PrivGene: Differentially Private Model Fitting Using Genetic Algorithms

Jun Zhang[1] **Xiaokui Xiao[1]** **Yin Yang[2]** **Zhenjie Zhang[2]** **Marianne Winslett[2,3]**

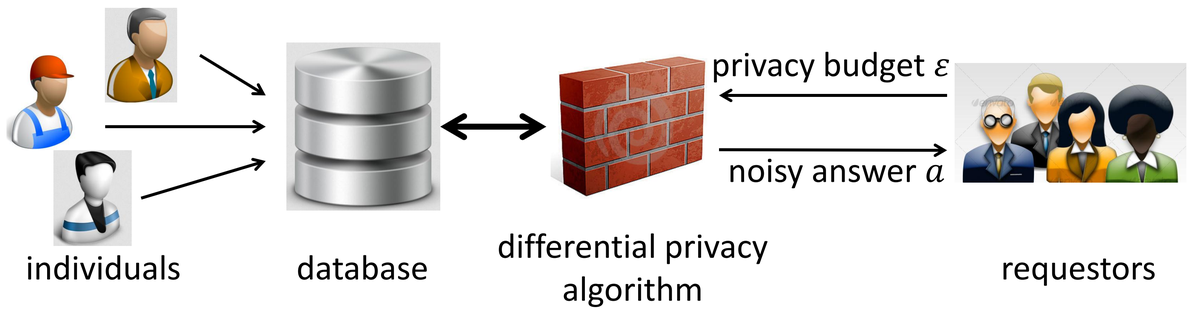
1.Nanyang Technological University
Singapore
{jzhang027, xkxiao}@ntu.edu.sg

2. Advanced Digital Sciences Center
Singapore
{yin.yang, zhenjie}@adsc.com.sg

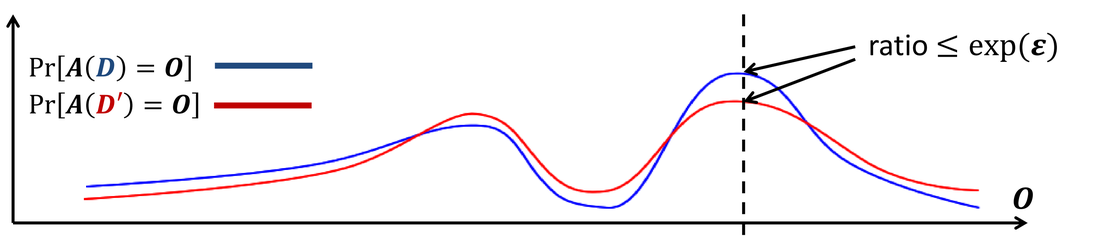
3. University of Illinois
USA
winslett@illinois.edu

1. Background

❑ Database system with differential privacy

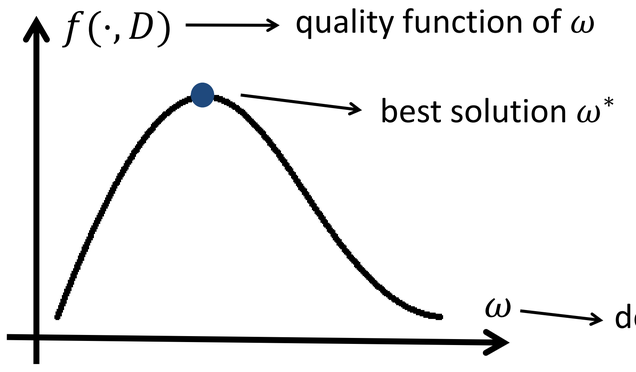


❑ Statistical model behind the wall



where D and D' are neighboring databases that differ by **at most one** tuple

❑ Model fitting problems



Our Target:
Return a solution with best possible quality under ε-differential privacy

