Project Report

Zhuonan Hao

November 2022

1 Offline Solution

1.1 Reformulation

Formulate the original problem

$$\arg\min_{\theta,b} \frac{1}{N}\|\mathcal{Y} - (\mathcal{X}\theta + b1)\|_1 + \alpha\|\theta\|_1$$

into the linear programming problem

$$\begin{array}{rl} \text{minimize} & \frac{1}{N}1^T t + \alpha u \\ \text{subject to} & -t \leq \mathcal{Y} - (\mathcal{X}\theta + b1) \leq t \\ & -u \leq \theta \leq u \end{array} \tag{1}$$

the inequality form

$$\begin{aligned} &\min_{\boldsymbol{x}} \boldsymbol{c}^T\boldsymbol{x} \\ &\text{s.t. } \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \end{aligned} \tag{2}$$

where $\boldsymbol{c} = [\frac{1}{N}1, \alpha 1, 0, 0]^T$, $\boldsymbol{x} = [t, u, \theta, b]^T$, $\boldsymbol{A} = \begin{bmatrix} -I_N & 0 & -\mathcal{X} & -1 \\ -I_N & 0 & \mathcal{X} & 1 \\ 0 & -I_M & I_M & -0 \\ 0 & -I_M & I_M & 0 \end{bmatrix}$, and $\boldsymbol{b} = [-\mathcal{Y}, \mathcal{Y}, 0, 0]^T$.

1.2 Train a Regressor

The optimal solution by *scipy.optimize.linprog*. The best α value is around 0.09.

1.3 Feature Selection

The feature selection by *sklearn.feature_selection.SelectKBest*.

