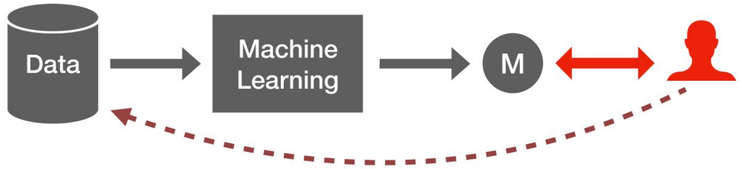
Evaluating Differentially Private Machine Learning In Practice

Bargav Jayaraman and David Evans
University of Virginia, Charlottesville, VA

Privacy Concerns in Machine Learning

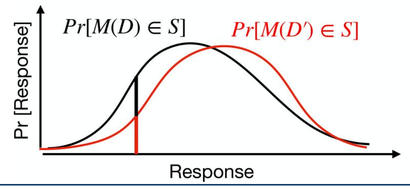


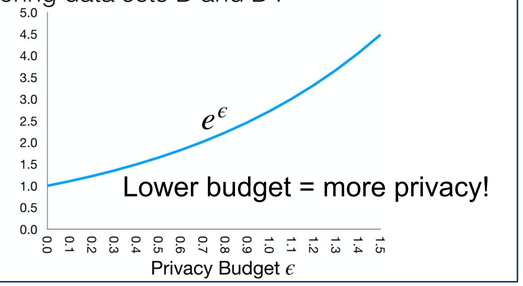
What can be inferred about the training data?

Solution: Differential Privacy

A randomized function M is ϵ-**differentially private** if it produces similar response over both neighboring data sets D and D':

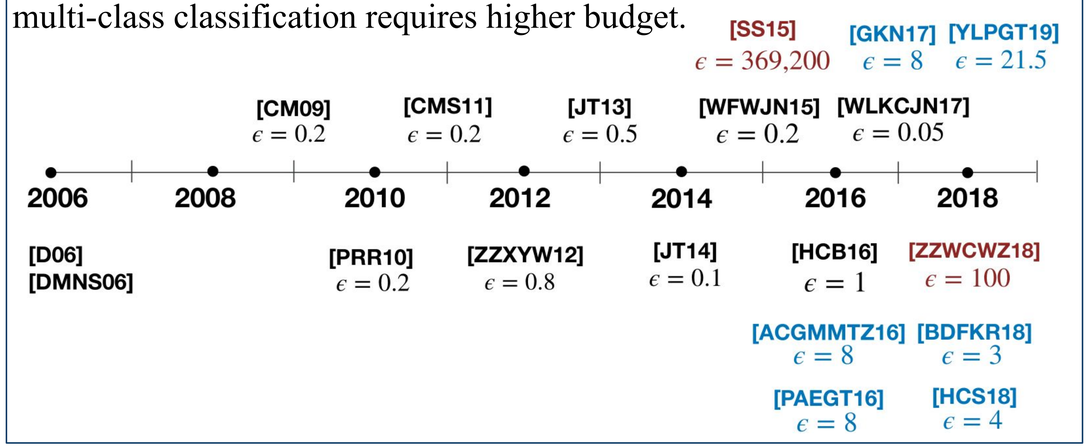
$$\frac{Pr[M(D) \in S]}{Pr[M(D') \in S]} \leq e^{\epsilon}$$



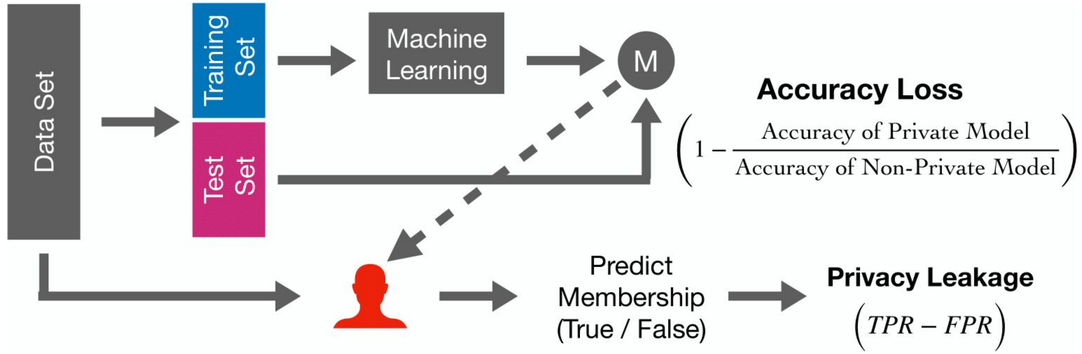


Existing Differentially Private ML Implementations

While private ERM algorithms for binary classification tasks can be achieved with low privacy budget, complex tasks such as deep learning for multi-class classification requires higher budget.