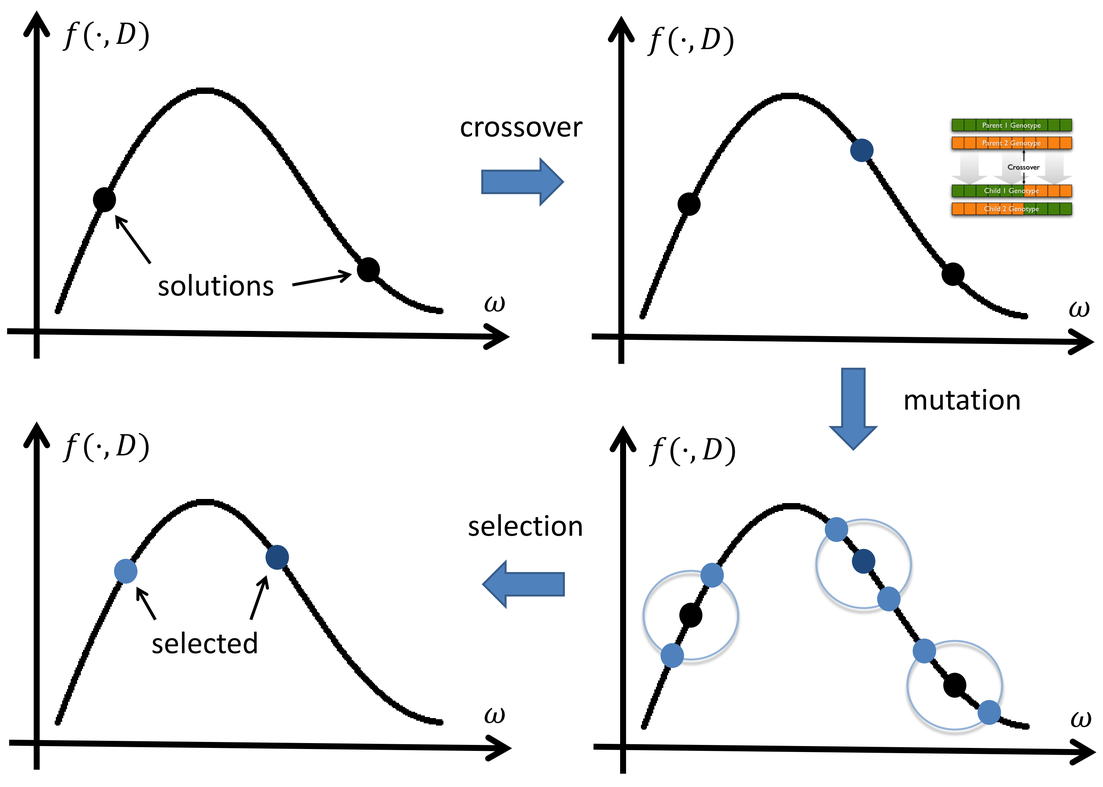
2. PrivGene

❑ A novel framework for differentially private model fitting based on genetic algorithms (GA)

❑ PrivGene is applicable to logistic regression, SVM classification and k-means clustering

❑ PrivGene runs iteratively to improve candidate solutions by following operations

- ➢ Crossover
- ➢ Mutation
- ➢ Selection with **Differential Privacy**



3. Enhanced Exponential Mechanism

❑ The assumption of enhanced exponential mechanism

$$f(\omega, D) = h(\omega) + \sum_{t \in D} q(\omega, t)$$

❑ The bound of exponential mechanism is

$$\Delta_1 = \max_{\omega \in \Omega,\, t, t' \in T} q(\omega, t) - q(\omega, t')$$

❑ New bound in enhanced exponential mechanism is

$$\Delta_2 = \max_{\omega, \omega' \in \Omega,\, t \in T} q(\omega, t) - q(\omega', t)$$

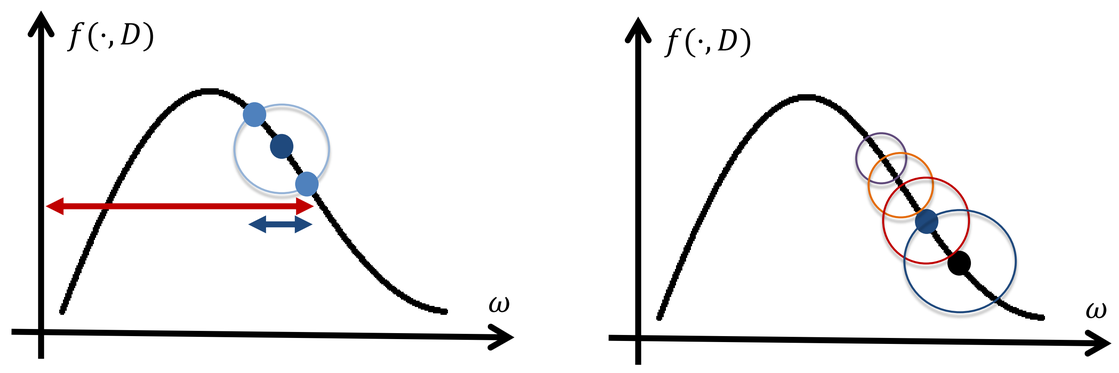
And the final bound is the minimum value

$$\Delta = \min(\Delta_1, \Delta_2)$$

❑ Case study: linear SVM

$$f(\omega, D) = -\left(\frac{1}{2}|\omega|^2 + C \cdot \sum_{(x,y) \in D} (1 - yx^T\omega)_+\right)$$

$$\Delta_1 = 2C \cdot \max_{\omega \in \Omega}(|\omega|_1 + 1) \qquad \Delta_2 = 2C \cdot \max_{\omega, \omega' \in \Omega} |\omega - \omega'|_1$$



4. Experiments