Differentially Private Normalizing Flows for Privacy-Preserving Density Estimation

Stanford University

Columbia University in the City of New York

Chris Waites and Rachel Cummings

Introduction

Density Estimation

- Want to estimate underlying probability distribution of observed data, enabling likelihood estimation and sampling

Data Privacy

- Learning and releasing such an estimate could leak potentially sensitive information if data is linked to individuals
- Solution: release estimate with differential privacy [1] guarantee, where change in distribution of our estimate due to removal of a single individual (D vs. D') is bounded:

$$\Pr[\mathcal{A}(D) \in \mathcal{S}] \leq \exp(\epsilon) \Pr[\mathcal{A}(D') \in \mathcal{S}] + \delta$$

Existing Baselines

- DP-EM [2], which enables privacy-preserving Gaussian mixture models (DP-MoG) through expectation maximization

Approach

Learning Invertible Transformations

- Idea: optimize a sequence of invertible transformations, mapping simple prior distribution to a complex distribution

$$\log p_{\boldsymbol{\theta}}(\boldsymbol{x}) = \log q(f_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x})) + \log \left| \det \left(\frac{\partial f_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x})}{\partial \boldsymbol{x}} \right) \right|$$

- Optimize parameters to minimize negative log likelihood of the data, achieving a privacy guarantee via DP-SGD [3]:

$$\mathcal{L}(\boldsymbol{\theta}) := -\frac{1}{N} \sum_{i=1}^{N} \log p_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})$$

Prior

- Simplistic choice of Gaussian can be improved upon by fitting a Gaussian mixture model using DP-EM to act as prior

Model Architecture

- Masked Autoregressive Flow [4], which composes a sequence of MADE and activation normalization layers

Results

Likelihood Estimation

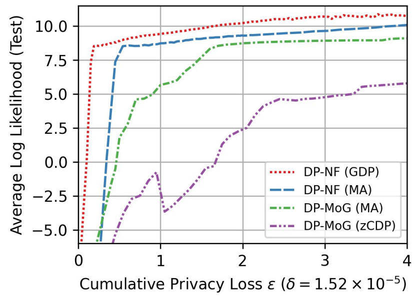

Figure 2: Average log likelihood on a held out test set from the Life Science dataset as a function of the cumulative privacy loss ε.

Synthetic Data Generation

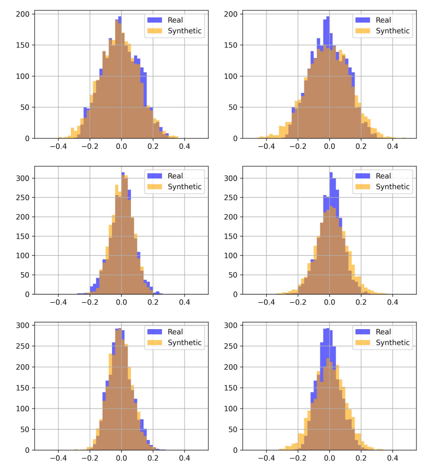

Figure 3: Dimension-wise histograms of synthetically generated Life Science data, superimposed over real data, for ε = 0.6 and δ = 1.52 × 10^{-5}. Left column is our algorithm; right column is our baseline DP-MoG.

Application: Anomaly Detection

Anomaly Detection as Density Estimation

- Can approach anomaly detection through a simple likelihood thresholding mechanism; predicts in-distribution or out-of-distribution depending on whether density exceeds some empirically derived threshold

Experiment

- Generated synthetic anomalies by uniformly sampling points at tail ends of the observed data distribution; proposed approach performs better than DP-MoG and comparably to a non-private mixture of Gaussians

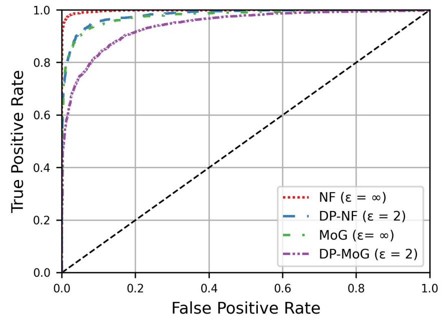

Figure 4: ROC curves displaying true positive rate and false positive rate for private and non-private likelihood threshold models. Privacy expenditure was calculated using the moments accountant with δ = 1.52 × 10–5.

References

[1] Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. in Halevi, S. & Rabin, T. (Eds.) Calibrating Noise to Sensitivity in Private Data Analysis Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4–7, 2006. Proceedings, Springer Berlin Heidelberg, 265-284, doi:10.1007/11681878 14.

[2] Park, M., Foulds, J., Choudhary, K., and Welling, M. (2017). DP-EM: Differentially Private Expectation Maximization. In Singh, A. and Zhu, J., editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, volume 54 of Proceedings of Machine Learning Research*, pages 896–904, Fort Lauderdale, FL, USA. PMLR.

[3] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security,* pages 308–318. ACM, 2016.

[4] Papamakarios, G., Murray, I., and Pavlakou, T. (2017). Masked autoregressive flow for density estimation. In *Advances in Neural Information Processing Systems*, pages 2335–2344.