The Impact of Subsampling on Differentially Private Fraud Detection

Author: Storm van Wassenaar <s.vanwassenaar@student.tudelft.nl>
Responsible Professor: Dr. Zeki Erkin

1. Introduction

- Fraud losses are increasing
- Sharing sensitive data is often not possible due to privacy regulations
- Differential privacy helps, but can affect the quality of the data
- Privacy amplification can help with achieving a stronger privacy guarantee

- $\varepsilon' = \log(1 + \gamma(e^{\varepsilon} - 1))$

- When privacy amplification by subsampling is added the noise added with differential privacy can be reduced

2. Research Question

How does privacy amplification by subsampling affect the privacy-utility trade-off in differentially private fraud detection?

3. Methodology

1. Dataset
   - "Credit Card Fraud Detection" by MLG at ULB
2. Subsampling
   - Poisson subsampling
3. Model training
   - Regularized Logistic Regression
4. Differential Privacy (Perturbation)
   - Output perturbation
5. Evaluation
   - AUPRC, Specificity and Sensitivity

5. Conclusion

- Improvement in AUPRC for small values of epsilon
- Better at distinguishing between fraudulent and legitimate transactions
- Sensitivity slightly decreased

- Depending on the application it can be an improvement

4. Results