# MLoN Computer Assignment 5

Group 1

Split "MNIST" dataset to 10 random disjoint subsets, each for one worker, and consider SVM classifier in the form of $min_w \frac{1}{N} \sum_{i \in [N]} f_i(w)$ with $N = 10$.

## A

### Problem

Run decentralized GD (from Lecture 6) with 10 workers. Characterize the convergence against the total number of signaling exchanges among all nodes, denoted by $T$.

### Solution

We run decentralized GD with 10 workers follow the computer assignment 4 with signal exchanges among nodes. The convergence against the total number of signaling exchanges among all nodes are shown in Fig. 1. The Grey scale model of each number is shown in Fig. 2

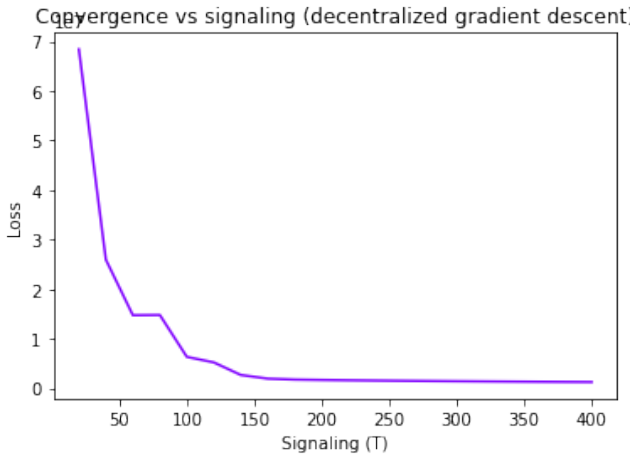


Figure 1: Convergence vs signaling (decentralized gradient descent)

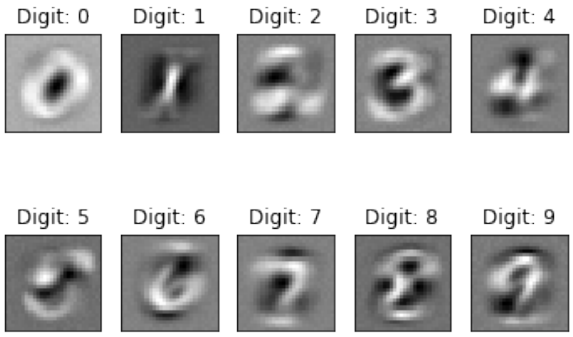


Figure 2: Grey scale model (decentralized gradient descent)

# B

## Problem

Consider a two-star topology with communication graph (1,2,3,4)-5-6-(7,8,9,10) and run decentralized subgradient method (from Lecture 6) and ADMM over the network (from Lecture 7). Characterize the convergence against $T$. Tune hyper-parameters to improve the convergence rate.

## Solution

For the decentralized subgradient method, we reuse the one in CA5. The convergence against the total number of signaling exchanges among all nodes are shown in Fig. 3. The Grey scale model of each number is shown in Fig. 4

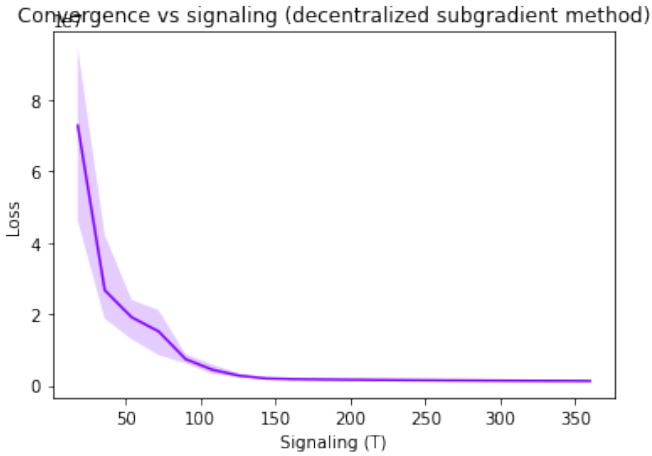


Figure 3: Convergence vs signaling (decentralized gradient descent)

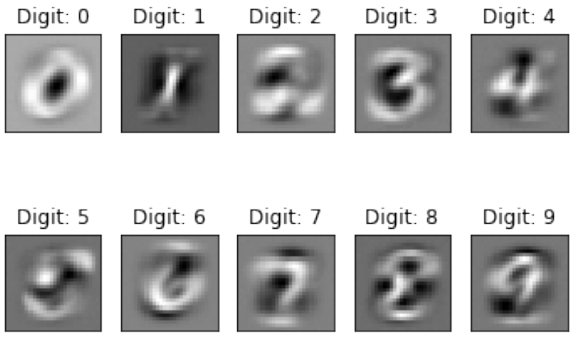


Figure 4: Grey scale model (decentralized gradient descent)

