Privacy-Preserving Machine Learning

CS 760: Machine Learning

Spring 2018

Mark Craven and David Page

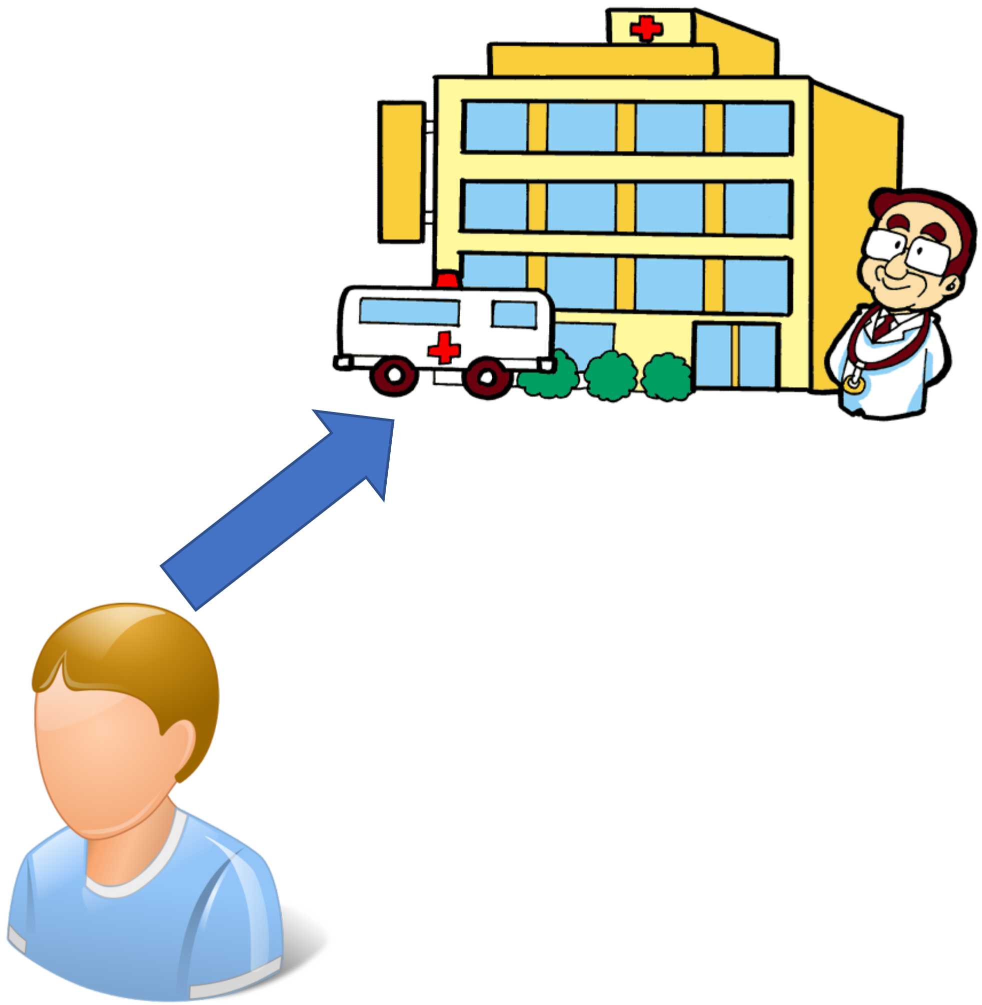
www.biostat.wisc.edu/~craven/cs760

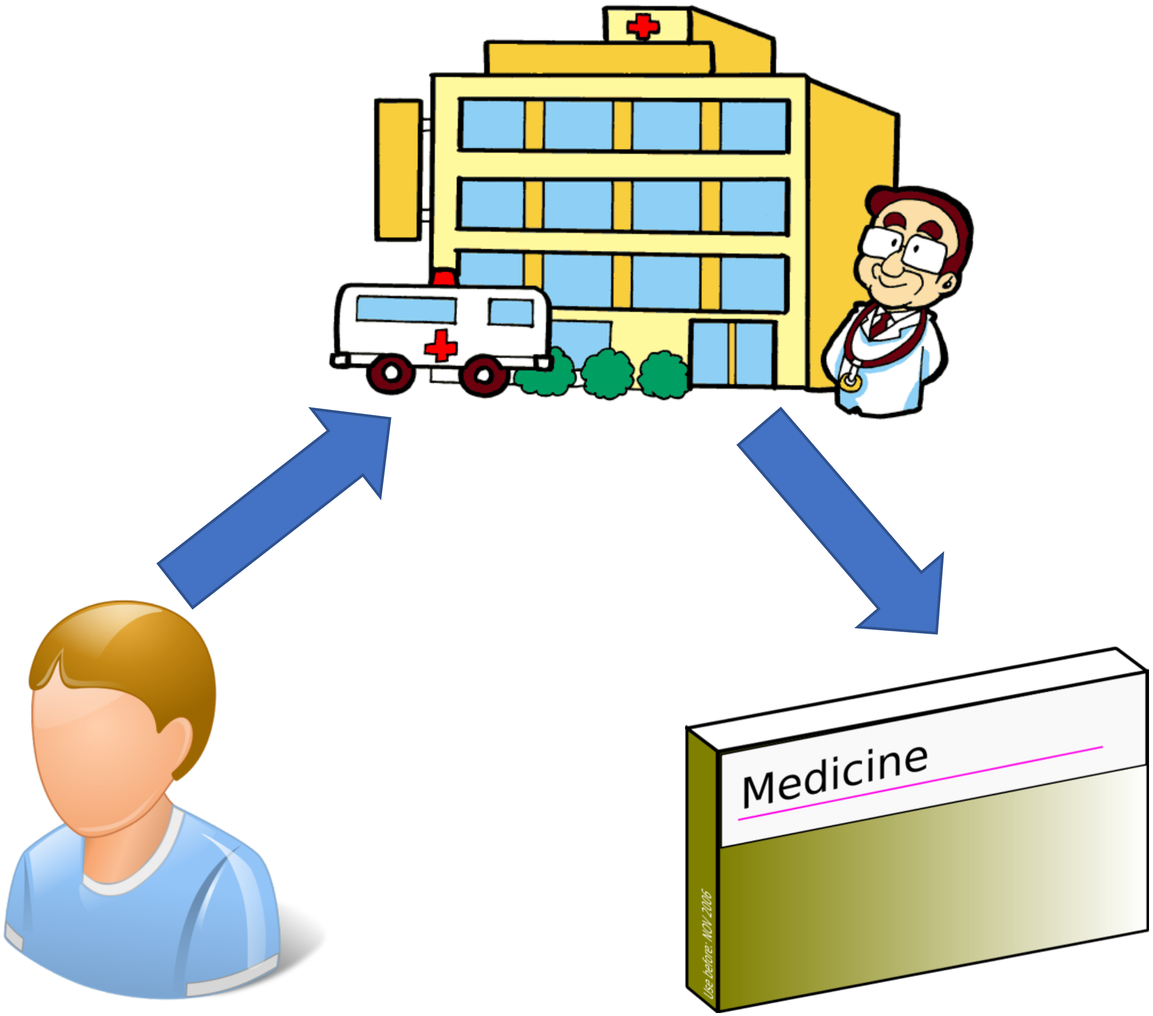
Goals for the Lecture

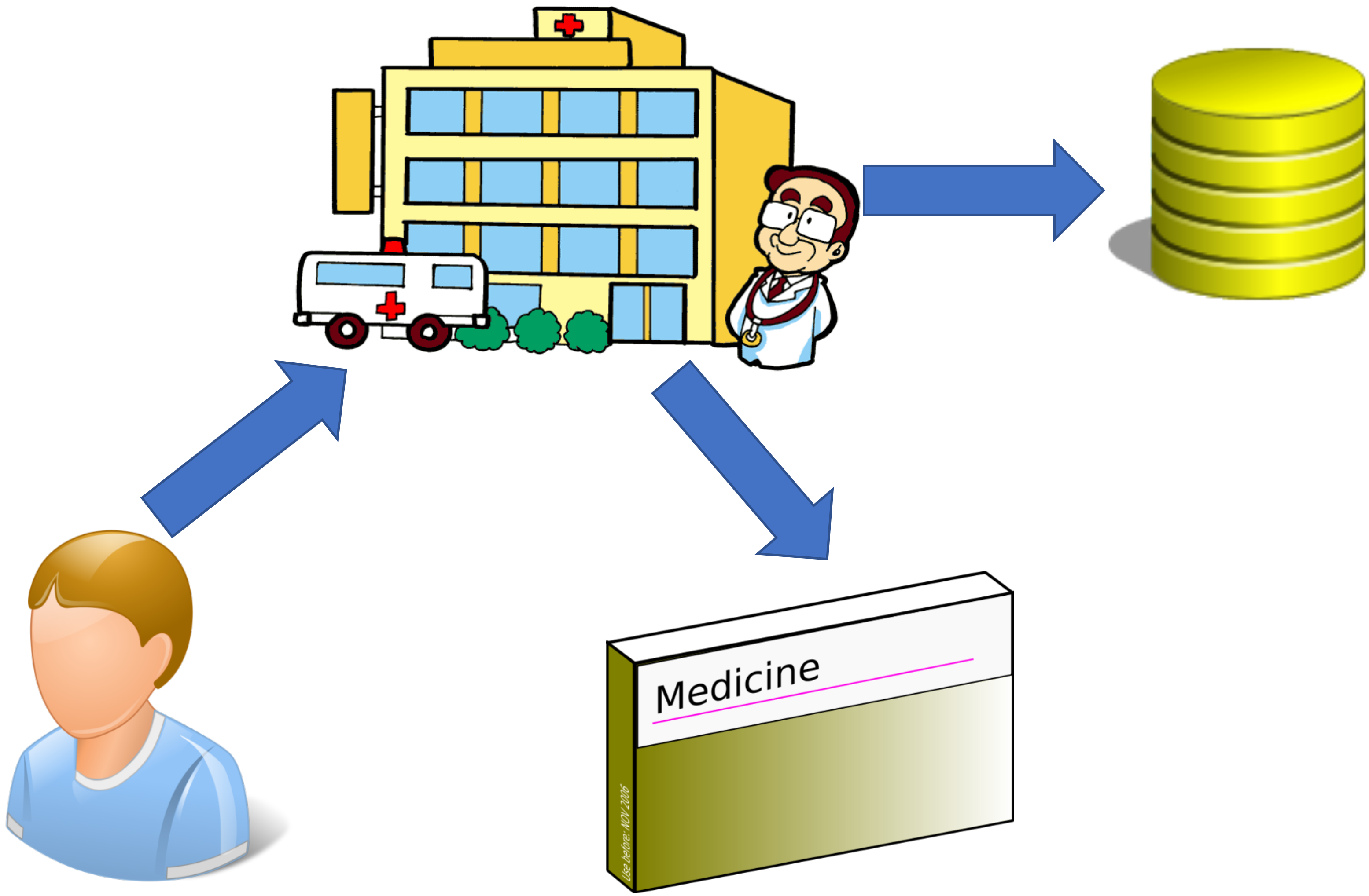
- You should understand the following concepts:

