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

Agenda

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)

Tumor:

ID	Tumor?	...
1	1	...
2	0	...
3	0	...
4	1	...

QUERY

TUMOR = 2

OUT

2

No Tumor:

ID	Tumor?	...
1	0	...
2	0	...
3	0	...
4	1	...

QUERY

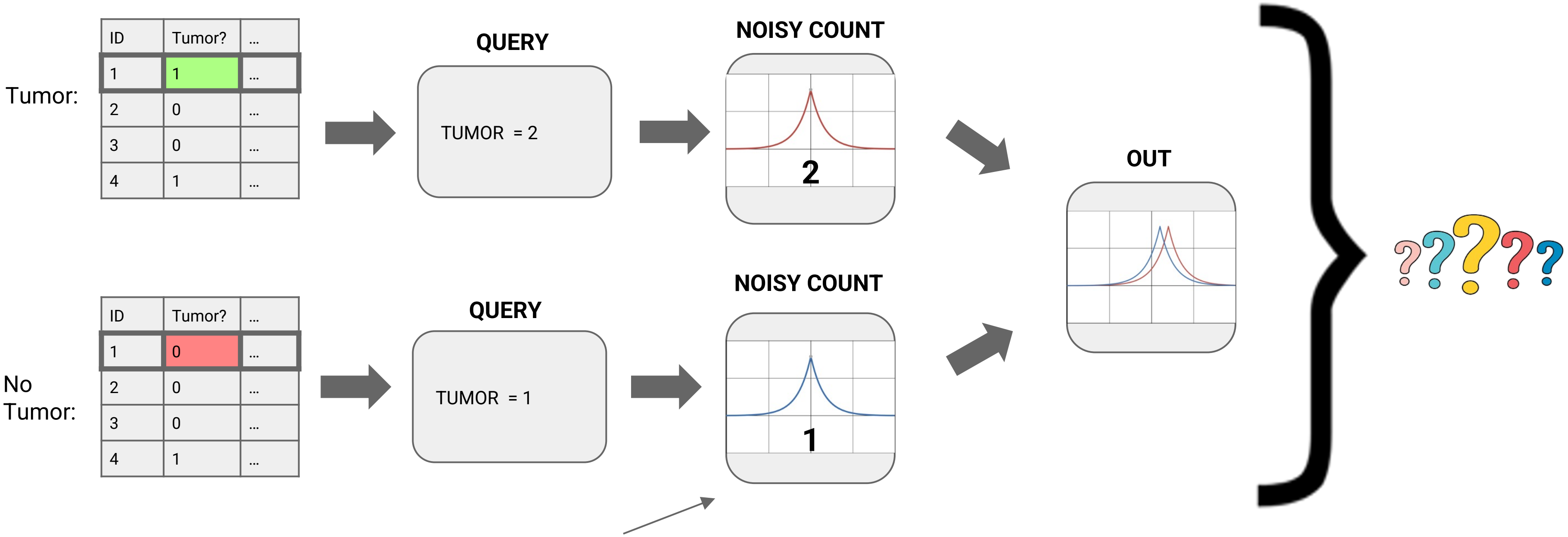
TUMOR = 1

OUT

1

Patient's Privacy Exposed

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 ε-differentially private (DP) if for all neighboring datasets *D* and *D′* and all sets of outcomes *S*:

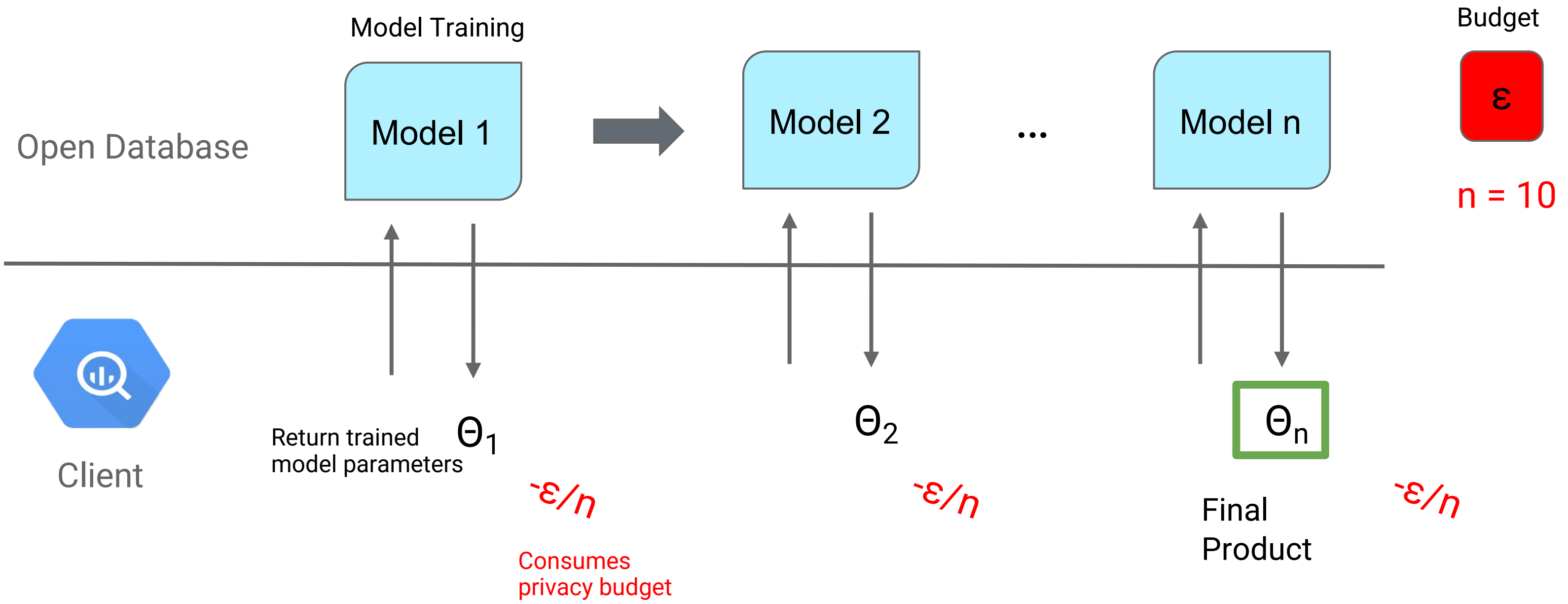
$$e^{-\varepsilon} \leq \frac{\Pr[\text{Outcome } \mathcal{M}(\mathcal{D}) \text{ is in } S]}{\Pr[\text{Outcome } \mathcal{M}(\mathcal{D}') \text{ is in } S]} \leq e^{\varepsilon}$$

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(epsilon).

Remarks:

- ε quantifies the **privacy cost** of the procedure.
  - If ε → 0, then no user information is leaked, so privacy cost is 0.
  - If ε 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.