For the Alternating direction method of multipliers (ADMM) algorithm, the convergence against the total number of signaling exchanges among all nodes are shown in Fig. 5. The Grey scale model of each number is shown in Fig. 6

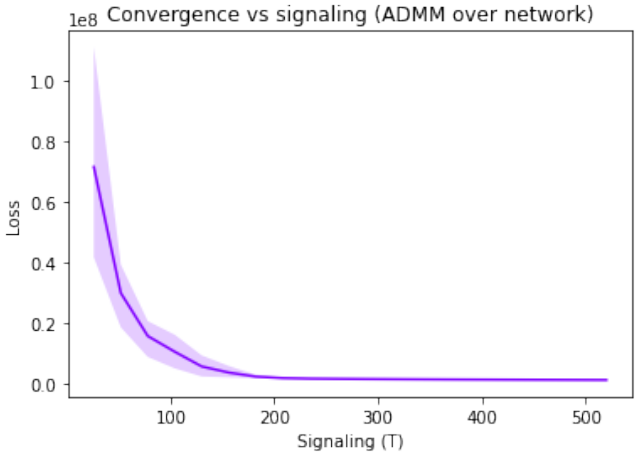


Figure 5: Convergence vs signaling (ADMM)

We also tune the hyper-parameter $\lambda$ to improve the convergence rate. The convergence curves are shown in Fig. 7. This parameter

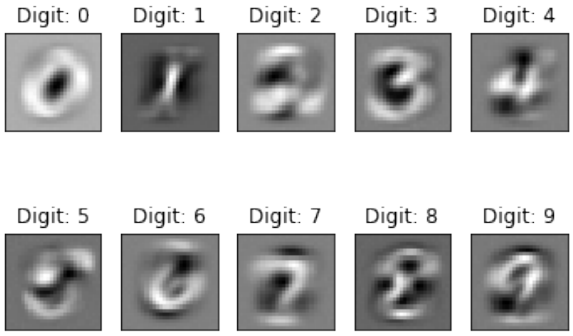


Figure 6: Grey scale model (ADMM)

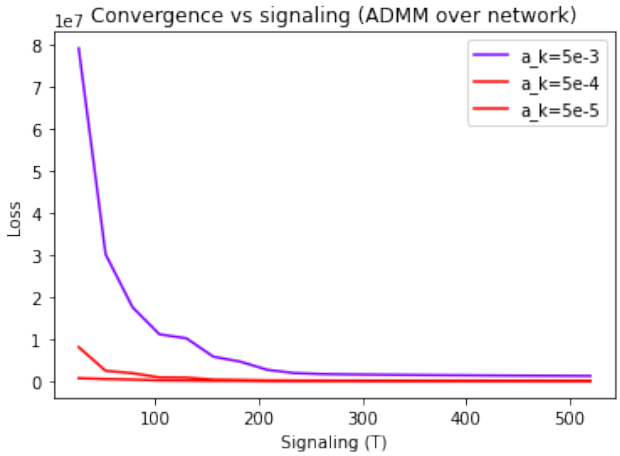


Figure 7: Convergence vs signaling with different $a_k$ (ADMM)

# C

## Problem

Propose an approach to reduce $T$ with a marginal impact on the convergence. Do not limit your imaginations and feel free to propose any solution. While being nonsense in some applications, your solution may actually make sense in some other applications. Discuss pros and cons of your solution and possibly provide numerical evidence that it reduces $T$.

## Solution

We propose to reduce T by increasing the number of stars. As shown in Fig. 8. With more stars, the number of signaling exchanges decrease. In this way, the model is more decentralized, but the number of master nodes increased, which makes the process more complex.

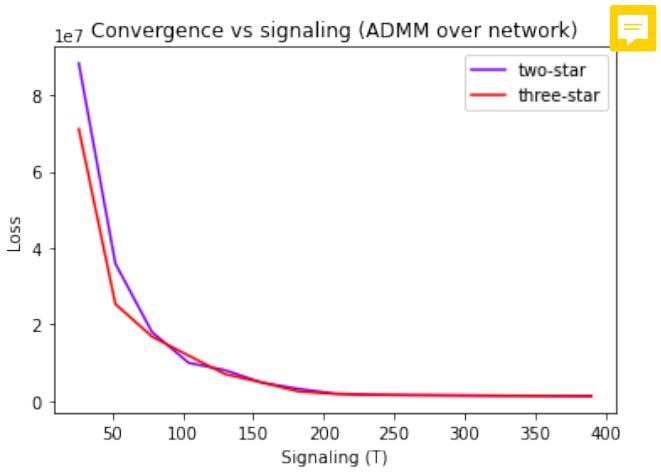


Figure 8: Convergence vs signaling with different $a_k$ (ADMM)