Privacy-preserving machine learning

@MLRS

Bangkok, Thailand

Aug 11, 2019

Mijung Park

Privacy & your data

- Increasingly more and more devices collect & stream your data

Internet & emails

[Lookeen]

Home appliances

[LG newsroom]

Drones

[Intogagets]

Privacy & your data

- Increasingly more and more devices collect & stream your data

Internet & emails

[Lookeen]

Home appliances

[LG newsroom]

Drones

[Intogagets]

[Electronic Frontier Foundation via Wikimedia Commons]

Worry:
A few corporations
own your data
& might abuse them!

My name is Cayla!

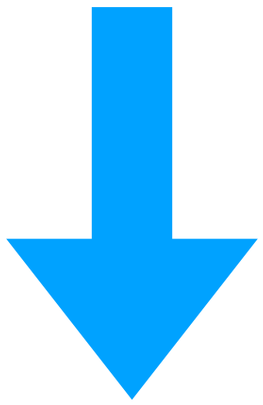
- Cayla is a children's smart toy, that can have interactive conversations with kids

My name is Cayla!

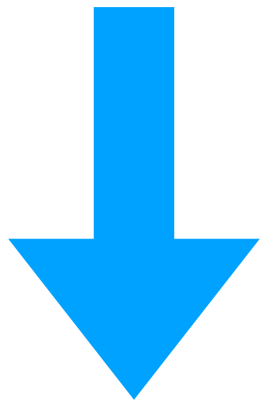
- Cayla is a children's smart toy, that can have interactive conversations with kids

- Issue 1: Conversations streamed to company's server
- Issue 2: anybody who has a bluetooth connection can hack this doll.

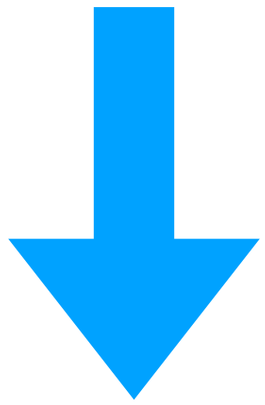
Threat!



hmm, maybe we should do something for privacy!

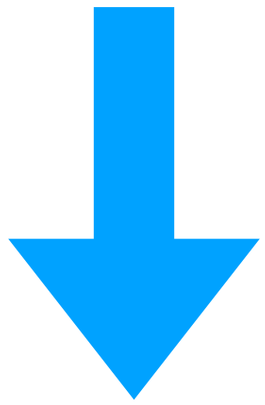


hmm, maybe we should do something for privacy!

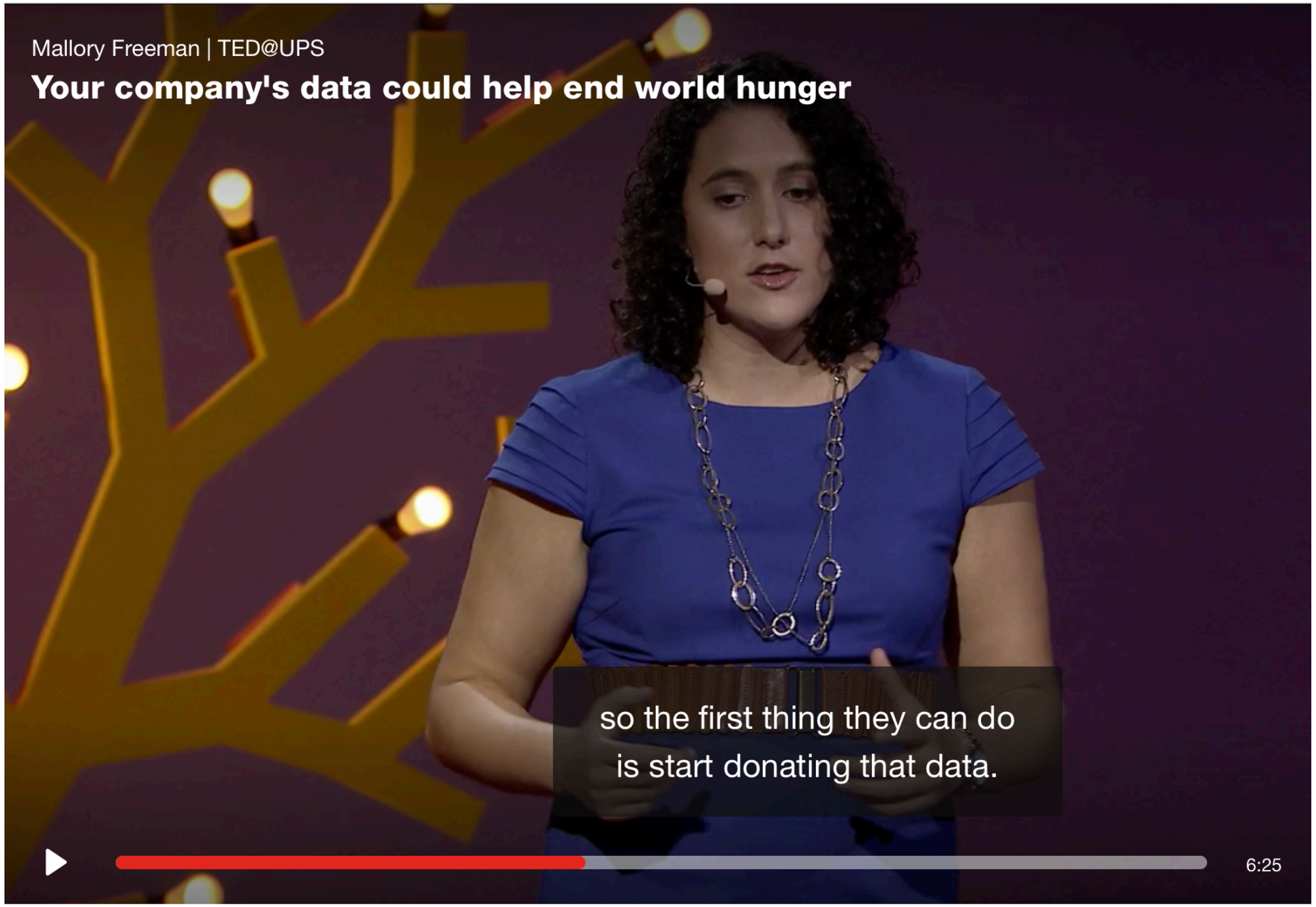


hmm, maybe we should do something for privacy!

Philanthropic reason!



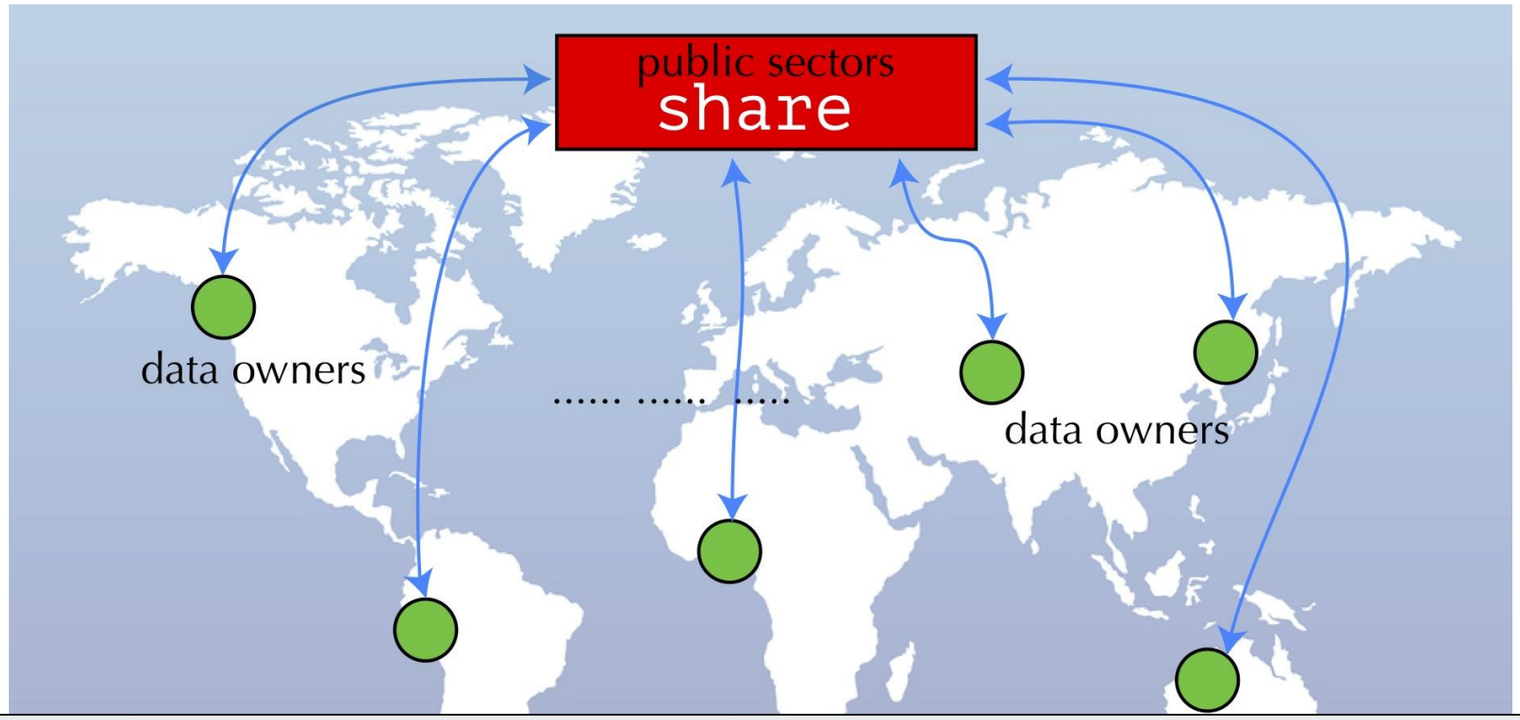
hmm, maybe we should do something for privacy!



- Emphasising the importance of sharing data for public good

Data philanthropy

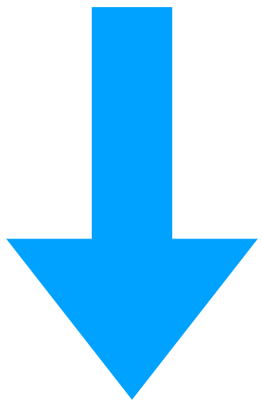
[United Nations Global Pulse]



"--- think of big data as a new kind of natural resource – infinitely renewable, increasingly ubiquitous – but … has fallen into the hands of… industry …Data has a social opportunity – and we have a social responsibility – …data reaches the people who need it most."

[Director of UN Global Pulse]

Threat!
Philanthropic reason!



hmm, maybe we should do something for privacy!

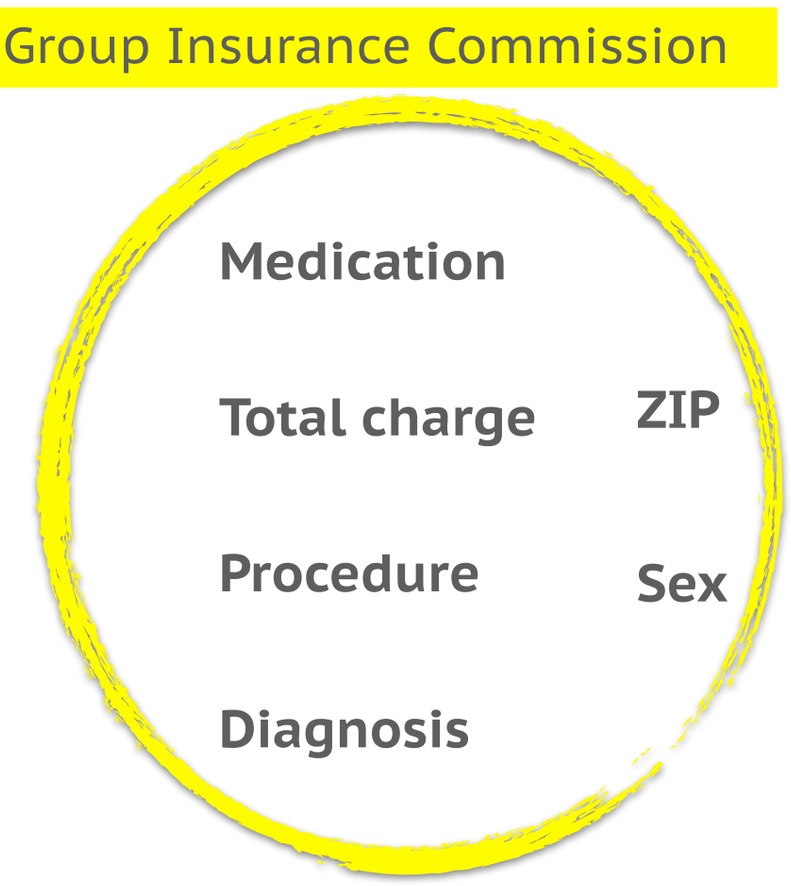
Let's anonymise data!

- Isn't anonymising, removing obviously identifiable information, enough?

Let's anonymise data!

- Isn't anonymising, removing obviously identifiable information, enough?

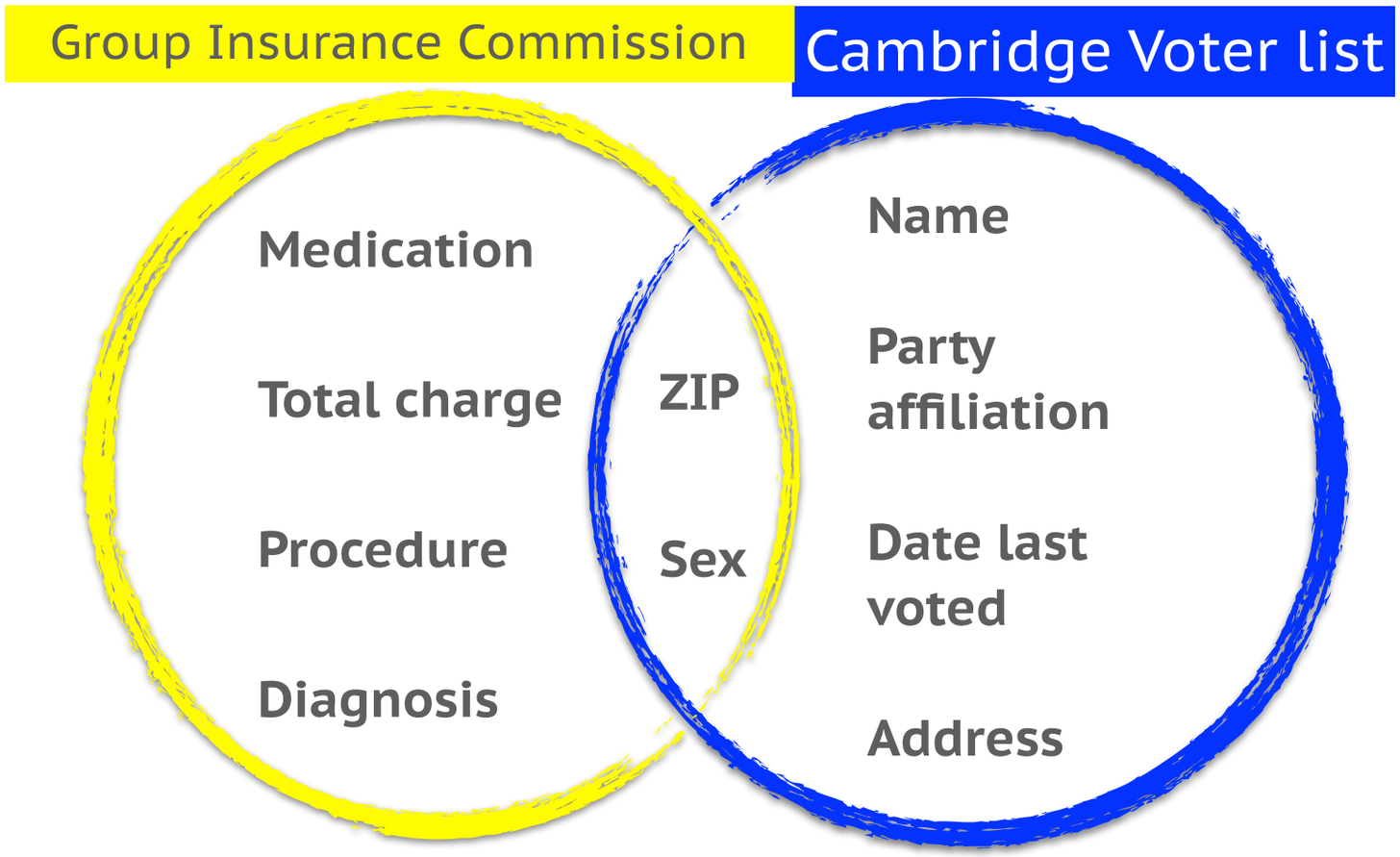
[Sweeny 02]



Let's anonymise data!

- Isn't anonymising, removing obviously identifiable information, enough?

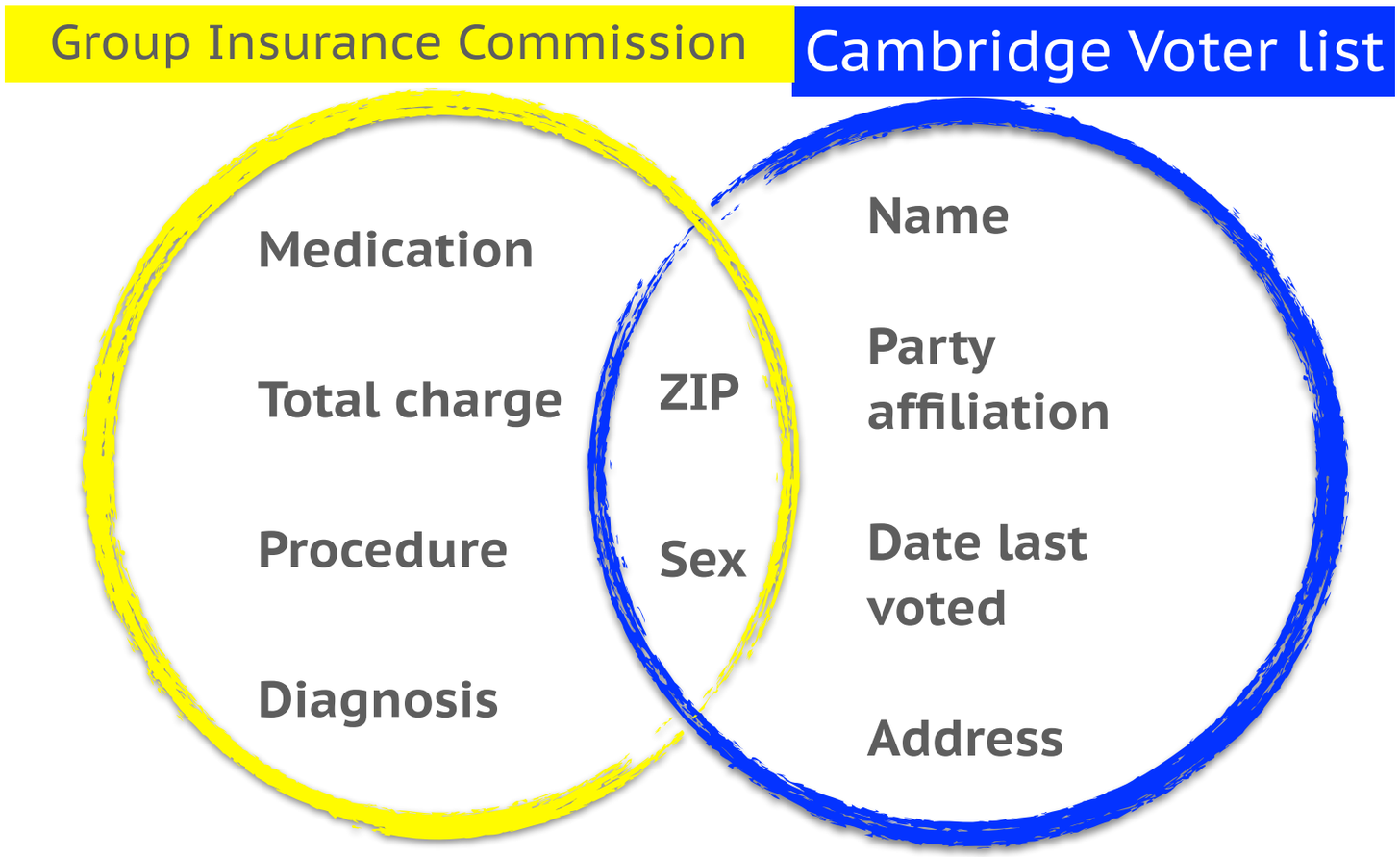
[Sweeny 02]



Let's anonymise data!

- Isn't anonymising, removing obviously identifiable information, enough?

[Sweeny 02]

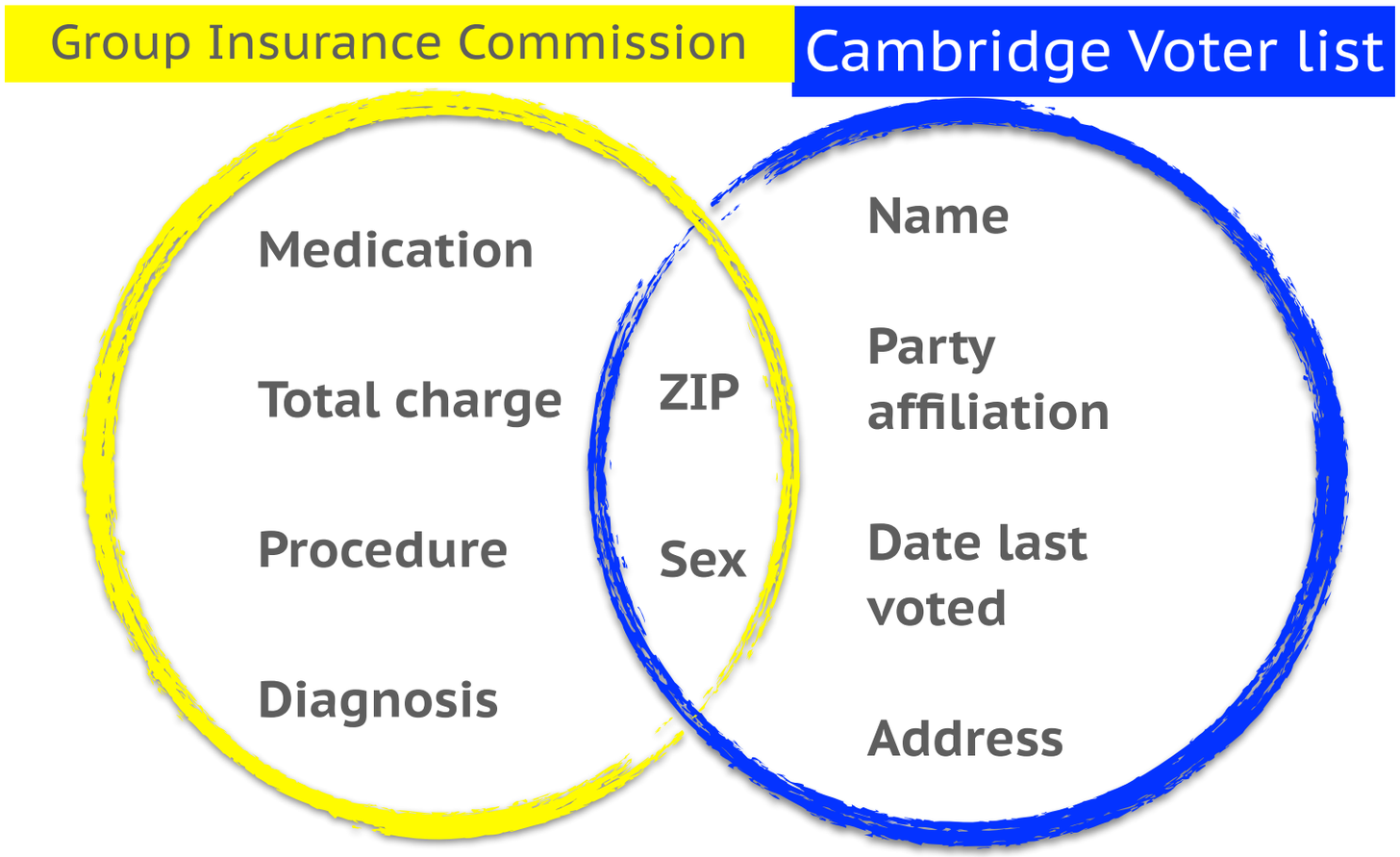


Six people listed, three men, but only one living in that particular zip-code

Let's anonymise data!

- Isn't anonymising, removing obviously identifiable information, enough?