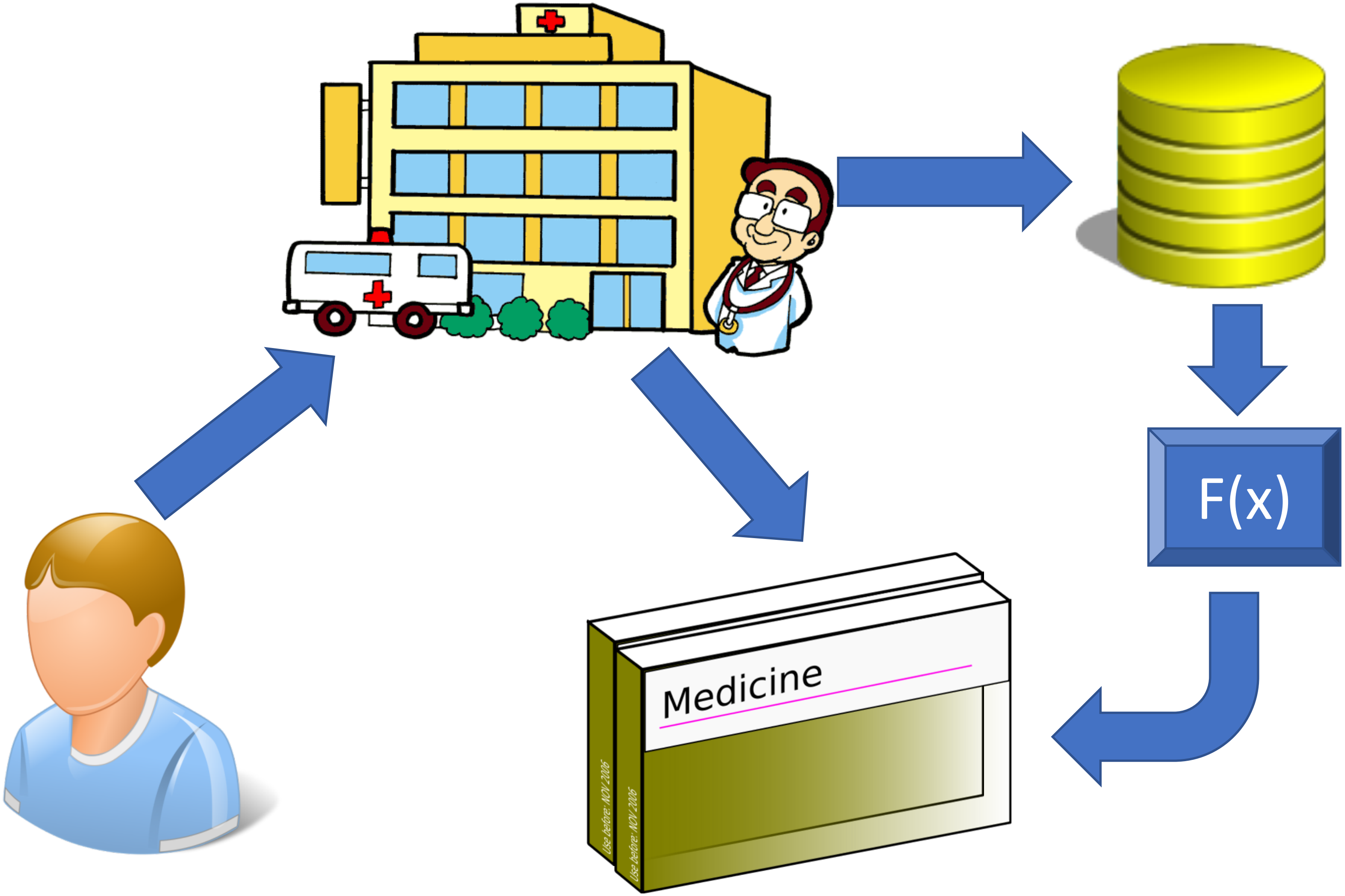
  - public key cryptography
  - linearly homomorphic encryption
  - fully homomorphic encryption
  - differential privacy
  - global sensitivity
  - Laplace mechanism

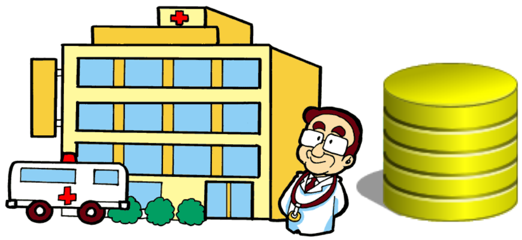
- Thanks Eric Lantz and Irene Giacomelli!

