Hermite Polynomial Features for Private Data Generation

by Mohammad-Amin Charusaie



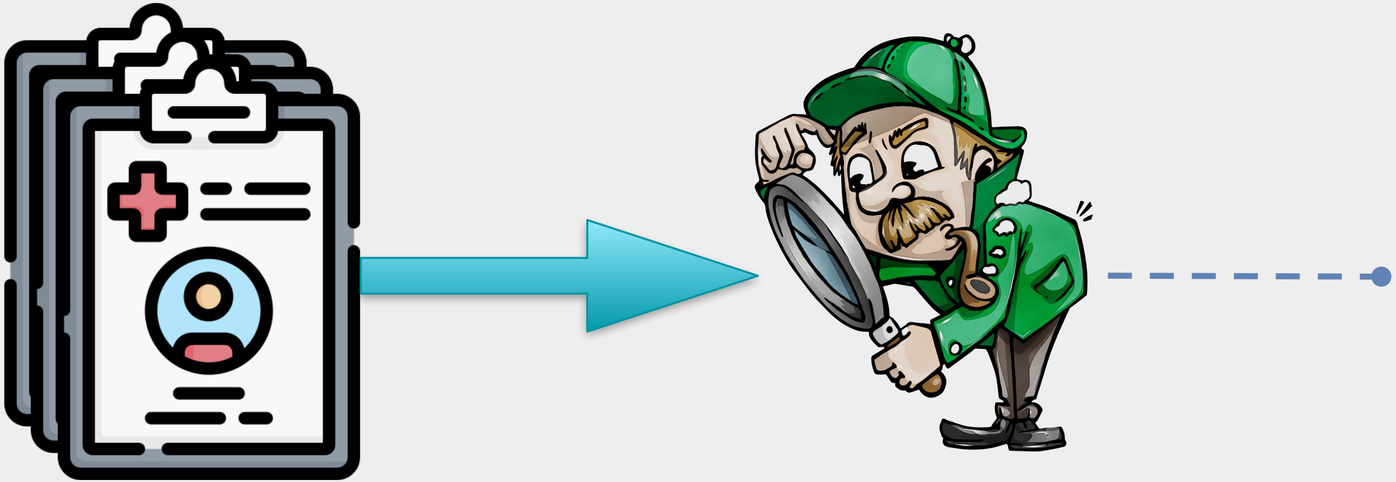

Margarita Vinaroz
@ MPI-IS

Frederik Harder