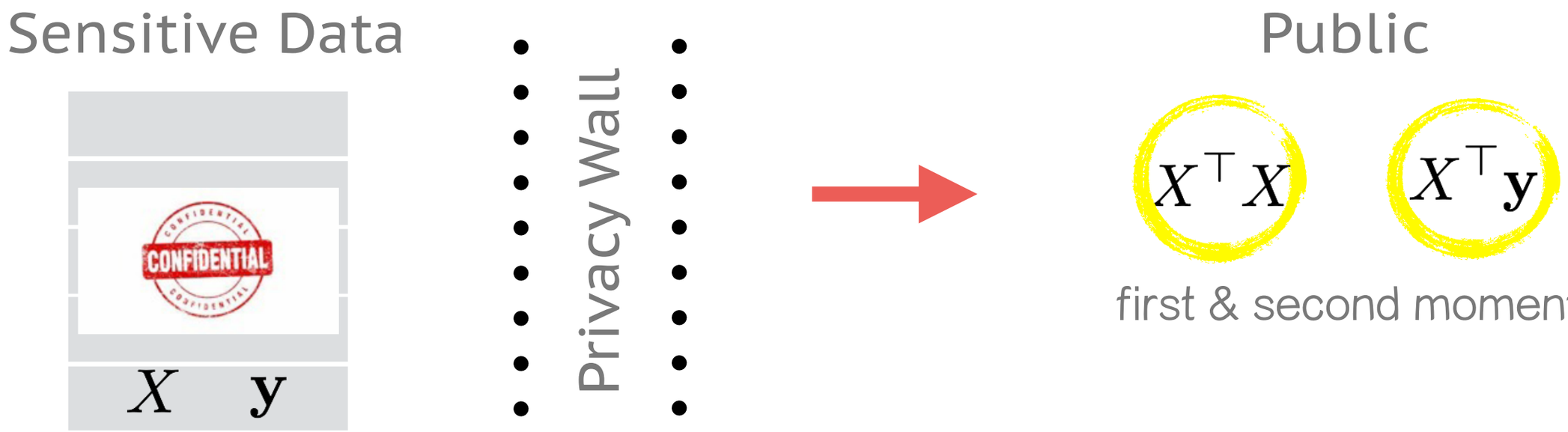
[Sweeny 02]



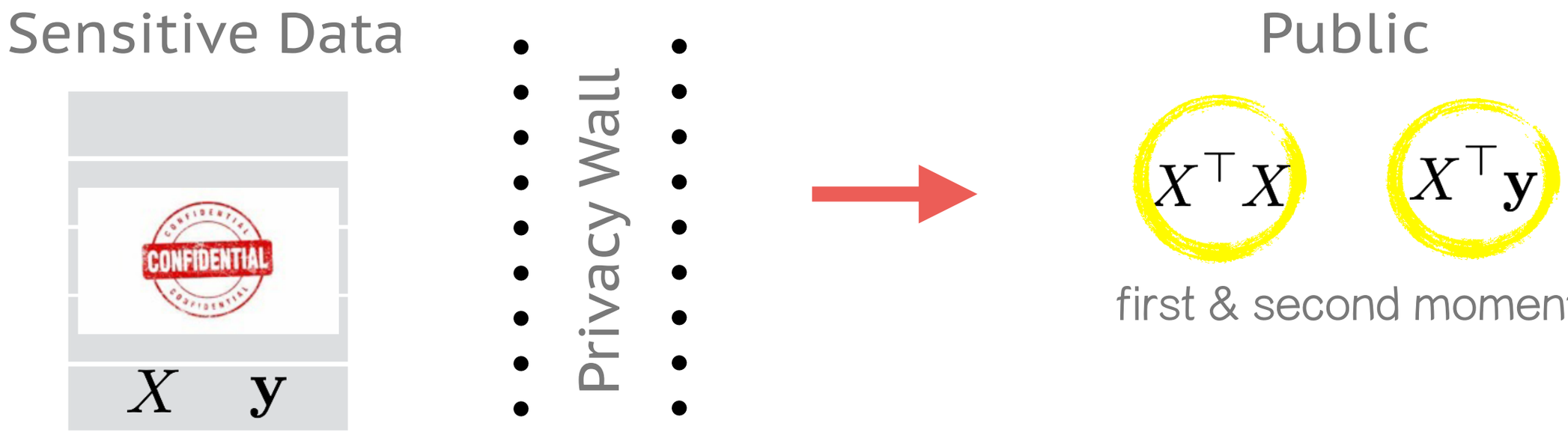
Six people listed, three men, but only one living in that particular zip-code

Governor William Weld!

What about releasing statistics?

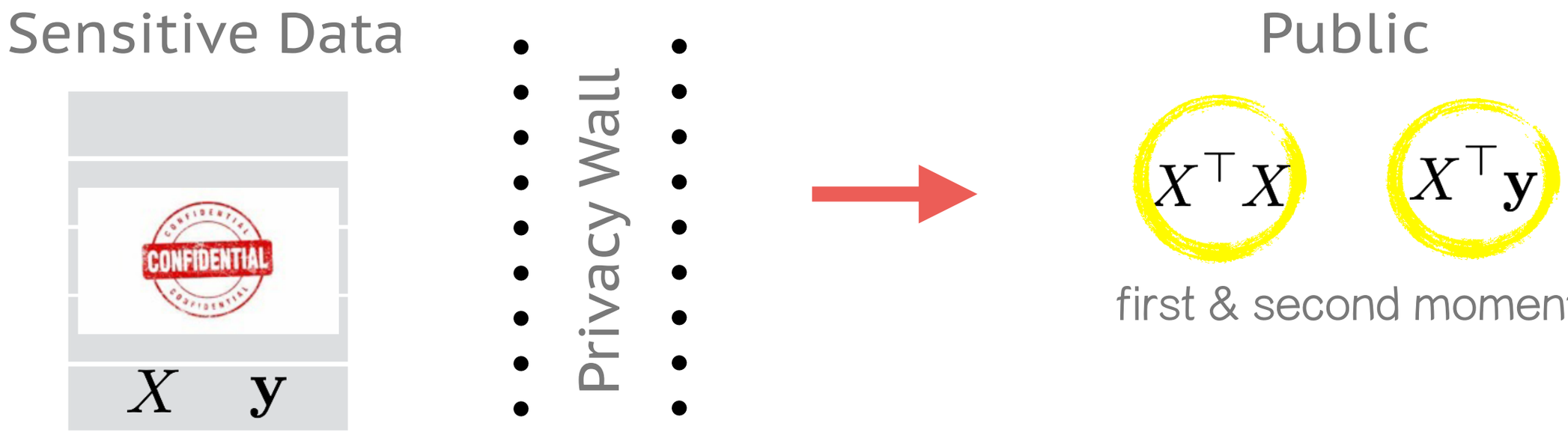


What about releasing statistics?



- Advanced attacking methods (e.g., model inversion attacks)!

What about releasing statistics?

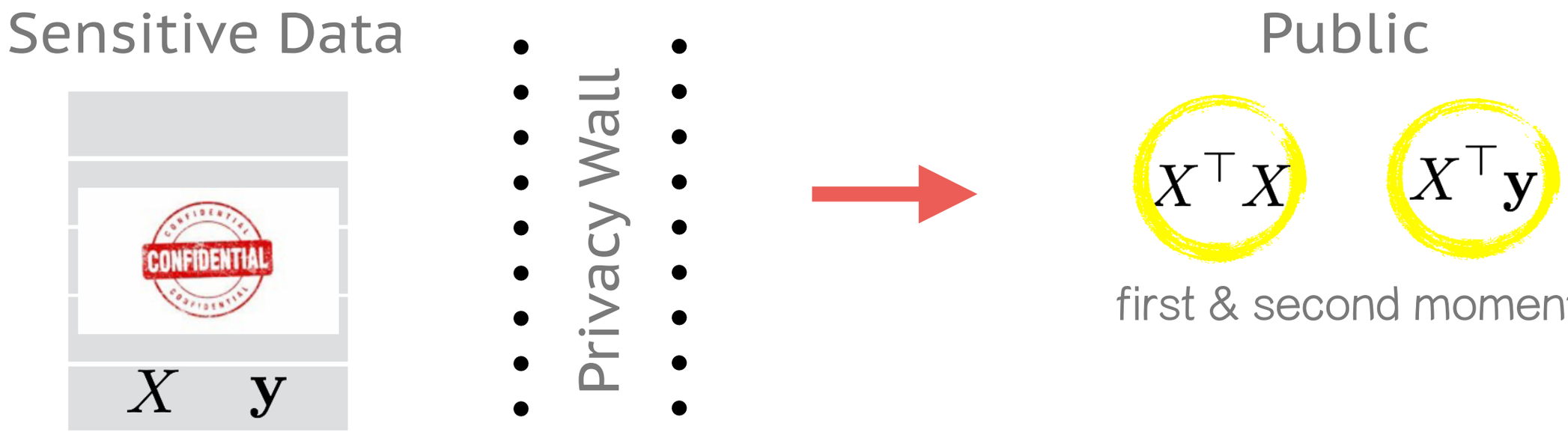


- Advanced attacking methods (e.g., model inversion attacks)!

prior knowledge

y_n $\mathbf{x}_n$ [known; unknown]

What about releasing statistics?



- Advanced attacking methods (e.g., model inversion attacks)!

prior knowledge

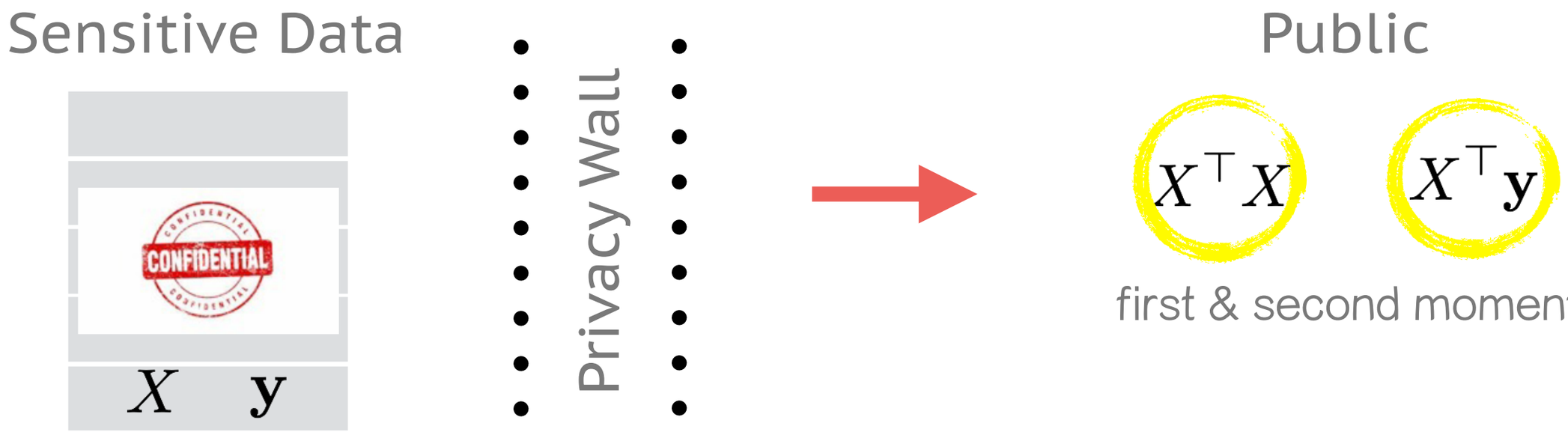
y_n $\mathbf{x}_n$ [known; unknown]

+

released stats

$$\hat{\boldsymbol{\theta}}_{ls} = (X^\top X)^{-1} X^\top \mathbf{y}$$

What about releasing statistics?



- Advanced attacking methods (e.g., model inversion attacks)!

prior knowledge

y_n $\mathbf{x}_n$

[known; unknown]

+

released stats

$\hat{\boldsymbol{\theta}}_{ls} = (X^\top X)^{-1} X^\top \mathbf{y}$

→

$y_n \approx \hat{\boldsymbol{\theta}}_{ls}{}^\top \mathbf{x}_n$

Statistically sound prediction of sensitive attributes

Simply anonymising data is unsafe!

Releasing statistics of data is unsafe!

Simply anonymising data is unsafe!

Releasing statistics of data is unsafe!

⋮

We need a provably strong privacy notion!

Privacy definition

Differential privacy!

Differential privacy!

[Dwork 06]

Differential privacy is the statistical science of trying to learn as much as possible about a group while learning as little as possible about any individual in it.

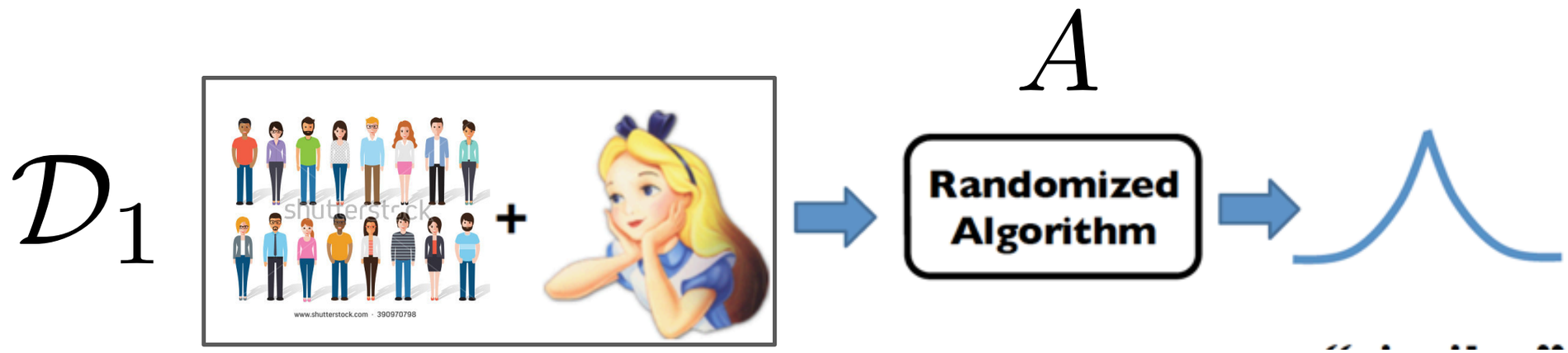
[Andy Greenberg, 2016]

Differential privacy!

[Dwork 06]

Differential privacy is the statistical science of trying to learn as much as possible about a group while learning as little as possible about any individual in it.

[Andy Greenberg, 2016]

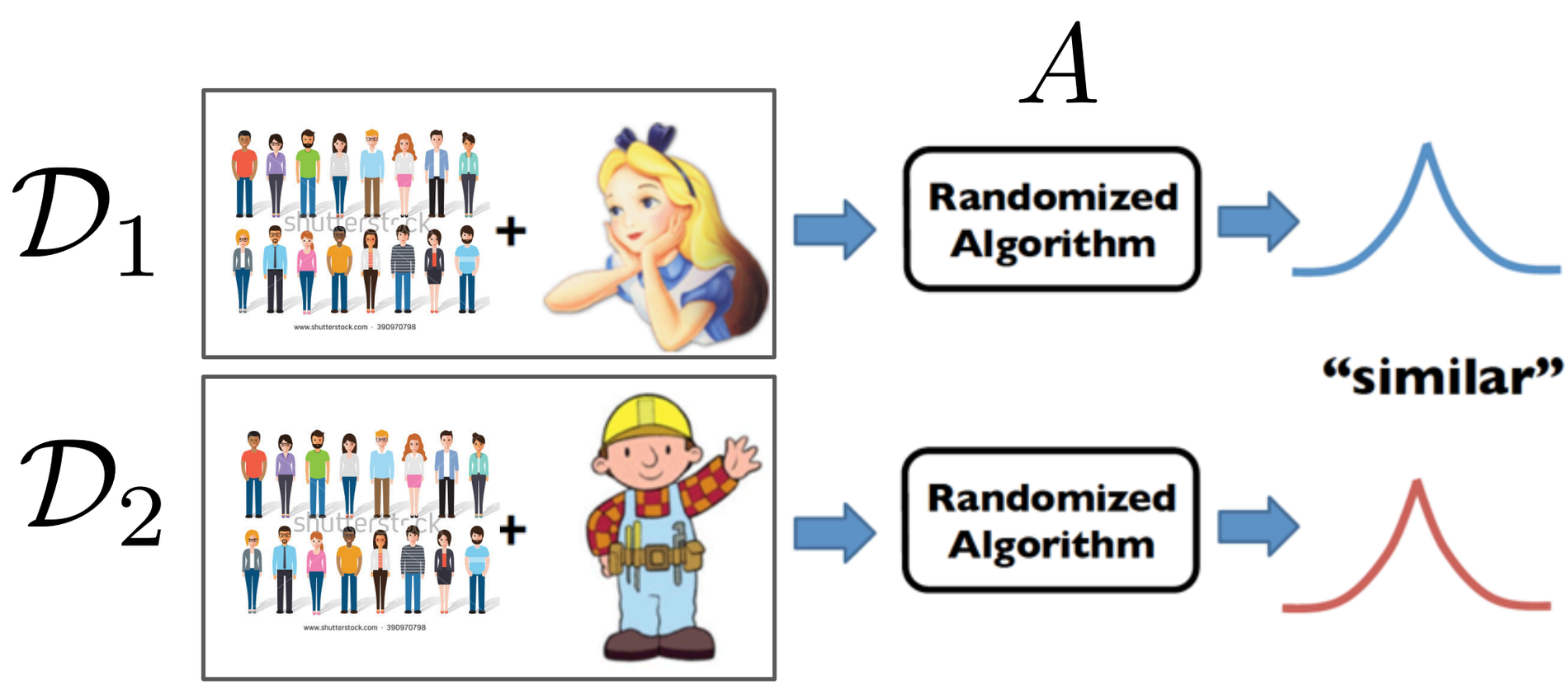


Differential privacy!

[Dwork 06]

Differential privacy is the statistical science of trying to learn as much as possible about a group while learning as little as possible about any individual in it.

[Andy Greenberg, 2016]



An algorithm is differentially private, if

A is epsilon-DP, if $|L^{(o)}| \leq \epsilon$

for all o and all pairs of datasets

An algorithm is differentially private, if

A is epsilon-DP, if $|L^{(o)}| \leq \epsilon$

for all o and all pairs of datasets

Privacy loss: $L^{(o)} = \log \frac{P(A(\mathcal{D}_1) = o)}{P(A(\mathcal{D}_2) = o)}$

An algorithm is differentially private, if

A is epsilon-DP, if $|L^{(o)}| \leq \epsilon$

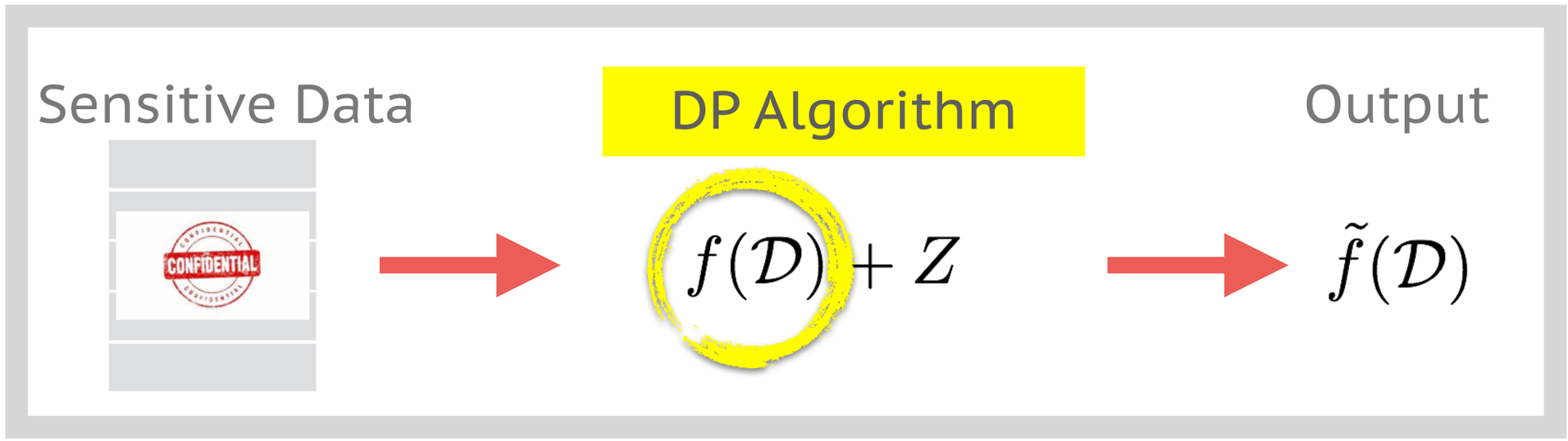
for all o and all pairs of datasets

Privacy loss: $L^{(o)} = \log \frac{P(A(\mathcal{D}_1) = o)}{P(A(\mathcal{D}_2) = o)}$

- Epsilon quantitatively measures how much the risk to an individual privacy may increase due to that individual's data inclusion in the inputs to the algorithm.

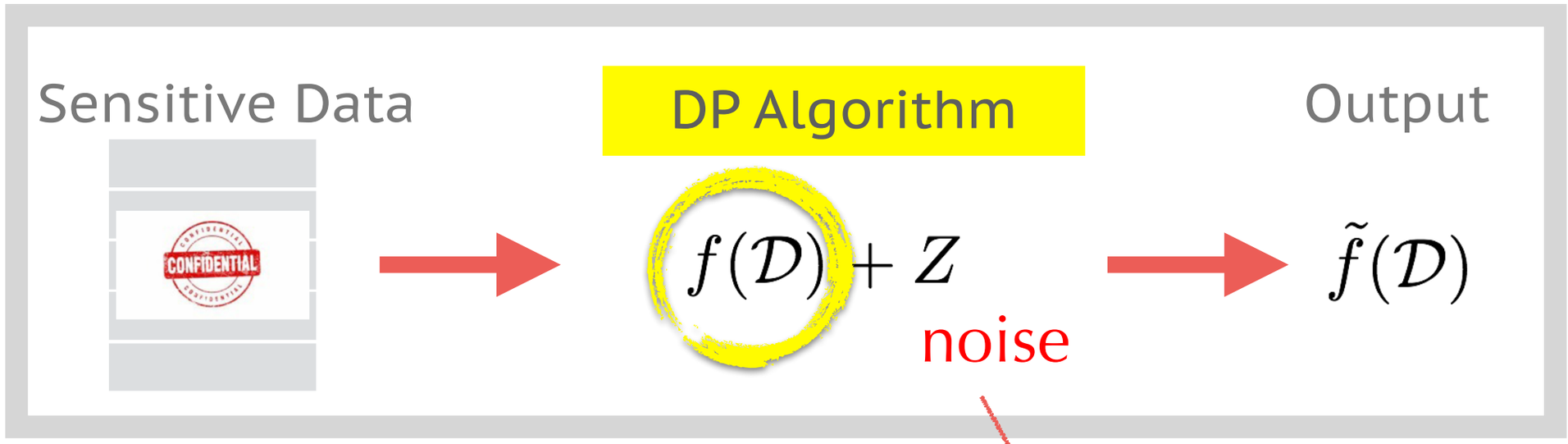
How to develop a DP Algorithm?

[Global Sensitivity Method, DMNS06]



How to develop a DP Algorithm?

