Google

AdKDD

Training Differentially Private Ad Prediction Models With Semi-Sensitive Features

Lynn Chua, Qiliang Cui, Badih Ghazi, Charlie Harrison,
Pritish Kamath, Walid Krichene, Ravi Kumar, Pasin Manurangsi,
Nicolas Mayoraz, Krishna Giri Narra, Steffen Rendle, Amer Sinha,
Avinash Varadarajan, Chiyuan Zhang

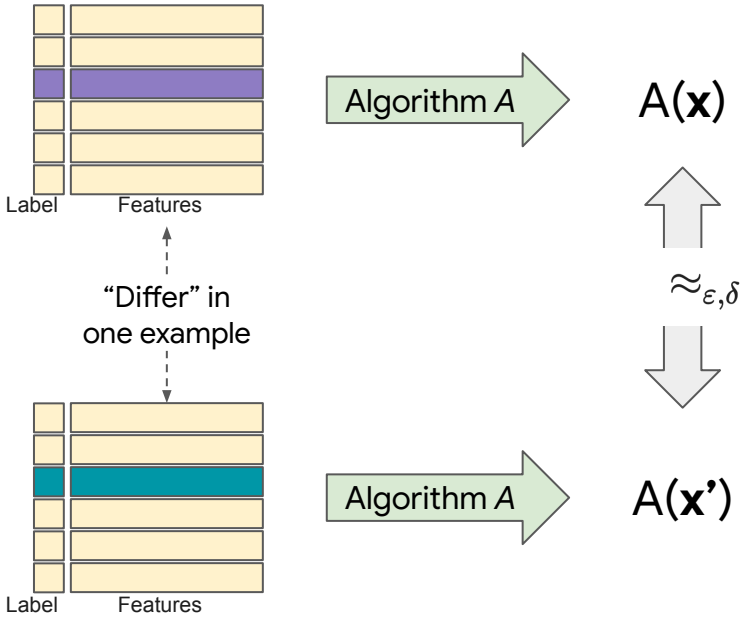
Google

Aug 2024

Motivation

- Ads modeling tasks: predict an ad pCTR or pCVR
- Deprecation of third-party cookies (3PC), which are cross-site identifiers that allow determining user features and labels from sites other than the publisher
- Study setting with **semi-sensitive features**, where some features depend on cross-site information and some do not
- Motivating example:
  - Non-sensitive features: publisher/ad-related features (e.g. publisher site, ad category)
  - Sensitive features: user-related features (e.g. demographics, or presence in a particular remarketing list)
  - User features are private, and mapping between users and publishers is also private.

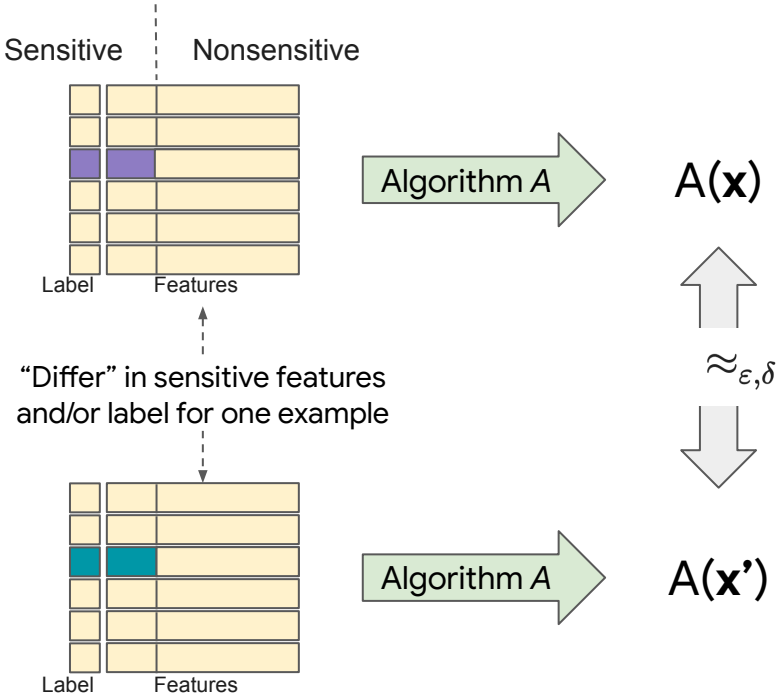
Differential privacy



(ε, δ) -Differential Privacy (DP) [Dwork et al.'06]
For all "adjacent" **x**, **x'** and for all E,