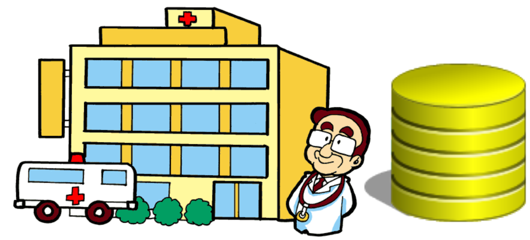


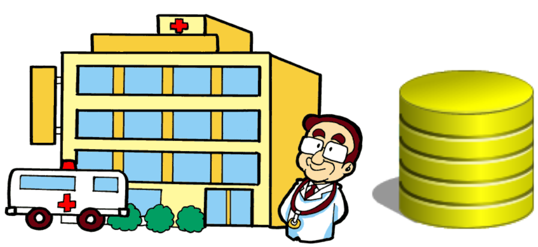

Medicine

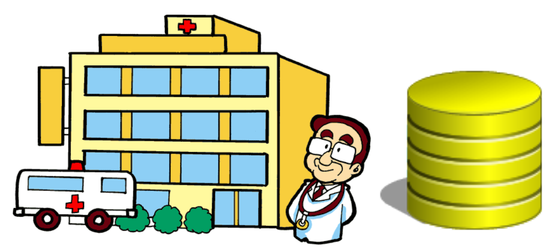

Medicine
Use before: NOV 2006

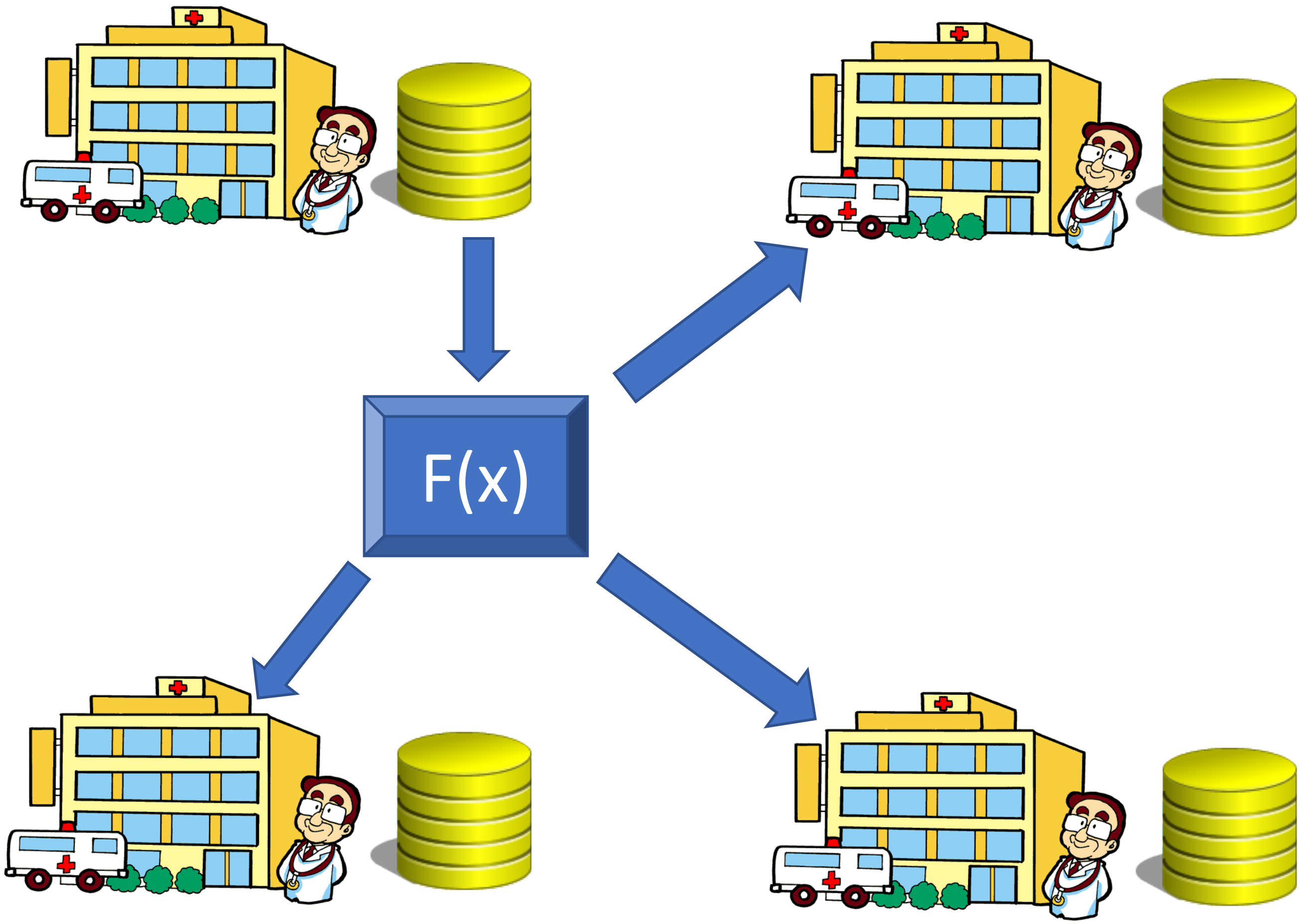

F(x)
Medicine

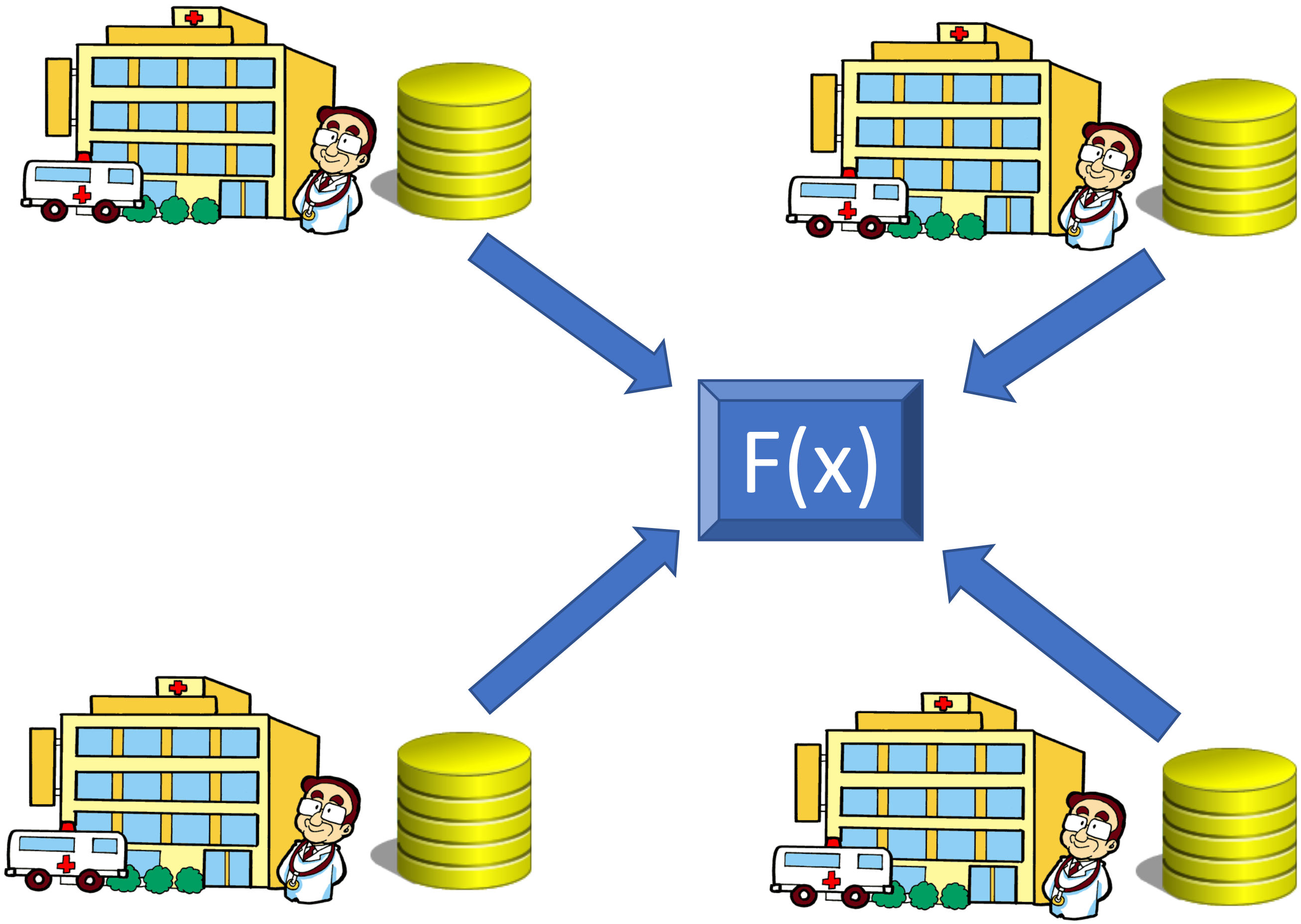


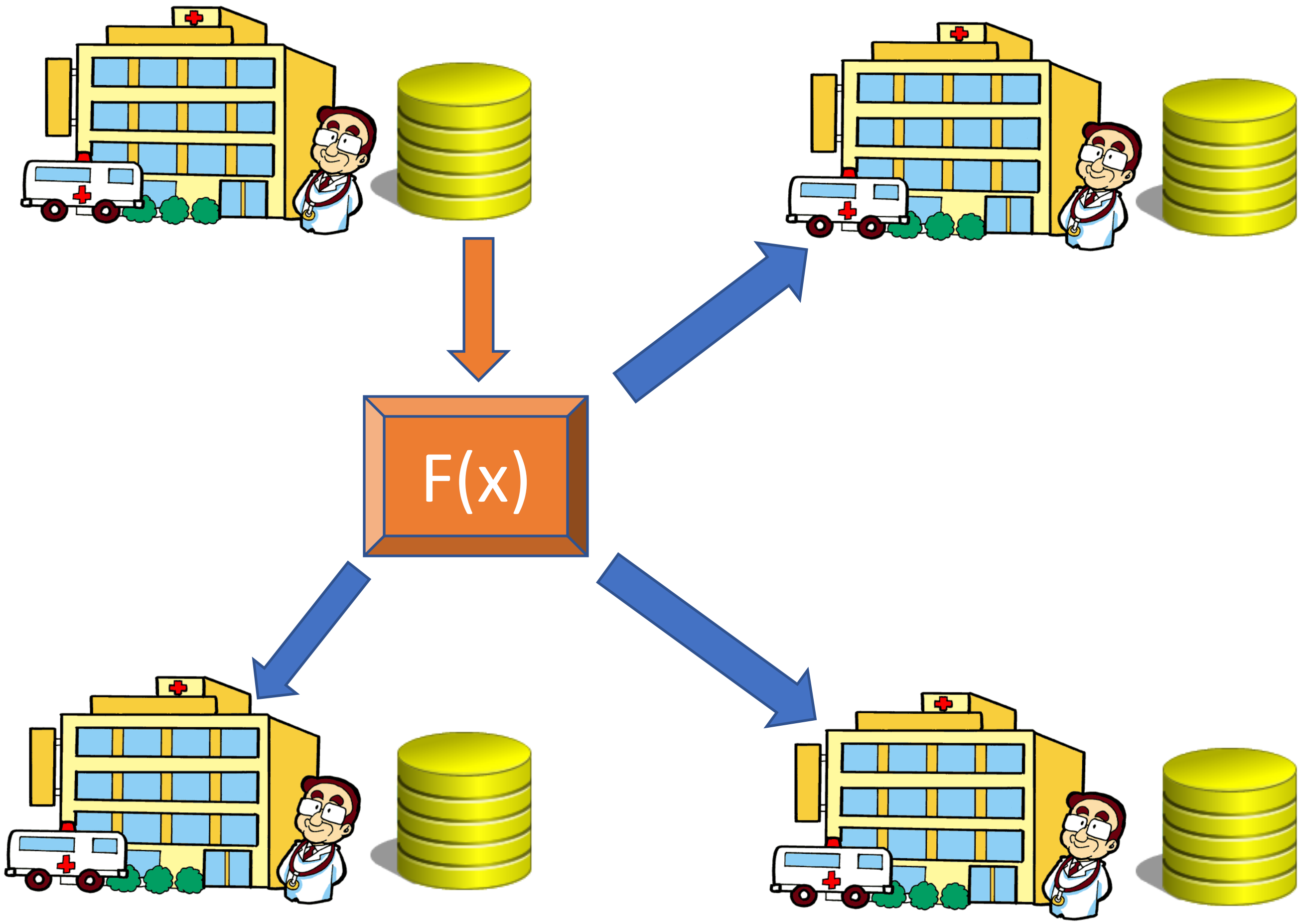


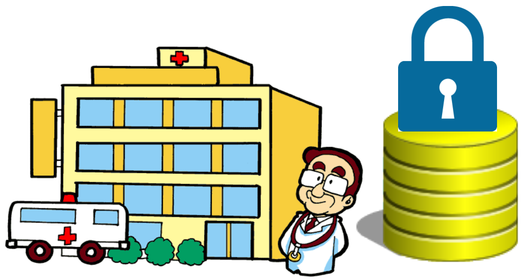


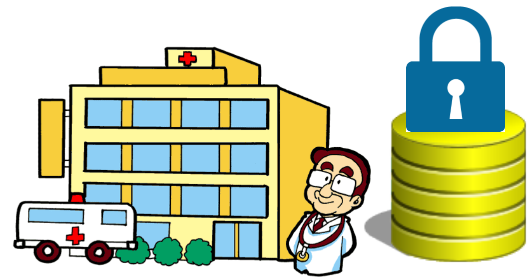


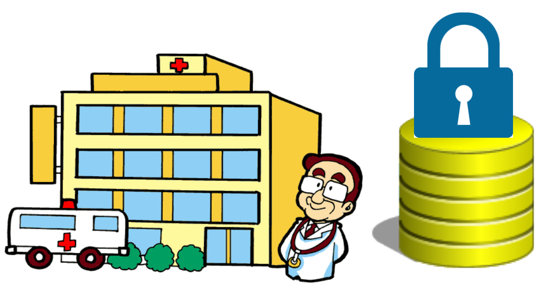

F(x)

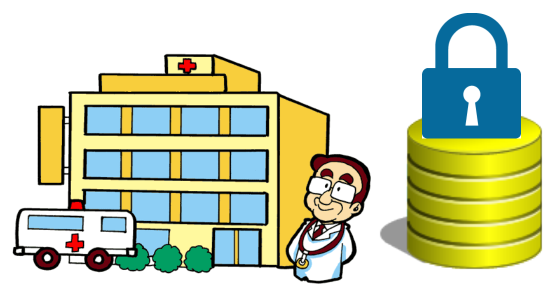

F(x)


F(x)









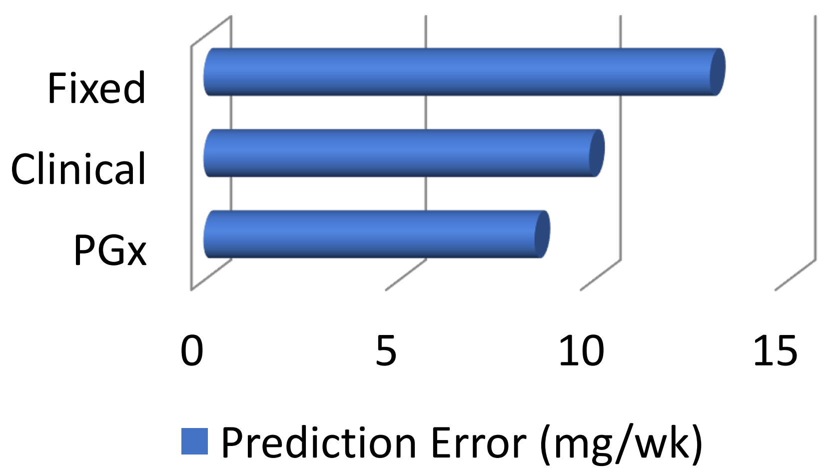
Need for Privacy

- Large databases of patient information
  - Regulations and expectations of privacy
  - Large potential gains from data mining
  - How to balance utility and privacy?

- Privacy approaches
  - k-anonymity (Sweeney, 2002), l-diversity (Machanavajjhala, 2007), t-closeness (Li, 2007)
  - Homomorphic encryption
  - Differential privacy (Dwork, 2006)

Recall: IWPC Warfarin dosing algorithm

- Over a dozen real-value prediction techniques were used
- Linear regression and support vector regression were the best performers



5.6044
-0.2614 Age in decades
+0.0087 Height in cm
+0.0128 Weight in kg
-0.8677 *VKORC1* A/G
-1.6974 *VKORC1* A/A
-0.4854 *VKORC1* genotype unknown
-0.5211 *CYP2C9* *1/*2
-0.9357 *CYP2C9* *1/*3
-1.0616 *CYP2C9* *2/*2
-1.9206 *CYP2C9* *2/*3
-2.3312 *CYP2C9* *3/*3
-0.2188 *CYP2C9* genotype unknown
-0.1092 Asian race
-0.2760 Black or African American
-0.1032 Missing or Mixed race
+1.1816 Enzyme inducer status
-0.5503 Amiodarone status
= square root of final dose

Recall: Ridge Regression

Data point: $(\mathbf{x}, y)$, $\mathbf{x} \in \mathbb{R}^d$ and $y \in \mathbb{R}$

Model: $\mathbf{w} \in \mathbb{R}^d$ vector of weights