[Global Sensitivity Method, DMNS06]



How much noise to add?

$$\text{noise variance} \propto \frac{S(f)^2}{\epsilon^2}$$

How much noise to add?

$$\text{noise variance} \propto \frac{S(f)^2}{\epsilon^2}$$

privacy loss

How much noise to add?

$$\text{noise variance} \propto \frac{S(f)^2}{\epsilon^2}$$

Sensitivity

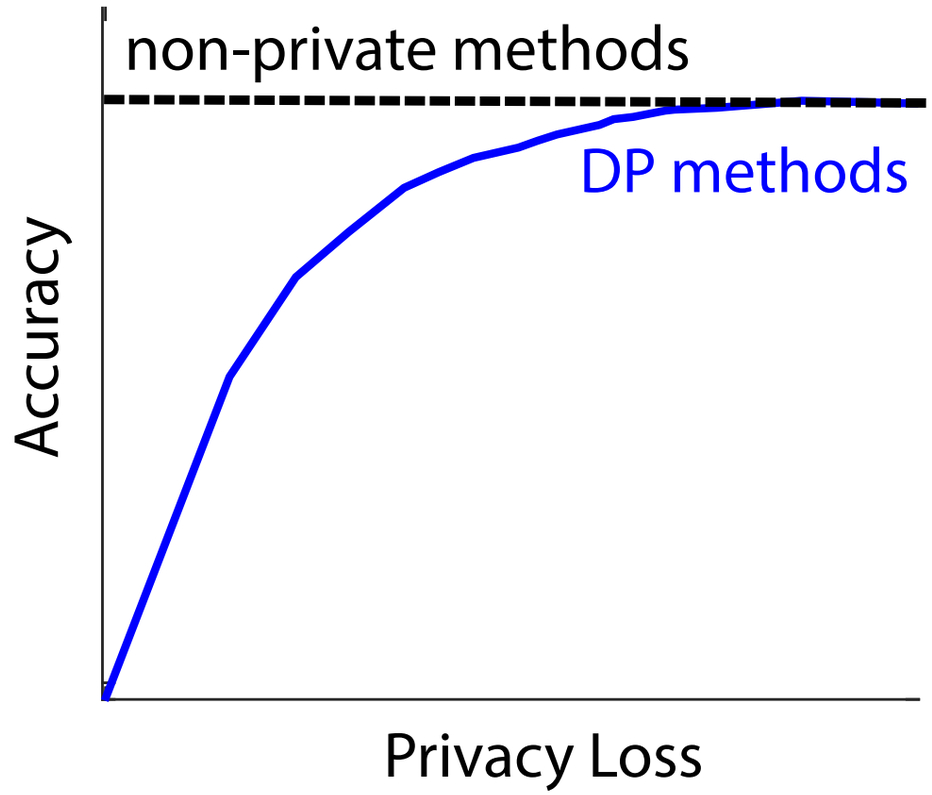
privacy loss

$$S(f) = \max_{D(\mathcal{D},\mathcal{D}')=1} |f(\mathcal{D}) - f(\mathcal{D}')|$$

maximum over all pairs of datasets

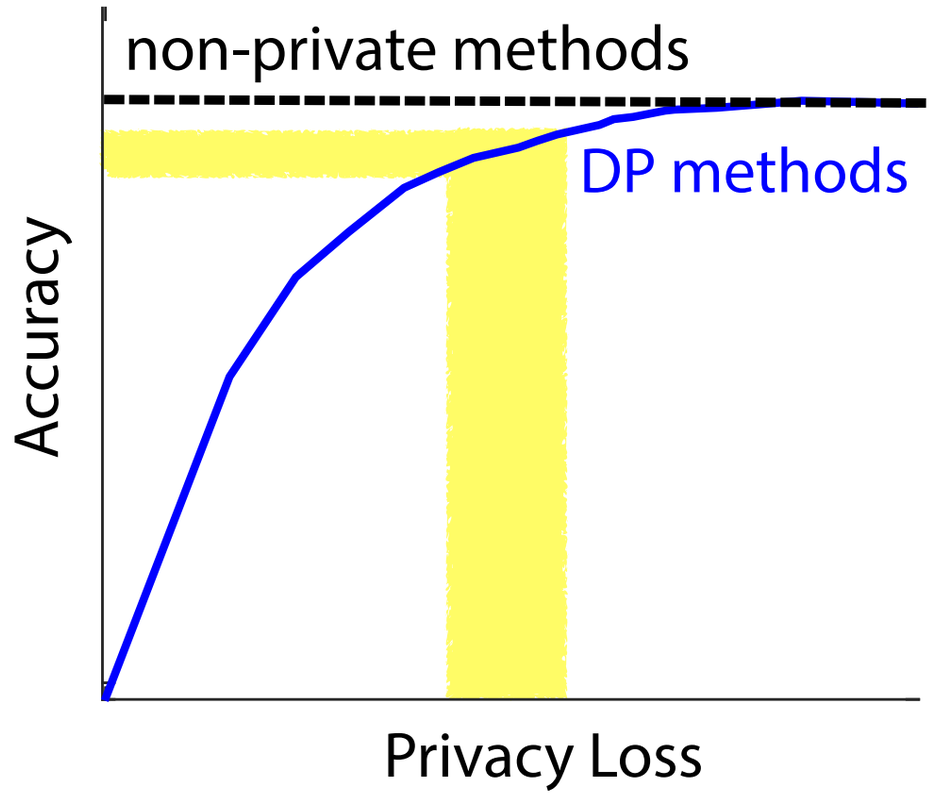
Privacy & Accuracy Trade-off

$$\text{noise} \propto \frac{\text{sensitivity}}{\text{privacy loss}}$$



Privacy & Accuracy Trade-off

$$\text{noise} \propto \frac{\text{sensitivity}}{\text{privacy loss}}$$



Good DP-ML algorithms:

provide accurate

& differentially private outputs

Example

DP mechanisms

Laplace Mechanism

Laplace Mechanism

Input: Monthly salary $x_1, x_2, \cdots, x_N$

Laplace Mechanism

Input: Monthly salary $x_1, x_2, \cdots, x_N$

$x_i \in [0, 300]$

Laplace Mechanism

Input: Monthly salary $x_1, x_2, \cdots, x_N$

$x_i \in [0, 300]$

Intended output: average $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$

Laplace Mechanism

Input: Monthly salary $x_1, x_2, \cdots, x_N$

$x_i \in [0, 300]$

Intended output: average $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$

Differentially private average:

$$\bar{x} + Z$$

Laplace Mechanism

Input: Monthly salary $x_1, x_2, \cdots, x_N$

$$x_i \in [0, 300]$$

Intended output: average $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$

Differentially private average:

$$\bar{x} + Z \longrightarrow Z \sim \text{Laplace}\left(\frac{S(\bar{x})}{\epsilon}\right)$$

Laplace Mechanism

Input: Monthly salary $x_1, x_2, \cdots, x_N$

$x_i \in [0, 300]$

Intended output: average $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$

Differentially private average:

$$\bar{x} + Z \longrightarrow Z \sim \text{Laplace}\left(\frac{S(\bar{x})}{\epsilon}\right)$$

Sensitivity: $S(\bar{x}) = \max_{\mathcal{D}, \mathcal{D}'} |\bar{x} - \bar{x}'|$

Laplace mechanism is DP

$$\mathcal{M}(\mathcal{D}) = \bar{x} + Z \qquad Z \sim \text{Laplace}\left(\frac{S(\bar{x})}{\epsilon}\right)$$

Laplace mechanism is DP

$$\mathcal{M}(\mathcal{D}) = \bar{x} + Z \qquad Z \sim \text{Laplace}\left(\frac{S(\bar{x})}{\epsilon}\right)$$

(exponentiated) Privacy loss:

$$\frac{p(\mathcal{M}(\mathcal{D}) = o)}{p(\mathcal{M}(\mathcal{D}') = o)}$$

Laplace mechanism is DP

$$\mathcal{M}(\mathcal{D}) = \bar{x} + Z \qquad Z \sim \text{Laplace}\left(\frac{S(\bar{x})}{\epsilon}\right)$$

(exponentiated) Privacy loss:

$$\frac{p(\mathcal{M}(\mathcal{D}) = o)}{p(\mathcal{M}(\mathcal{D}') = o)} = \frac{\exp\left(-\frac{\epsilon}{S(\bar{x})}|\bar{x} - o|\right)}{\exp\left(-\frac{\epsilon}{S(\bar{x}\ )}|\bar{x}' - o|\right)}$$

Laplace mechanism is DP

$$\mathcal{M}(\mathcal{D}) = \bar{x} + Z \qquad Z \sim \text{Laplace}\left(\frac{S(\bar{x})}{\epsilon}\right)$$

(exponentiated) Privacy loss:

$$\frac{p(\mathcal{M}(\mathcal{D}) = o)}{p(\mathcal{M}(\mathcal{D}') = o)} = \frac{\exp\left(-\frac{\epsilon}{S(\bar{x})}|\bar{x} - o|\right)}{\exp\left(-\frac{\epsilon}{S(\bar{x}\ )}|\bar{x}' - o|\right)}$$

$$\leq \exp\left(\epsilon \frac{|\bar{x} - \bar{x}'|}{S(\bar{x})}\right)$$

Laplace mechanism is DP

$$\mathcal{M}(\mathcal{D}) = \bar{x} + Z \qquad Z \sim \text{Laplace}\left(\frac{S(\bar{x})}{\epsilon}\right)$$

(exponentiated) Privacy loss:

$$\frac{p(\mathcal{M}(\mathcal{D}) = o)}{p(\mathcal{M}(\mathcal{D}') = o)} = \frac{\exp\left(-\frac{\epsilon}{S(\bar{x})}|\bar{x} - o|\right)}{\exp\left(-\frac{\epsilon}{S(\bar{x}\ )}|\bar{x}' - o|\right)}$$

$$\leq \exp\left(\epsilon \frac{|\bar{x} - \bar{x}'|}{S(\bar{x})}\right)$$

$$S(\bar{x}) = \max_{\mathcal{D},\mathcal{D}'} |\bar{x} - \bar{x}'|$$

Laplace mechanism is DP

$$\mathcal{M}(\mathcal{D}) = \bar{x} + Z \qquad Z \sim \text{Laplace}\left(\frac{S(\bar{x})}{\epsilon}\right)$$

(exponentiated) Privacy loss:

$$\frac{p(\mathcal{M}(\mathcal{D}) = o)}{p(\mathcal{M}(\mathcal{D}') = o)} = \frac{\exp\left(-\frac{\epsilon}{S(\bar{x})}|\bar{x} - o|\right)}{\exp\left(-\frac{\epsilon}{S(\bar{x}\ )}|\bar{x}' - o|\right)}$$

$$\leq \exp\left(\epsilon \frac{|\bar{x} - \bar{x}'|}{S(\bar{x})}\right)$$

$$\leq \exp(\epsilon)$$

$$S(\bar{x}) = \max_{\mathcal{D},\mathcal{D}'} |\bar{x} - \bar{x}'|$$

Gaussian Mechanism

Add Gaussian noise $f(\mathcal{D}) + Z$

$Z \sim \mathcal{N}(0, \sigma^2)$

Gaussian Mechanism

Add Gaussian noise $f(\mathcal{D}) + Z$

$$Z \sim \mathcal{N}(0, \sigma^2)$$

noise standard deviation: $\sigma \geq \dfrac{c(\delta)S(f)}{\epsilon}$

Gaussian Mechanism

Add Gaussian noise $f(\mathcal{D}) + Z$

$Z \sim \mathcal{N}(0, \sigma^2)$

noise standard deviation: $\sigma \geq \frac{c(\delta)S(f)}{\epsilon}$

$P(L^{(o)} > \epsilon) \leq \delta$

Privacy loss

Gaussian Mechanism

Add Gaussian noise $f(\mathcal{D}) + Z$

$$Z \sim \mathcal{N}(0, \sigma^2)$$

noise standard deviation: $\sigma \geq \dfrac{c(\delta)S(f)}{\epsilon}$

Approximate DP: $P(L^{(o)} > \epsilon) \leq \delta$

Privacy loss

$(\epsilon, \delta) - \mathrm{DP}$

Pure DP holds w.p. at least $1 - \delta$

DP Mechanisms

- Laplace mechanism: provides a pure DP guarantee

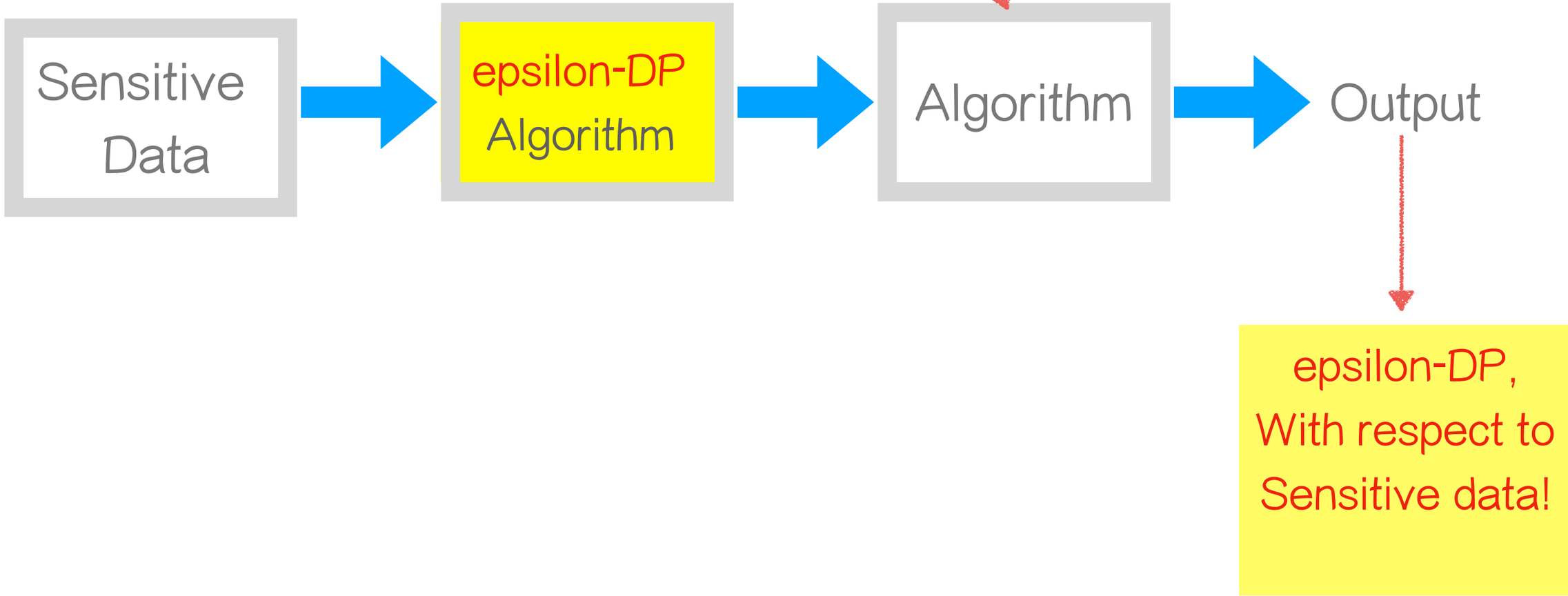
- Gaussian mechanism: provides an approximate DP guarantee

- There are more mechanisms. See [Algorithmic foundations of differential privacy, Dwork & Roth 2014].

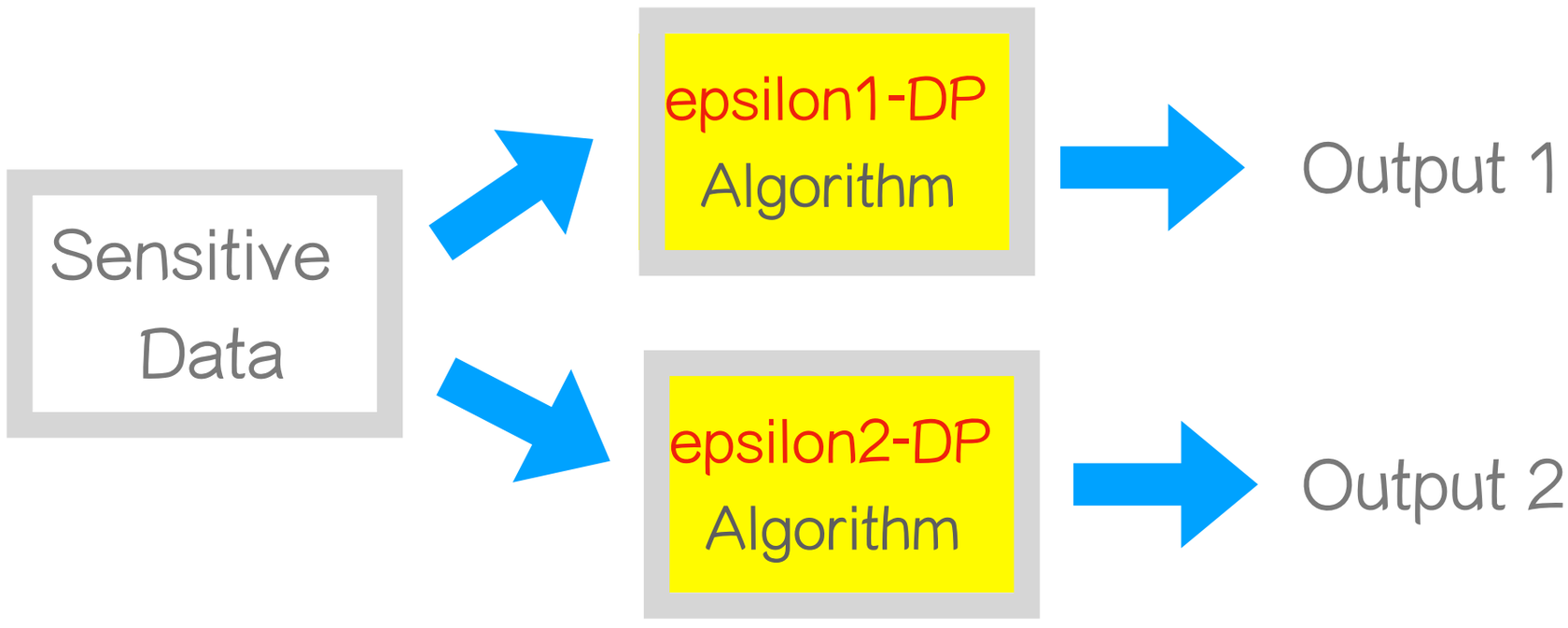
Properties of DP

1. Postprocessing invariance

- Differential privacy is immune to post-processing:



2. Composition



- Union of output 1 (from epsilon1-DP algorithm) & output 2 (from epsilon2-DP algorithm) is (epsilon1+epsilon2)-DP!
- More refined composition methods do exist! e.g., [Abadi et al,16]

Key properties of DP

- Post-processing invariance

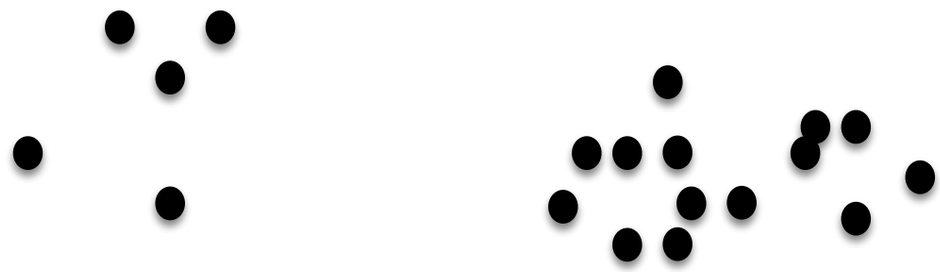
- Composition : introduces challenges to make iterative algorithms in ML differentially private!

Examples of DP-ML algorithms

Clustering & density estimation

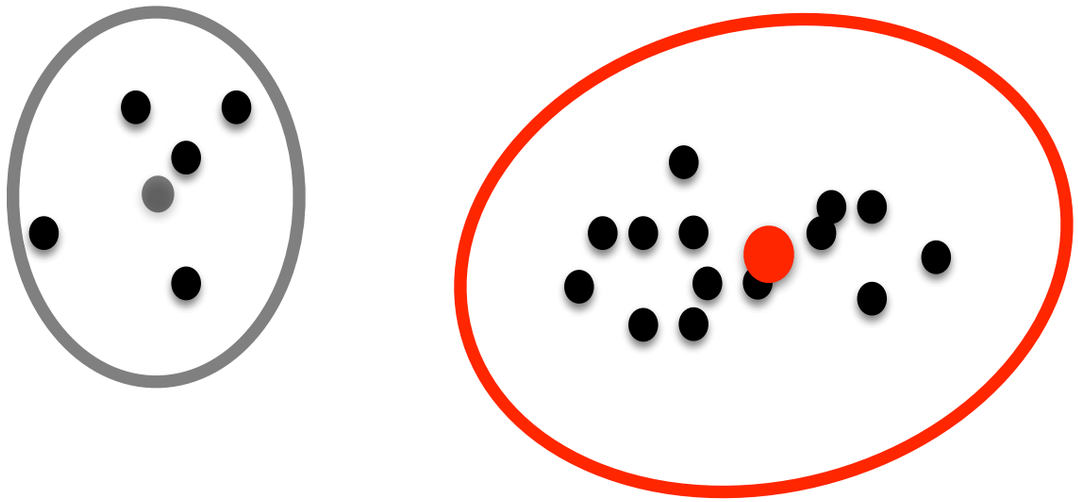
Mixture of Gaussians

For clustering & density estimation



Mixture of Gaussians

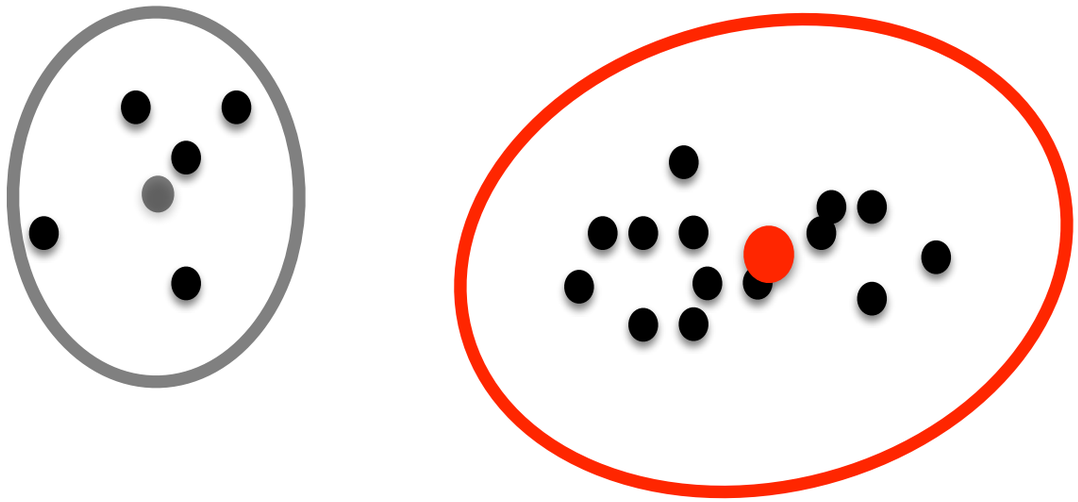
For clustering & density estimation



- Assign a Gaussian to learn each cluster

Mixture of Gaussians

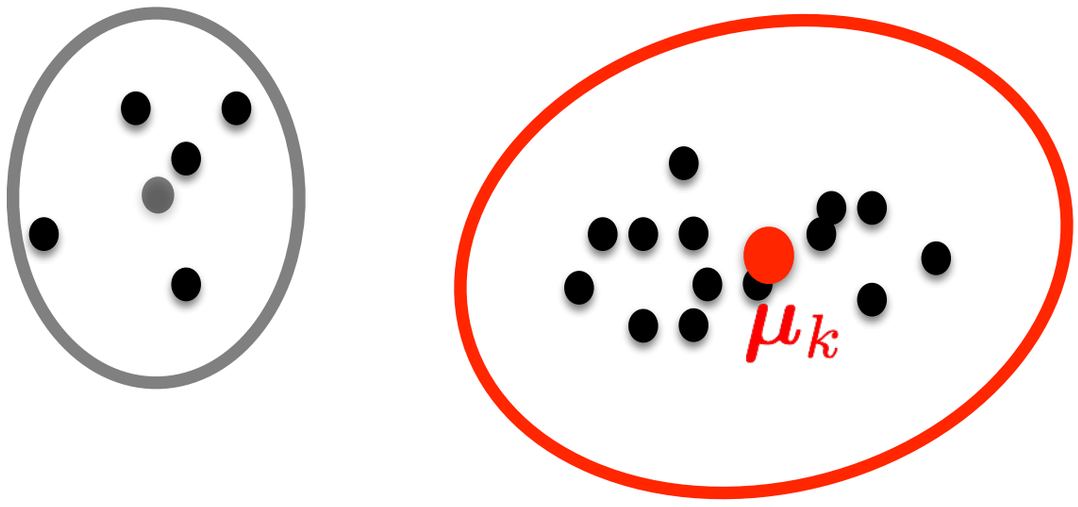
For clustering & density estimation



- Assign a Gaussian to learn each cluster
- Each Gaussian is parameterised by $\boldsymbol{m} = \{\boldsymbol{\mu}_k, \Sigma_k\}_{k=1}^{K}$

Mixture of Gaussians

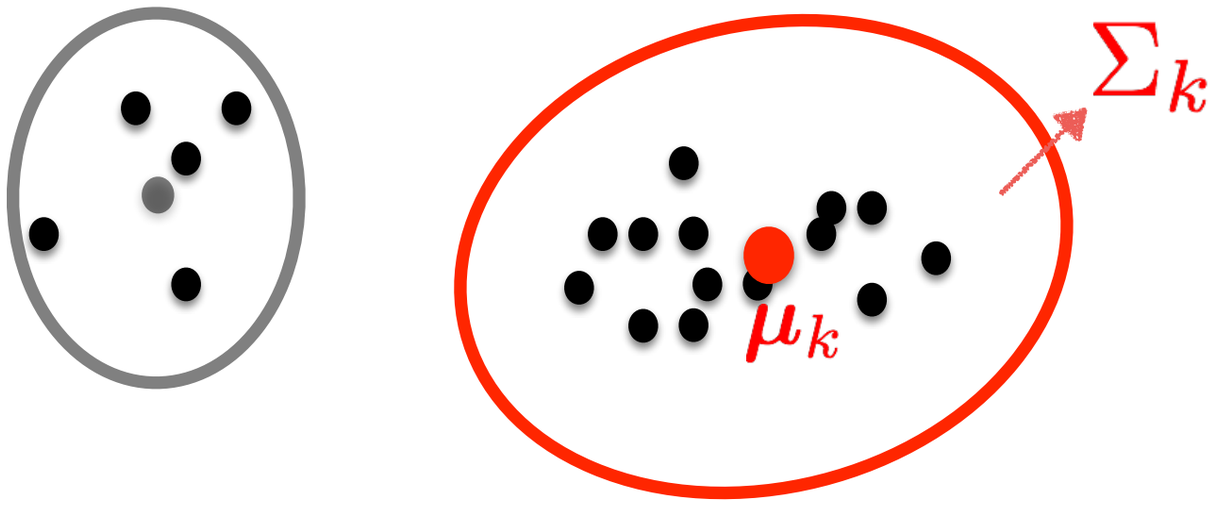
For clustering & density estimation



- Assign a Gaussian to learn each cluster
- Each Gaussian is parameterised by $\boldsymbol{m} = \{\boldsymbol{\mu}_k, \Sigma_k\}_{k=1}^{K}$