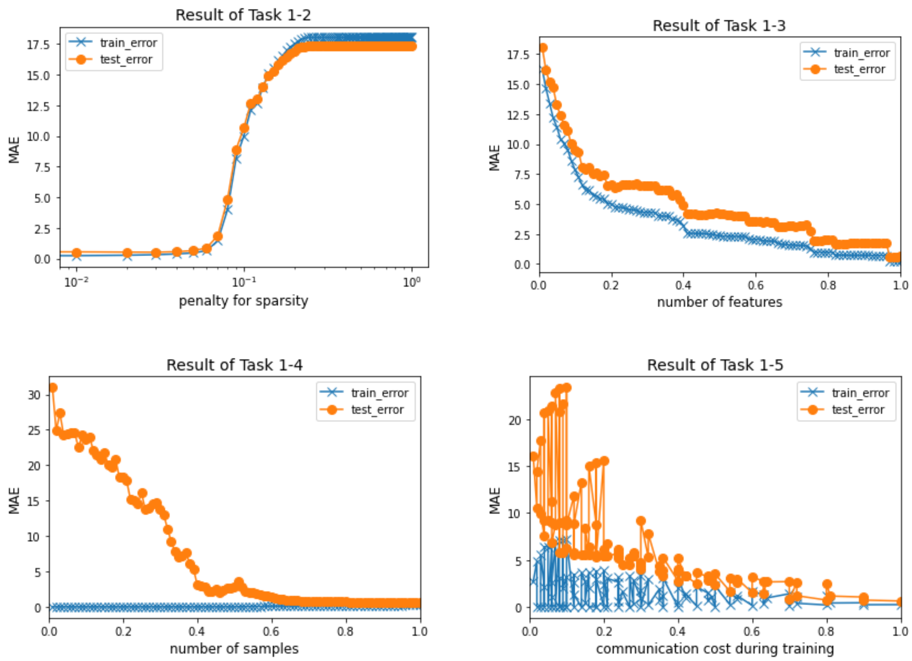

Figure 1: Offline solution of **Synthetic Dataset**

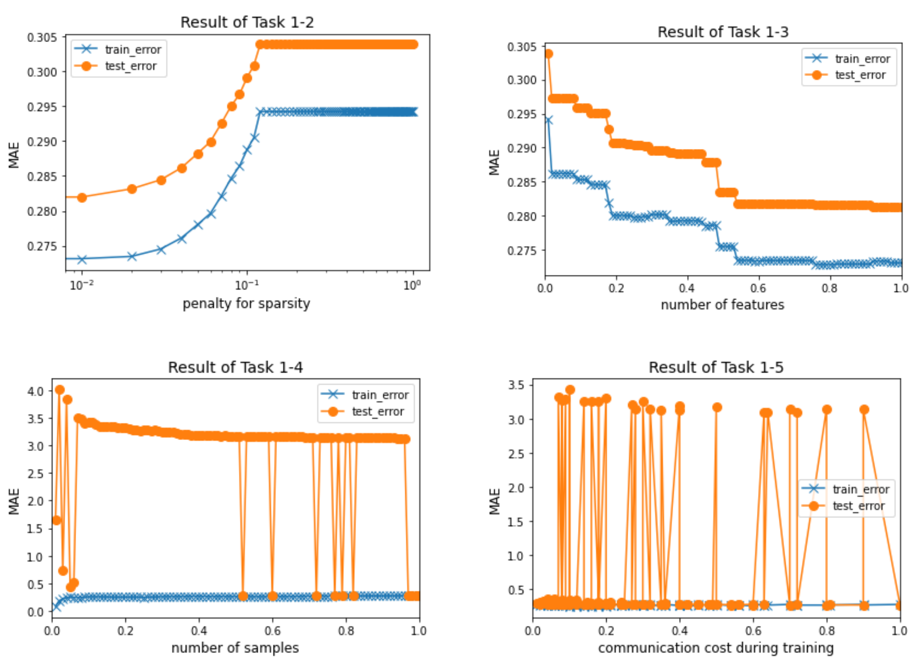

Figure 2: Offline solution of **Online News Popularity Dataset**

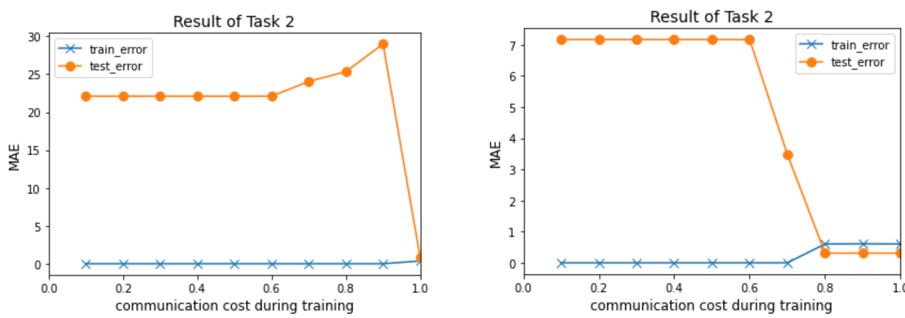

Figure 3: Online solution of **Synthetic Dataset** (left) and **Online News Popularity Dataset** (right)

1.4 Sample Selection

Each feature is generated from i.i.d. standard norm. The "preferred" sample should locate around 0, i.e., low standard derivation. We formulate the sample selection process by solving below linear programming problem

$$\begin{aligned} &\min_{\boldsymbol{x}} \boldsymbol{c}^T \boldsymbol{x} \\ &\text{s.t. } \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \end{aligned} \tag{3}$$

where $\boldsymbol{c} = [\|\mathcal{X}_1\|_1, \|\mathcal{X}_2\|_1, ..., \|\mathcal{X}_N]\|_1$, $\boldsymbol{x} = [x_1, x_2, ..., x_N]^T$, $\boldsymbol{A} = \begin{bmatrix} 1 \\ -1 \\ I_N \\ -I_N \end{bmatrix}$, and $\boldsymbol{b} = [-k, k, 0, 0]^T$. The variable is $x_i = \{0,1\}$ and $\|\mathcal{X}_i\|_1$ is the 1-norm of each row. The linear programming problem is to pick k maximum 1-norm from $\mathcal{X}$.

1.5 Sample and Feature Selection

Combining aforementioned sample and feature selection, the experiments include all the results of percentage of sample and feature in range of (0,1), respectively.

2 Online Solution

The weight and bias are initialized in sensor node with random values. The error with current weight and bias is then evaluated at every time step. The data sample is sent to central node only when the error is greater than a prescribed threshold. The central node updates the weight and bias via linear programming by accepted data and feed the new weight and bias to the sensor node.