$$y \approx f_{\mathbf{w}}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle = \sum_{j=1}^{d} \mathbf{w}(j)\mathbf{x}(j)$$

Training: find argmin of $F(\mathbf{w}) = \underbrace{\sum_{i=1}^{n} (y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle)^2}_{\text{mean squared error}} + \lambda \underbrace{||\mathbf{w}||_2^2}_{\text{regularization}}$

Public-Key Encryption

$sk \rightarrow$ secret key
$pk \rightarrow$ public key

Encryption:

Decryption:

Public-Key Encryption

$sk \rightarrow$ secret key

$pk \rightarrow$ public key

<u>Encryption</u>: $\mathbf{c} = \mathrm{Enc}_{pk}(\mathbf{m})$

$\mathbf{c} \longrightarrow$ hides **m** to everyone that does NOT have sk

<u>Decryption</u>:

$$m = \text{hello!} \xrightarrow{pk} \boxed{\text{Enc}} \longrightarrow c = 6a7\#87t$$

Public-Key Encryption

$sk \rightarrow$ secret key
$pk \rightarrow$ public key

<u>Encryption</u>: $\mathbf{c} = \text{Enc}_{pk}(\mathbf{m})$

$\mathbf{c} \rightarrow$ hides $\mathbf{m}$ to everyone that does NOT have sk

<u>Decryption</u>:

$\mathbf{c} \rightarrow$ reveals $\mathbf{m}$ to everyone that has sk

$$m = \text{hello!} \xrightarrow{pk} \boxed{\text{Enc}} \rightarrow c = \text{6a7\#87t} \rightarrow \boxed{\text{Dec}} \xrightarrow{sk} \text{hello!}$$

Linearly-Homomorphic Encryption

Addition of ciphertexts

$$\mathrm{Enc}_{pk}(\mathbf{m}_1) \boxplus \mathrm{Enc}_{pk}(\mathbf{m}_2) = \mathrm{Enc}_{pk}(\mathbf{m}_1 + \mathbf{m}_2)$$

Multiplication of a ciphertext by a plaintext

$$\mathbf{m}_1 \boxplus \mathrm{Enc}_{pk}(\mathbf{m}_2) = \mathrm{Enc}_{pk}(\mathbf{m}_1 \times \mathbf{m}_2)$$

Linearly-Homomorphic Encryption

Addition of ciphertexts

$$\mathrm{Enc}_{pk}(\mathbf{m}_1) \boxplus \mathrm{Enc}_{pk}(\mathbf{m}_2) = \mathrm{Enc}_{pk}(\mathbf{m}_1 + \mathbf{m}_2)$$

