1. Membership Inference

Adversary wants to test whether data of a target victim has been used to train a model

Serious problem if inclusion in training set is privacy-sensitive

E.g., main task is: predict whether a smoker gets cancer

[Shokri et al., S&P'17] show it first for discriminative models

[Hayes et al. PETS'19] for generative models (later)

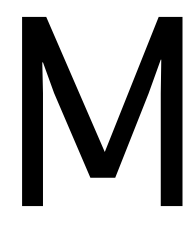
Membership inference is a very active research area, not only in machine learning

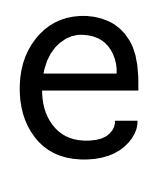
3

has been used to train a model

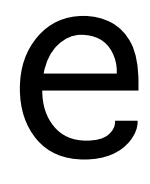
Adversary wants to test whether data of a target victim

only in machine learning

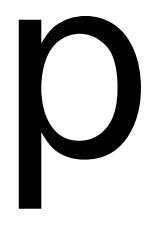

M

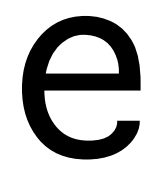


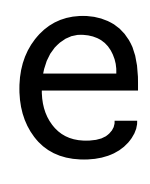

b

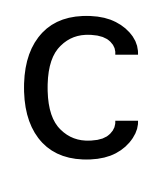


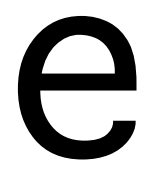
S

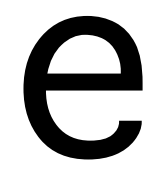


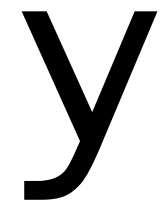




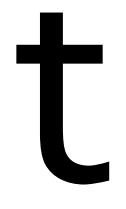

C



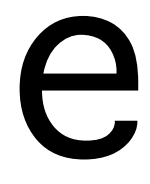


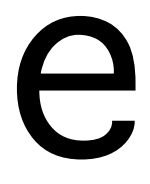


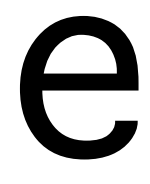

C



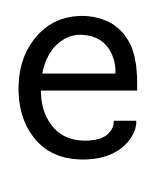
V

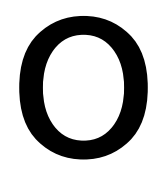


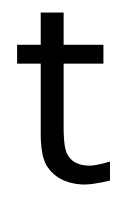




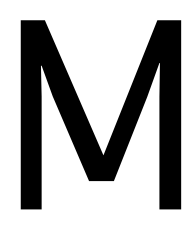
C

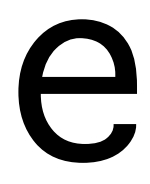




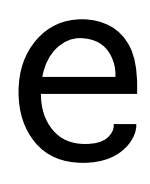


only in machine learning

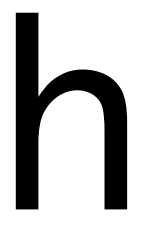


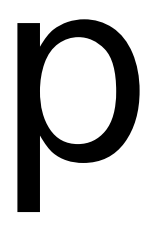


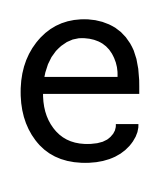

b

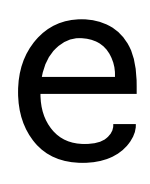


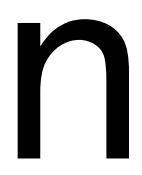
S



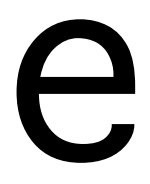


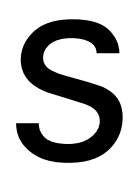


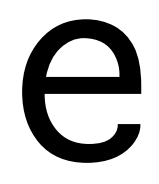


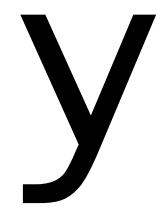


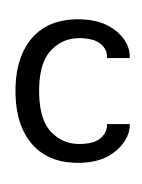
C

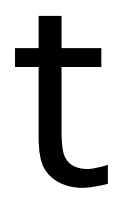




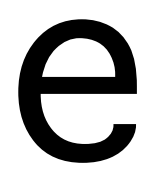


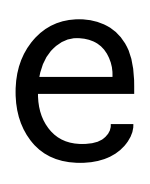


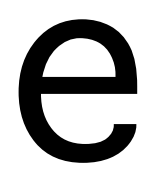

C

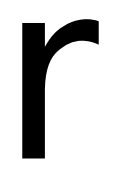


V





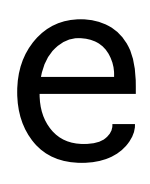


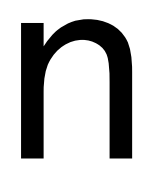


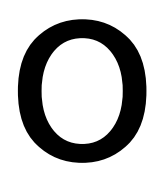
C

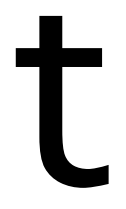

h

r










4