z}_0 : \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}} (\boldsymbol{\theta}_k) \right) \geqslant \gamma \right\}$$

The set $\mathcal{A}_\gamma$ is the acceptance region of $H_0 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0} P_{\boldsymbol{Z}_0}$.

The set $\mathcal{A}^{\mathrm{c}}_\gamma$ is the acceptance region of $H_1 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k} P_{\boldsymbol{Z}_0}$.

Consider $\gamma > 0$,

$$\mathcal{A}_\gamma = \left\{ (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \in \mathcal{M}_k \times \mathcal{Z}_0 : \log \left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0 = \boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_k = \boldsymbol{z}_k}} (\boldsymbol{\theta}_k) \right) \geqslant \gamma \right\}$$

The set $\mathcal{A}_\gamma$ is the acceptance region of $H_0 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0} P_{\boldsymbol{Z}_0}$.

The set $\mathcal{A}_\gamma^{\mathrm{c}}$ is the acceptance region of $H_1 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_k} P_{\boldsymbol{Z}_0}$.

$P^{(m)}_{\boldsymbol{\Theta}_k} P_{\boldsymbol{Z}_0} (\mathcal{A}_\gamma)$

Consider $\gamma > 0$,

$$\mathcal{A}_\gamma = \left\{ (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \in \mathcal{M}_k \times \mathcal{Z}_0 : \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0 = \boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_k = \boldsymbol{z}_k}} (\boldsymbol{\theta}_k) \right) \geqslant \gamma \right\}$$

The set $\mathcal{A}_\gamma$ is the acceptance region of $H_0 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0} P_{\boldsymbol{Z}_0}$.

The set $\mathcal{A}_\gamma^{\mathrm{c}}$ is the acceptance region of $H_1 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_k} P_{\boldsymbol{Z}_0}$.

$$P^{(m)}_{\boldsymbol{\Theta}_k} P_{\boldsymbol{Z}_0} (\mathcal{A}_\gamma)$$

$$P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0} P_{\boldsymbol{Z}_0} (\mathcal{A}_\gamma)$$

Consider $\gamma > 0$,

$$\mathcal{A}_\gamma = \left\{ (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \in \mathcal{M}_k \times \mathcal{Z}_0 : \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0 = \boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_k = \boldsymbol{z}_k}} (\boldsymbol{\theta}_k) \right) \geqslant \gamma \right\}$$

The set $\mathcal{A}_\gamma$ is the acceptance region of $H_0 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0} P_{\boldsymbol{Z}_0}$.

The set $\mathcal{A}^{\mathrm{c}}_\gamma$ is the acceptance region of $H_1 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_k} P_{\boldsymbol{Z}_0}$.

$$P^{(m)}_{\boldsymbol{\Theta}_k} P_{\boldsymbol{Z}_0} (\mathcal{A}_\gamma)$$

$$P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0} P_{\boldsymbol{Z}_0} (\mathcal{A}_\gamma)$$

$$P^{(m)}_{\boldsymbol{\Theta}_k} P_{\boldsymbol{Z}_0} (\mathcal{A}^{\mathrm{c}}_\gamma)$$

Consider $\gamma > 0$,

$$\mathcal{A}_\gamma = \left\{ (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \in \mathcal{M}_k \times \mathcal{Z}_0 : \log \left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0 = \boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_k = \boldsymbol{z}_k}} (\boldsymbol{\theta}_k) \right) \geqslant \gamma \right\}$$

The set $\mathcal{A}_\gamma$ is the acceptance region of $H_0 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0} P_{\boldsymbol{Z}_0}$.

The set $\mathcal{A}^{\mathrm{c}}_\gamma$ is the acceptance region of $H_1 : (\boldsymbol{\theta}_k, \boldsymbol{z}_0) \sim P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_k} P_{\boldsymbol{Z}_0}$.

$P^{(m)}_{\boldsymbol{\Theta}_k} P_{\boldsymbol{Z}_0} (\mathcal{A}_\gamma)$	$P^{(m)}_{\boldsymbol{\Theta}_k} P_{\boldsymbol{Z}_0} (\mathcal{A}^{\mathrm{c}}_\gamma)$
$P^{(m)}_{\boldsymbol{\Theta}_k \| \boldsymbol{Z}_0} P_{\boldsymbol{Z}_0} (\mathcal{A}_\gamma)$	$P^{(m)}_{\boldsymbol{\Theta}_k \| \boldsymbol{Z}_0} P_{\boldsymbol{Z}_0} (\mathcal{A}^{\mathrm{c}}_\gamma)$