Multiplication of a ciphertext by a plaintext ($\mathbf{m}_1$ is public!)

$$\mathbf{M}_1 \boxplus \mathrm{Enc}_{pk}(\mathbf{m}_2) = \mathrm{Enc}_{pk}(\mathbf{m}_1 \times \mathbf{m}_2)$$

Linearly-Homomorphic Encryption

Addition of ciphertexts

$$\mathrm{Enc}_{pk}(\mathbf{m}_1) \boxplus \mathrm{Enc}_{pk}(\mathbf{m}_2) = \mathrm{Enc}_{pk}(\mathbf{m}_1 + \mathbf{m}_2)$$

Multiplication of a ciphertext by a plaintext ($\mathbf{m}_1$ is public)

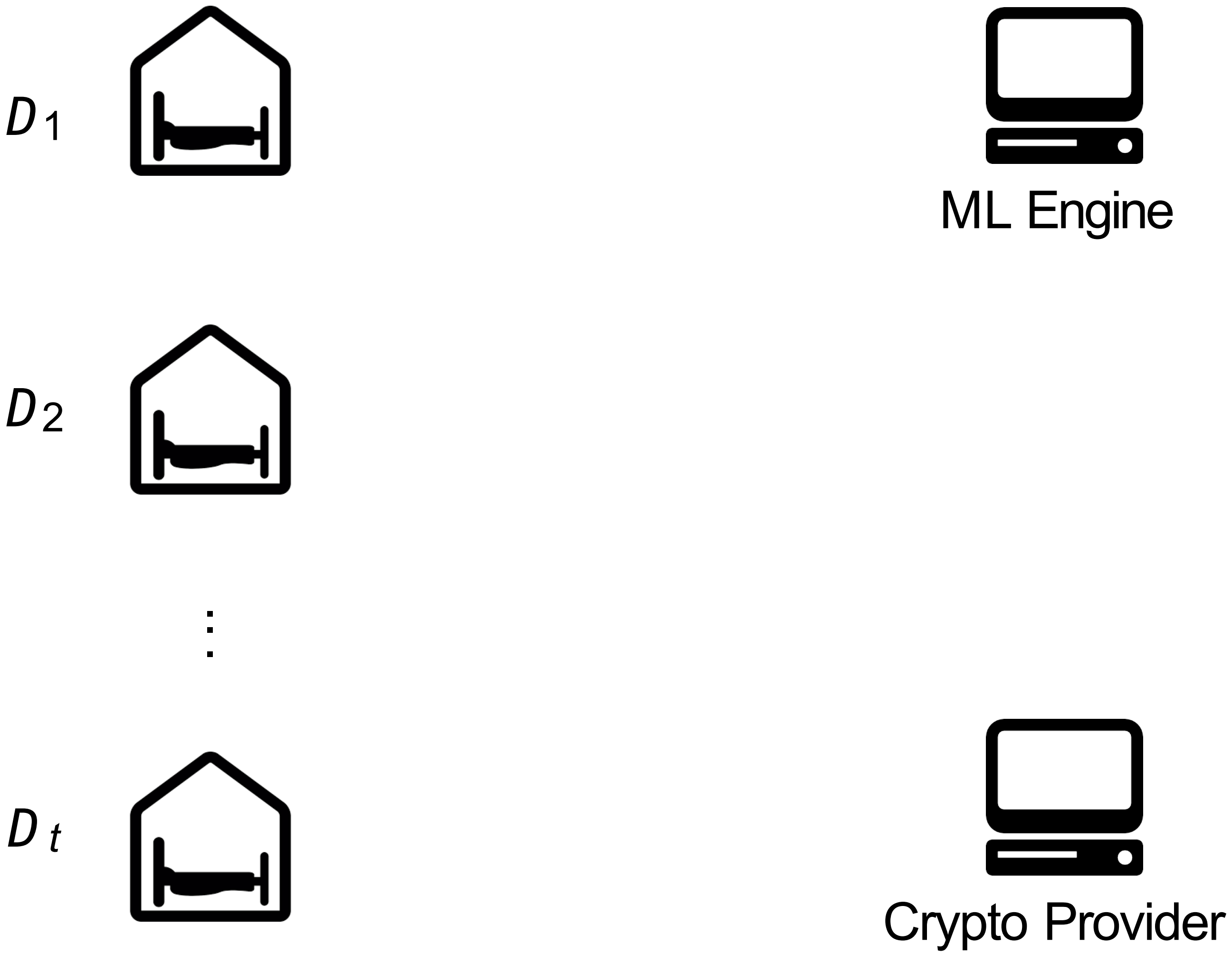
$$\mathbf{M}_1 \boxplus \mathrm{Enc}_{pk}(\mathbf{m}_2) = \mathrm{Enc}_{pk}(\mathbf{m}_1 \times \mathbf{m}_2)$$

Fully homomorphic requires *multiplication* analog of $\boxplus$ and currently is ***much*** slower.

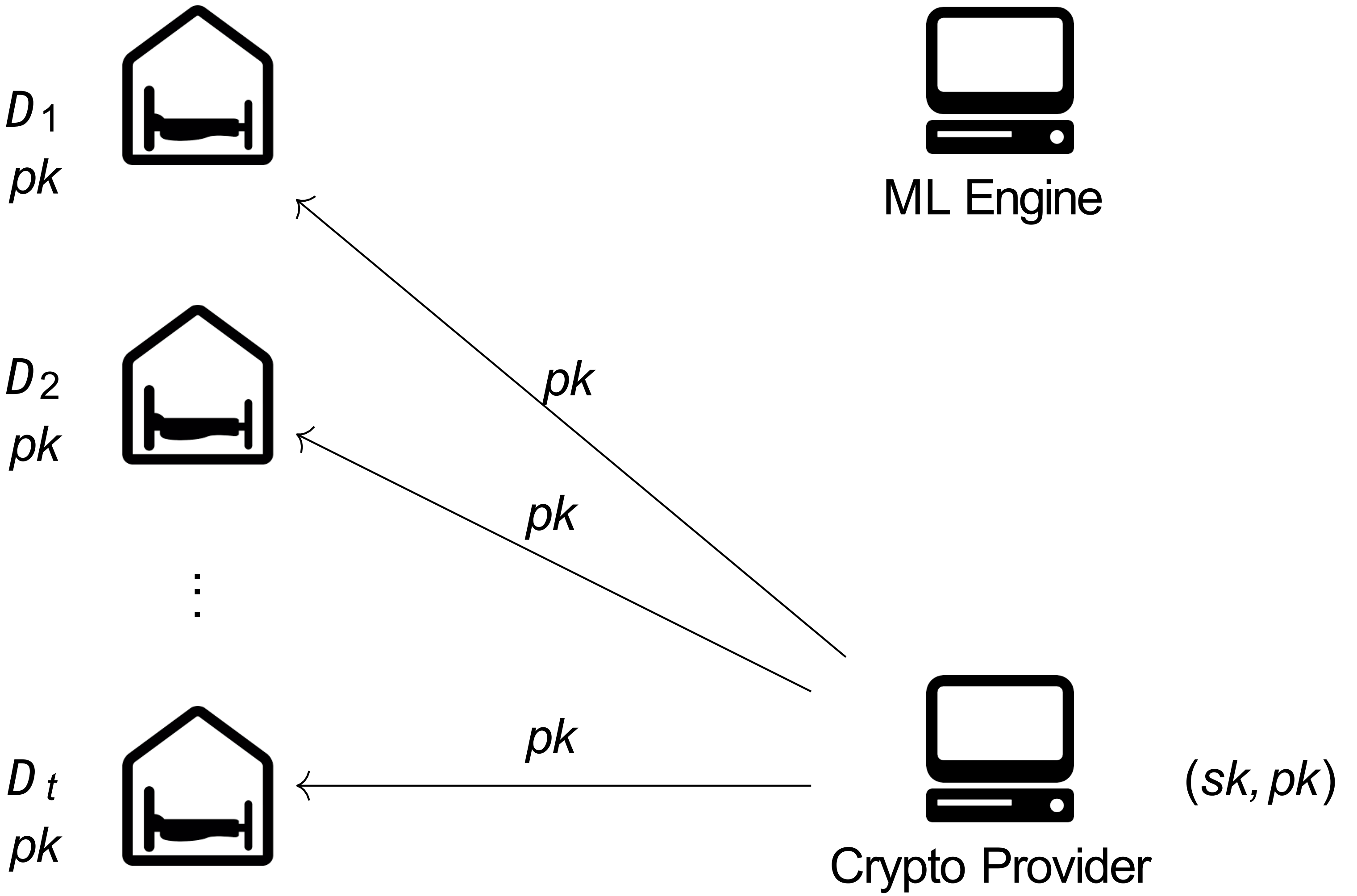
Database (DB): $10^5 \times 10^2$ real numbers in $[-2000, 2000]$ with 3 digits in the fractional part. Times using linearly-homomorphic encryption:

- encrypt the DB: 40 minutes
- sum of two DBs: 3 seconds
- mult. by a constant: 25 mins

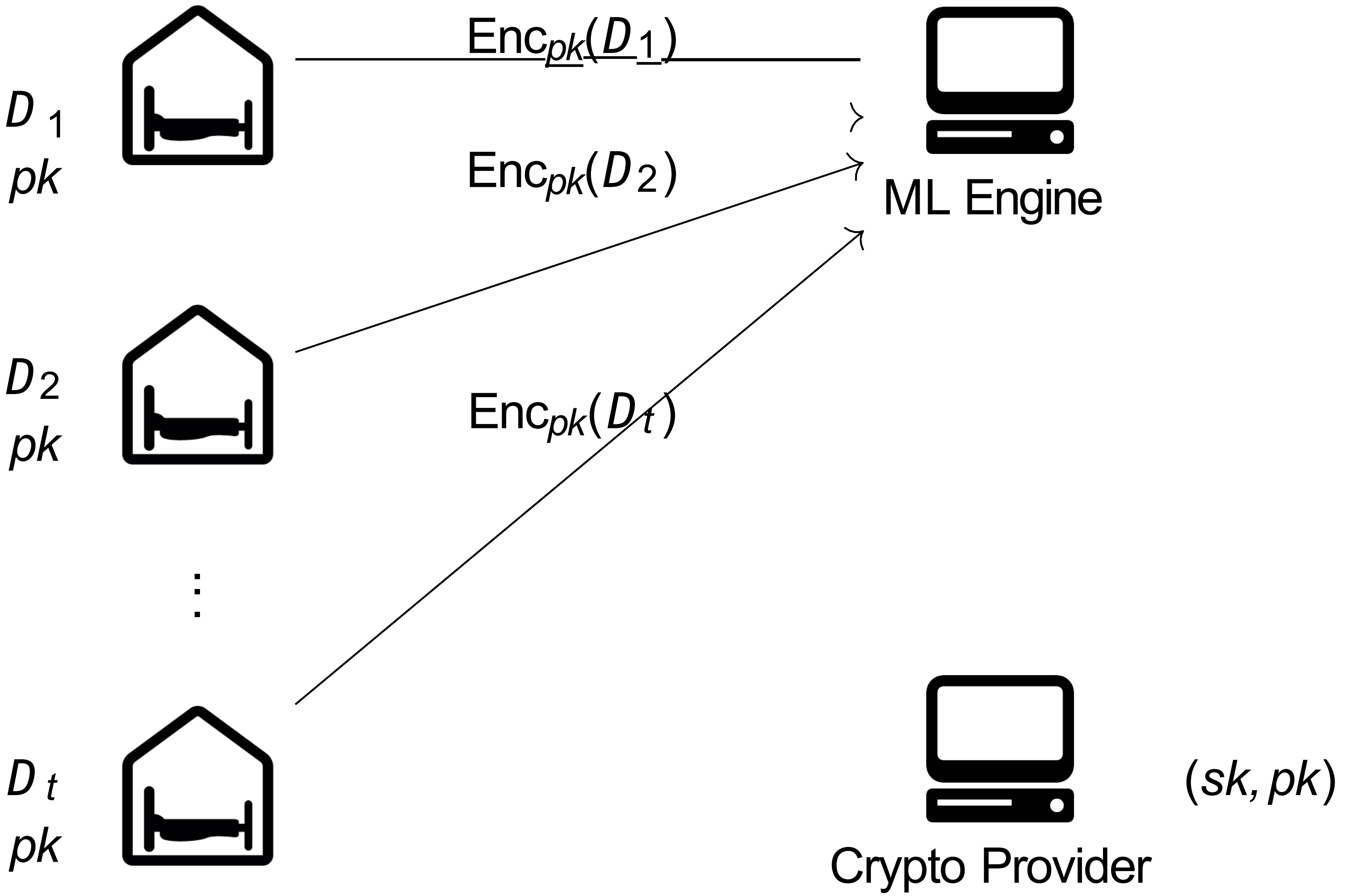
Illustration



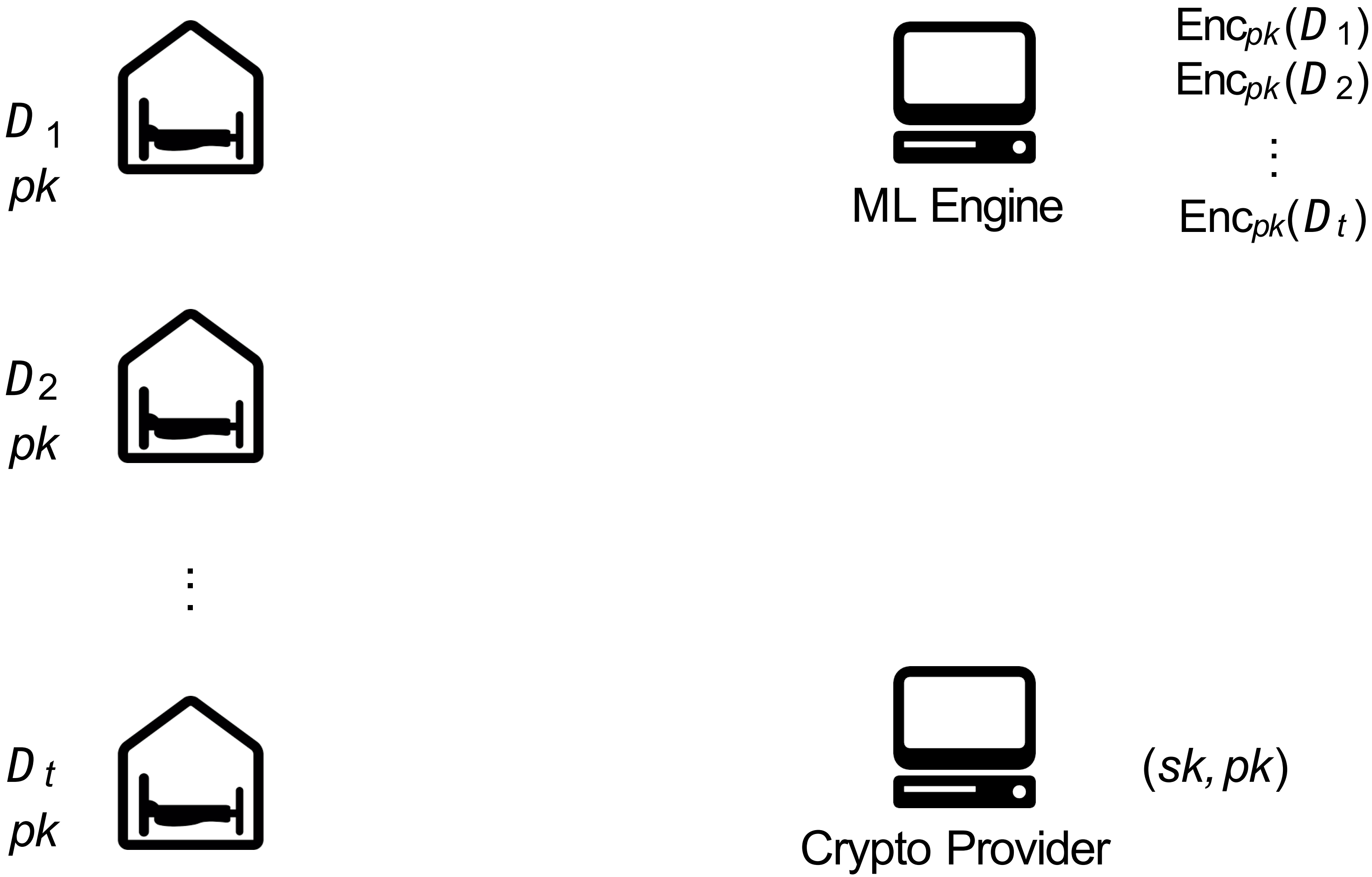
Illustration



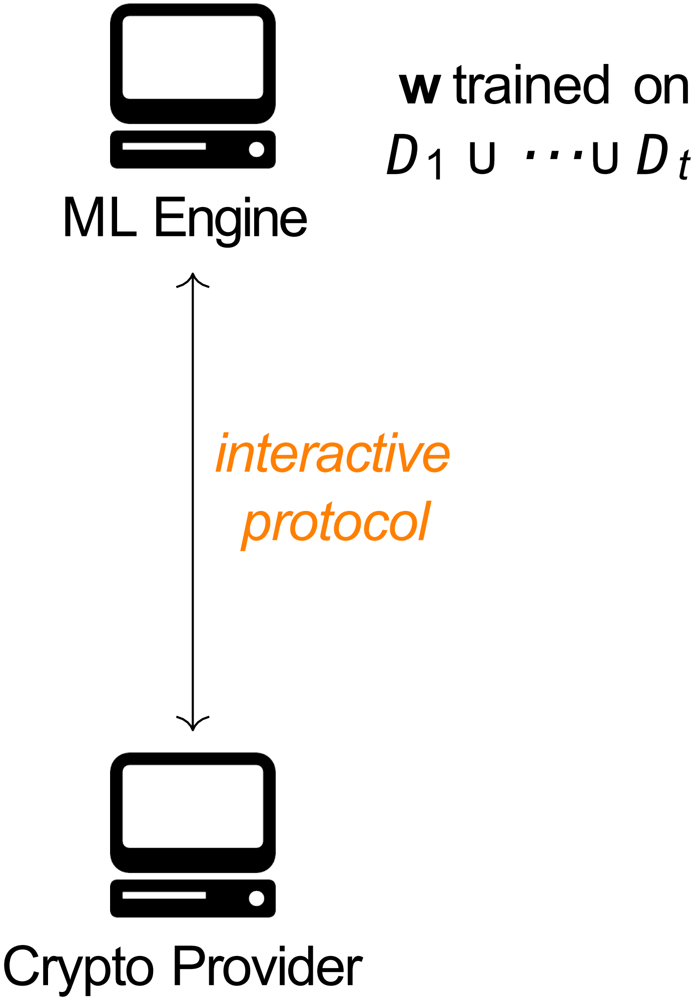
Illustration



Illustration



Illustration



Illustration

Interactive protocol:

1. the ML engine "masks inside the encryption"
   $\mathrm{Enc}_{pk}(D) \rightarrow \mathrm{Enc}_{pk}(\tilde{D})$

2. the crypto provider decrypts, gets $\tilde{D}$ and computes a "masked model", $\tilde{\mathbf{w}}$

3. the ML engine computes the real model $\mathbf{w}$ from the masked one

Illustration

Results for seven UCI datasets (time in seconds):
(phase 1 = encryption, phase 2 = interactive protocol)

Dataset	n	d	ℓ	$\log_2(N)$	R_{MSE}	Phase 1		Phase 2	
						Time	kB	Time	kB
air	6252	13	1	2048	4.15E-09	1.99	53.24	3.65	96.51
beijing	37582	14	2	2048	5.29E-07	2.37	60.93	4.26	110.10
boston	456	13	4	2048	2.34E-06	2.00	53.24	3.76	96.51
energy	17762	25	3	2724	5.63E-07	12.99	238.26	37.73	451
forest	466	12	3	2048	3.57E-09	1.66	46.08	2.81	82.94
student	356	30	1	2048	4.63E-07	9.36	253.44	30.40	483.84
wine	4409	11	4	2048	2.62E-05	1.71	39.42	2.38	70.40

n = training data (number of data points)
d = number of features

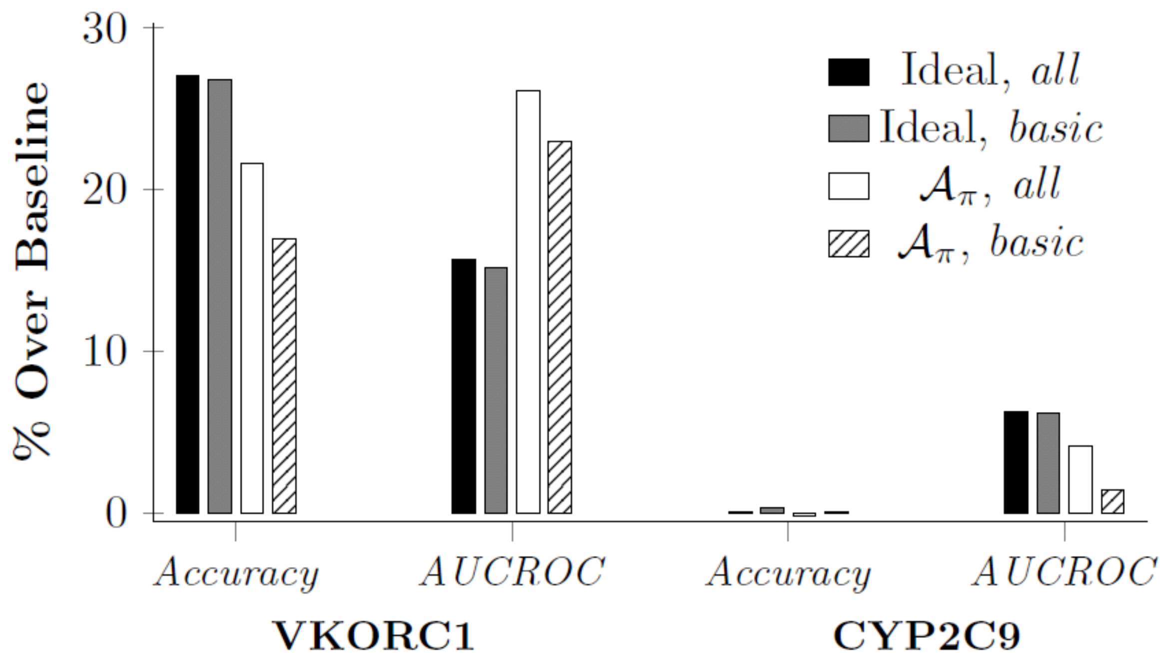
Comments on Homomorphic Encryption

- Benefits
  - High utility – because No Noise!!!
  - No one sees data “in the clear”

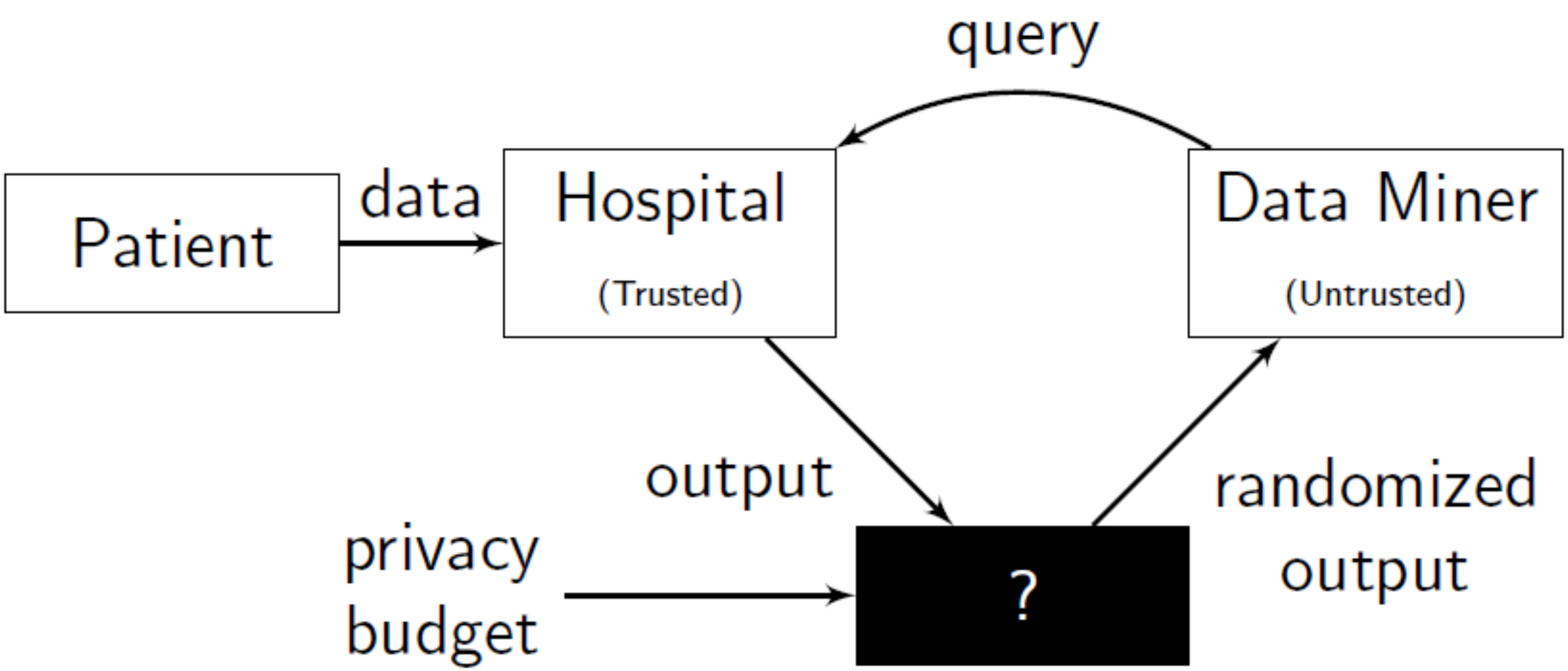
- Disadvantages
  - Models (or even just predictions) may still give away more information about training examples (e.g., patients) than about other examples (patients)
  - Very high (as of now, completely impractical) runtimes for some methods (fully homomorphic encryption)
  - Feasible approaches (e.g., linearly homomorphic encryption) require re-developing each learning algorithm (e.g., ridge regression) from scratch with limited operations
  - Protections may be lost if/when Quantum Computers become available

Just Releasing a Learned *Model* Can Violate Privacy

- IWPC Warfarin Model
- Can we predict genotype of training set better than others?