Mixture of Gaussians

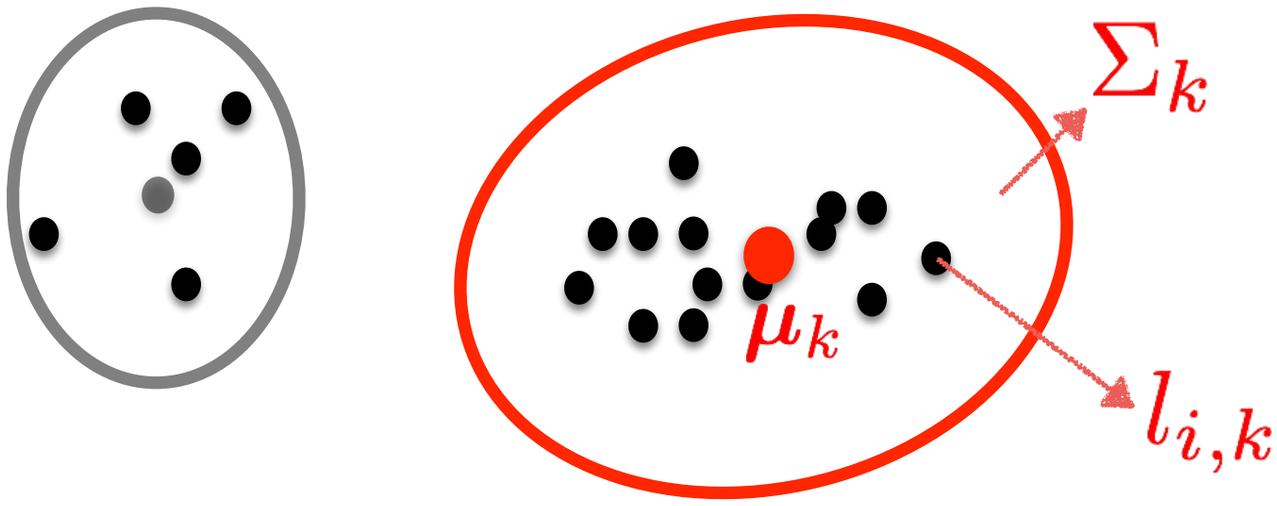
For clustering & density estimation



- Assign a Gaussian to learn each cluster
- Each Gaussian is parameterised by $\boldsymbol{m} = \{\boldsymbol{\mu}_k, \Sigma_k\}_{k=1}^{K}$

Mixture of Gaussians

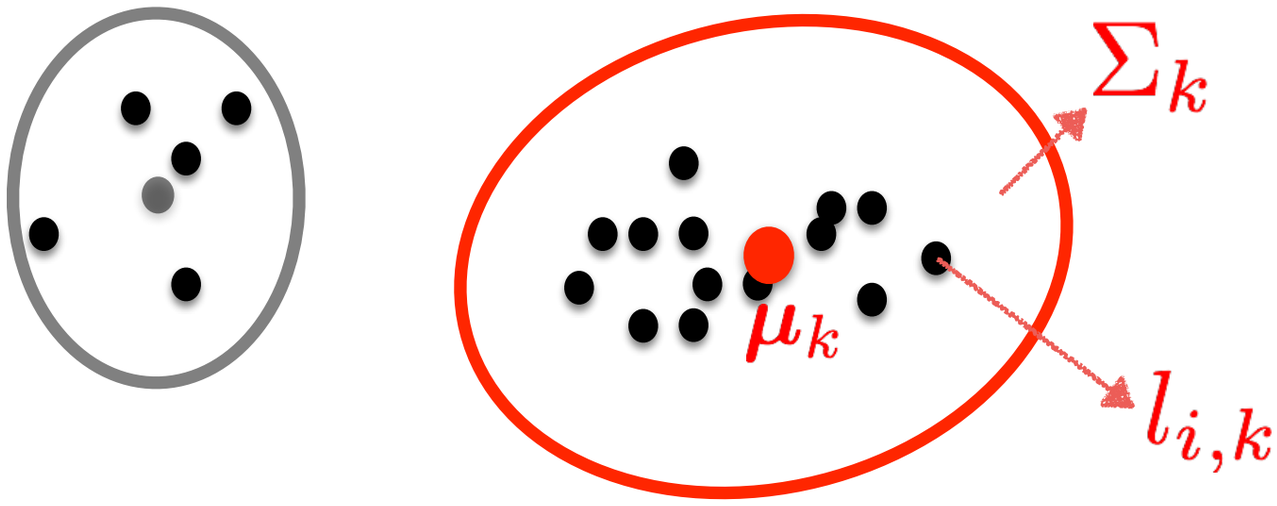
For clustering & density estimation



- Assign a Gaussian to learn each cluster
- Each Gaussian is parameterised by $\boldsymbol{m} = \{\boldsymbol{\mu}_k, \Sigma_k\}_{k=1}^{K}$
- Introduce **latent** variables for cluster assignment

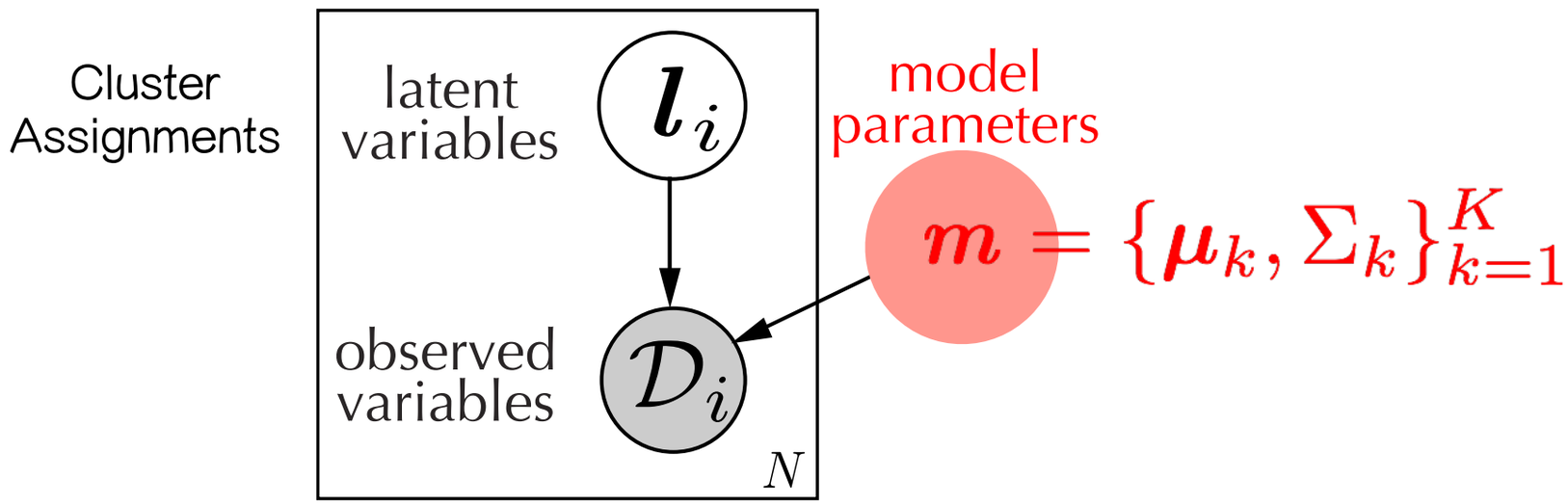
Mixture of Gaussians

For clustering & density estimation

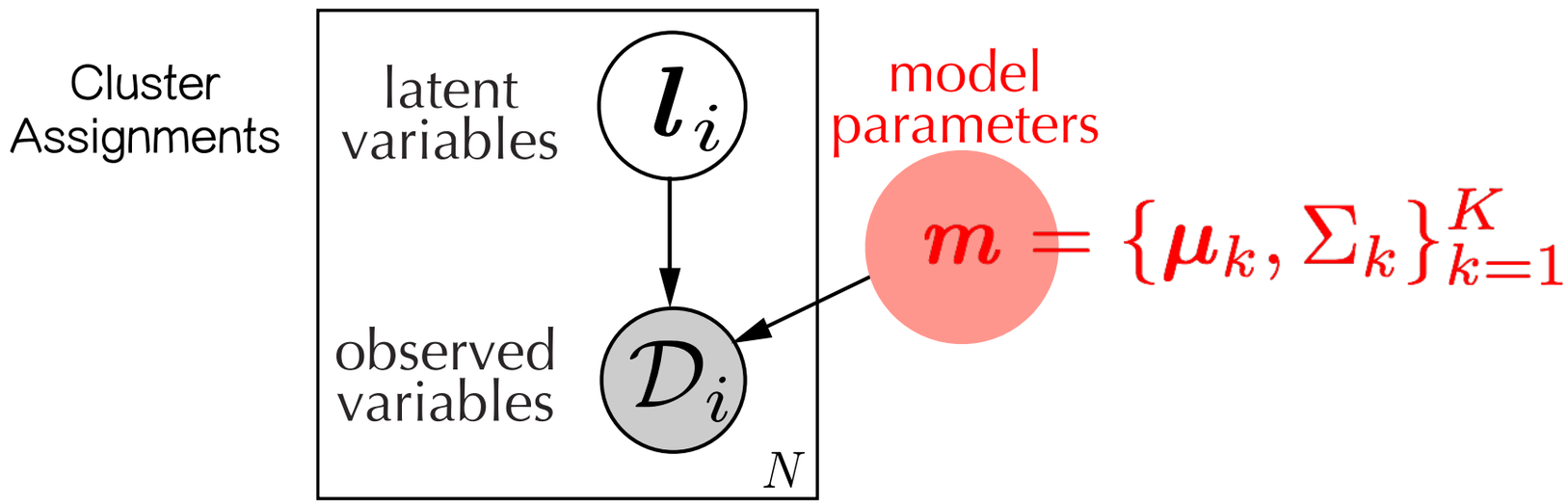


- Assign a Gaussian to learn each cluster
- Each Gaussian is parameterised by $\boldsymbol{m} = \{\boldsymbol{\mu}_k, \Sigma_k\}_{k=1}^{K}$
- Introduce **latent** variables for cluster assignment
- **Goal:** estimate the parameters

Expectation Maximization



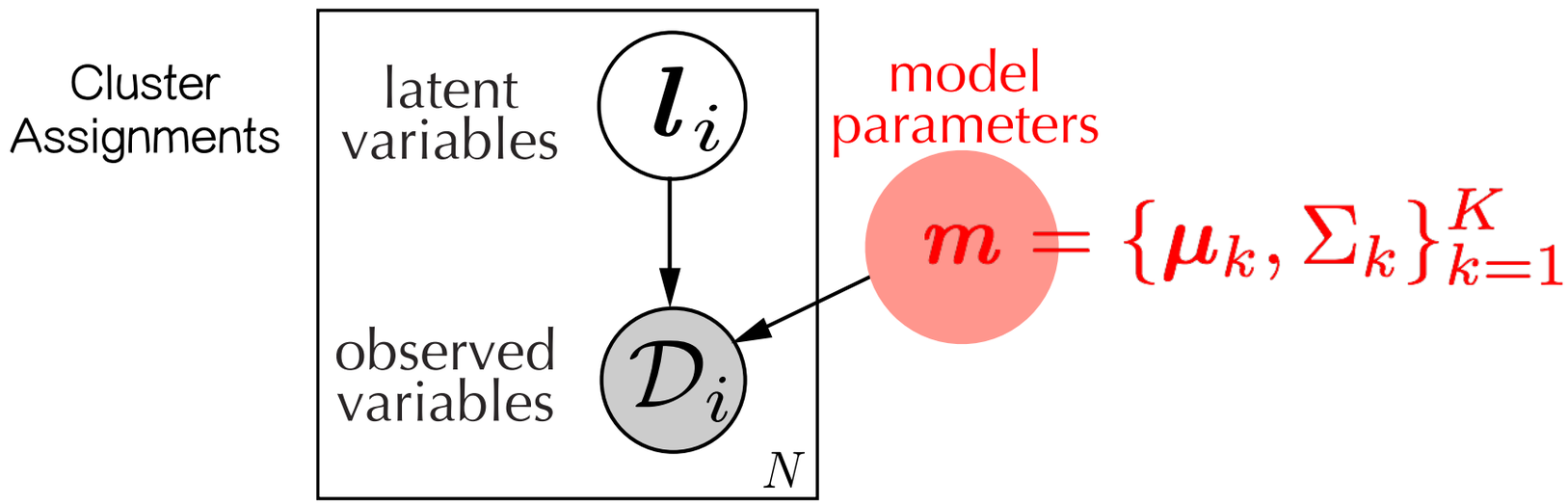
Expectation Maximization



log-likelihood

$$\log p(\mathcal{D}|\mathbf{m}) = \log \int d\mathbf{l}\; p(\mathcal{D}, \mathbf{l}|\mathbf{m})$$

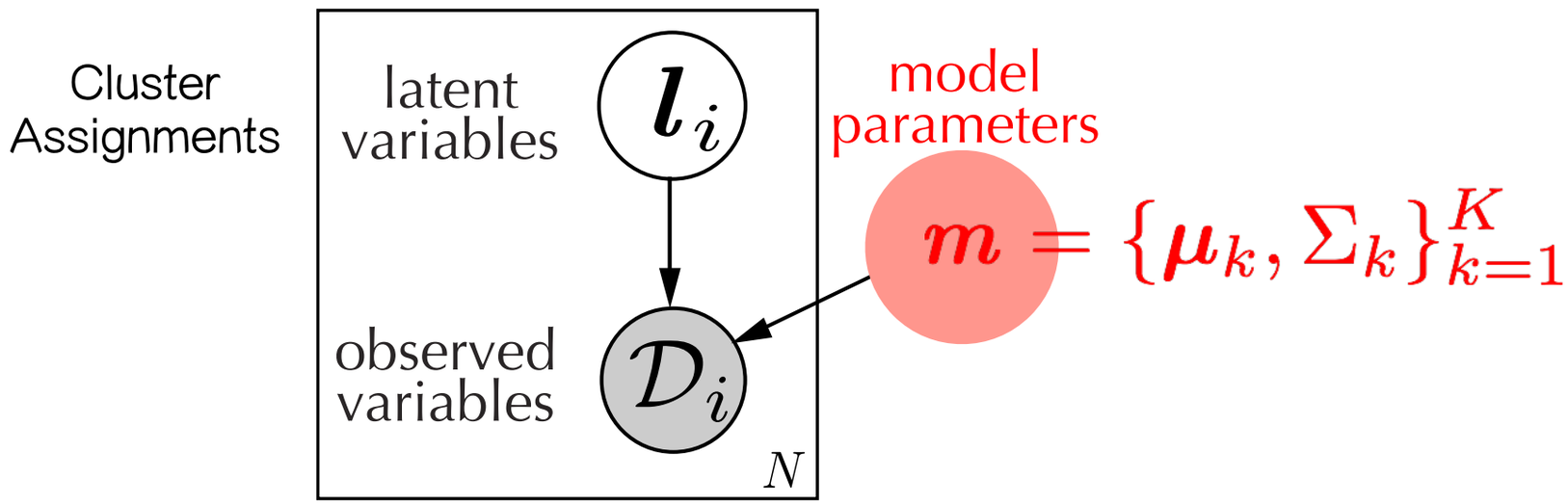
Expectation Maximization



log-likelihood

$$\log p(\mathcal{D}|\mathbf{m}) = \log \int d\mathbf{l}\; p(\mathcal{D}, \mathbf{l}|\mathbf{m}) \;\geq \int d\mathbf{l}\; q(\mathbf{l}) \log \frac{p(\mathcal{D}, \mathbf{l}|\mathbf{m})}{q(\mathbf{l})}$$

Expectation Maximization

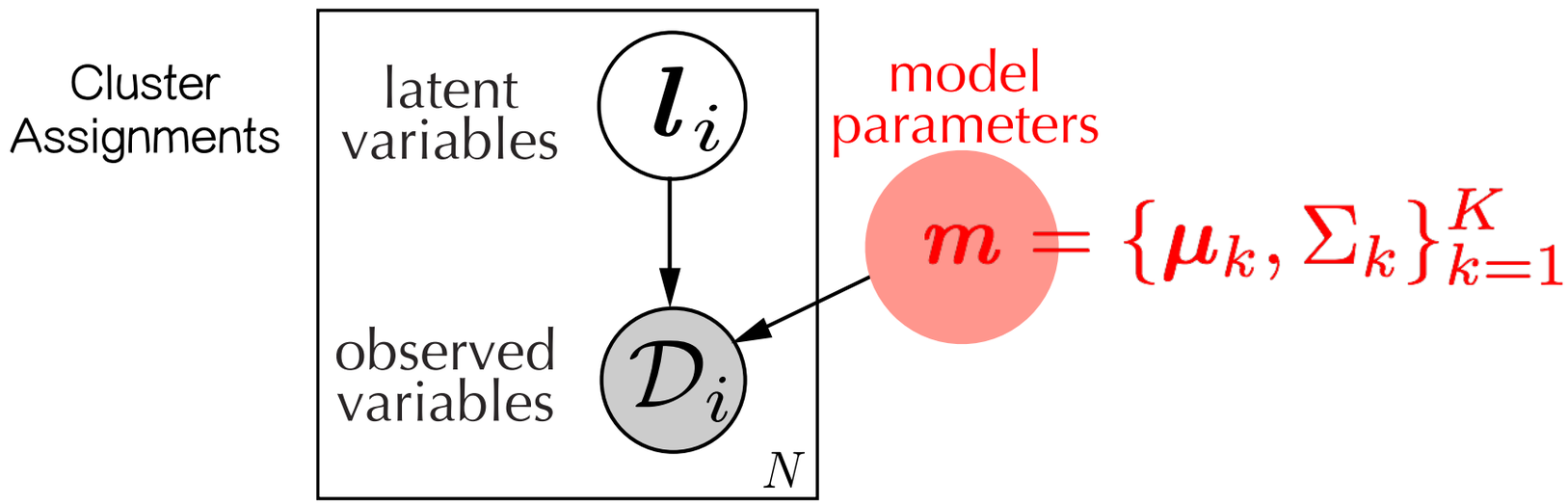


log-likelihood

$$\log p(\mathcal{D}|\mathbf{m}) = \log \int d\mathbf{l}\; p(\mathcal{D}, \mathbf{l}|\mathbf{m}) \;\geq \int d\mathbf{l}\; q(\mathbf{l}) \log \frac{p(\mathcal{D}, \mathbf{l}|\mathbf{m})}{q(\mathbf{l})}$$

$\mathcal{F}(q, \mathbf{m})$ Free-energy

Expectation Maximization



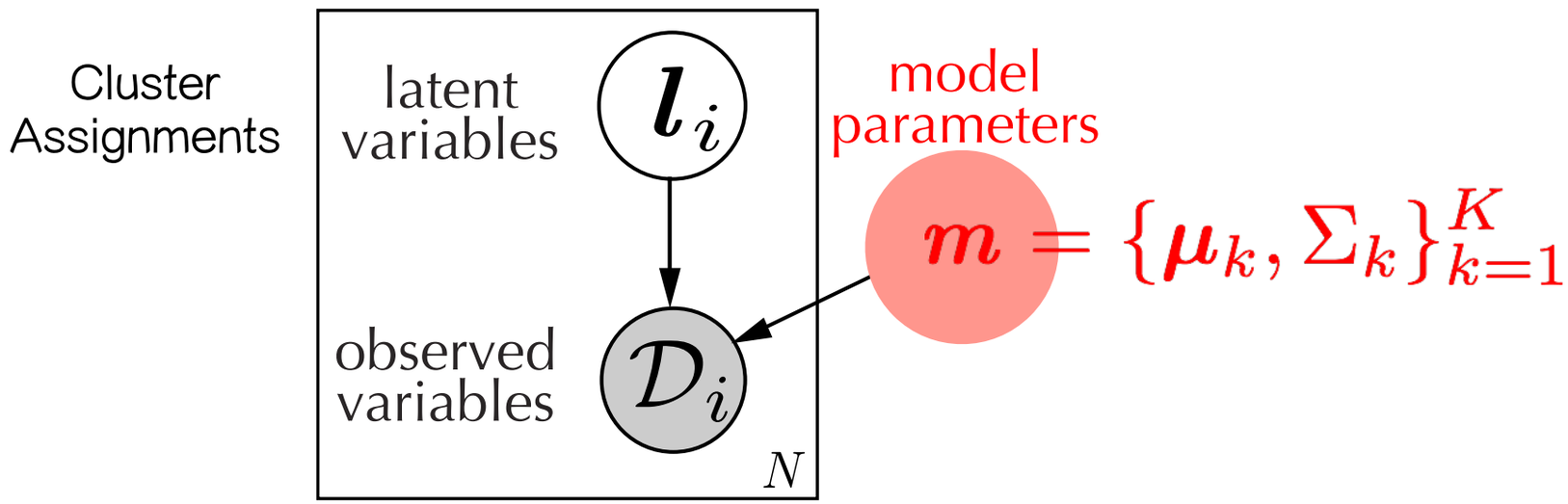
log-likelihood

$$\log p(\mathcal{D}|\mathbf{m}) = \log \int d\mathbf{l}\; p(\mathcal{D}, \mathbf{l}|\mathbf{m}) \;\geq \int d\mathbf{l}\; q(\mathbf{l}) \log \frac{p(\mathcal{D}, \mathbf{l}|\mathbf{m})}{q(\mathbf{l})}$$

$\mathcal{F}(q, \mathbf{m})$ Free-energy

E-step: $q^{(t)}(\mathbf{l}) = \arg\max_{q(\mathbf{l})} \mathcal{F}(q(\mathbf{l}), \mathbf{m}^{(t-1)})$

Expectation Maximization



log-likelihood

$$\log p(\mathcal{D}|\mathbf{m}) = \log \int d\mathbf{l}\; p(\mathcal{D}, \mathbf{l}|\mathbf{m}) \;\geq \int d\mathbf{l}\; q(\mathbf{l}) \log \frac{p(\mathcal{D}, \mathbf{l}|\mathbf{m})}{q(\mathbf{l})}$$

$\mathcal{F}(q, \mathbf{m})$ Free-energy

E-step: $q^{(t)}(\mathbf{l}) = \arg\max_{q(\mathbf{l})} \mathcal{F}(q(\mathbf{l}), \mathbf{m}^{(t-1)})$

M-step: $\mathbf{m}^{(t)} = \arg\max_{\mathbf{m}} \mathcal{F}(q^{(t)}(\mathbf{l}), \mathbf{m})$

DP-EM

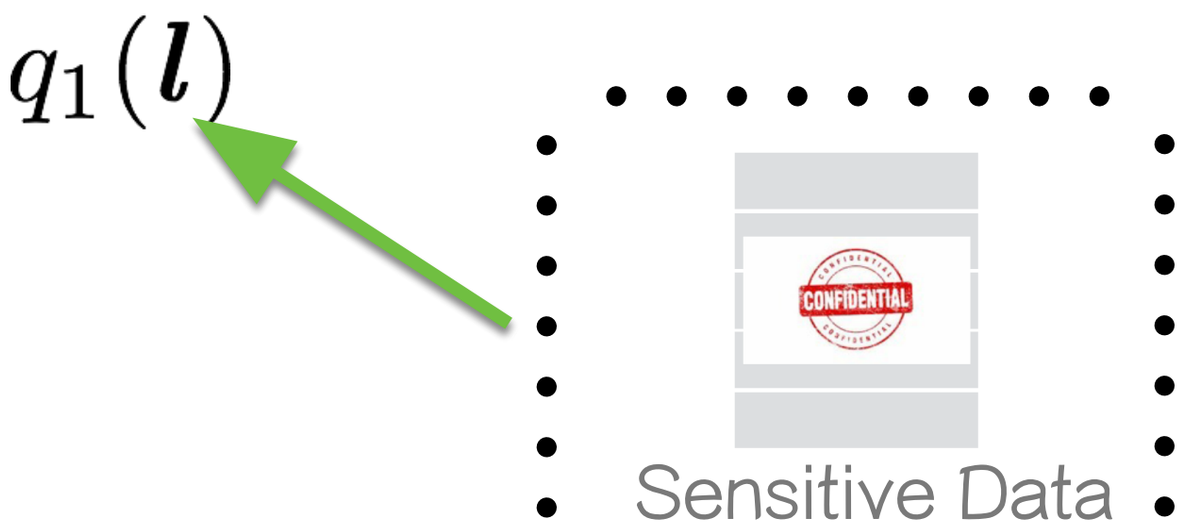
[Park et al, 2017]

DP-EM

[Park et al, 2017]

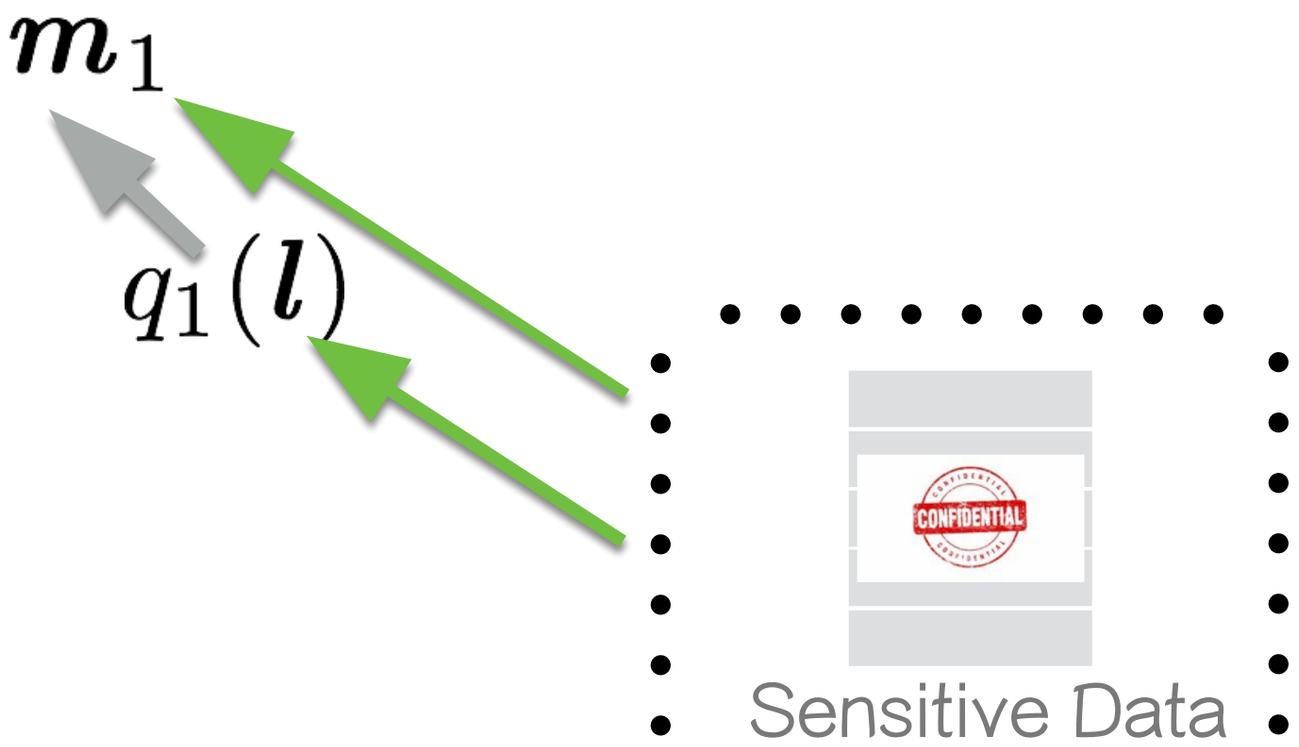
DP-EM

[Park et al, 2017]