Bounds in Probability

Theorem

Assume that for all $\boldsymbol{z}_0 \in \mathcal{Z}_0$, $P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0} \ll Q_{\boldsymbol{\Theta}_k}$*; consider a free parameter* $\gamma > 0$*; the following constants*

$$\underline{\gamma} \triangleq \min_{(\boldsymbol{\theta}_k,\boldsymbol{z}_0)\in\mathcal{M}_k\times\mathcal{Z}_0} \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right) \text{ and } \bar{\gamma} \triangleq \max_{(\boldsymbol{\theta}_k,\boldsymbol{z}_0)\in\mathcal{M}_k\times\mathcal{Z}_0} \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right);$$

and the set $\mathcal{A}_\gamma = \left\{(\boldsymbol{\theta}_k,\boldsymbol{z}_0)\in\mathcal{M}_k\times\mathcal{Z}_0 : \log\left(\frac{\mathrm{d}P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}}{\mathrm{d}P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}}(\boldsymbol{\theta}_k)\right) \geqslant \gamma\right\}$. *Then, the value* $J_{\mathsf{L}_k,Q_{\boldsymbol{\Theta}_k},\lambda_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right)$*, satisfies*

$$\begin{aligned}
&P^{(m)}_{\boldsymbol{\Theta}_k}P_{\boldsymbol{Z}_0}(\mathcal{A}_\gamma)(\bar{\gamma}-\gamma) + P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}(\mathcal{A}^c_\gamma)\left(\gamma-\underline{\gamma}\right)\\
&\geqslant J_{\mathsf{L}_k,Q_{\boldsymbol{\Theta}_k},\lambda_k}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) \geqslant\\
&P^{(m)}_{\boldsymbol{\Theta}_k}P_{\boldsymbol{Z}_0}(\mathcal{A}^c_\gamma)\left(\underline{\gamma}-\gamma\right) + P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0}P_{\boldsymbol{Z}_0}(\mathcal{A}_\gamma)(\gamma-\bar{\gamma}).
\end{aligned}$$

Table of Contents

Generalization Error: The Second Closed Form Expression

Theorem

Consider the generalization error $\mathbf{G}_k\left(P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right)$ *and assume that for all* $\boldsymbol{z}_0 \in \mathcal{Z}_0$, $P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0 = \boldsymbol{z}_0} \ll Q_{\boldsymbol{\Theta}_k}$. *Then,*

$$\begin{aligned} &\mathbf{G}_k\left(P_{\underline{\boldsymbol{\Theta}}, \underline{\boldsymbol{U}}^{(m)}, \underline{\boldsymbol{V}}^{(m)} | \boldsymbol{Z}_0}; P_{\boldsymbol{Z}_0}\right) \\ &= \lambda_k \iint \left(\mathrm{D}\left(P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0 = \boldsymbol{z}_0} \| P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_k = \hat{\boldsymbol{z}}_k}\right) - \mathrm{D}\left(P^{(m)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_0 = \boldsymbol{z}_0} \| P^{(Q_{\boldsymbol{\Theta}_k}, \lambda_k)}_{\boldsymbol{\Theta}_k | \boldsymbol{Z}_k = \boldsymbol{z}_k}\right)\right) \mathrm{d}P_{\boldsymbol{Z}_k}(\hat{\boldsymbol{z}}_k) \, \mathrm{d}P_{\boldsymbol{Z}_0}(\boldsymbol{z}_0). \end{aligned}$$

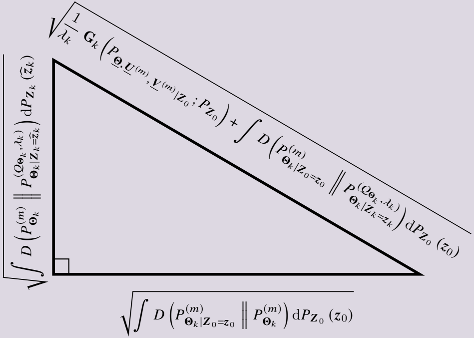
Table of Contents

Generalization Error: A Pythagorean Identity

Theorem

Consider the generalization error $\mathbf{G}_k\left(P_{\underline{\boldsymbol{\Theta}},\underline{\boldsymbol{U}}^{(m)},\underline{\boldsymbol{V}}^{(m)}|\boldsymbol{Z}_0};P_{\boldsymbol{Z}_0}\right)$. *Assume that for all* $\boldsymbol{z}_0 \in \mathcal{Z}_0$, $P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0} \ll P^{(m)}_{\boldsymbol{\Theta}_k} \ll Q_{\boldsymbol{\Theta}_k}$. *Then,*

$$
\begin{aligned}
&\frac{1}{\lambda_k}\mathbf{G}_k\left(P_{\underline{\boldsymbol{\Theta}},\underline{\boldsymbol{U}}^{(m)},\underline{\boldsymbol{V}}^{(m)}|\boldsymbol{Z}_0};P_{\boldsymbol{Z}_0}\right)\\
&\quad+\int \mathrm{D}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}\middle\|P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\boldsymbol{z}_k}\right)\mathrm{d}P_{\boldsymbol{Z}_0}(\boldsymbol{z}_0)\\
&=\int \mathrm{D}\left(P^{(m)}_{\boldsymbol{\Theta}_k}\middle\|P^{(Q_{\boldsymbol{\Theta}_k},\lambda_k)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_k=\hat{\boldsymbol{z}}_k}\right)\mathrm{d}P_{\boldsymbol{Z}_k}(\hat{\boldsymbol{z}}_k)\\
&\quad+\int \mathrm{D}\left(P^{(m)}_{\boldsymbol{\Theta}_k|\boldsymbol{Z}_0=\boldsymbol{z}_0}\middle\|P^{(m)}_{\boldsymbol{\Theta}_k}\right)\mathrm{d}P_{\boldsymbol{Z}_0}(\boldsymbol{z}_0),
\end{aligned}
$$



Thank you for your attention!

Questions/Comments/Typos: yaiza.bermudez@inria.fr