Model Building Pipeline



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

Ensemble Accuracy

GOAL

Make DP model building more practical

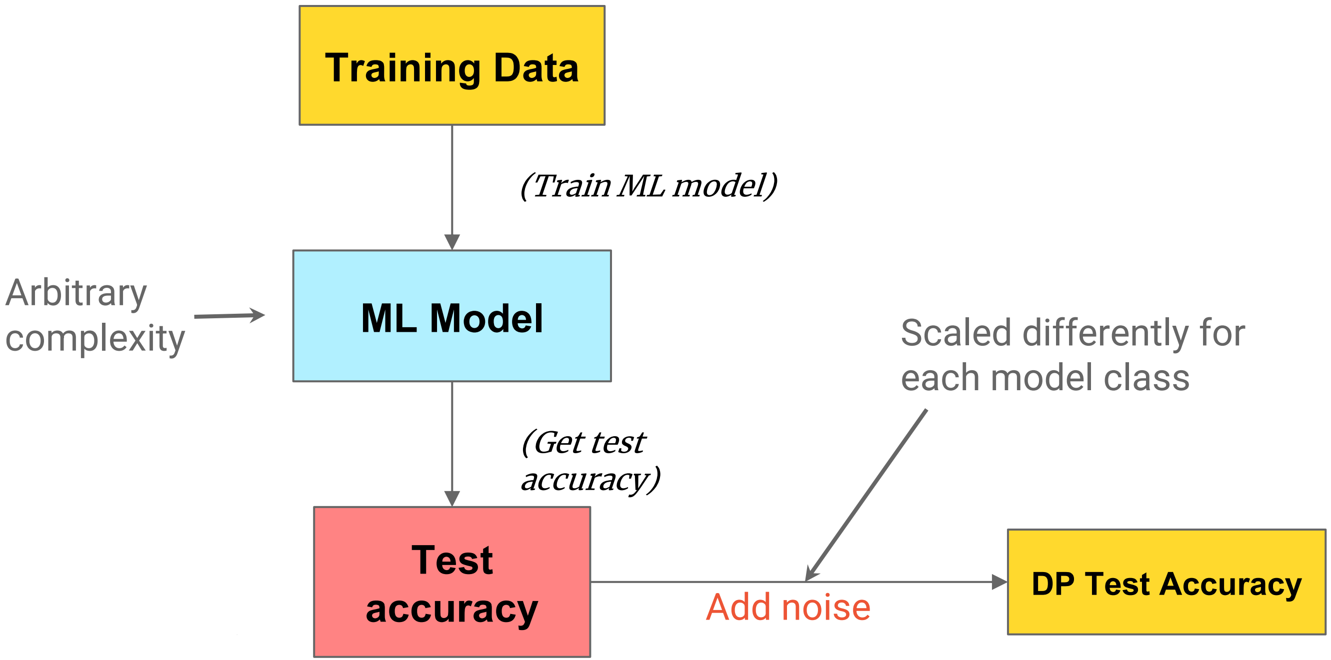
Make model training **cheap**!

Only release performance metrics.

