The Role of Adaptive Optimizers for Honest Private Hyperparameter Selection

Shubhankar Mohapatra[1], Sajin Sasy[1], Xi He[1], Gautam Kamath[1], Om Thakkar[2]
University of Waterloo[1], Google[2]

Problem

Consider a sensitive dataset D split into a training set and validation set. A trusted curator with a total privacy budget of (ϵ_f, δ_f) wants to train a model which achieves high accuracy on the validation set. This budget must account for the cost of any queries performed for the sake of hyper parameter selection.

The most popular DP optimizer, DPSGD, has five hyperaparameters :

1. Iterations (T)
2. Lot size (L)
3. Learning Rate (α)
4. Clipping Threshold (C)
5. Noise Scale (σ)

DPMomentum requires additional momentum tuning

Relationship of α and C

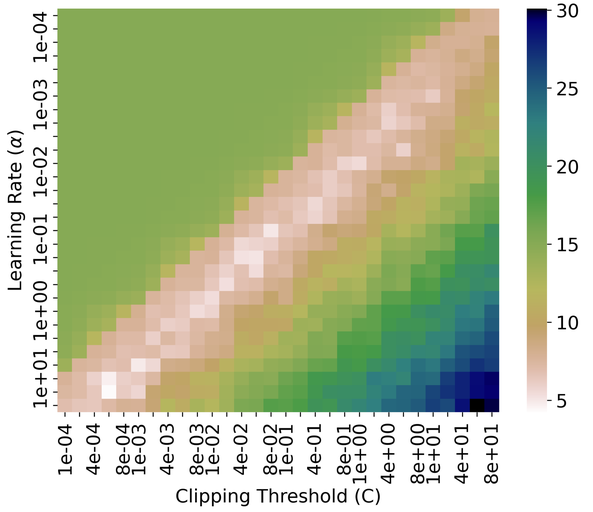

Figure 1: Log of training loss for simulation at σ = 4. The white pixels (lowest loss) lie on a diagonal expressing an inverse relationship between LR and C.

Cost of privately tuning DP optimizers

We investigate hyper parameter tuning using Moments Accountant (MA) and Liu and Talwar (LT).

LT is a randomly stopping approach which allows to privately select the best candidate seen by the algorithm from a selected pool of candidates. The random stopping is controlled by the parameter Gamma.

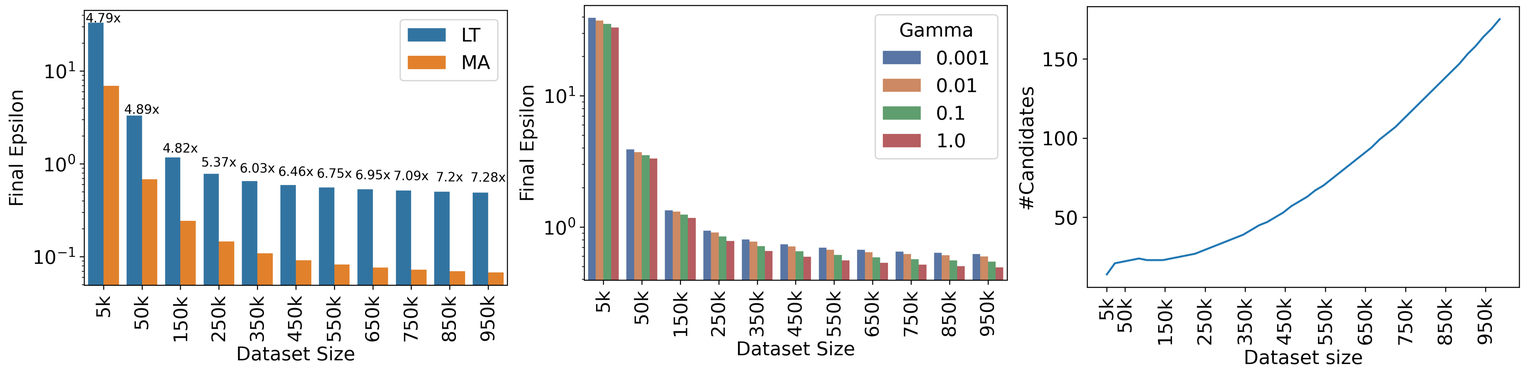

Figure 2: Comparing the privacy cost of LT vs MA. The minimal privacy overhead incurred by LT is at least ~5x, and increases with the dataset size (left). However, as we allow LT to test more hyper parameters, the privacy cost barely increases (middle). MA is able to test a significant number of candidates at the same cost as the minimal overhead cost of LT (right).

Comparing private optimizers

We compare various private optimizers and show that adaptive optimizer, DPAdam requires the least tuning and achieves consistent performance. Our experiments are performed with constant values of T, L and σ.

1. Tuning using DPSGD and DPMomentum requires larger grid search. DPAdam saves tuning of one hyper parameter (α).
2. The defaults of non-private Adam work for its private counterpart, DPAdam.
3. The best candidates of DPSGD, DPMomentum and DPAdam perform similarly.
4. DPSGD requires higher privacy cost for tuning due to larger candidate pool.
5. DPAdam due to fewer candidates can be tuned using MA for a lesser privacy cost.
6. The learning rate of DPAdam converges after few iterations of training, which we call effective step size (ESS).
7. DPAdamWOSM, a novel optimiser whose learning rate is initialised with ESS starts with this converged learning avoids the second moment computation and enjoys better accuracy at earlier iterations.

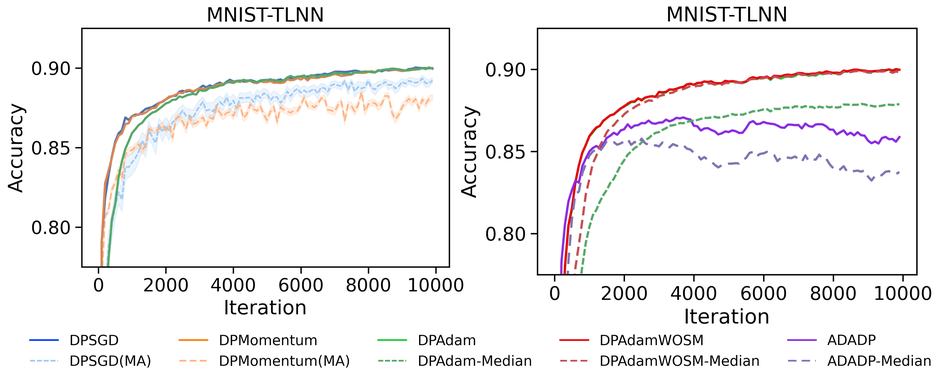

Figure 3: Comparing the testing accuracy curves of different optimisers across their hyper parameter grids with σ = 4.

References:
1.Jingcheng, Liu and Talwar, Kunal. "Private selection from private candidates." *ACM STOC*. 2019.
2.Abadi, Martin, et al. "Deep learning with differential privacy." ACM CCS 2016.