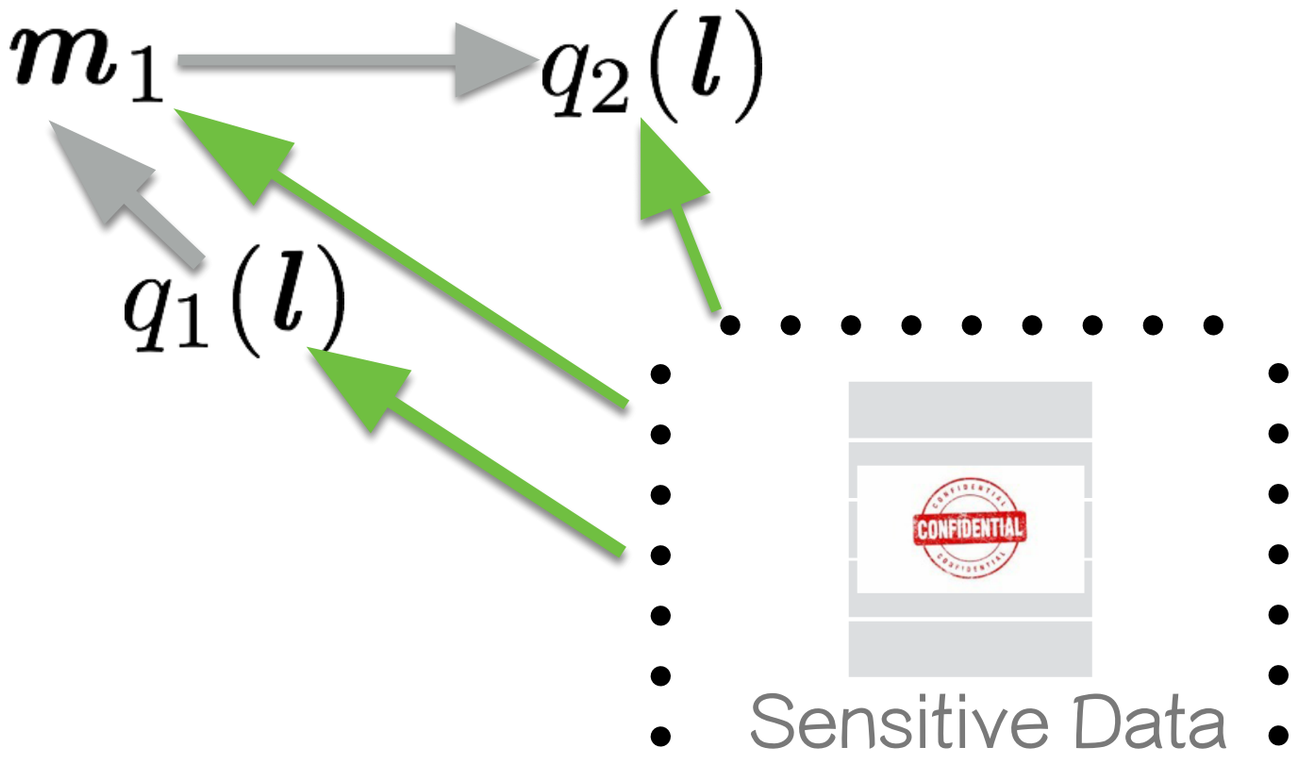
DP-EM

[Park et al, 2017]



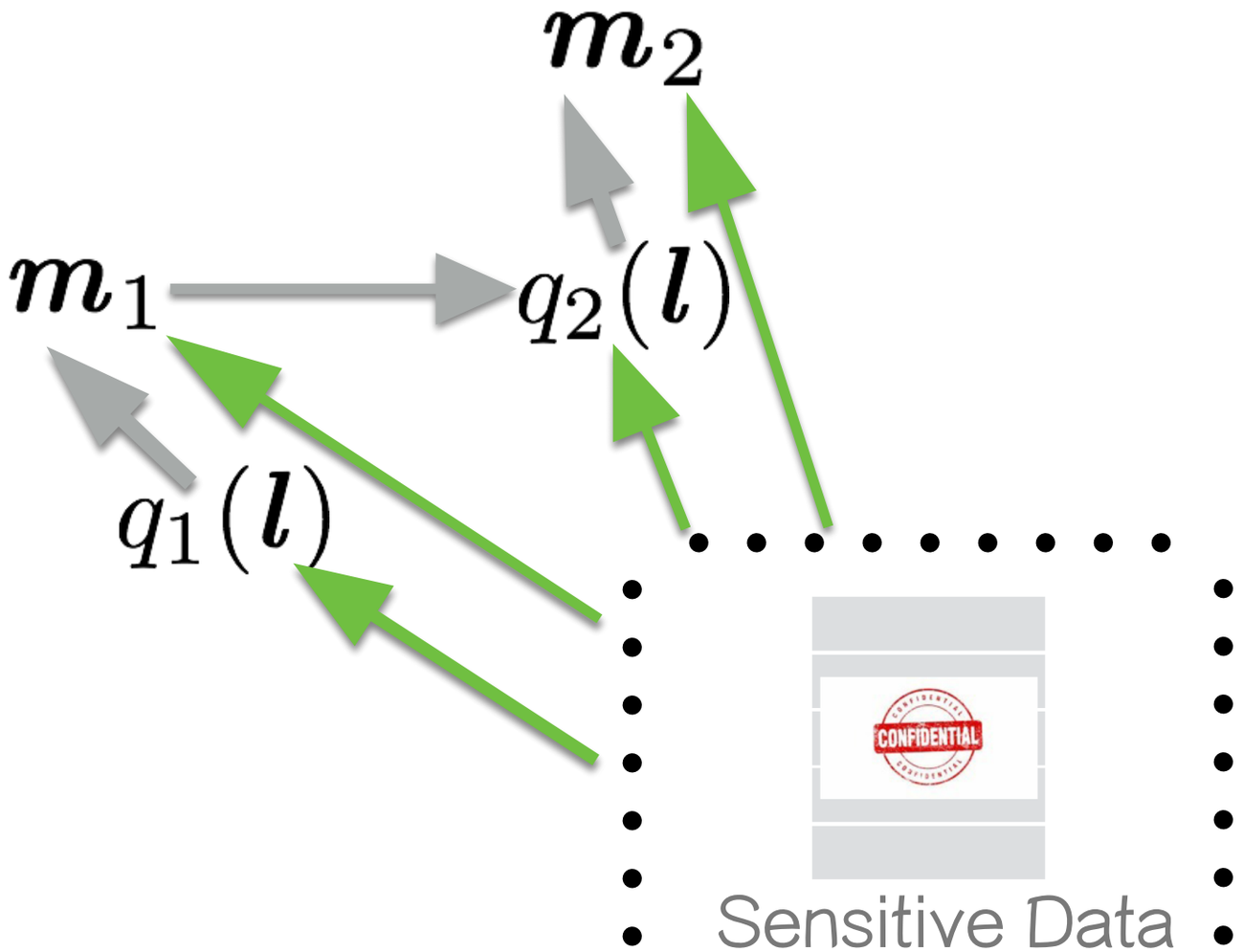
DP-EM

[Park et al, 2017]



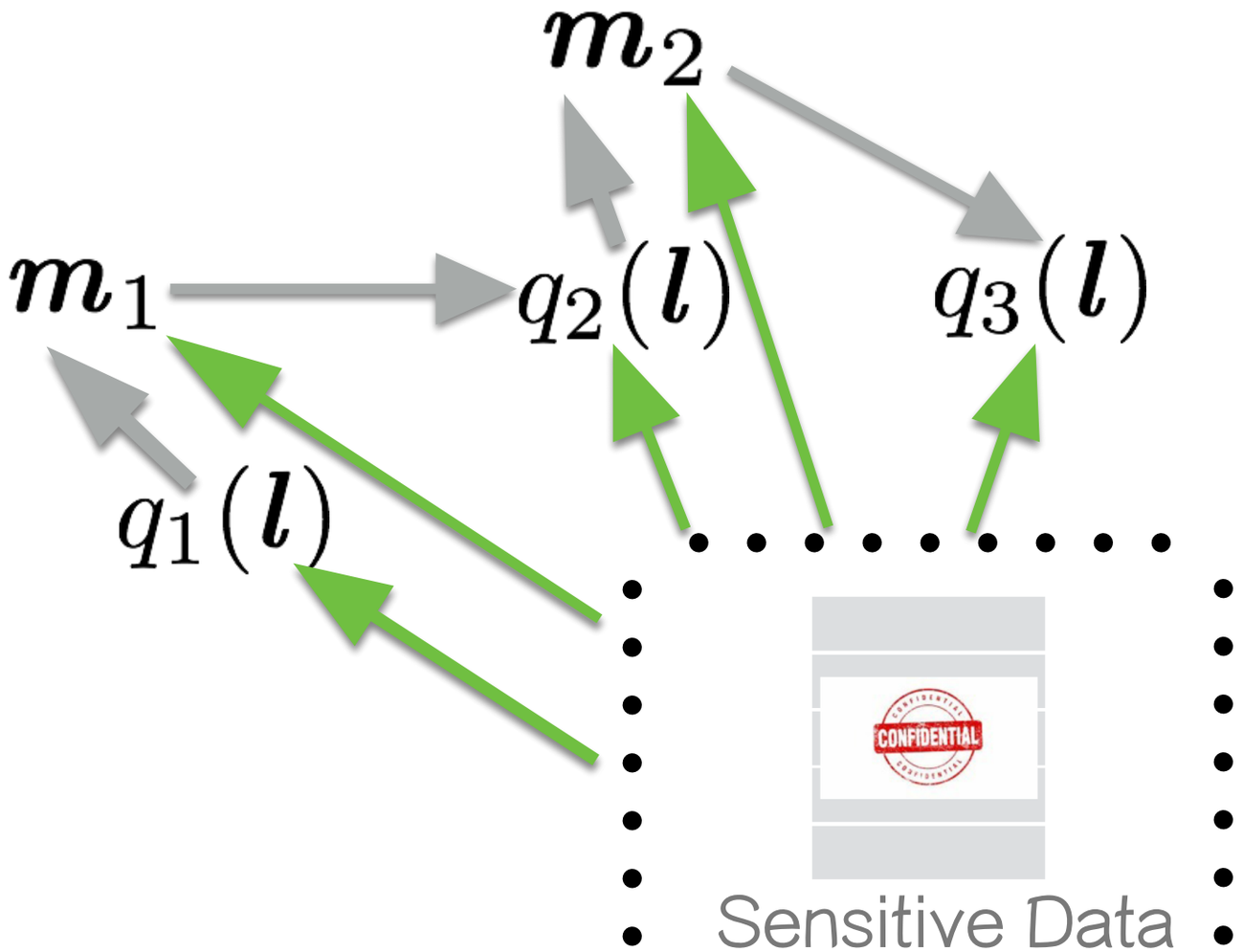
DP-EM

[Park et al, 2017]



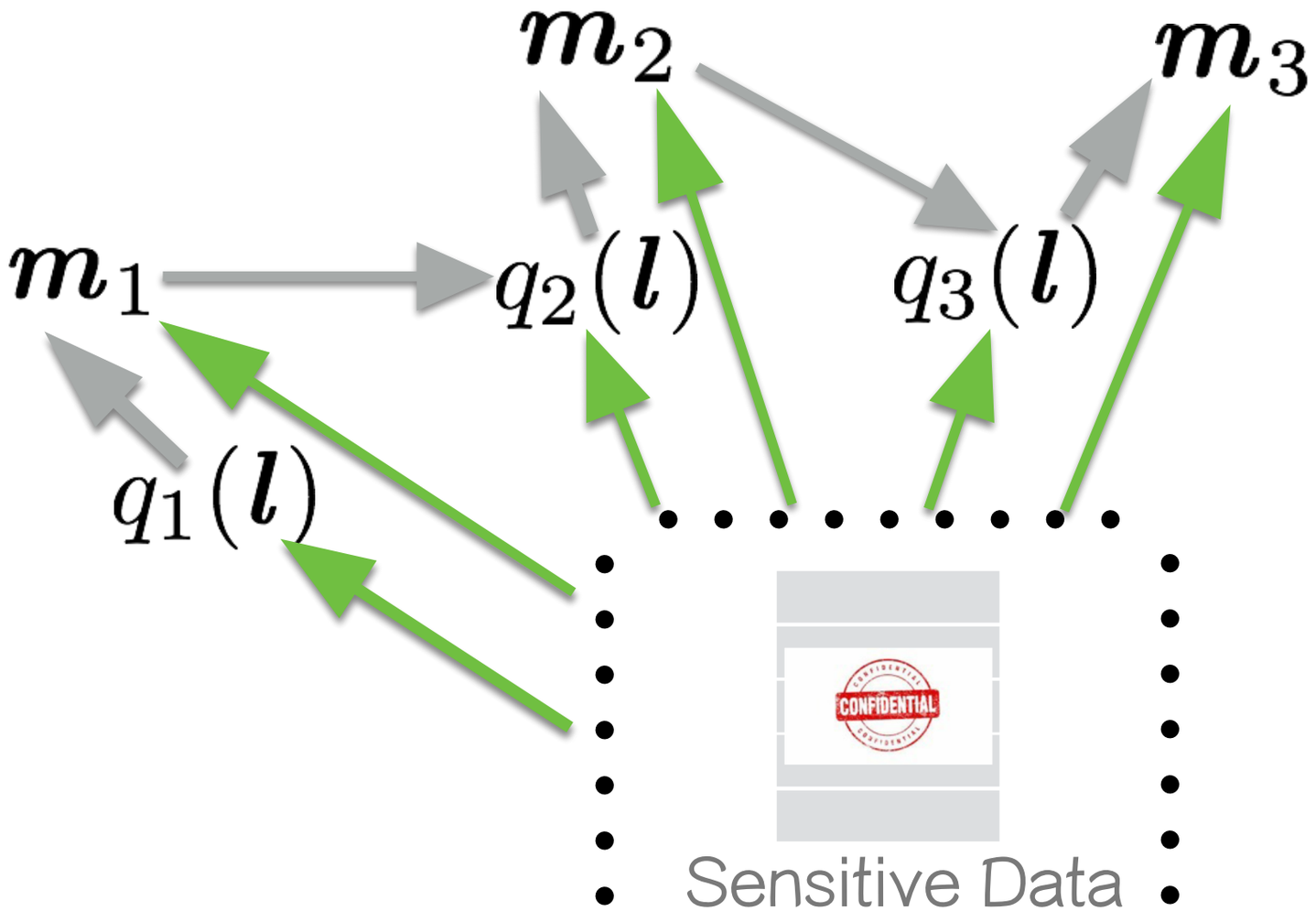
DP-EM

[Park et al, 2017]



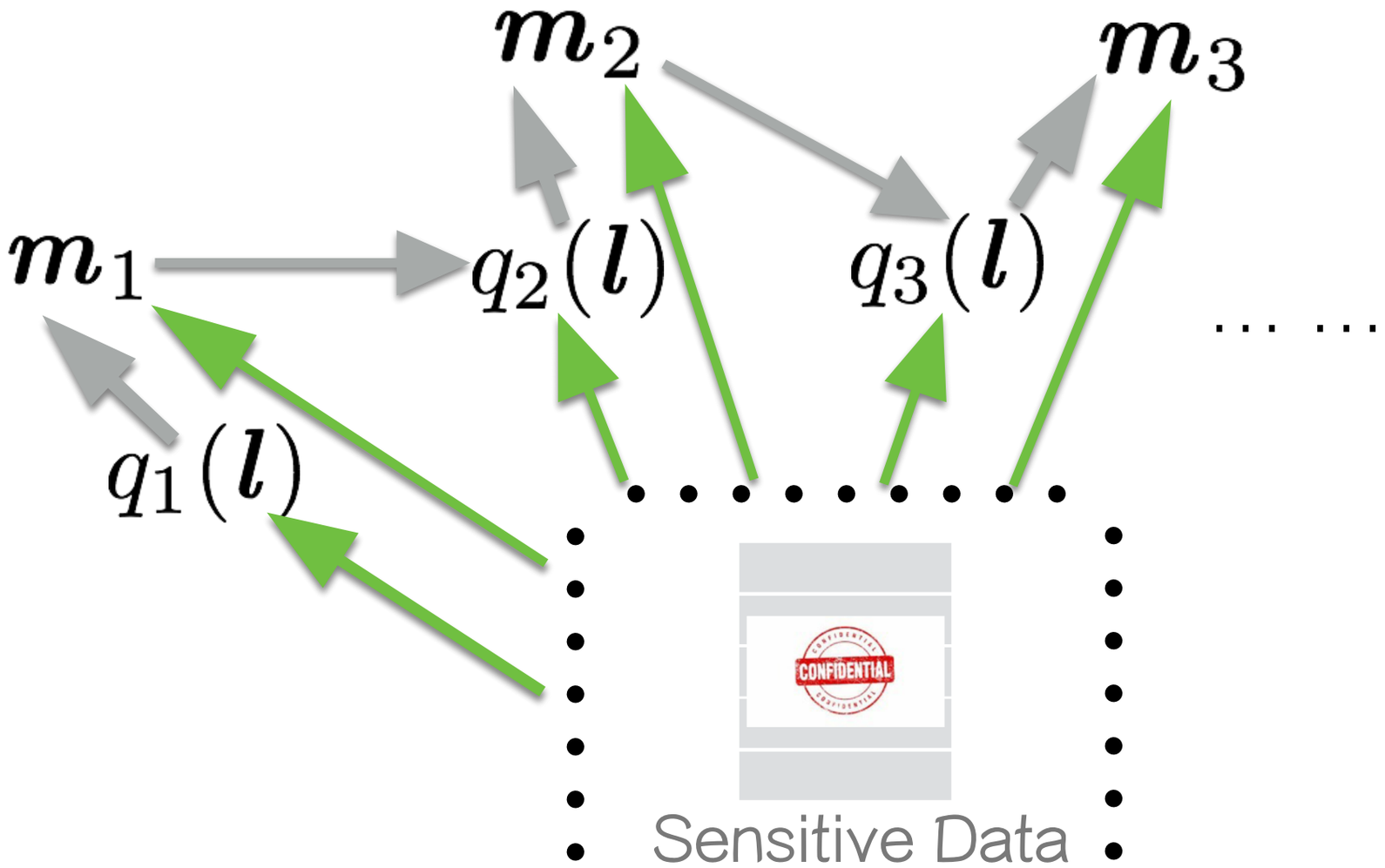
DP-EM

[Park et al, 2017]



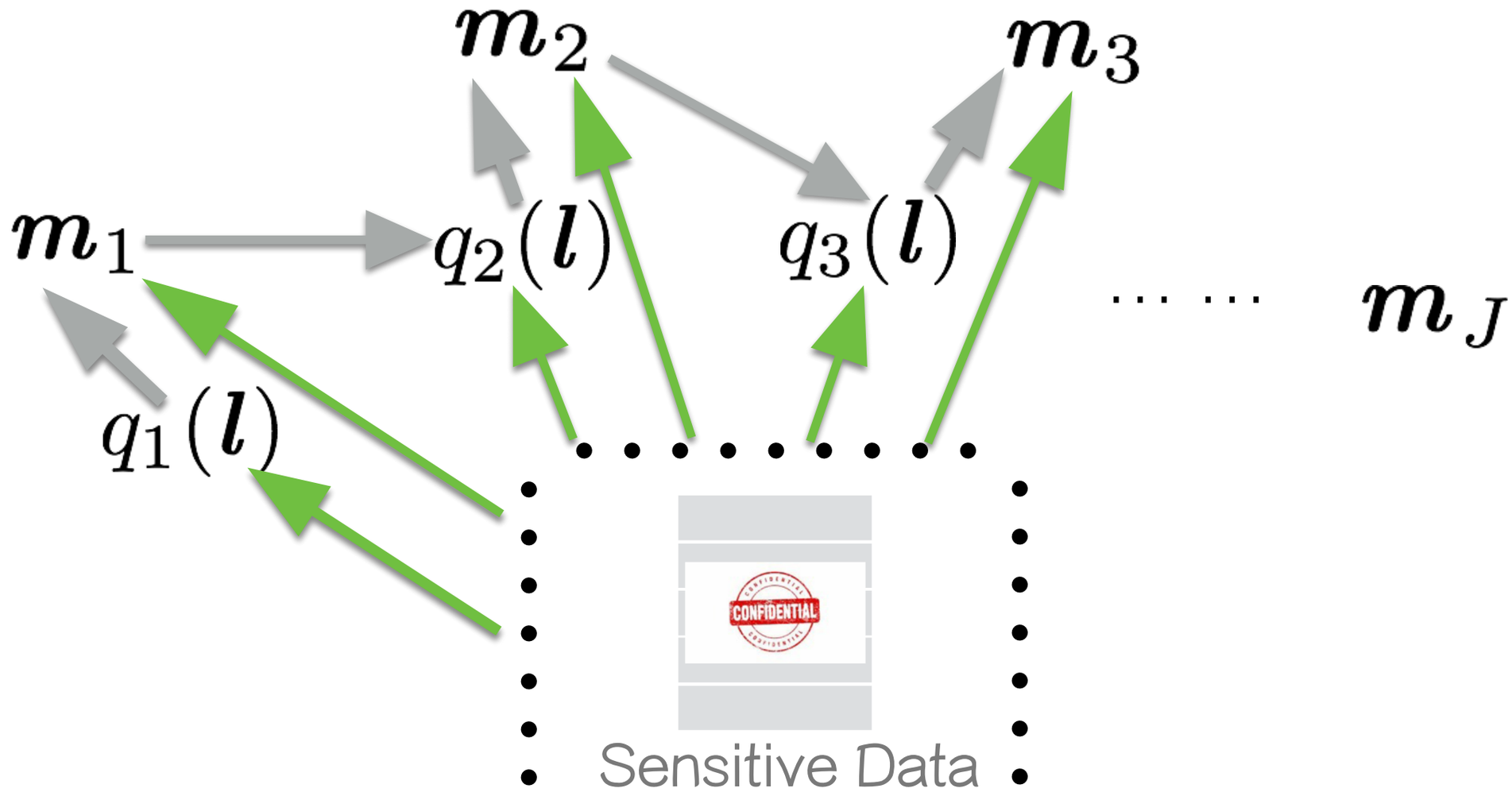
DP-EM

[Park et al, 2017]



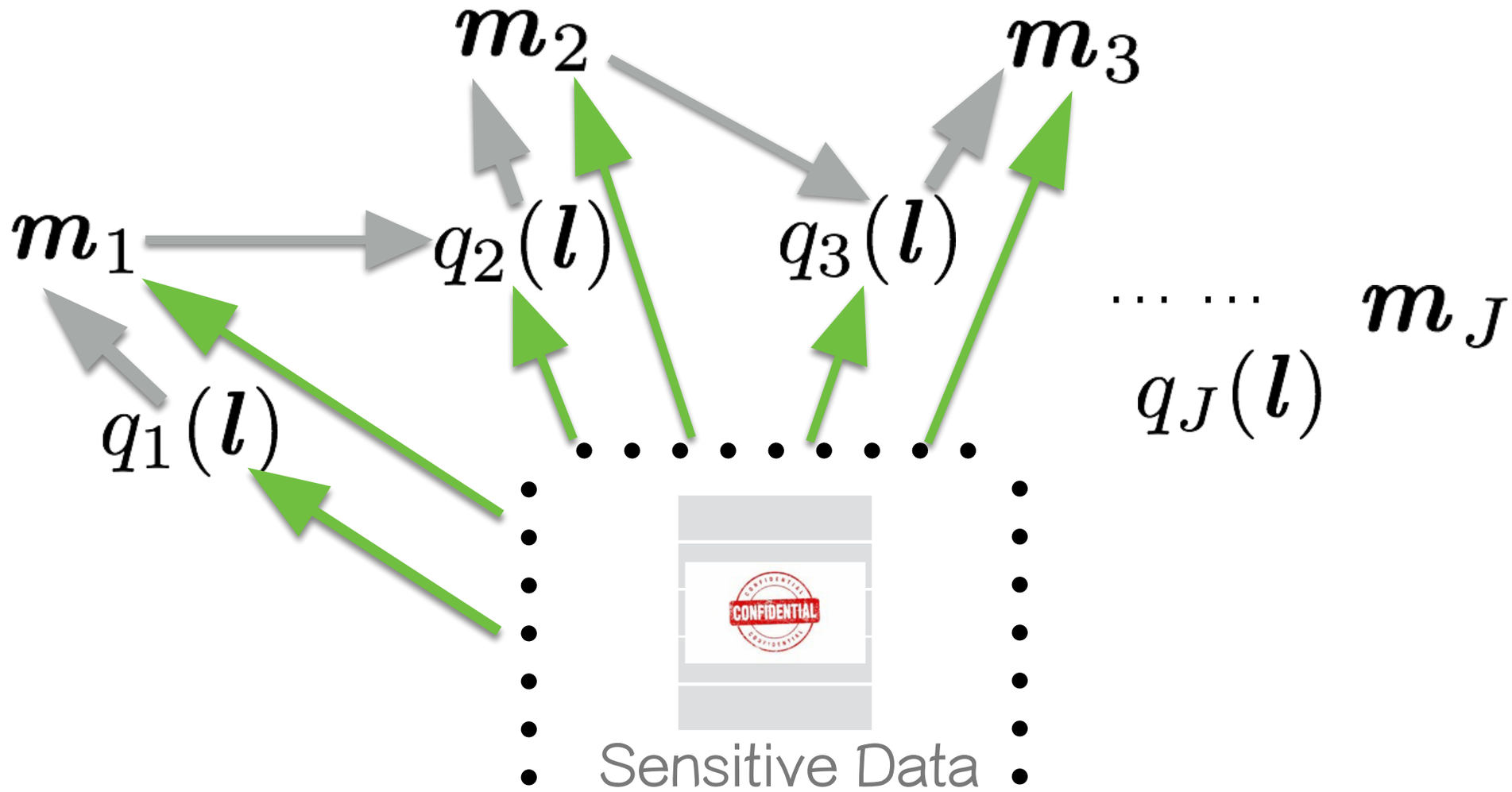
DP-EM

[Park et al, 2017]