$$\Pr[A(\mathbf{x}) \in E] \leq e^{\varepsilon} \cdot \Pr[A(\mathbf{x}') \in E] + \delta$$

DP with semi-sensitive features

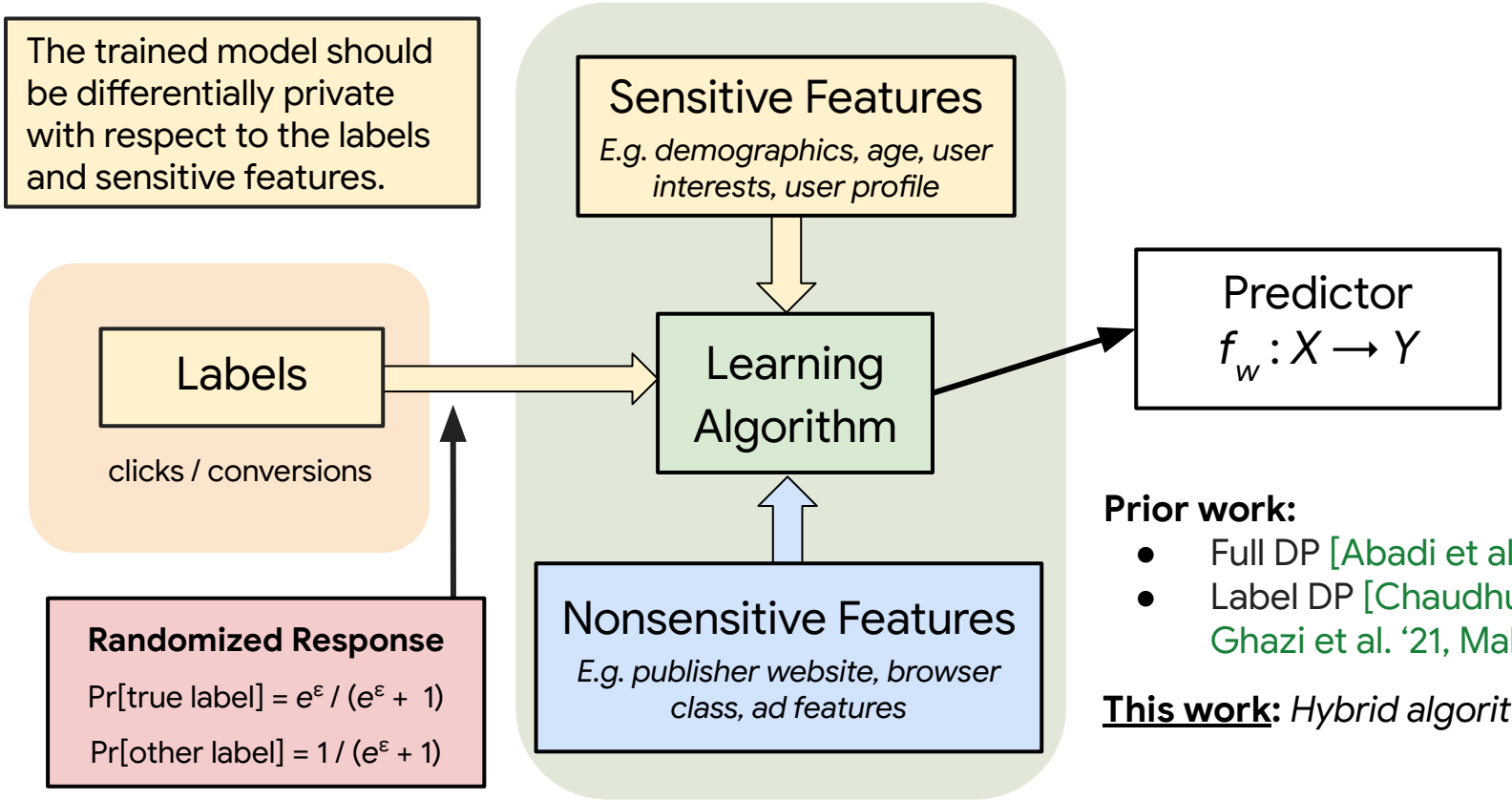


(ε, δ) -Differential Privacy (DP) [Dwork et al.'06]
For all "adjacent" $\mathbf{x}, \mathbf{x}'$ and for all E,

$$\Pr[A(\mathbf{x}) \in E] \leq e^{\varepsilon} \cdot \Pr[A(\mathbf{x}') \in E] + \delta$$

DP Learning with Semi-Sensitive Features

The trained model should be differentially private with respect to the labels and sensitive features.

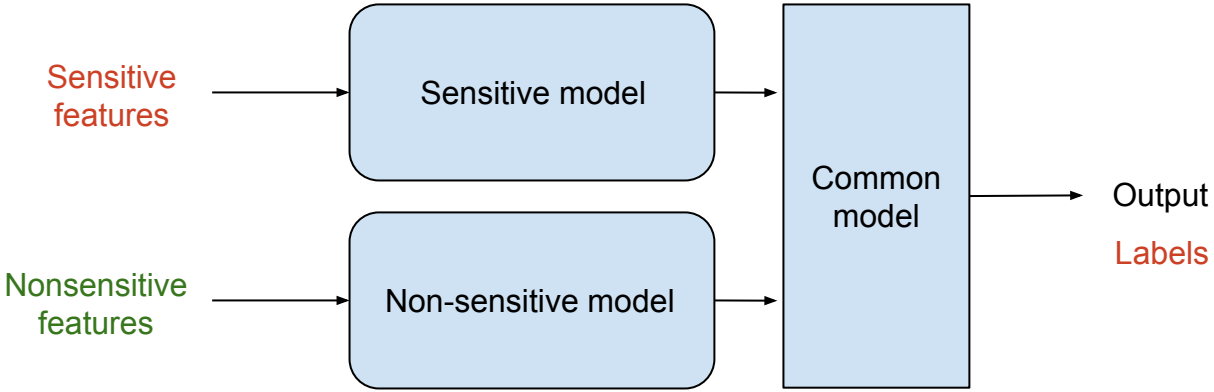


Prior work:

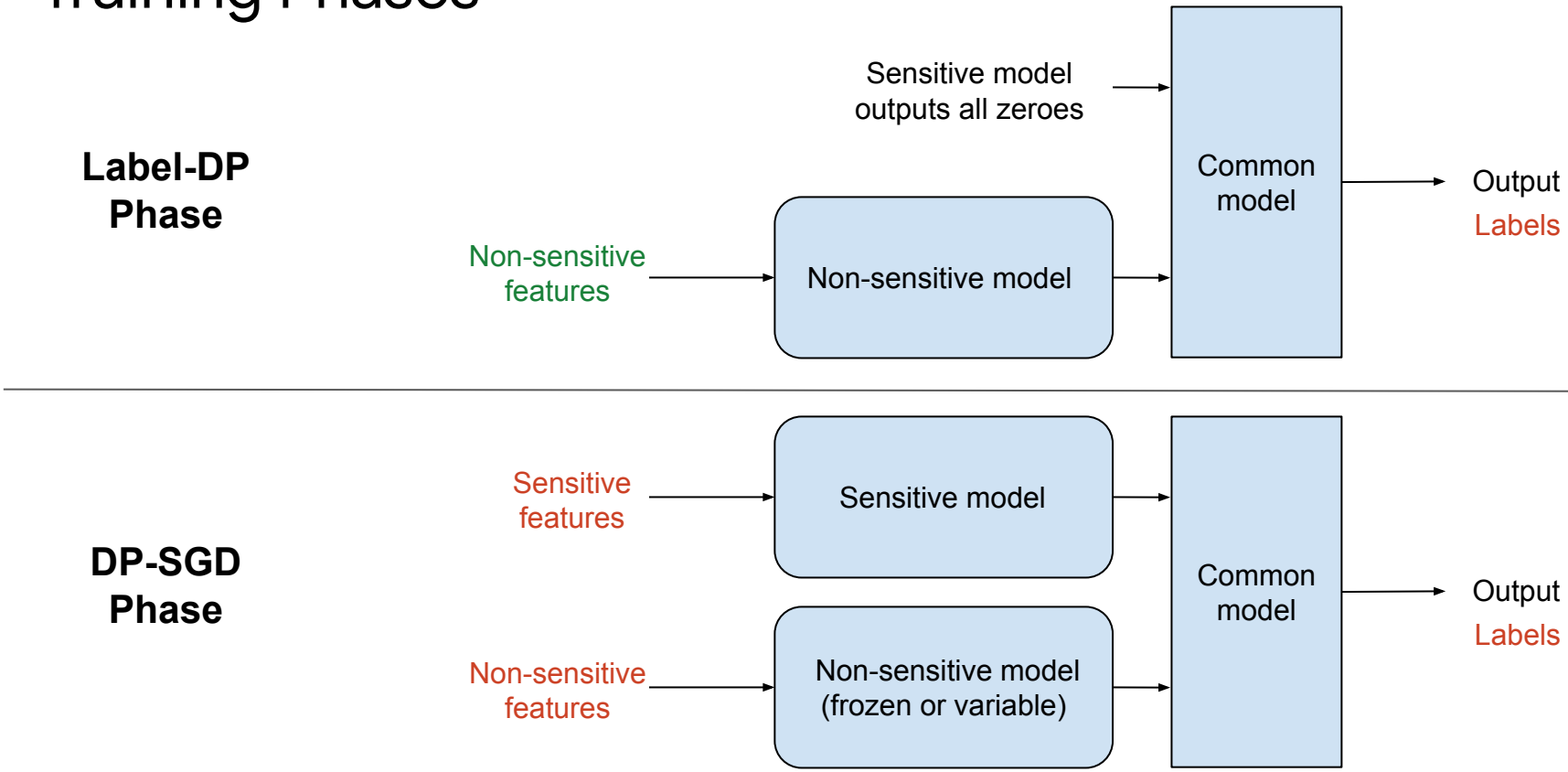
- Full DP [Abadi et al. '16]
- Label DP [Chaudhuri-Hsu '11, Ghazi et al. '21, Malek et al. '21]

This work: *Hybrid algorithm*

Model architecture



Training Phases



Hybrid Algorithm

Total privacy budget (ε, δ) is split between two phases as $\varepsilon = \varepsilon_1 + \varepsilon_2$

Label-DP Phase: Train truncated model with randomized response labels and sensitive embeddings set to 0, with $(\varepsilon_1, 0)$-DP

DP-SGD Phase: Train entire model with (ε_2, δ)-DP with frozen or variable non-sensitive tower

Two baselines:

$\varepsilon_1 = 0$ ⟷ $\varepsilon_1 = \varepsilon$

DP-SGD: all features treated as sensitive

RR: labels privatized, sensitive features discarded

Datasets

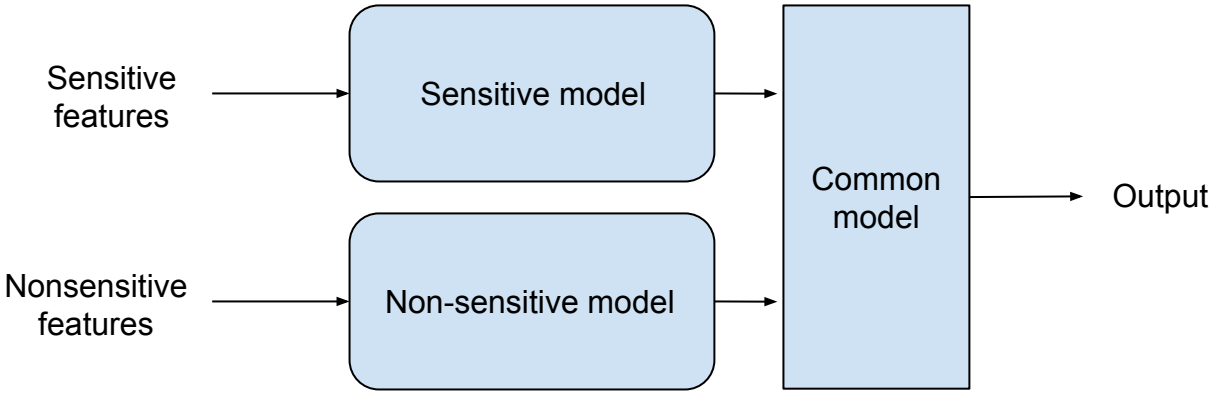
Criteo Display Ads pCTR Dataset

- 40M examples over 7 days of Criteo traffic
  kaggle.com/c/criteo-display-ad-challenge/overview
- Treat even-numbered features as sensitive and odd-numbered features as non-sensitive
- Goal: Predict click

Criteo Sponsored Search Conversion Log Dataset

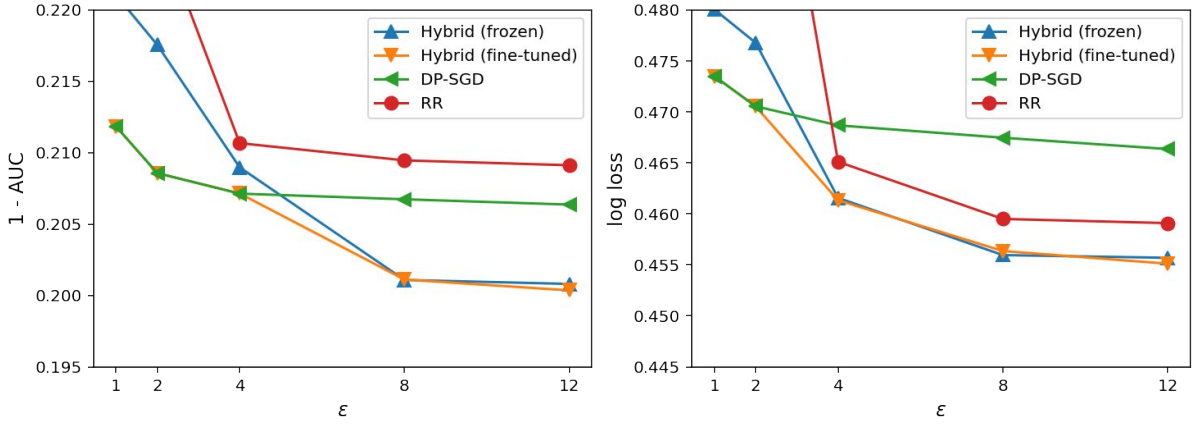
- 16M examples
  ailab.criteo.com/criteo-sponsored-search-conversion-log-dataset
- Sensitive features are device_type, audience_id, user_id
  - Outcome/labels and product_price are omitted
- Goal: Predict sale