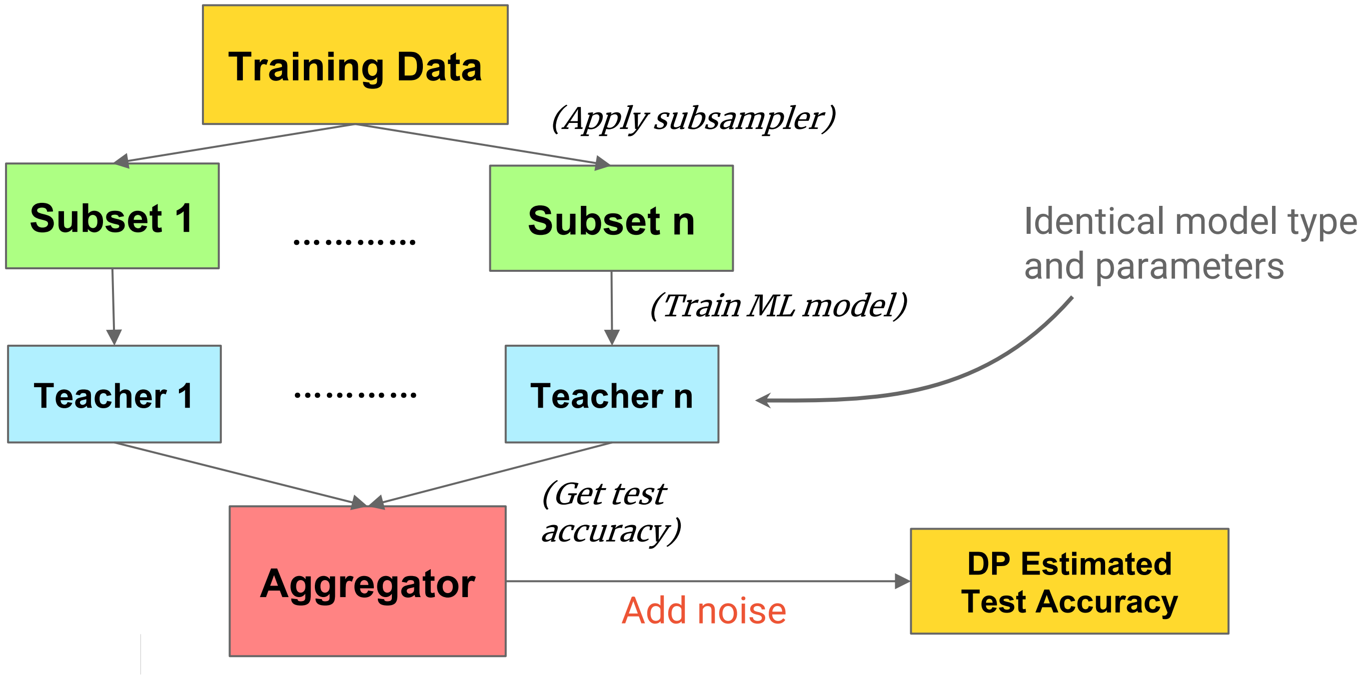
- Restrict our attention to classification models.
- Only release performance metrics instead of all model parameters
    - Focus on test accuracy
- Subsample & Aggregate → uses an ensemble vote to estimate test accuracy.