Privacy Blueprint



Differential Privacy (Dwork, 2006)

- Goal
  - Small added risk of adversary learning (private) information about an individual if his/her data in the private database versus not in the database
- Informally
  - Query output does not change much between neighboring databases
  - E.g.: what is fraction of people in clinic with diabetes?

Name	Has Diabetes (X)
Ross	1
Monica	1
Joey	0
Phoebe	0
Chandler	1

Differential Privacy Definition

- Given
  - Input database *D*
  - Randomized algorithm *f* : *D* -> *Range*(*f*)
  - *f* is (ϵ , δ)-differentially private iff

$$\Pr(f(D) \in S) \leq e^{\epsilon} \Pr(f(D') \in S) + \delta$$

- For any $S \in$ *Range*(*f*) and *D*' where d(*D*,*D*')=1
  - ε and δ are privacy budget
    - Smaller means more private

Obtaining Differential Privacy

- Note: Definition requires stochastic output… how to achieve?

- Perturbation {Laplace Mechanism} (Dwork, 2006)
  - Calculate correct answer f(D)
  - Add noise f(D) + η

- Soft-max {Exponential Mechanism} (McSherry and Talwar, 2007)
  - Quality function q(D,s)
  - Exponential weighting exp(ϵ q(D,s))

- In both cases, noise is proportional to the *sensitivity* of the function

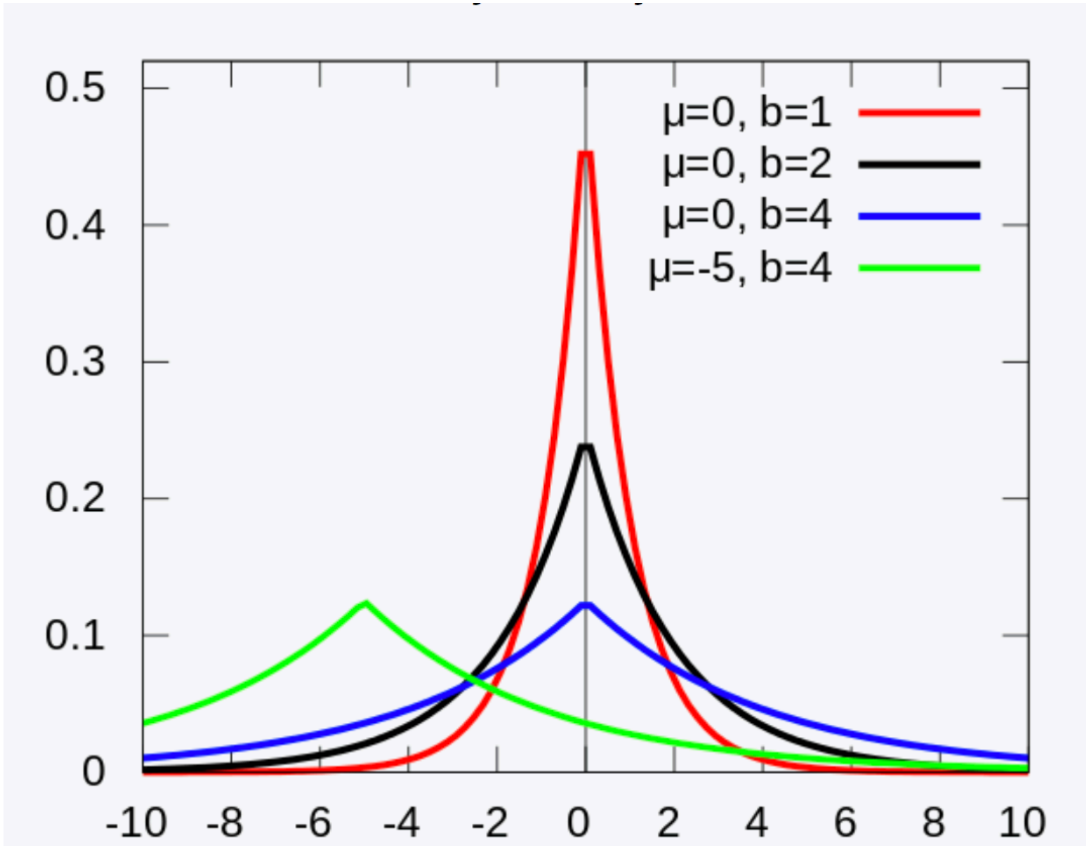
Global Sensitivity

- Given $f : D \rightarrow \mathbb{R}$, global sensitivity of f is

$$GS_f = \max_{d(D,D')=1} |f(D) - f(D')|$$

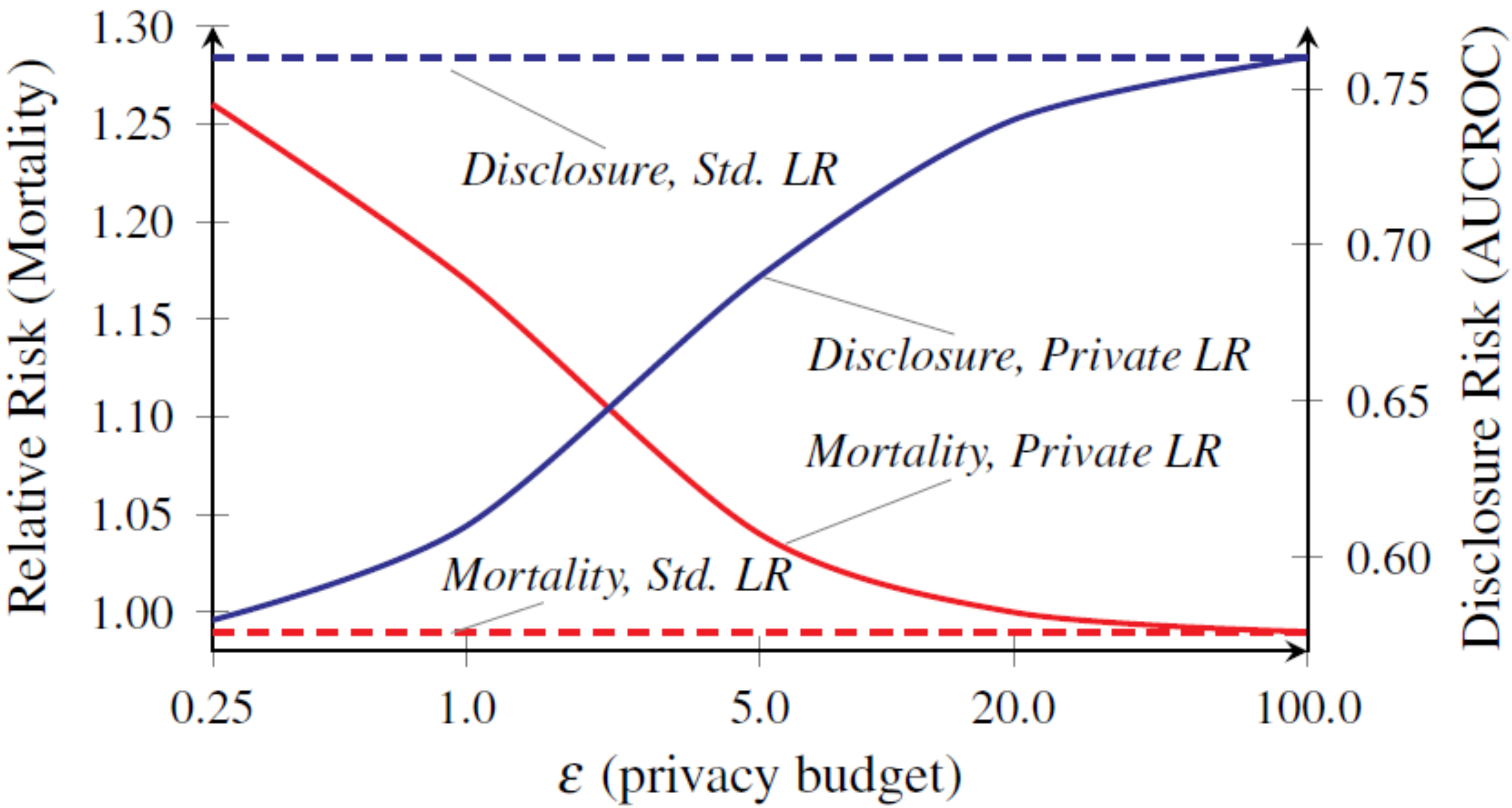
- Worst case
- Once f and the domain of D are chosen, global sensitivity is fixed

Add Laplace Noise, μ=0, *b* a function of sensitivity and ε



$$f(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

Privacy-Utility Tradeoff for Private Warfarin Model



Comments on Differential Privacy

- Provable guarantees, regardless of side information adversary has
- Elegant formulation that leads to many attractive algorithms
- Has insights for other areas such as fairness
- Poor intuition for how to select ε
- Can kill utility (e.g., accuracy, AUC) unless we have very many examples… so good fit for age of Big Data but not for medium data
- How to set privacy budget? If release DP dataset, can update with new release without adding to previous ε, so must plan far ahead