DP-EM

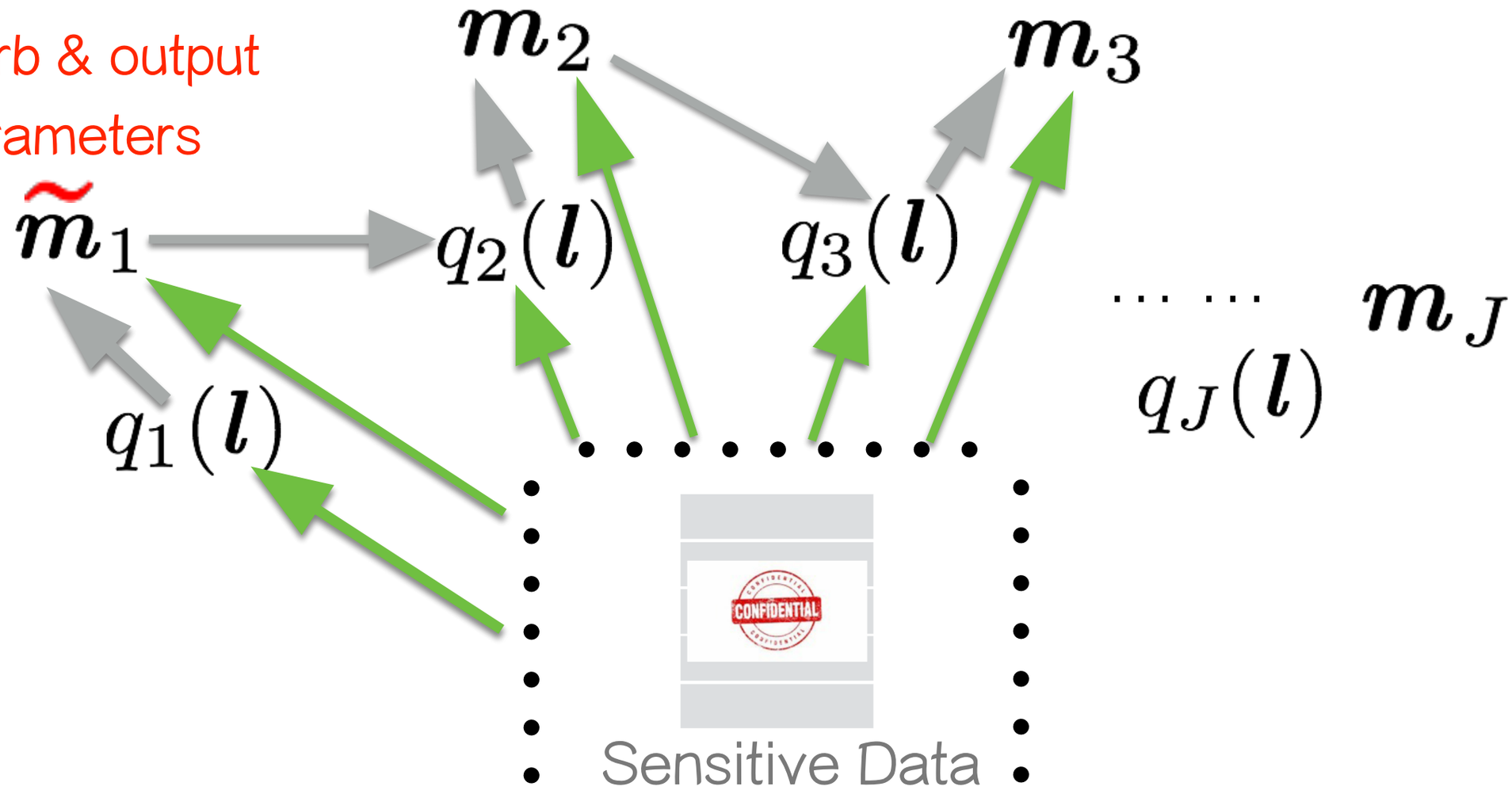
[Park et al, 2017]



DP-EM

[Park et al, 2017]

Perturb & output parameters

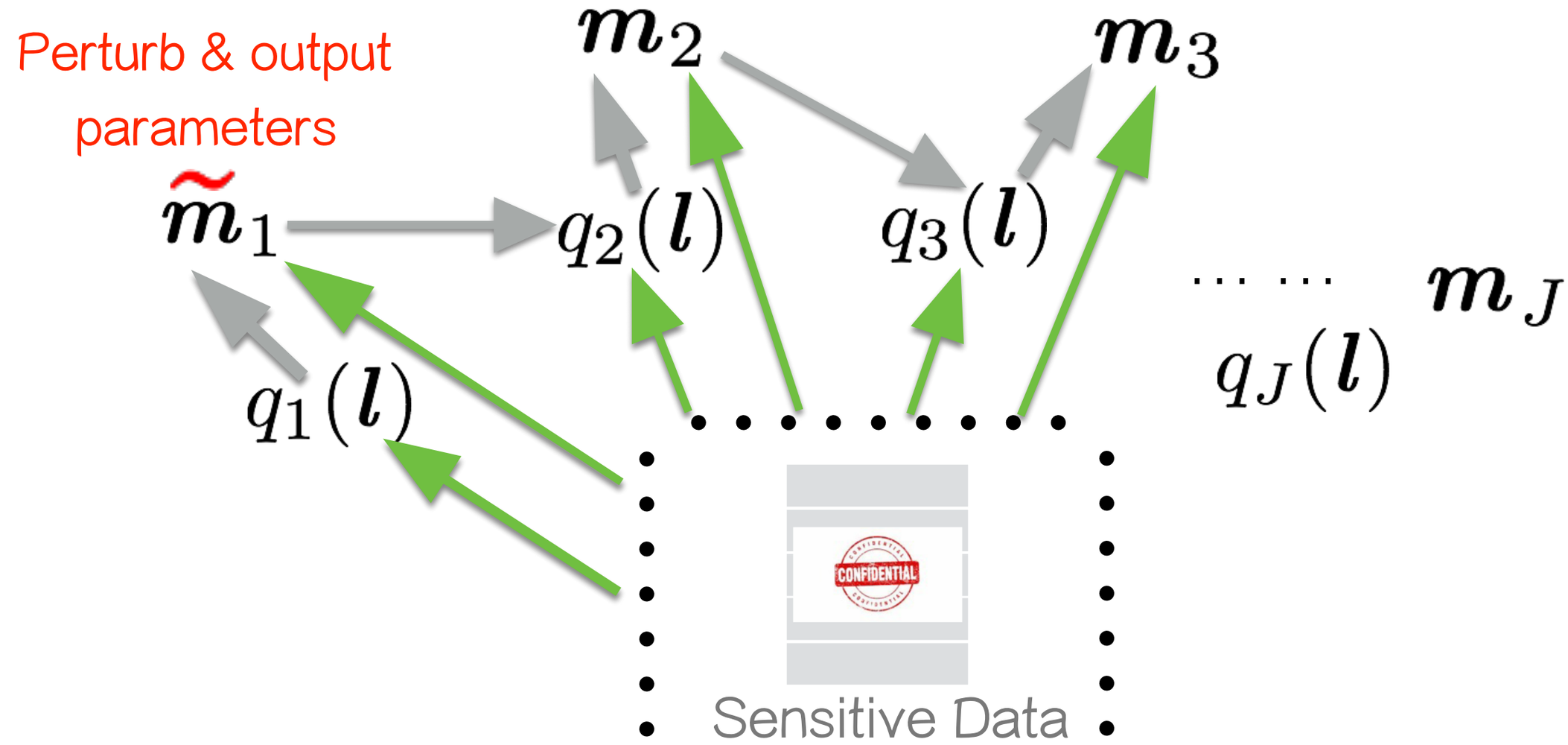


DP-EM

[Park et al, 2017]

Sensitivity: one-datapoint's worth directly from dataset & indirectly through q(l)

Perturb & output parameters

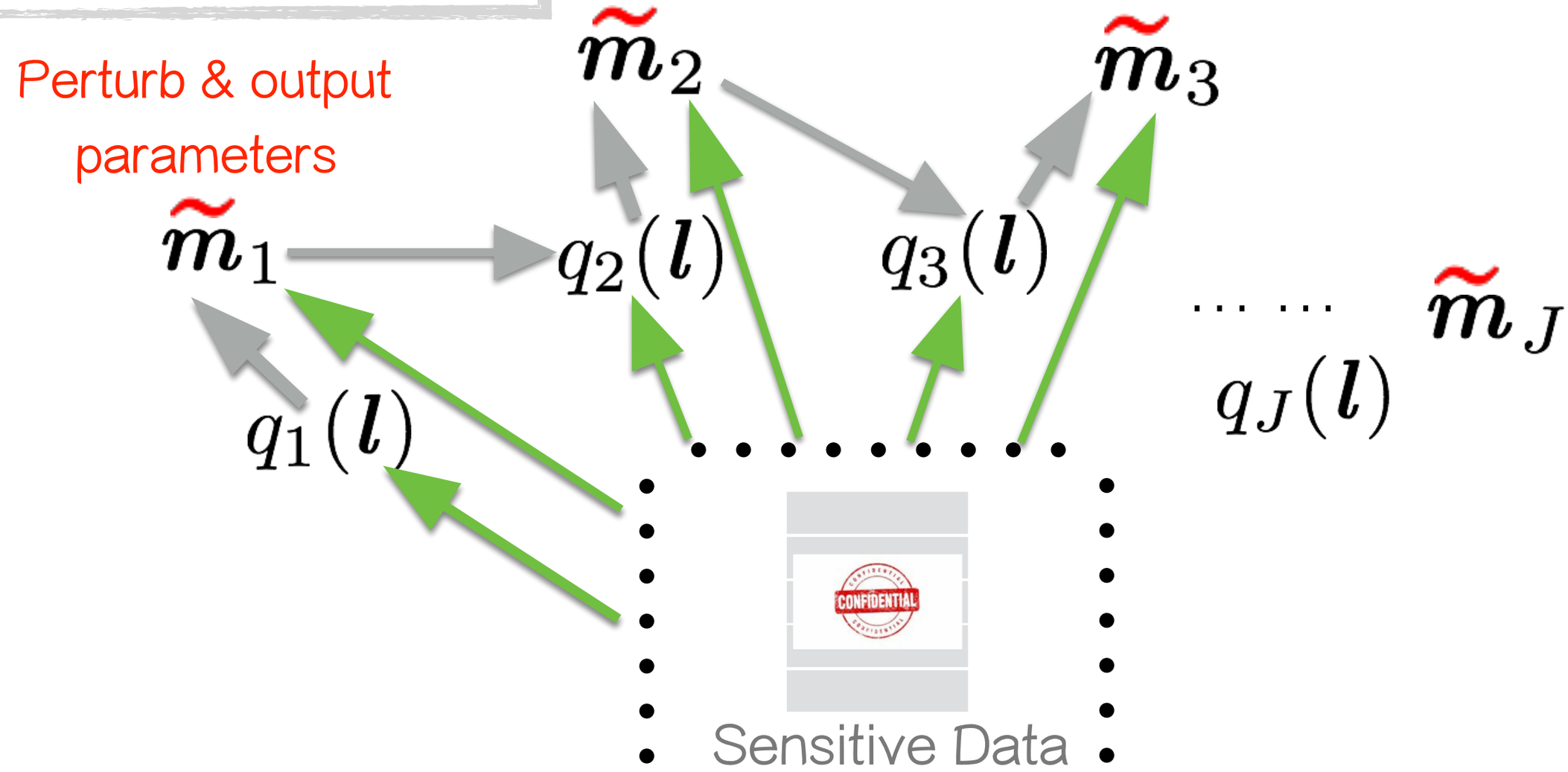


DP-EM

[Park et al, 2017]

Sensitivity: one-datapoint's worth directly from dataset & indirectly through q(l)

Perturb & output parameters

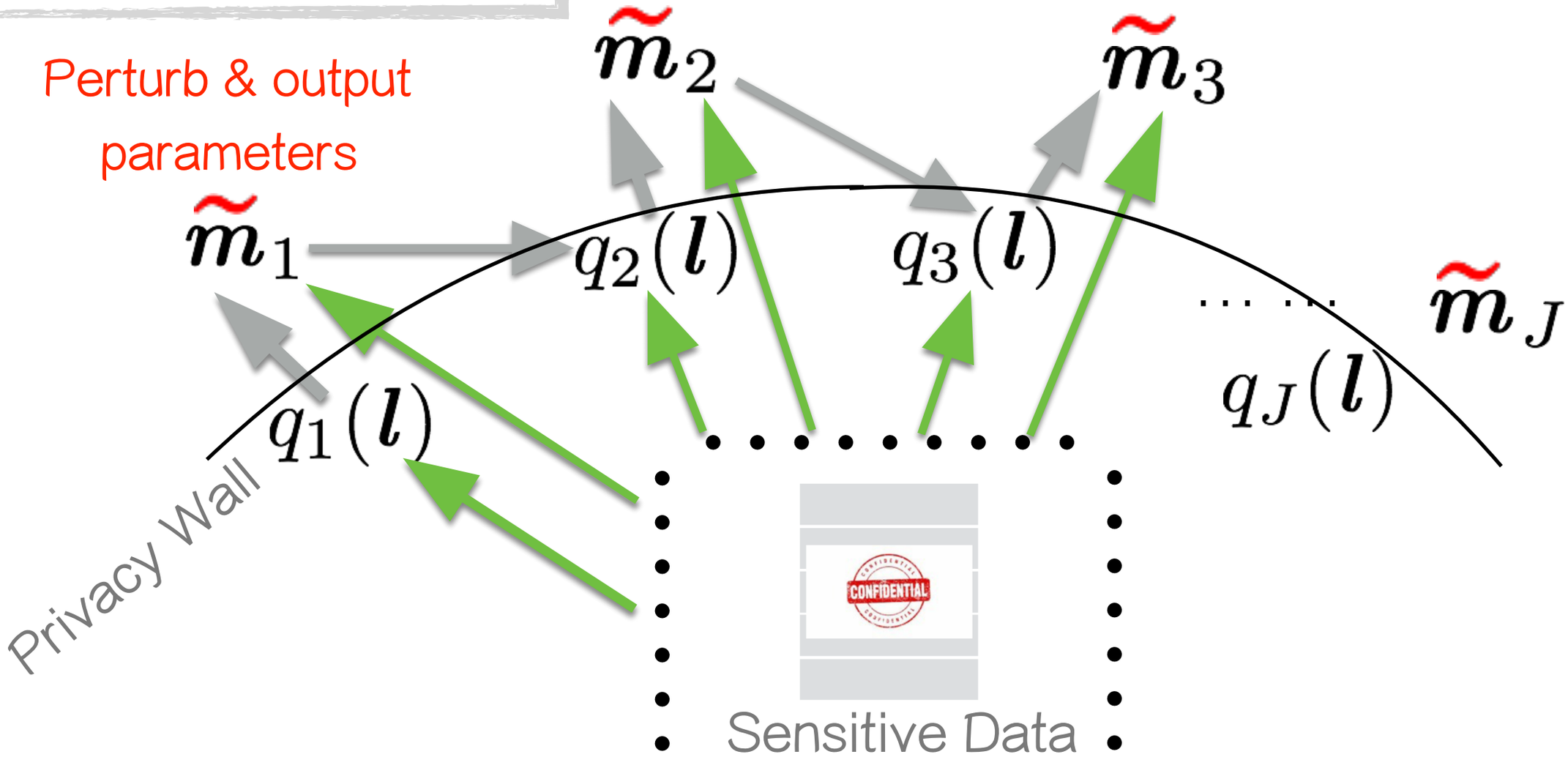


DP-EM

[Park et al, 2017]

Sensitivity: one-datapoint's worth directly from dataset & indirectly through q(l)

Perturb & output parameters



Repetitive access to data, large privacy loss

Repetitive access to data, large privacy loss

Moments accountant

[Abadi et al, 2016]

Repetitive access to data, large privacy loss

Moments accountant

[Abadi et al, 2016]

- Treat privacy loss (PL) as

random variable

$$L^{(o)} = \log \frac{p(A(\mathcal{D}) = o)}{p(A(\mathcal{D}') = o)}$$

Repetitive access to data, large privacy loss

Moments accountant

[Abadi et al, 2016]

- Treat privacy loss (PL) as

random variable

$$L^{(o)} = \log \frac{p(A(\mathcal{D}) = o)}{p(A(\mathcal{D}') = o)}$$

keep track of MGF of L,
convert **moment bound to tail bound**

→ **strong privacy guarantee**

Repetitive access to data, large privacy loss

Moments accountant

[Abadi et al, 2016]

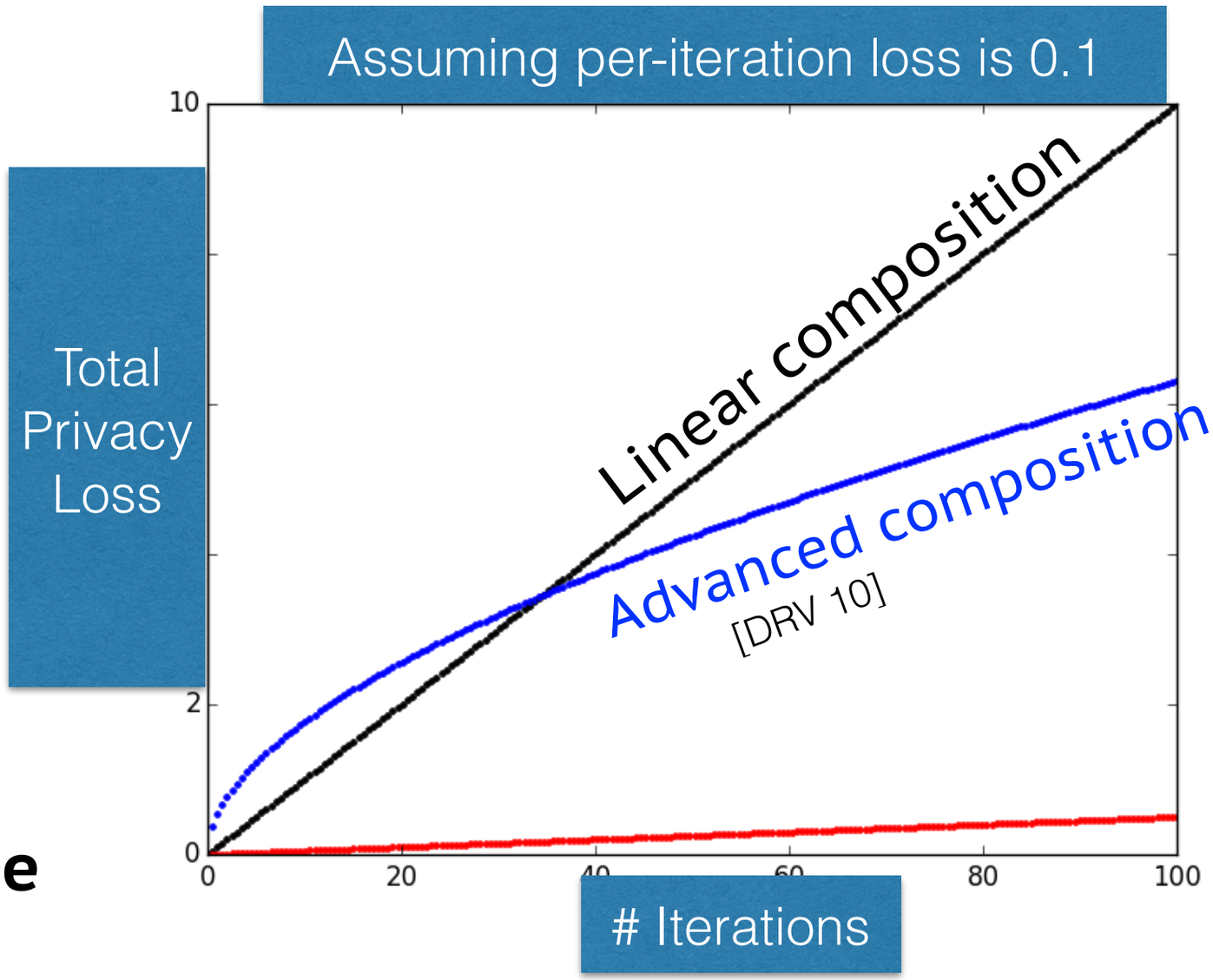
- Treat privacy loss (PL) as

random variable

$$L^{(o)} = \log \frac{p(A(\mathcal{D}) = o)}{p(A(\mathcal{D}') = o)}$$

keep track of MGF of L, convert **moment bound to tail bound**

→ **strong privacy guarantee**



Repetitive access to data, large privacy loss

Moments accountant

[Abadi et al, 2016]