II. Ensemble Accuracy

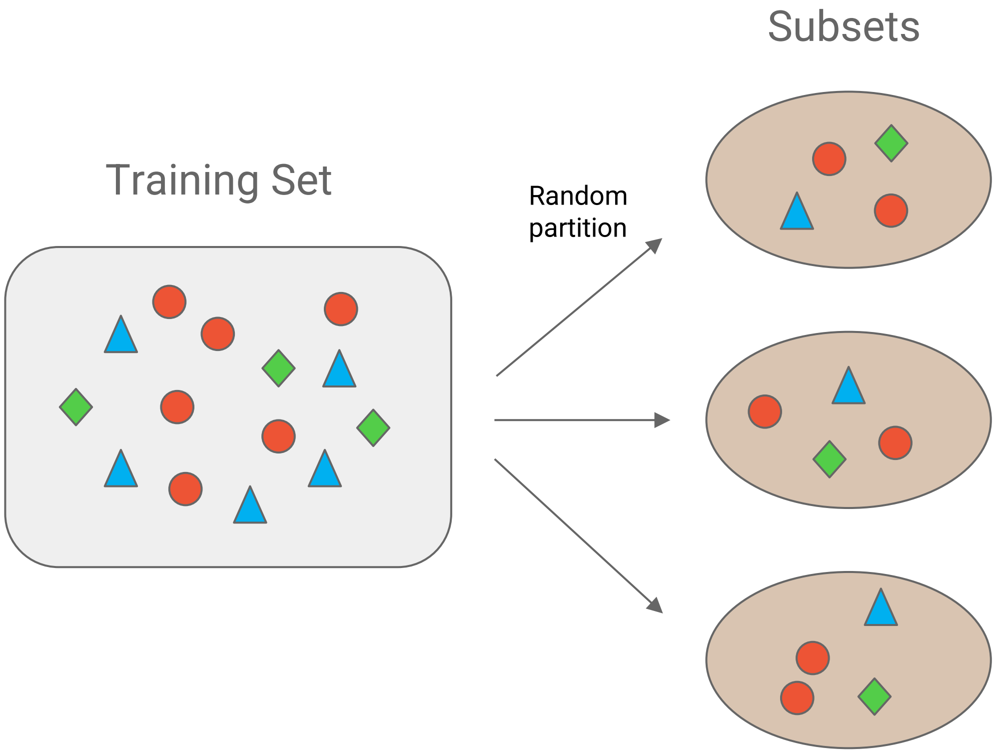
Releasing Test Accuracy



Ensemble Accuracy: Applying Subsample and Aggregate



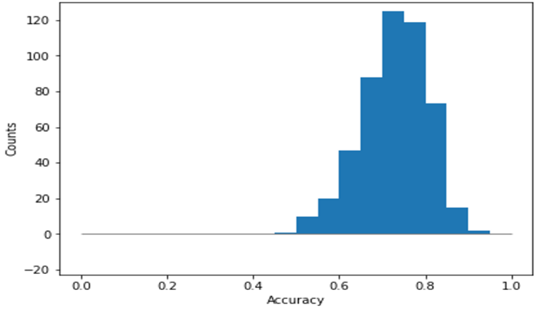
Subsampler: Randomized Class-Balanced Partition



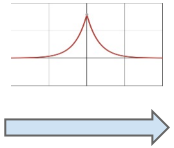
Disjoint subsamples preserve class balance of training set.

Aggregator: Report Noisy Arg Max

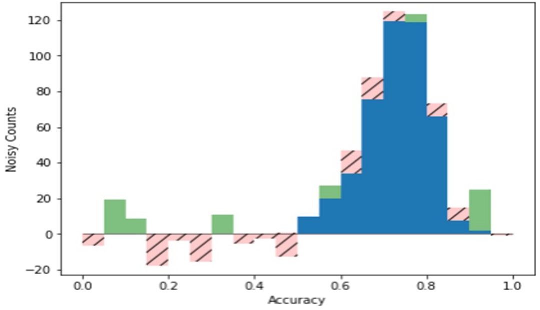
Works for any machine learning model.



Histogram of teacher accuracies



Add Lap(1/ε) noise to each bin count.