Models

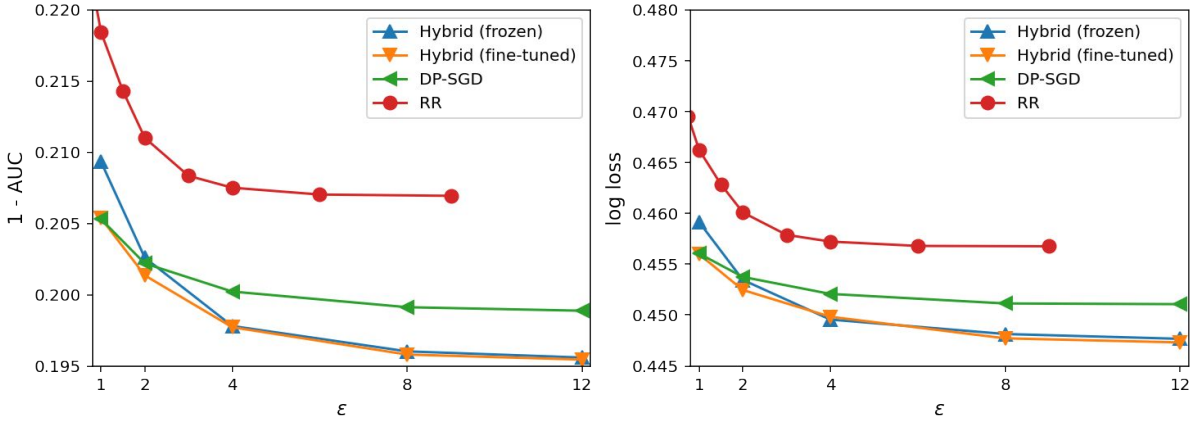


- Multilayer perceptron (MLP)
  - Each model is an MLP
  - Concatenated output layers of Sensitive and Non-sensitive models are input to Common model
- Factorization Machine (FM)
  - Sensitive and Non-sensitive models are embedding lookups
  - Common model is a sum of pairwise dot products between all input embeddings
  - No dense layers