@ MPI-IS

Kamil Adamczewski
@ MPI-IS

Mijung Park
@ UBC

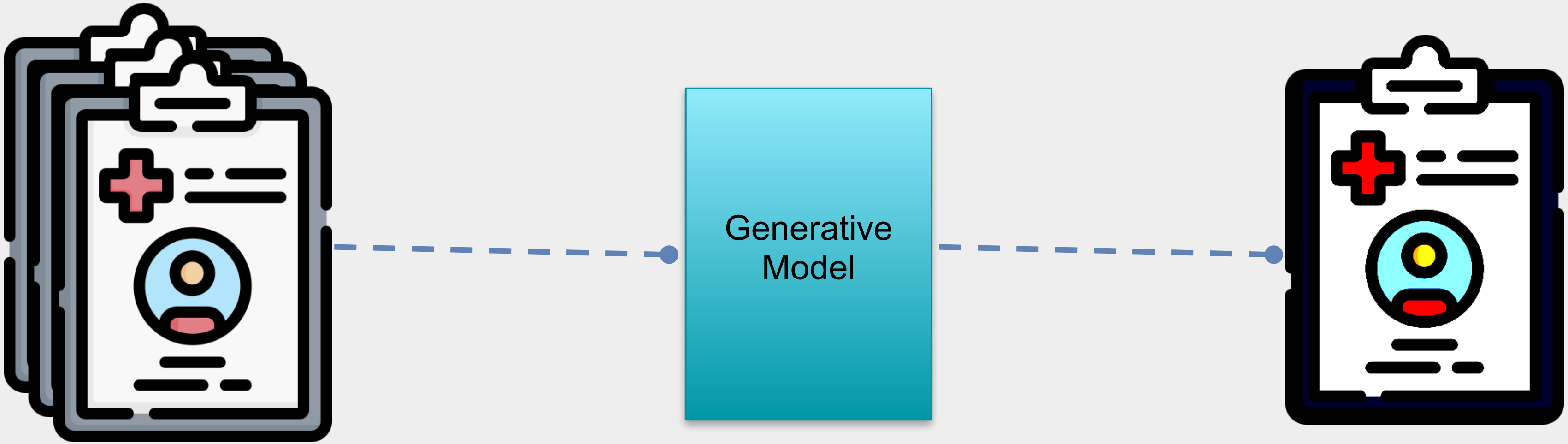
Data Release



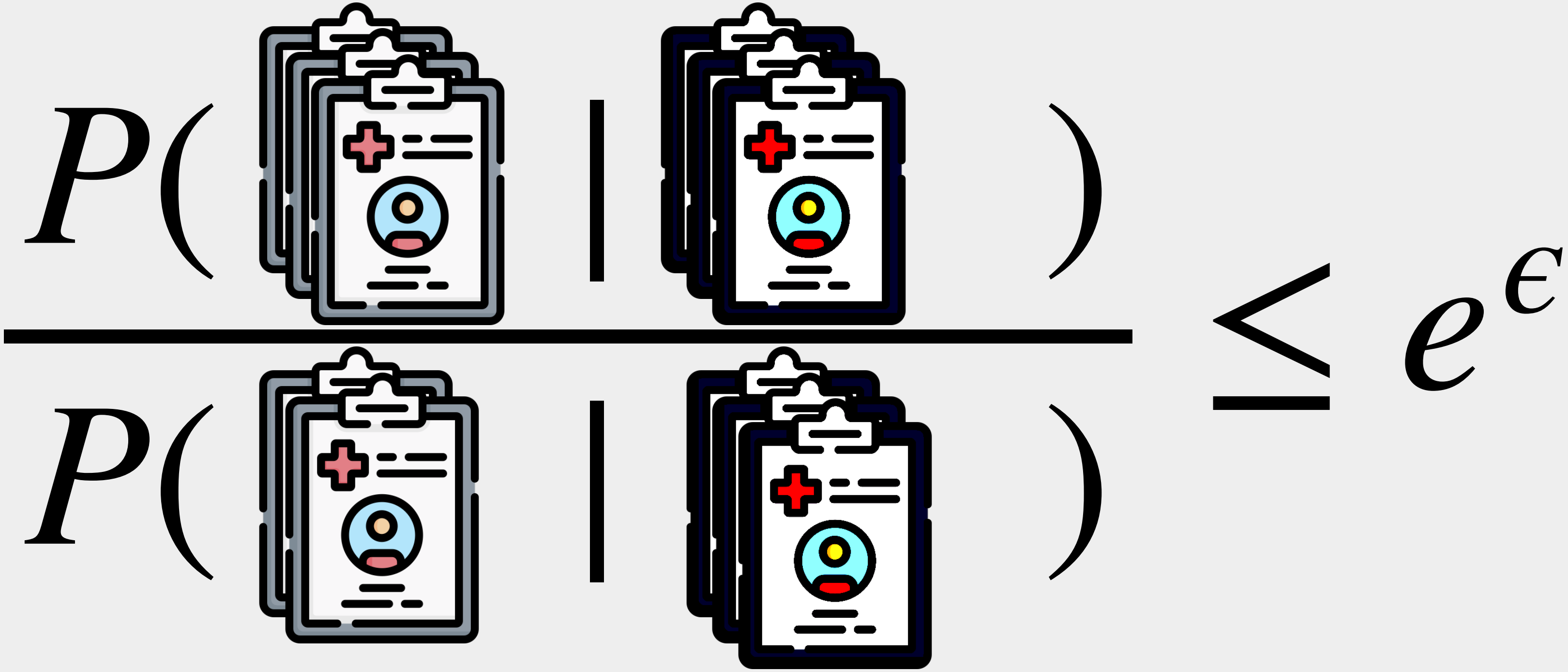
Smoking causes cancer?