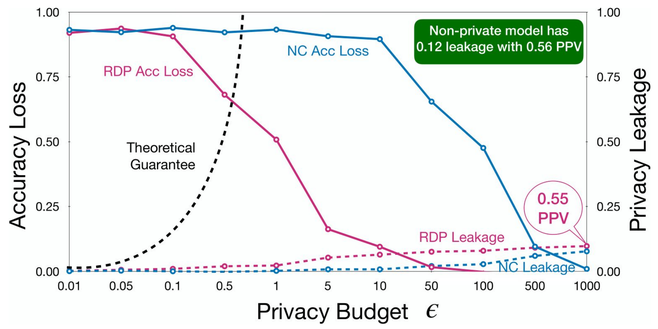


Experimental Setup: Evaluating Accuracy Loss and Privacy Leakage

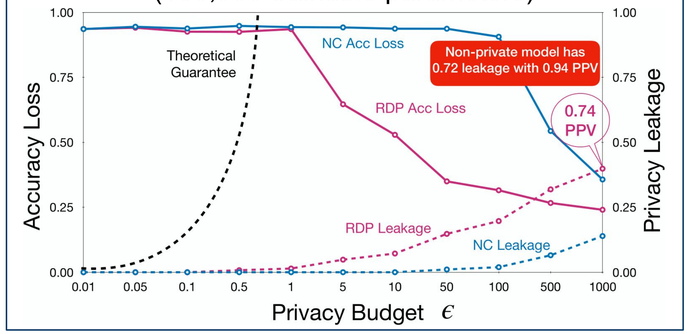


Privacy-Utility Comparison of Models trained with DP

Results for Logistic Regression (5,000 trainable parameters)

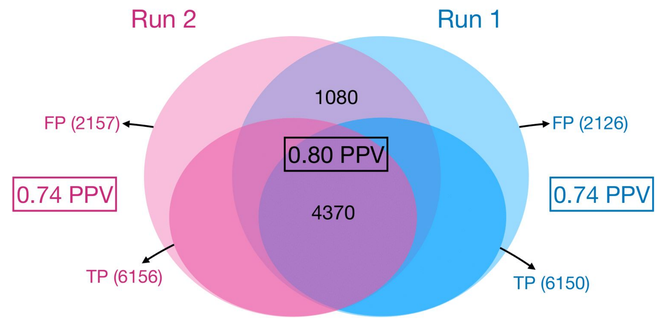


Results for Neural Network (103,936 trainable parameters)

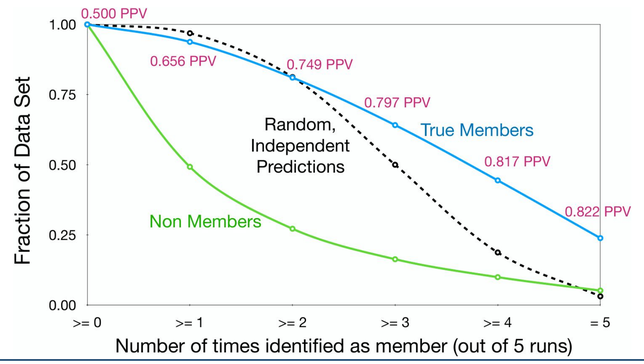


Analyzing Multiple Runs of Neural Network trained with RDP ($\epsilon = 1000$)

Membership Predictions across Two Runs



Membership Predictions across Multiple Runs



Bargav Jayaraman and David Evans. Evaluating Differentially Private Machine Learning in Practice. In USENIX Security 2019.
Link to Code: https://github.com/bargavj/EvaluatingDPML

Conclusion: Privacy doesn't come for free.
For a fixed privacy budget, improving composition increases model utility but at the cost of privacy leakage.