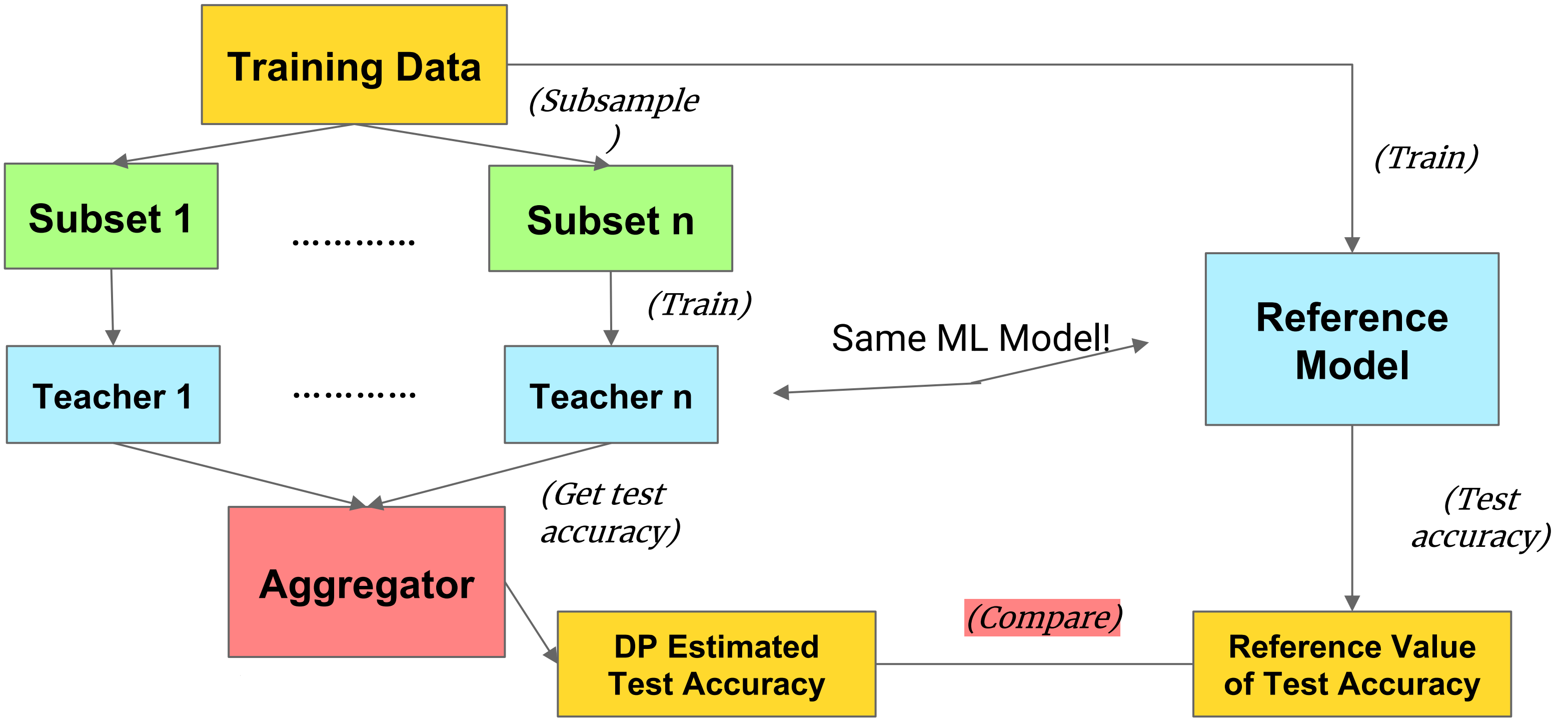
Noisy Histogram

Return bin with highest noisy count.

<u>Proposition:</u> The Report Noisy Arg Max algorithm is ε-differentially private.

Source: C. Dwork and A. Roth, *The Algorithmic Foundations of Differential Privacy* (2014).

Ensemble Accuracy: Evaluation



Experimental Setup: Effect of Number of Teachers

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

Models:

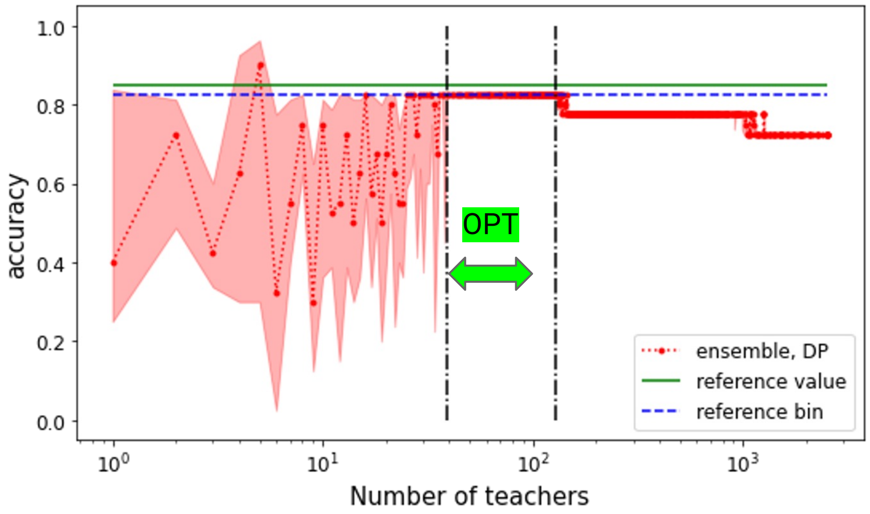
- Logistic Regression (LR)
- Random Forest (RF)
- Multilayer Perceptron (MLP)

$\varepsilon = \ln(3)/10 \cong 0.11$

Uniform histogram bins for aggregator: 0.05

Ensemble Accuracy Results on UCI Adult Logistic Regression

Accuracy v.s. Number of Teachers



- 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 ε, histogram bins.
  - Good-quality predictions of model test accuracy.
- Suggestions for future work:
  - Investigate empirical relationship between optimal number of teachers, ε, and the width of histogram bins.
  - Improve performance with alternative subsamplers and aggregators.
    - E.g. Non-disjoint subsampler, median aggregator.

Q&A