Information Leakage

Synthetic Data Generation

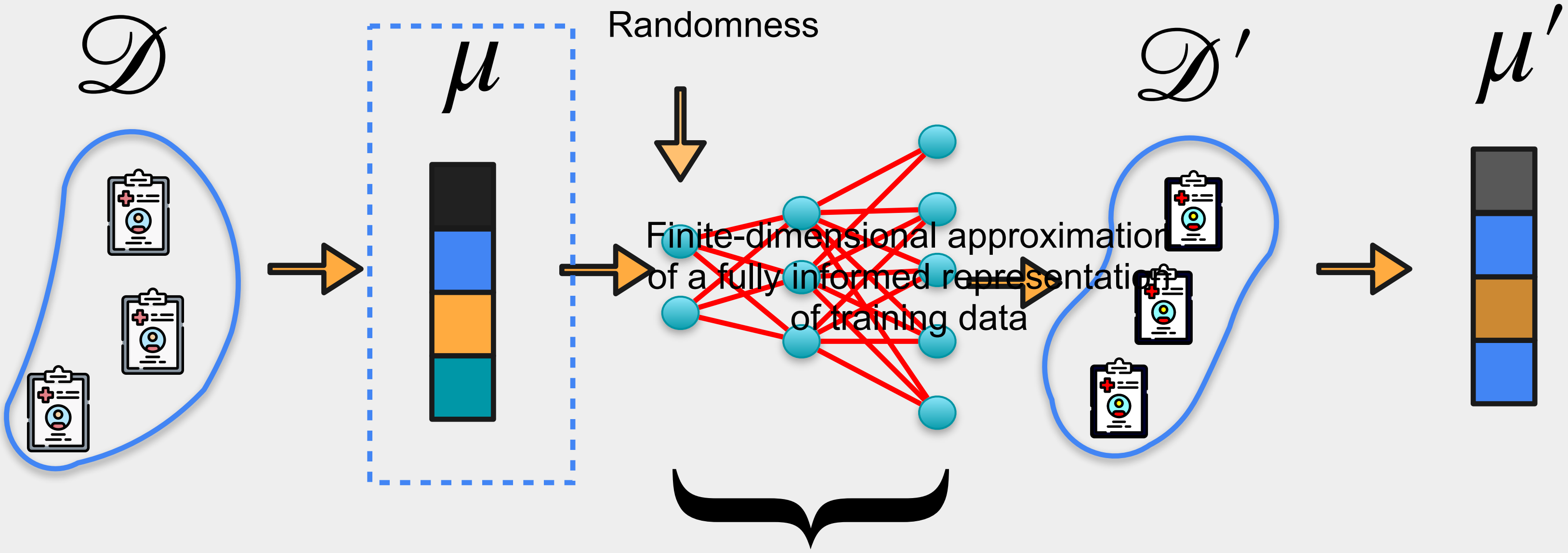


Differential Privacy



Data Generation via Kernel Mean Embedding

Privatization via perturbation



$$\min \|\mu - \mu'\|_2$$

Approximations of Gaussian Kernel

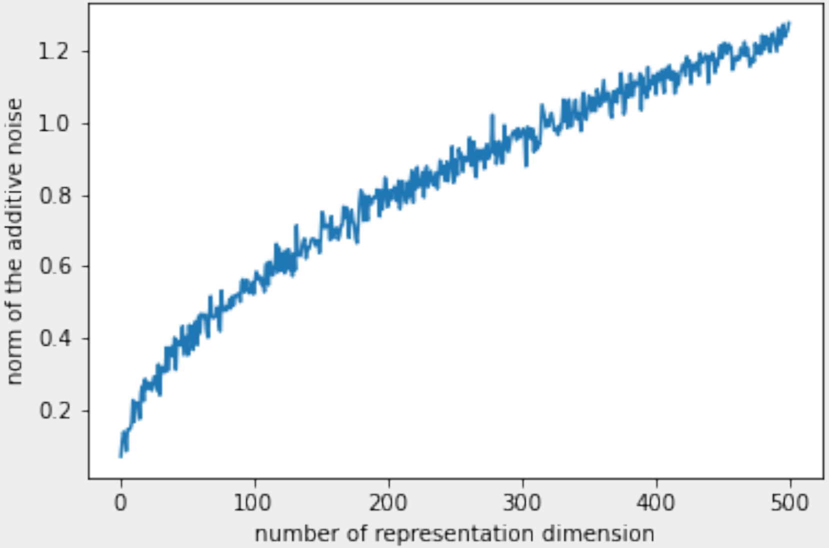
1. **Random Features** (Harder et al. 2021): $k(x, y) \simeq \frac{1}{d}\sum_{i=1}^{d} \cos(\omega_i x)\cos(\omega_i y)$ for $\omega_i \sim \mathcal{N}(0,1)$

$$\mu = \frac{1}{m}\sum_{x\in\mathcal{D}} [\cos(\omega_1 x), \ldots, \cos(\omega_d x)]$$

