Background

c-differential privacy (DP) bounds the influence of any single input on the output of a computation. DP machine learning bounds the leakage of training data from a trained model. The c-parameter controls this bound and thus the tradeoff between "privacy" and accuracy of the model.

But the drop in accuracy is not equal across different groups.

Experiments

- MNIST: combining common regularizers clipping and noise together significantly reduces accuracy on smaller classes.

- Real datasets: fair models turn unfair, unfair models become more unfair.

- Naturalist dataset: accuracy for DP model significantly drops on smaller-size classes, but it's not only the size that causes this drop.

Federated Learning

- Federated learning approach trains individual models on participants’ data and aggregates produced models into a single global model
- DP preserves participant-level privacy
- We use Reddit dataset of 80,000 participants
- DP-models outperform simpler vocabularies:

Hyperparameters + size effects:

- Accuracy drop depends on (a) clipping and noise values, (b) batch size, (c) number of epochs, and depends the most on (d) class size.

Differential Privacy Has Disparate Impact on Model Accuracy

Eugene Bagdasaryan, Omid Poursaeed, Vitaly Shmatikov @ Cornell Tech
eugene@cs.cornell.edu