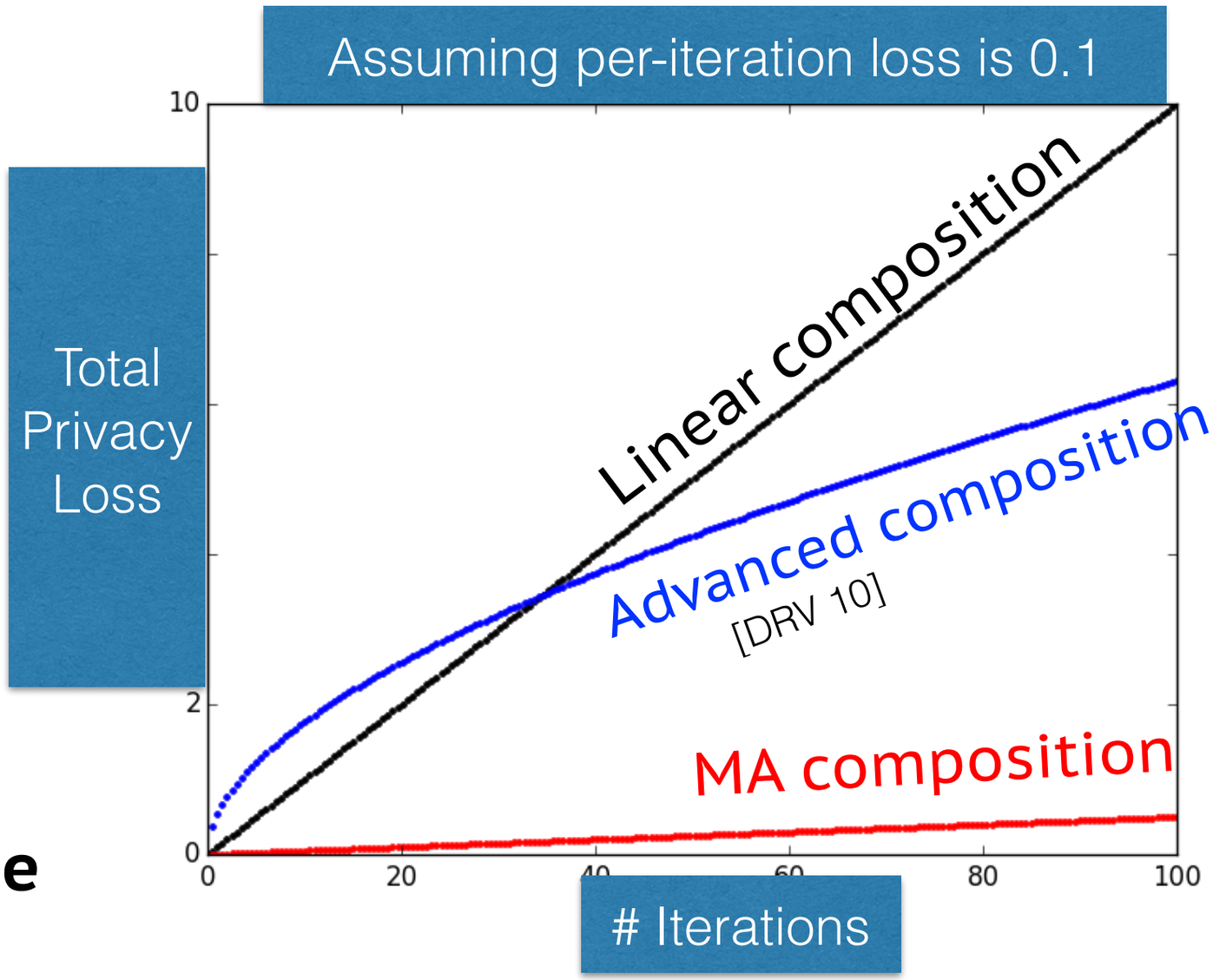
- Treat privacy loss (PL) as

random variable

$$L^{(o)} = \log \frac{p(A(\mathcal{D}) = o)}{p(A(\mathcal{D}') = o)}$$

keep track of MGF of L, convert **moment bound to tail bound**

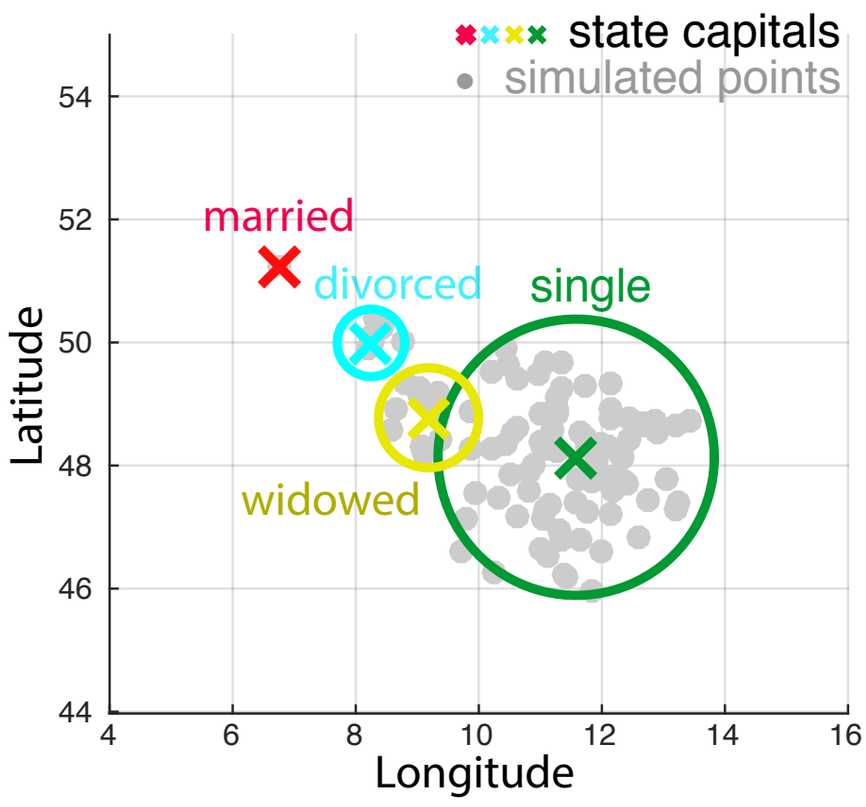
→ **strong privacy guarantee**



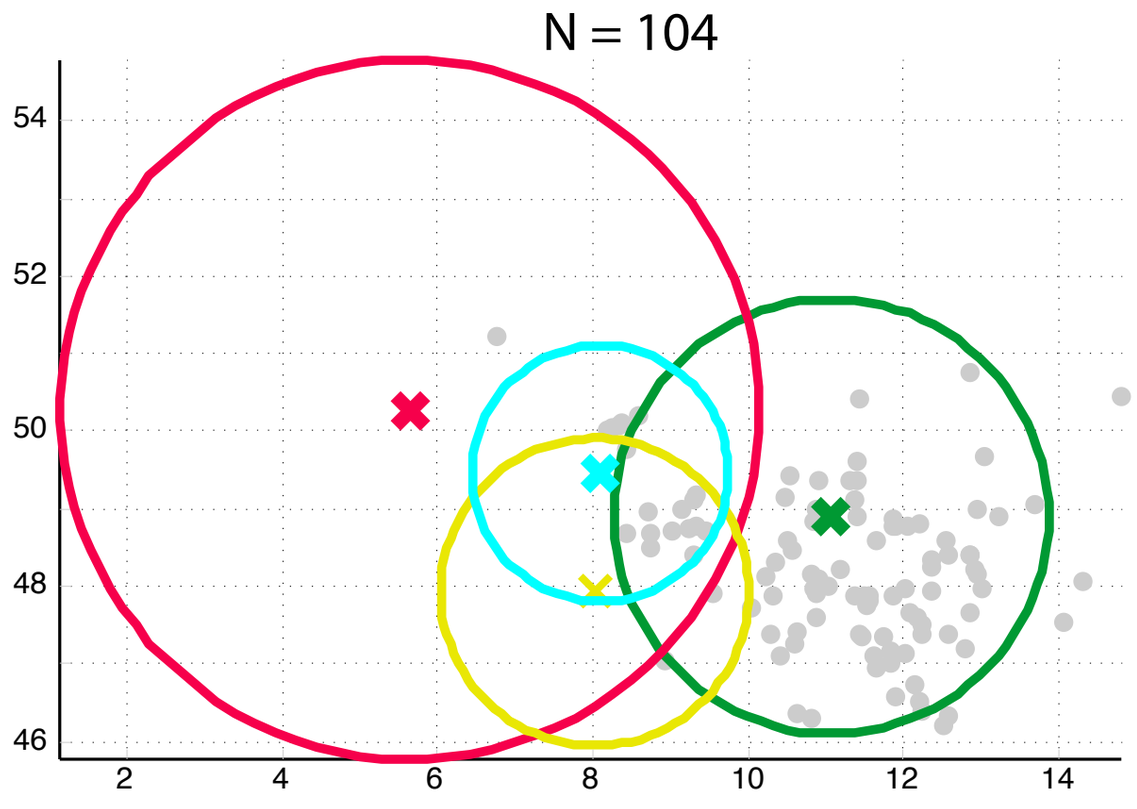
DP-EM: Abortion data

[destatis.de]

Raw data



Privatized MoG using DP-EM

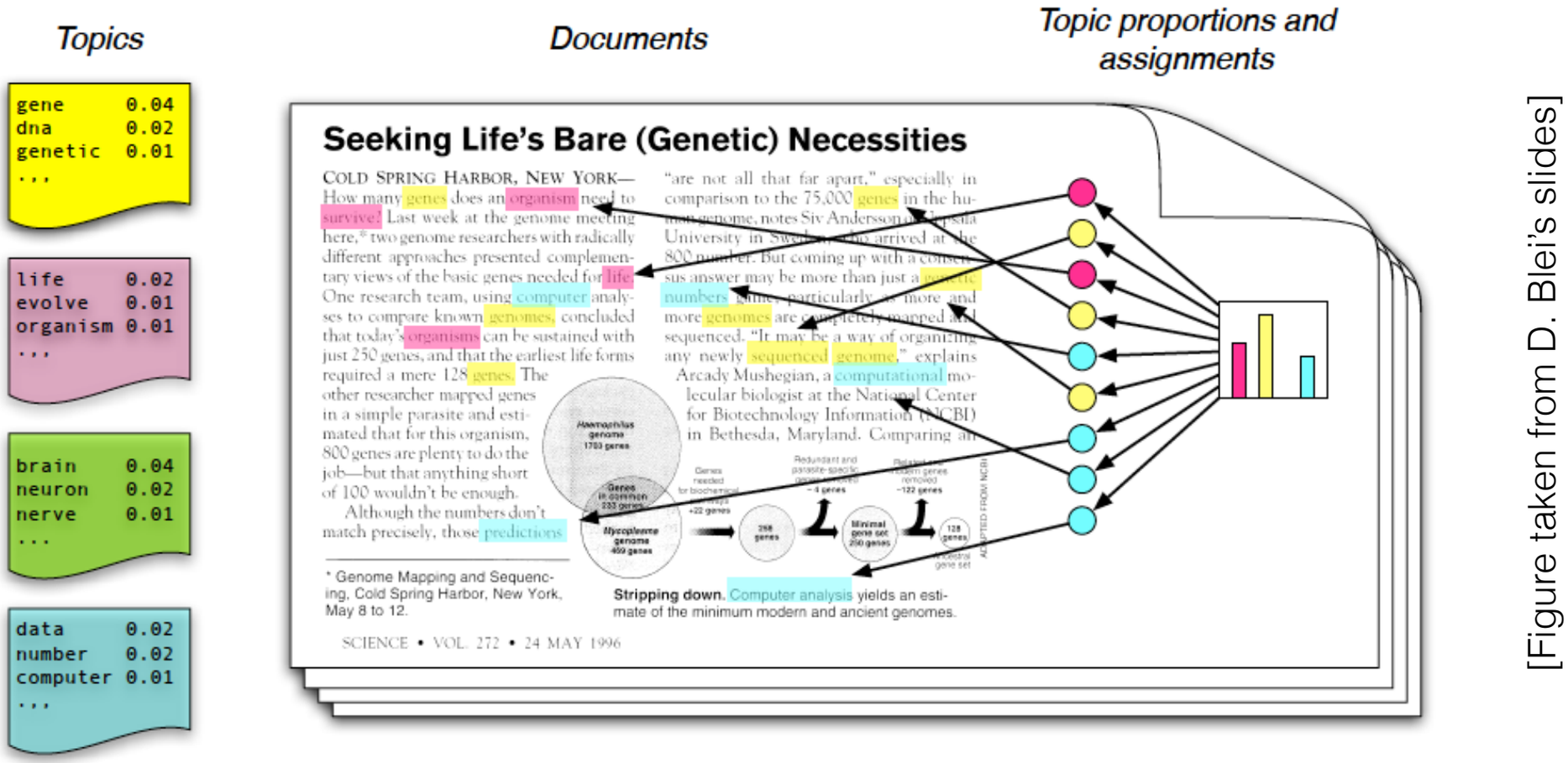


Even for the cluster with a single datapoint (red cluster), one can't accurately identify the geographical location of that individual from outputs of DP-EM

Topic modeling

Topic modeling

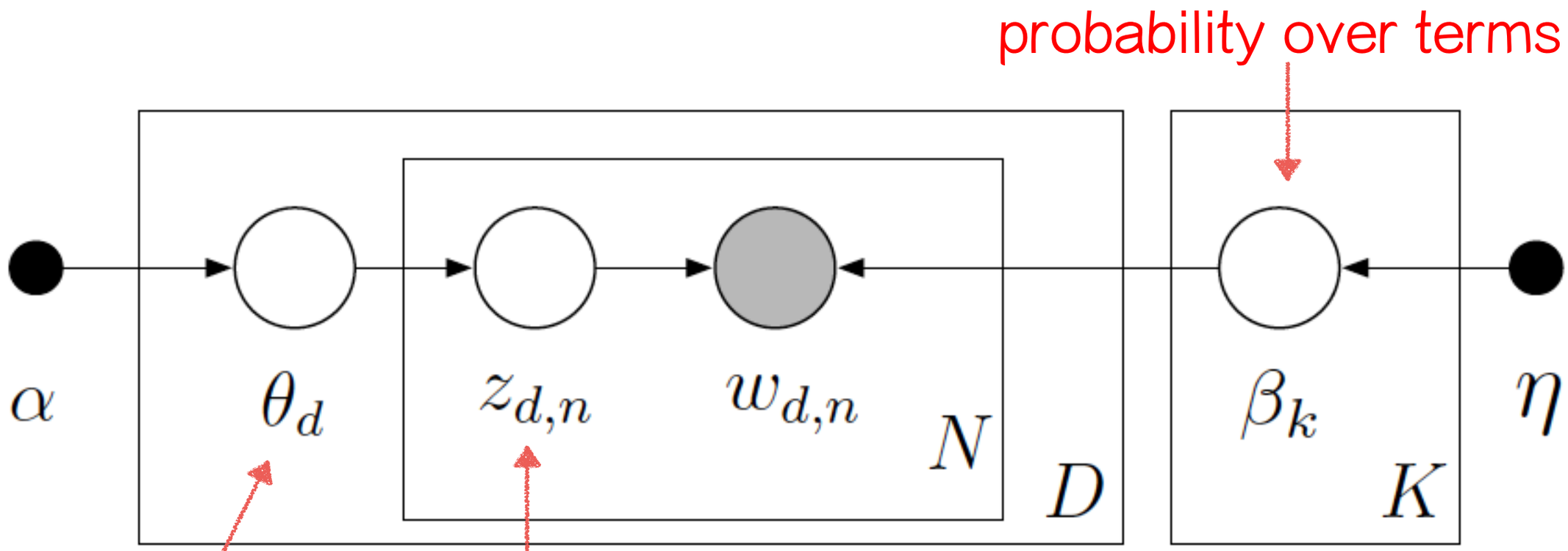
- Given documents, **describe how** documents are organised!



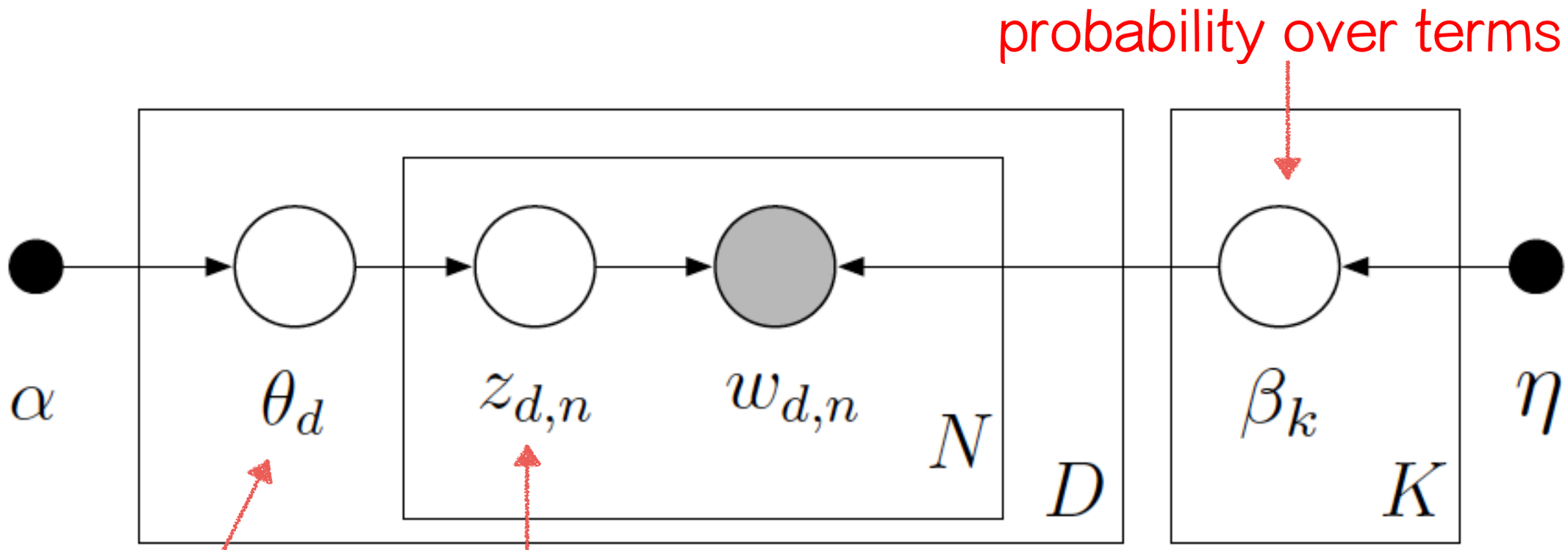
[Figure taken from D. Blei's slides]

Goal : allocate words to topics & assign probability to terms in each topic.

Latent Dirichlet Allocation (LDA)



Latent Dirichlet Allocation (LDA)



topic proportion of each document

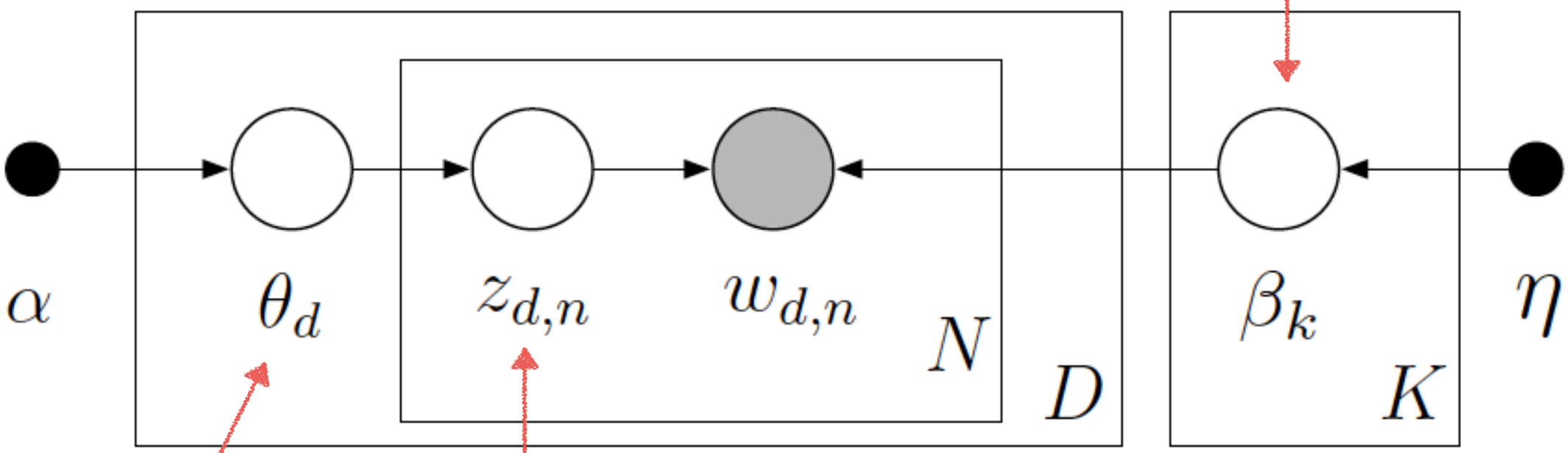
topic assignment of each word

Latent variables

Latent Dirichlet Allocation (LDA)

parameters

probability over terms in voc

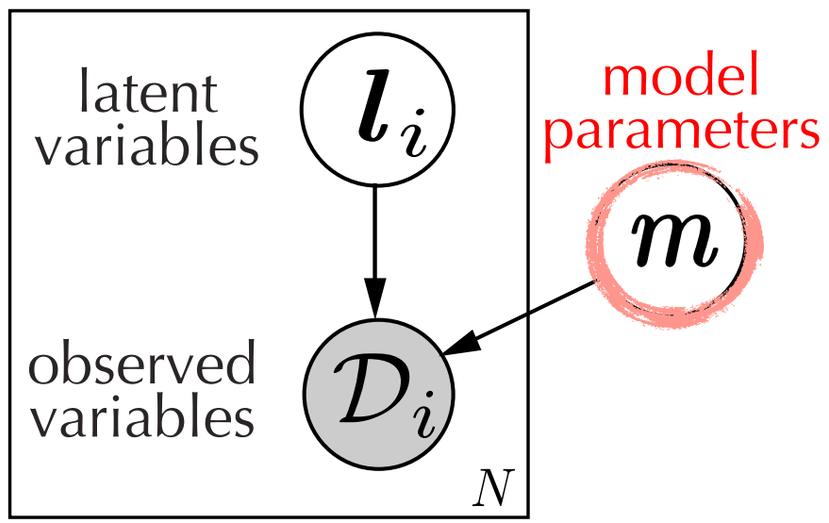


topic proportion of each document

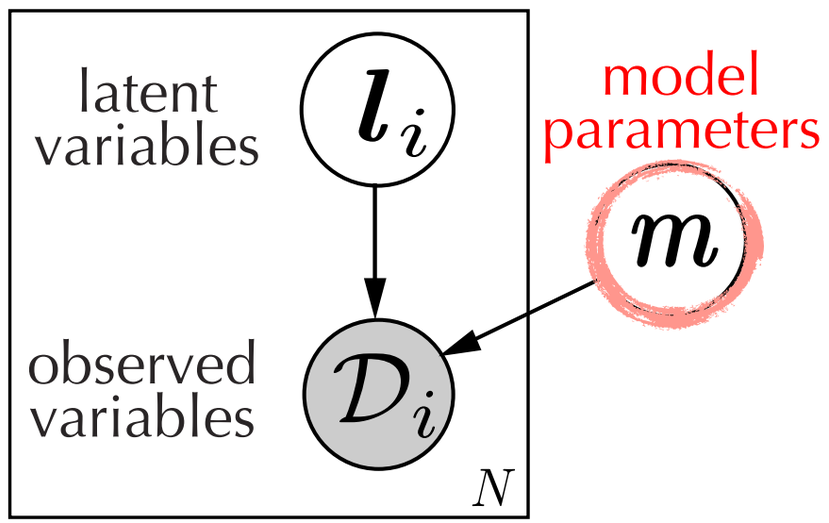
topic assignment of each word

Latent variables

Variational inference

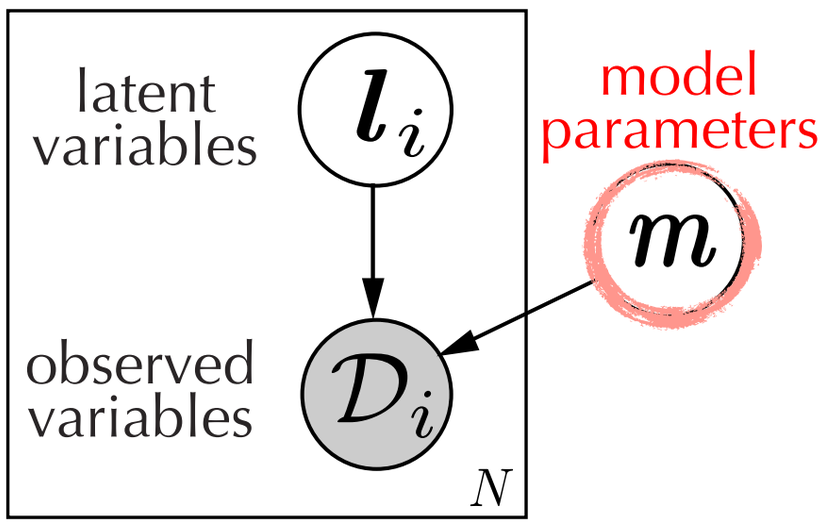


Variational inference



log-marginal-likelihood

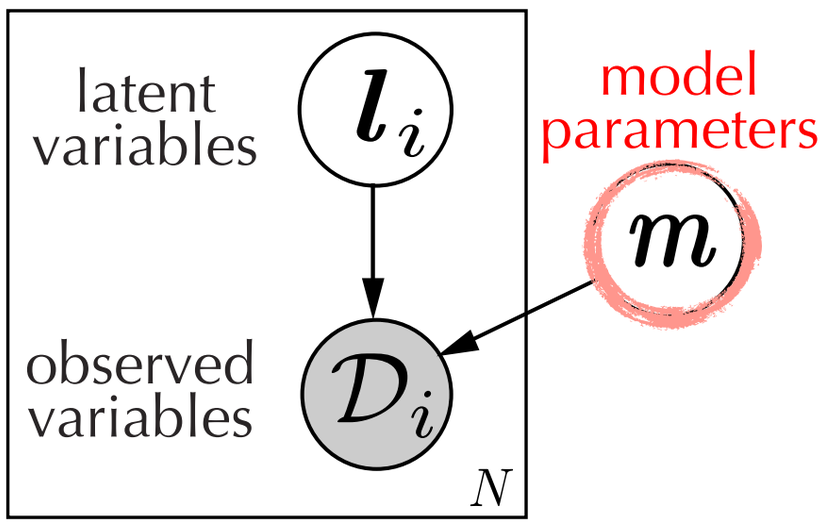
Variational inference



log-marginal-likelihood

$$\log p(\mathcal{D}) \geq \int q(\boldsymbol{l})q(\boldsymbol{m}) \log \frac{p(\mathcal{D}, \boldsymbol{l}|\boldsymbol{m})p(\boldsymbol{m})}{q(\boldsymbol{l})q(\boldsymbol{m})}$$

