A Simple Model for Personalized Machine Learning

A. Jung

April 9, 2025

Figure 1 depicts a simple mathematical model for a personalized machine learning application. The model is parametrized by:

- a natural number B (batch size),
- a natural number R (the number of distinct probability distributions),
- a collection of probability distributions $p^{(r)}(x)$, for $r = 1, \ldots, R$, each absolutely continuous with respect to (w.r.t.) the Lebesgue measure on $\mathbb{R}$.

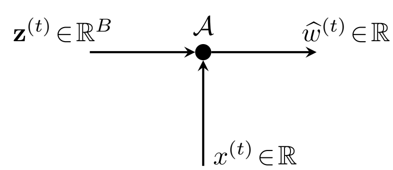

Figure 1: A learner $\mathcal{A}$ generates an estimate $\widehat{w}^{(t)}$ based on data streams $x^{(t)}$ and $\mathbf{z}^{(t)}$.

The data stream $x^{(t)}$ is an i.i.d. sequence of random variables (RVs) drawn from the *correct* distribution $p^{(1)}$. The second data stream $\mathbf{z}^{(t)}$ is generated at each time instant t as follows:

1. Draw an index $r^{(t)} \in \{1, \ldots, R\}$ uniformly at random and independently over time.
2. Generate the vector $\mathbf{z}^{(t)} = \left(z_1^{(t)}, \ldots, z_B^{(t)}\right)^T$, where entries are i.i.d. RVs from the distribution $p^{(r^{(t)})}(z)$.

   Note that the correct distribution $p^{(1)}$ is selected with probability $1/R$ at each time t.

We may measure the performance of the learning algorithm $\mathcal{A}$ in various ways. One natural choice is to treat $\widehat{w}^{(t)}$ as an estimator of a true parameter $\overline{w}$ (e.g., the mean or variance) associated with the correct distribution $p^{(1)}$. This allows defining the estimation error as $\widehat{w}^{(t)} - \overline{w}$ and subsequently the mean squared error (MSE):

$$\varepsilon_t = \mathbb{E}\big[(\widehat{w}^{(t)} - \overline{w})^2\big].$$

We are particularly interested in addressing the following questions:

- What are the fundamental limits (e.g., minimax MSE) achievable by any learning algorithm?
- What practical algorithms exist that approach these theoretical limits?