Criteo Display Ads pCTR dataset



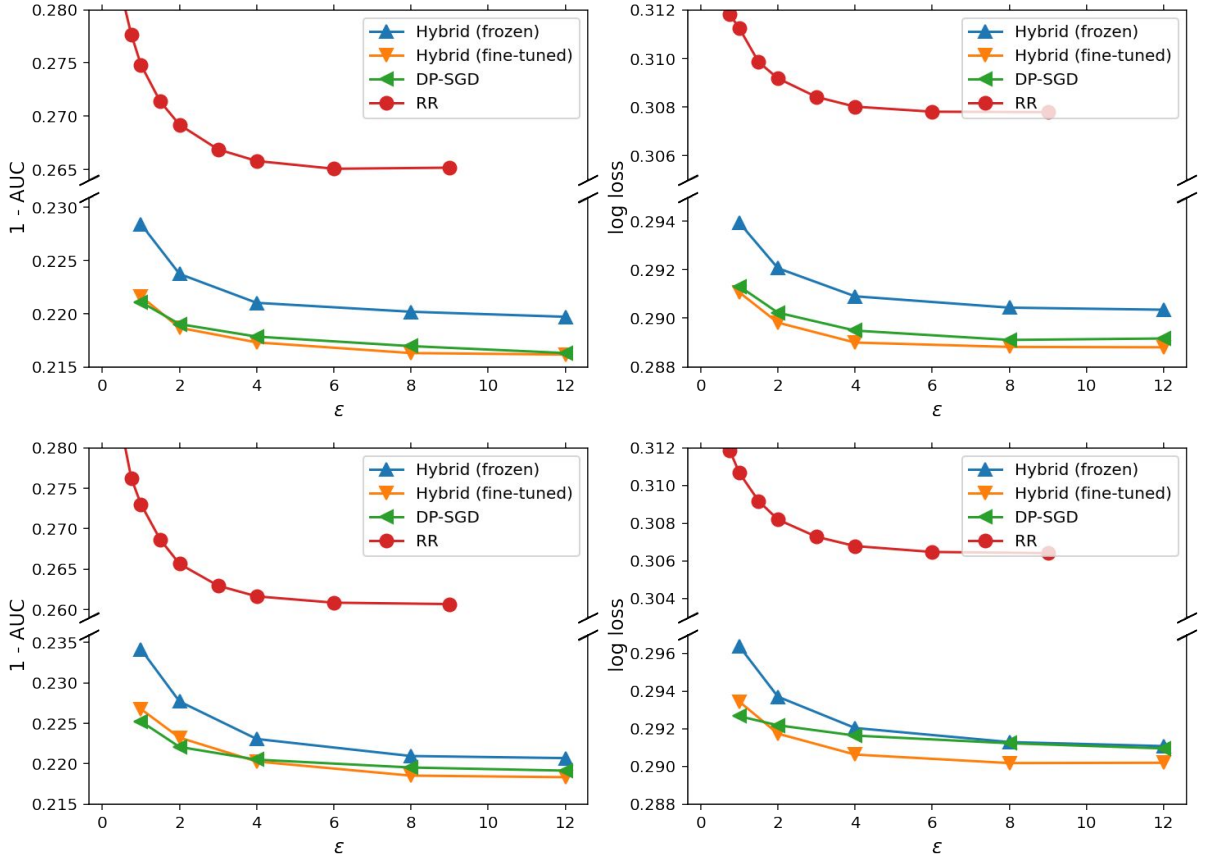
MLP



FM

Hybrid improves over RR and DP-SGD when $\varepsilon \geq 4$

Criteo Sponsored Search Conversion Log dataset

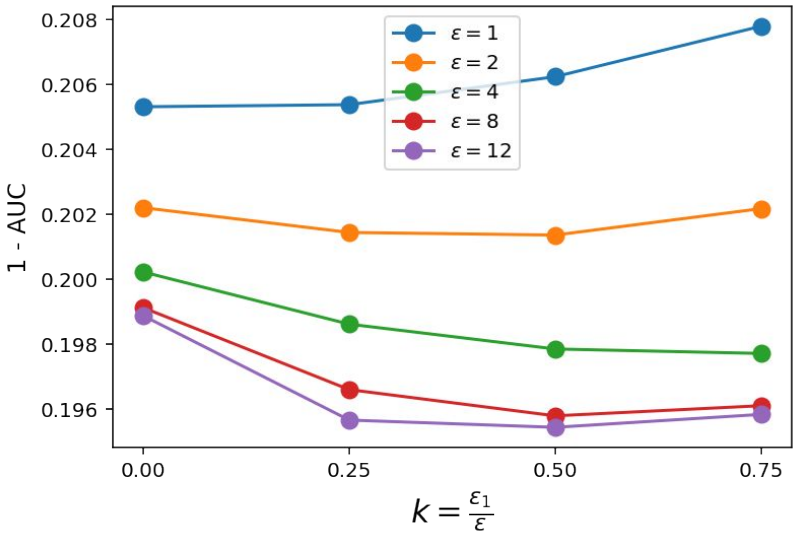


MLP

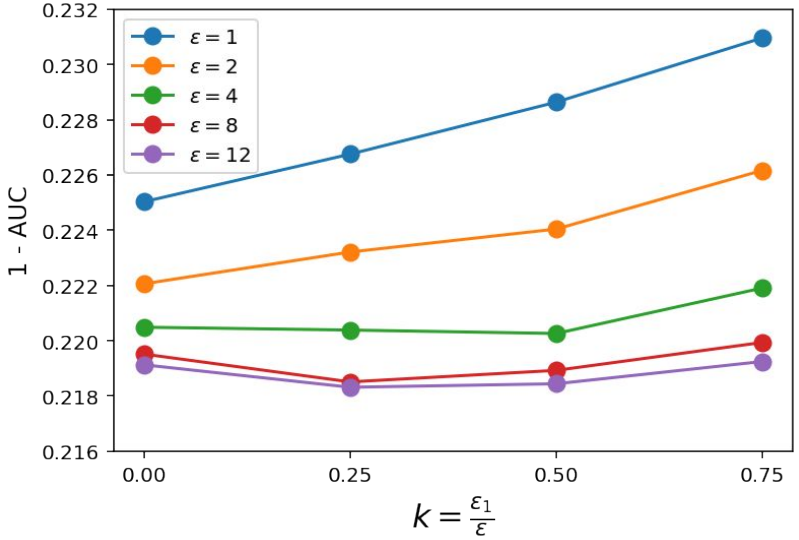
FM

Fine-tuning generally achieves higher utility than freezing