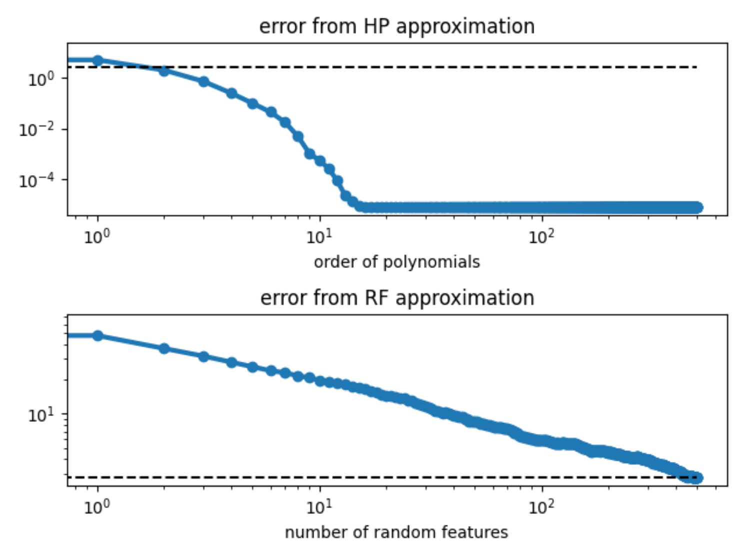
2. **Hermite Features** (DP-HP): $k(x, y) \simeq \frac{1}{d}\sum_{i=1}^{d} B_d H_d(x) e^{-\alpha x^2} H_d(y) e^{-\alpha y^2}$ for d-th Hermite polynomials $H_d(\cdot)$

$$\mu = \frac{1}{m}\sum_{x\in\mathcal{D}} \sqrt{B_d} e^{-\alpha x^2} [H_1(x), \ldots, H_d(x)]$$

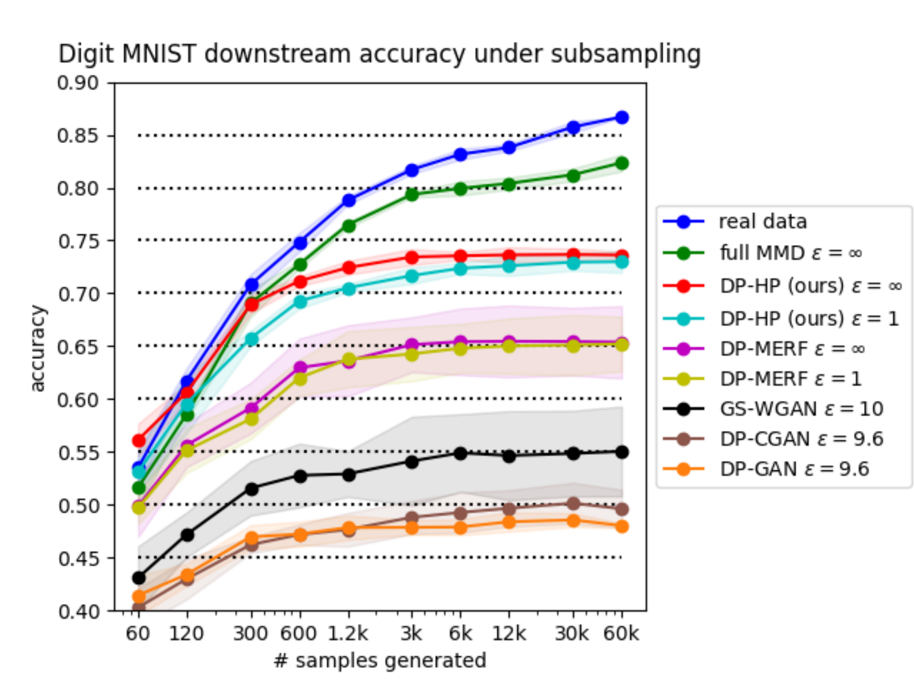
Representation Dimension vs. Perturbation Effect



Representation Dimension vs. Approximation Power



Experiments



Thank you!