Budget Conservation in the Training of Differential Private Models

Amer Elsheikh, Hanfei Lin, Baha Topbas, Yulan Zhang

Academic Mentor: Siting Liu

Final Presentation for Research in Industrial Projects for Students (RIPS) 2022
I. Background
II. Ensemble Accuracy
   A. Algorithm
   B. Results
   C. Summary
III. Questions
I. Background
Consider...

- Open dataset for training models, i.e. medical record, political opinion survey ...
- Protect Respondents’ Privacy!
Motivating Example for Differential Privacy (DP)

<table>
<thead>
<tr>
<th>ID</th>
<th>Tumor?</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>...</td>
</tr>
</tbody>
</table>

Tumor: No

<table>
<thead>
<tr>
<th>ID</th>
<th>Tumor?</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>...</td>
</tr>
</tbody>
</table>

No Tumor: Patient's Privacy Exposed

QUERY

TUMOR = 2

OUT

2

QUERY

TUMOR = 1

OUT

1
Motivating Example for Differential Privacy (DP)

The output for the count query will be drawn from a random distribution centered at the true value.
Differential Privacy (DP)

**Definition: (ε-differential privacy)**
Randomized algorithm $M$ is $\varepsilon$-differentially private (DP) if for all neighboring datasets $D$ and $D'$ and all sets of outcomes $S$:

$$e^{-\varepsilon} \leq \frac{\Pr[\text{Outcome } M(D) \text{ is in } S]}{\Pr[\text{Outcome } M(D') \text{ is in } S]} \leq e^{\varepsilon}$$

Intuitive: $M$ is epsilon differentially private if for all neighboring datasets $D$ and $D'$, their probabilities of observing any outcomes under $M$ differ by a factor of at most $\exp(\varepsilon)$.

**Remarks:**
- $\varepsilon$ quantifies the **privacy cost** of the procedure.
  - If $\varepsilon \to 0$, then no user information is leaked, so privacy cost is 0.
  - If $\varepsilon$ is large, more user information is leaked, so privacy cost is high.
- Composition Rule: the cumulative privacy cost of DP procedures applied in sequence is at worst additive in epsilon.
  - Let us set and track **privacy budget** for iterative DP procedures.
At present, the amount of privacy consumed for each model trained is too high to support a practical number of model-building iterations.
Towards Practical Model Building

**GOAL**

Make DP model building more practical

Make model training **cheap!**

Only release performance metrics.

- Restrict our attention to classification models.
- Only release performance metrics instead of all model parameters
  - Focus on test accuracy
- Subsample & Aggregate → uses an ensemble vote to estimate test accuracy.
II. Ensemble Accuracy
Releasing Test Accuracy

- Training Data
  - (Train ML model)
  - ML Model
    - Arbitrary complexity
    - (Get test accuracy)
  - Test accuracy
    - Add noise
    - DP Test Accuracy

- Scaled differently for each model class
Ensemble Accuracy: Applying Subsample and Aggregate

1. Training Data
2. Apply subsampler
3. Train ML model
4. Get test accuracy
5. Add noise
6. DP Estimated Test Accuracy

Teacher 1
- Subset 1
- Teacher 1

Teacher n
- Subset n
- Teacher n

Identical model type and parameters
Subsampler: Randomized Class-Balanced Partition

Disjoint subsamples preserve class balance of training set.
Aggregator: Report Noisy Arg Max

Proposition: The Report Noisy Arg Max algorithm is $\varepsilon$-differentially private.

Works for any machine learning model.

Histogram of teacher accuracies

Noisy Histogram

Add Lap($1/\varepsilon$) noise to each bin count.

Return bin with highest noisy count.

Ensemble Accuracy: Evaluation

- **Training Data**
  - (Subsample)
  - Subset 1
  - Subset n
  - Teacher 1
  - Teacher n

- **Aggregator**
  - (Get test accuracy)

- **Reference Model**
  - (Train)
  - Same ML Model!

- **DP Estimated Test Accuracy**
  - (Compare)

- **Reference Value of Test Accuracy**
  - (Test accuracy)
## Experimental Setup: Effect of Number of Teachers

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Classes</th>
<th>Features</th>
<th>Training/Test Samples</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCI Adult (Census)</td>
<td>2</td>
<td>14</td>
<td>30162 / 16281</td>
<td>Tabular</td>
</tr>
<tr>
<td>MNIST (Digits)</td>
<td>10</td>
<td>784</td>
<td>60000 / 10000</td>
<td>Image</td>
</tr>
<tr>
<td>KDD CUP 99 (downsampled)</td>
<td>4</td>
<td>41</td>
<td>78544 / 28017</td>
<td>Tabular</td>
</tr>
</tbody>
</table>

**Models:**
- Logistic Regression (LR)
- Random Forest (RF)
- Multilayer Perceptron (MLP)

\[
\varepsilon = \frac{\ln(3)}{10} \approx 0.11
\]

Uniform histogram bins for aggregator: 0.05
Ensemble Accuracy Results on UCI Adult Logistic Regression

- 10 trials per data point.
- Plot midpoint of median histogram bin.
- Shaded region is IQR of bin midpoints.

- **Noisy beginning**: if number of teacher is too small, add too much noise.
- **Bad prediction in tail**: if number of teachers is too large, training set for each teacher is too small.
- **Optimal value** of teacher number (~35) falls in between these two regions.

We observed that all 9 experiments has optimal region!
Ensemble Accuracy: Summary

- Consistent behavior across several model classes and real-life datasets.
  - Optimal number of teachers is consistent for fixed $\epsilon$, histogram bins.
  - Good-quality predictions of model test accuracy.

- Suggestions for future work:
  - Investigate empirical relationship between optimal number of teachers, $\epsilon$, and the width of histogram bins.
  - Improve performance with alternative subsamplers and aggregators.
    - E.g. Non-disjoint subsampler, median aggregator.
Q&A