Effect of budget split



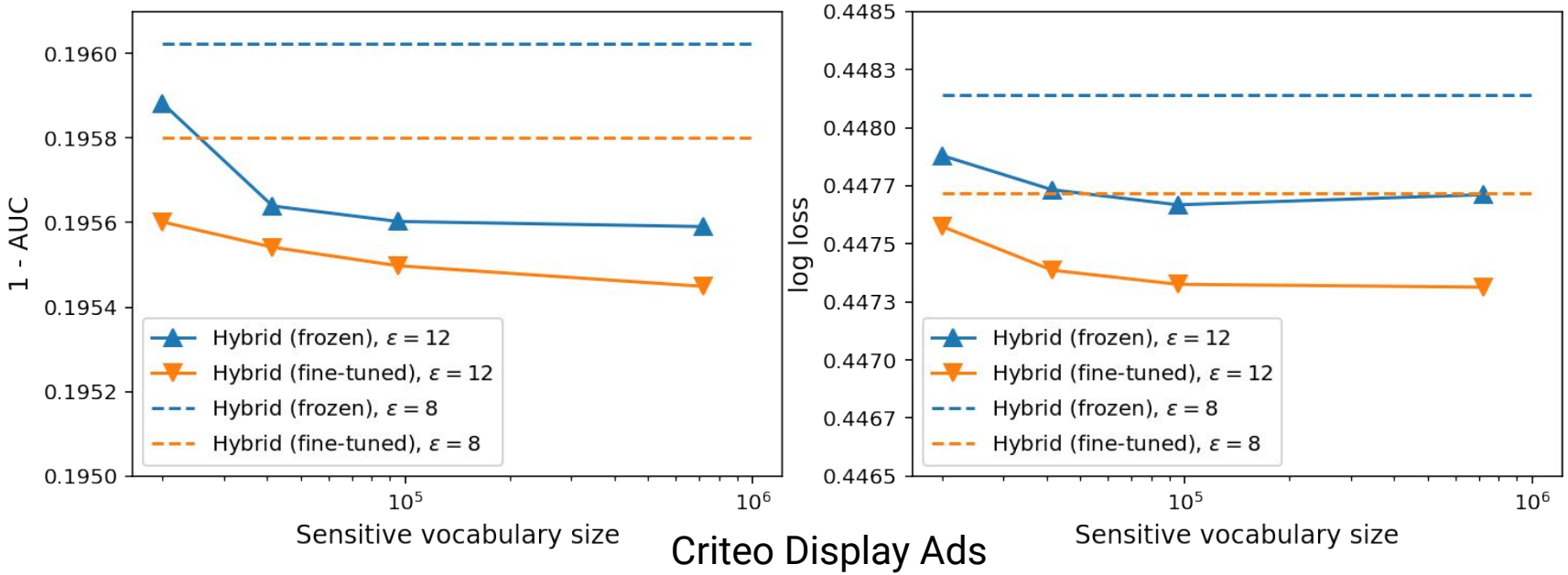
Criteo Display Ads



Criteo Sponsored Search

Budget split *k* should be tuned separately for each ε

Model-size utility trade-off



Criteo Display Ads

Significantly smaller models can be trained without largely sacrificing utility

Conclusion

- Presence of non-sensitive features can improve model quality (compared to treating all as sensitive)
- Hybrid DP algorithm for semi-sensitive features improves over baselines across range of privacy budgets and model sizes
- Requires careful tuning of the budget split

- Future directions:
  - Improving on DP-SGD in the high privacy regime
  - Applying these methods on datasets of different scales