

# DP-SGD vs PATE: Which Has Less Disparate Impact on Model Accuracy?



Archit Uniyal \* Rakshit Naidu \* Sasikanth Kotti  
 Sahib Singh Patrik Joslin Kenfack Fatemehsadat Mireshghallah Andrew Trask

## ① We aim to..

Compare PATE with DP-SGD in terms of fairness.

## ③ What we observe

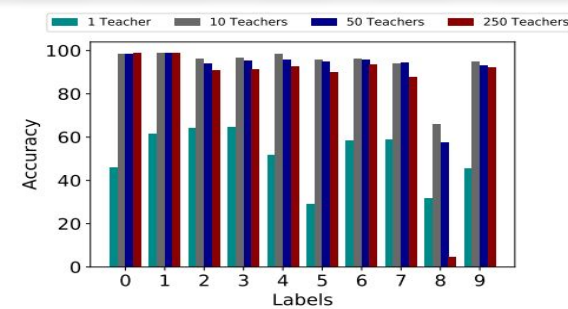
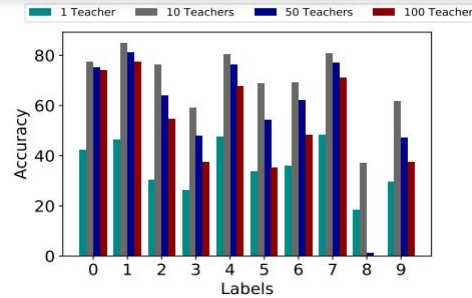
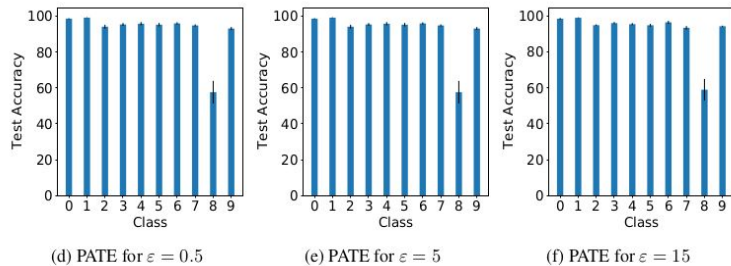
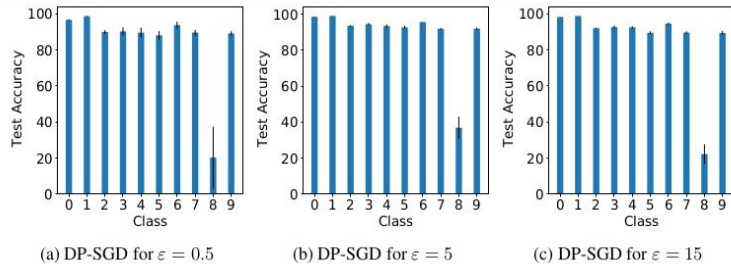
## ② BACKGROUND

### ➤ DP-SGD

A modification of SGD which bounds the sensitivity of each gradient and uses a moments accountant algorithm to amplify and track the privacy loss across weight updates.

### ➤ PATE

Private Aggregation of Teacher Ensembles (PATE) minimizes the leakage of sensitive information from a model on careful analysis.



	Does not need access to a public dataset	Does not need compute to train multiple teacher models	Provide better privacy-utility trade-off [cite]	Has less disparate impact on underrepresented subgroups
DP-SGD	✓	✓	✗	✗
PATE	✗	✗	✓	✓

- PATE has significantly less disproportionate impact on utility compared to DP-SGD. The standard deviation of the accuracy for each class over 5 runs was much lower in PATE compared to DP-SGD.
- Having multiple teachers often provides a higher utility than a single teacher for underrepresented groups. However beyond the tipping point of this ensemble (10 teachers in our case), the utility stagnates and then starts dropping significantly