Variational inference

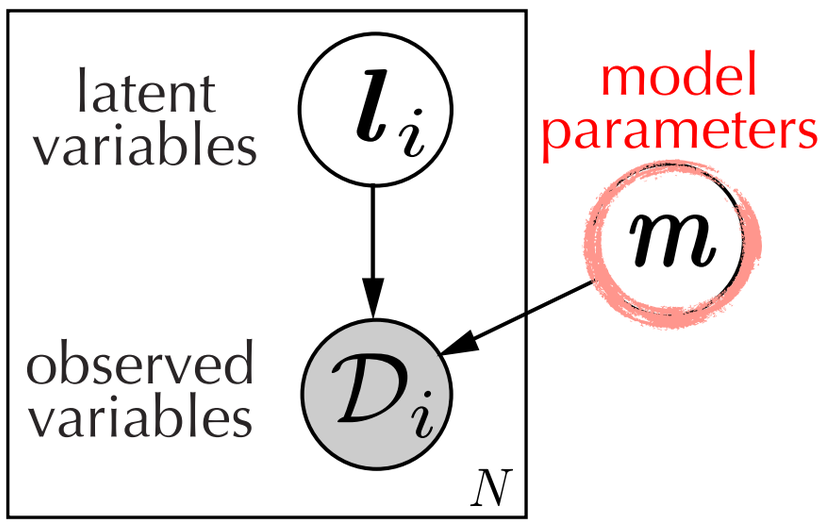


log-marginal-likelihood

$$\log p(\mathcal{D}) \geq \int q(\boldsymbol{l})q(\boldsymbol{m}) \log \frac{p(\mathcal{D}, \boldsymbol{l}|\boldsymbol{m})p(\boldsymbol{m})}{q(\boldsymbol{l})q(\boldsymbol{m})}$$

variational posterior
over latent variables

Variational inference



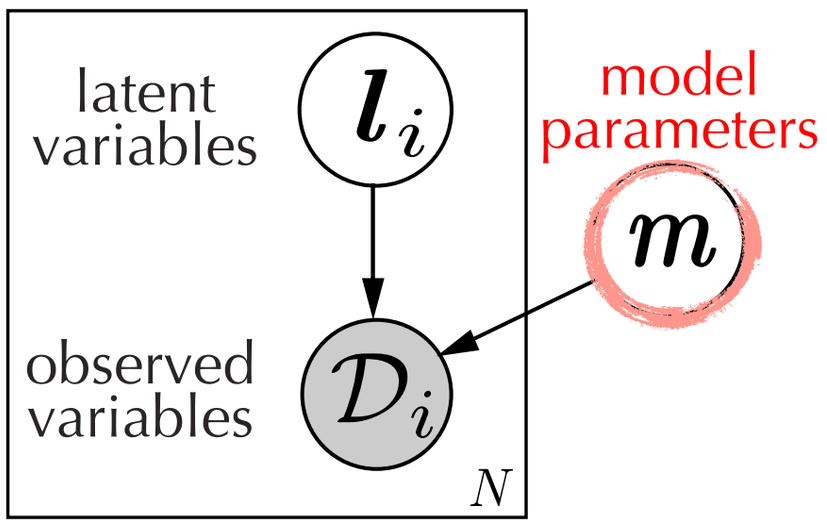
log-marginal-likelihood

$$\log p(\mathcal{D}) \geq \int q(\boldsymbol{l})q(\boldsymbol{m}) \log \frac{p(\mathcal{D}, \boldsymbol{l}|\boldsymbol{m})p(\boldsymbol{m})}{q(\boldsymbol{l})q(\boldsymbol{m})}$$

variational posterior
over latent variables

variational posterior
over model params

Variational inference



Variational lower bound

log-marginal-likelihood

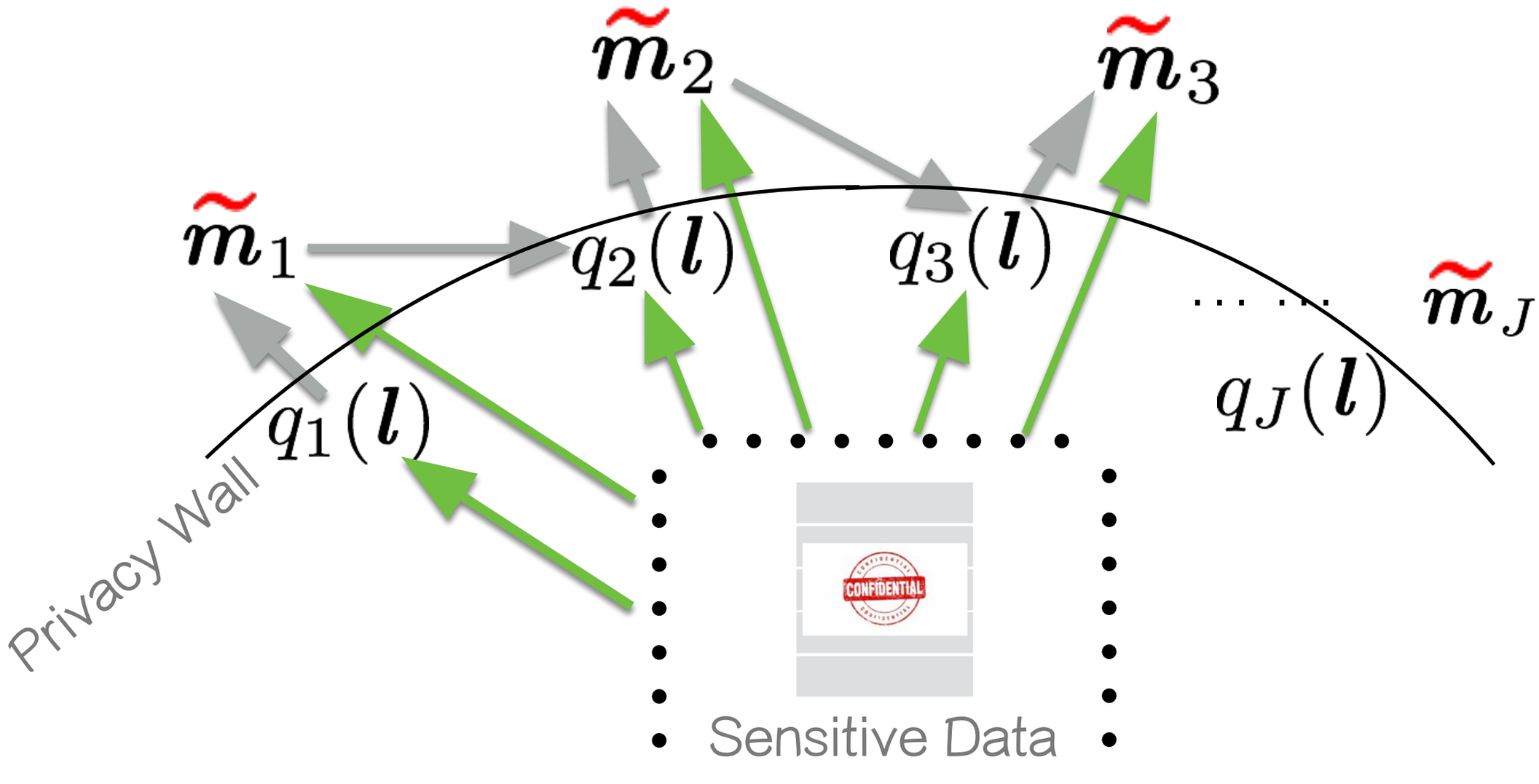
$$\log p(\mathcal{D}) \geq \mathcal{L} \int q(\boldsymbol{l})q(\boldsymbol{m}) \log \frac{p(\mathcal{D}, \boldsymbol{l}|\boldsymbol{m})p(\boldsymbol{m})}{q(\boldsymbol{l})q(\boldsymbol{m})}$$

variational posterior
over latent variables

variational posterior
over model params

DP-VI

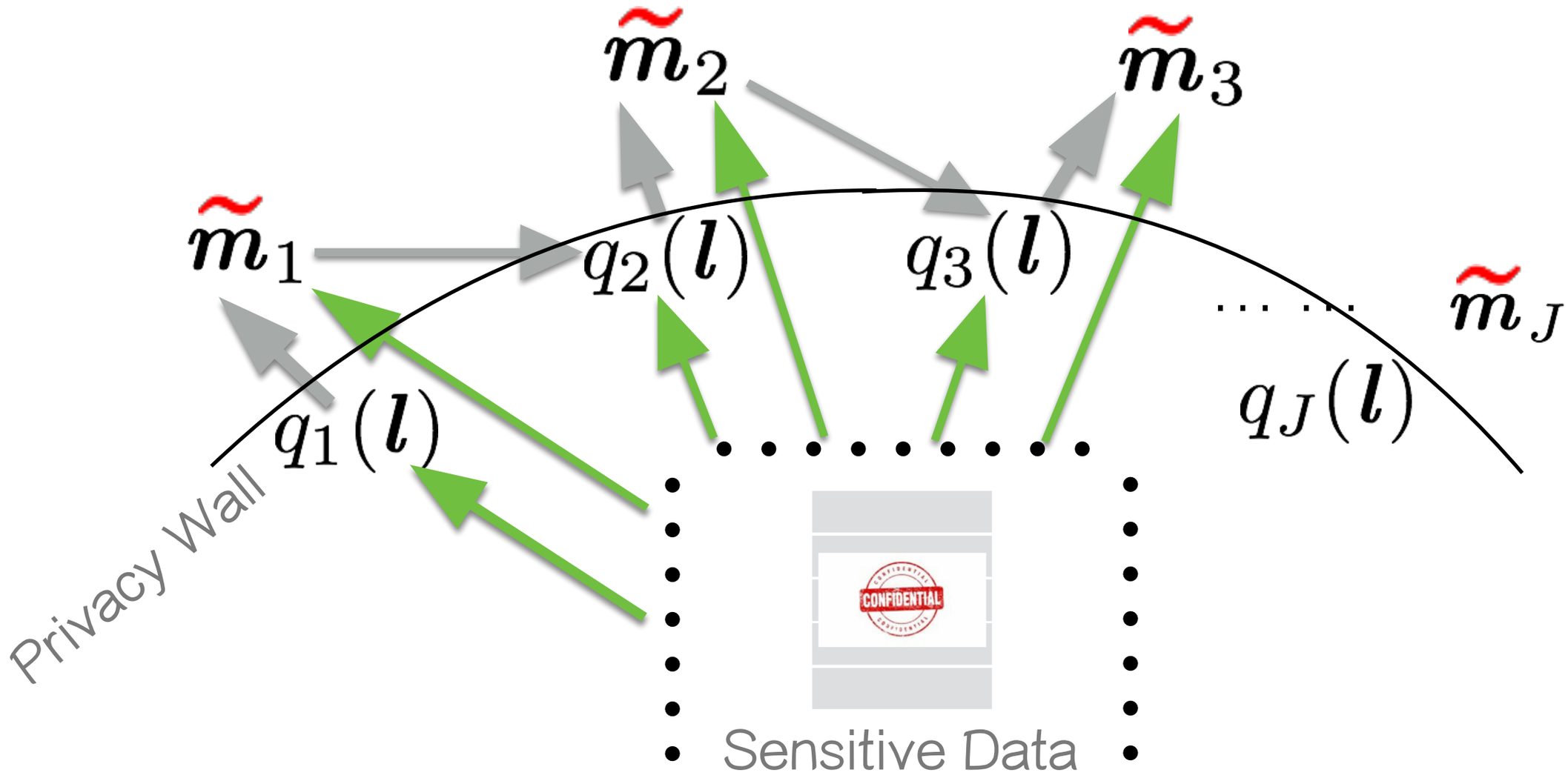
[Park et al, 2019]



DP-VI

[Park et al, 2019]

exponential family: $p(\mathcal{D}_n, \boldsymbol{l}_n|\boldsymbol{m}) \propto \exp(\mathbf{n}(\boldsymbol{m})^\top \mathbf{s}(\mathcal{D}_n, \boldsymbol{l}_n))$

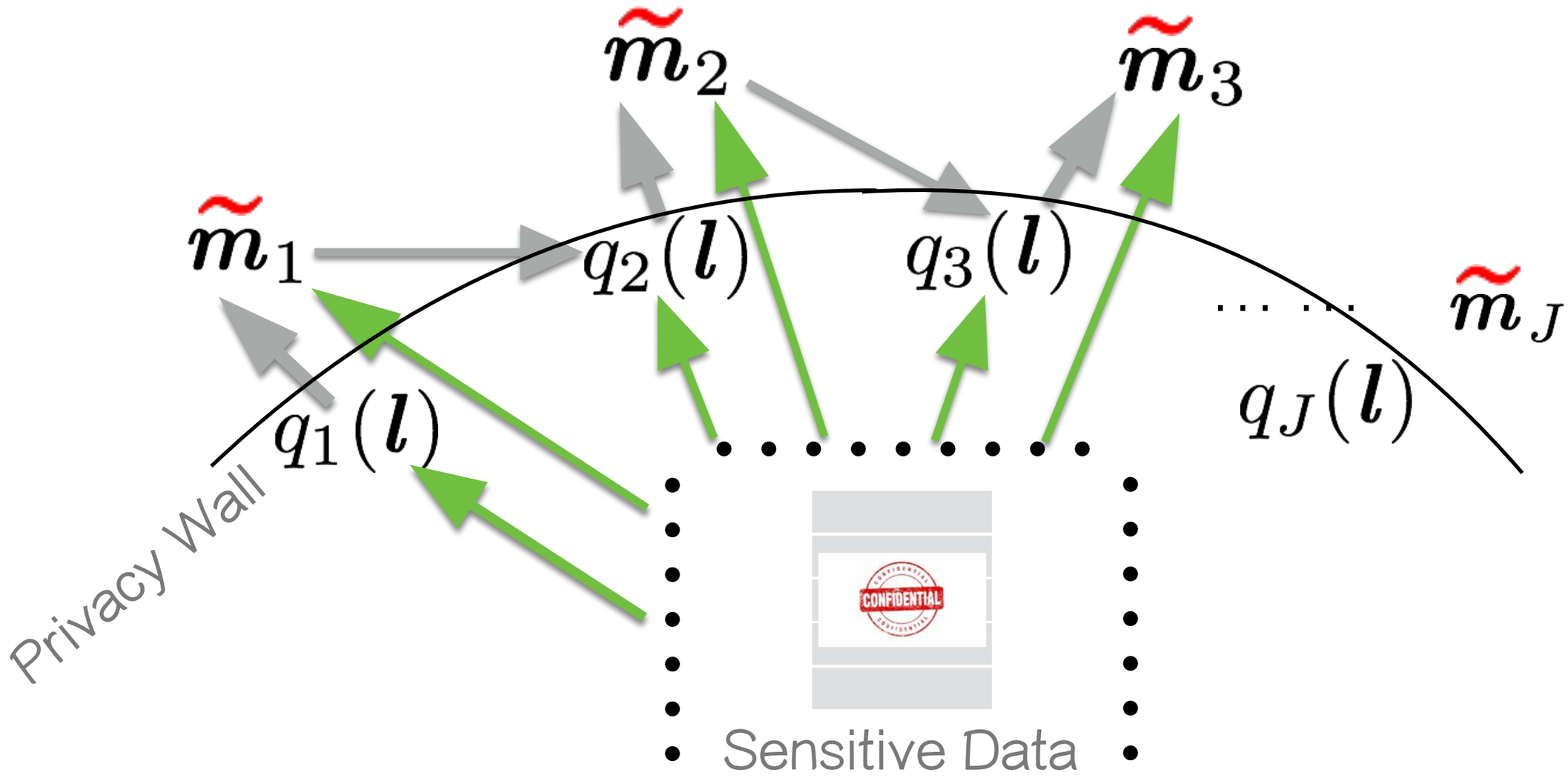


DP-VI

[Park et al, 2019]

exponential family: $p(\mathcal{D}_n, \boldsymbol{l}_n | \boldsymbol{m}) \propto \exp(\mathbf{n}(\boldsymbol{m})^\top \mathbf{s}(\mathcal{D}_n, \boldsymbol{l}_n))$

natural params

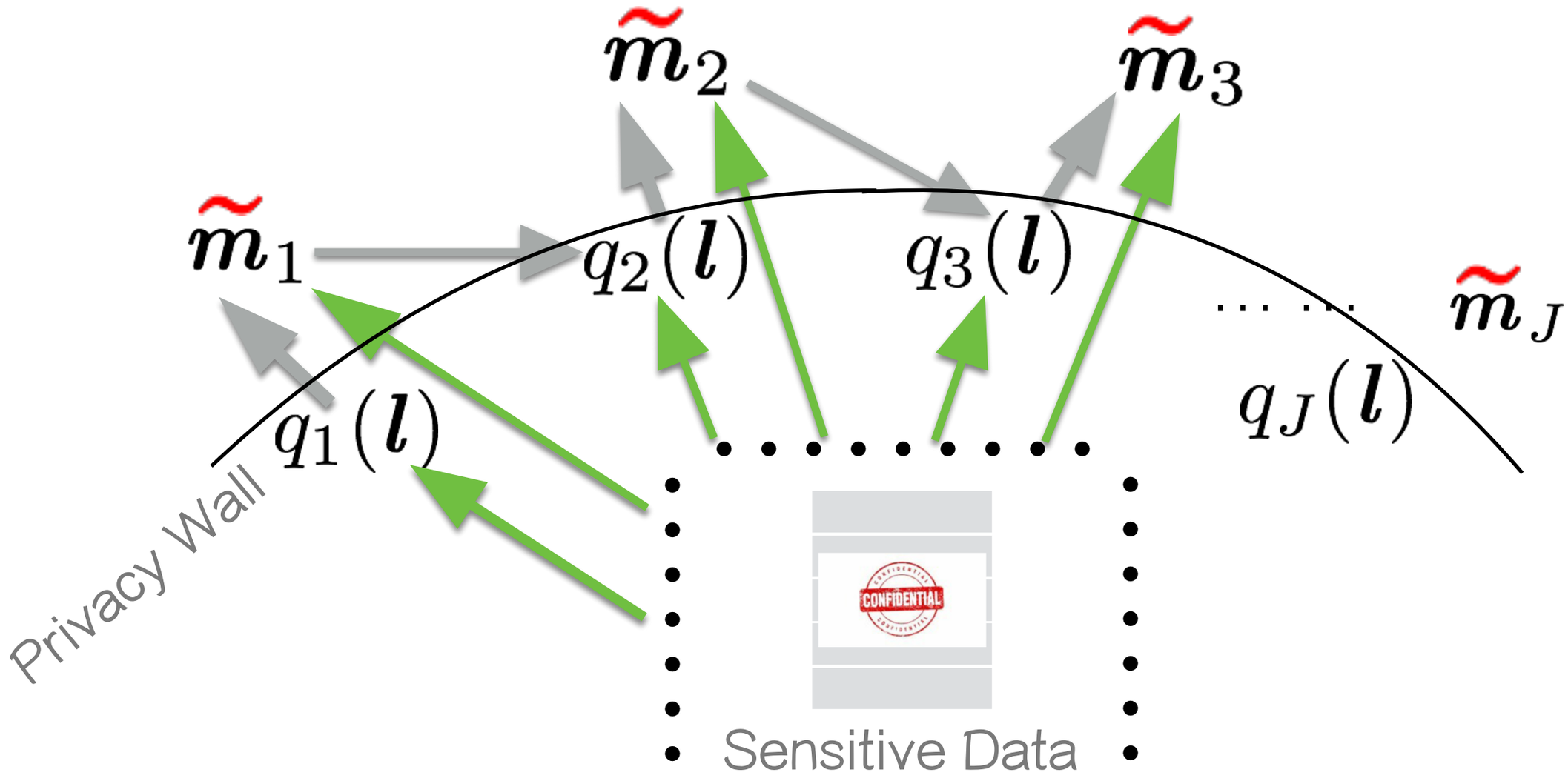


DP-VI

[Park et al, 2019]

exponential family: $p(\mathcal{D}_n, \boldsymbol{l}_n|\boldsymbol{m}) \propto \exp(\mathbf{n}(\boldsymbol{m})^\top \mathbf{s}(\mathcal{D}_n, \boldsymbol{l}_n))$

natural params suff stats



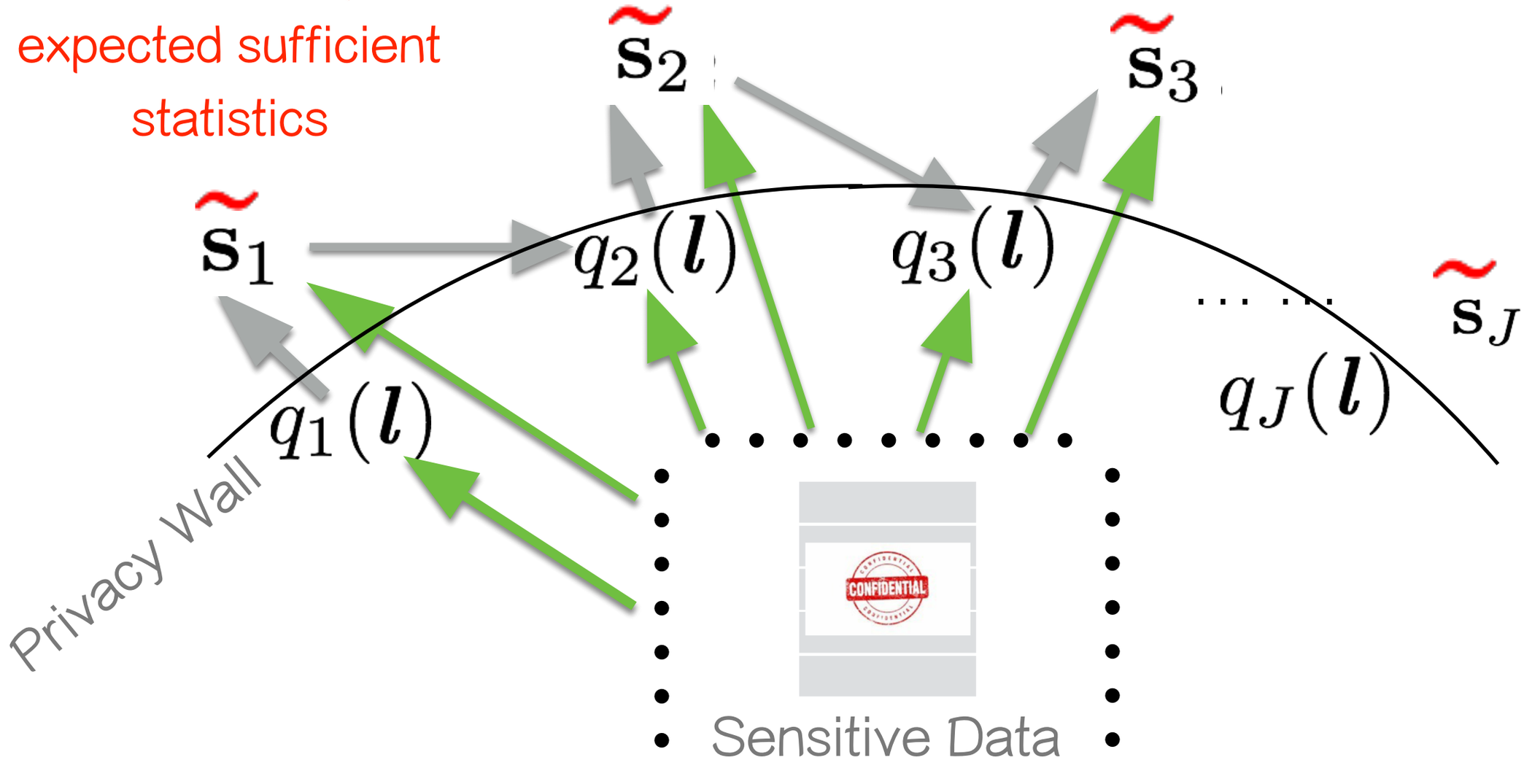
DP-VI

[Park et al, 2019]

exponential family: $p(\mathcal{D}_n, \boldsymbol{l}_n|\boldsymbol{m}) \propto \exp(\mathbf{n}(\boldsymbol{m})^\top \mathbf{s}(\mathcal{D}_n, \boldsymbol{l}_n))$

natural params suff stats

Perturb & output expected sufficient statistics



Large document datasets : stochastic learning

Privacy Amplification Effect due to subsampling

Theorem 1 *(Theorem 1 in (Li et al., 2012)) Any* $(\epsilon_{iter}, \delta_{iter})$*-DP mechanism running on a uniformly sampled subset of data with a sampling ratio* ν *guarantees* $(\log(1+\nu(\exp(\epsilon_{iter})-1)), \nu\delta_{iter})$*-differential privacy.* [Li et al,12]

Large document datasets : stochastic learning

Privacy Amplification Effect due to subsampling

Theorem 1 *(Theorem 1 in (Li et al., 2012)) Any* $(\epsilon_{iter}, \delta_{iter})$*-DP mechanism running on a uniformly sampled subset of data with a sampling ratio* ν *guarantees* $(\log(1+\nu(\exp(\epsilon_{iter})-1)), \nu\delta_{iter})$*-differential privacy.* [Li et al,12]

sampling rate	total privacy loss
1	1
0,1	0,159
0,01	0,017

Posterior topics by DP-VI

Non-private LDA

topic 3:	
david	0.0667
king	0.0318
god	0.0304
son	0.0197
israel	0.0186
bible	0.0156
hebrew	0.0123
story	0.0102
book	0.0095
adam	0.0092

Private LDA (Moments Accountant)

topic 83:	
david	0.0410
jonathan	0.0199
king	0.0188
samuel	0.0186
israel	0.0112
saul	0.0075
son	0.0068
dan	0.0067
god	0.0053
story	0.0048

Private LDA (Advanced Composition)

topic 73:	
mount	0.0034
display	0.0011
animal	0.0011
equipment	0.0011
cynthia	0.0009
position	0.0008
systems	0.0008
support	0.0008
software	0.0008
heavy	0.0008

DP-VI with MA composition + privacy amplification outputs more coherent topics (red box) than DP-VI with AC (blue box) + privacy amplification

Deep learning

DP-Stochastic gradient descent

[Abadi et al, 2016]

Given a mini-batch,

(A) $\mathbf{g}_t(x_i) \leftarrow \nabla_{\theta_t} \mathcal{L}(\theta_t, x_i)$

(B) $\mathbf{g}_t = \frac{1}{L} \sum_i \mathbf{g}_t(x_i)$

(C) $\theta_{t+1} \leftarrow \theta_t - \eta_t \mathbf{g}_t$

DP-Stochastic gradient descent

[Abadi et al, 2016]

Given a mini-batch,

(A) $\mathbf{g}_t(x_i) \leftarrow \nabla_{\theta_t} \mathcal{L}(\theta_t, x_i)$

(B) $\mathbf{g}_t = \frac{1}{L} \sum_i \mathbf{g}_t(x_i)$

Learning rate

(C) $\theta_{t+1} \leftarrow \theta_t - \eta_t \mathbf{g}_t$

DP-Stochastic gradient descent

[Abadi et al, 2016]

Given a mini-batch,

(A) $\mathbf{g}_t(x_i) \leftarrow \nabla_{\theta_t} \mathcal{L}(\theta_t, x_i)$

(A.1) $\bar{\mathbf{g}}_t \leftarrow \dfrac{\mathbf{g}_t(x_i)}{\max(1, \|\mathbf{g}_t(x_i)\|_2 / C)}$

(B) $\mathbf{g}_t = \dfrac{1}{L} \sum_i \mathbf{g}_t(x_i)$

Learning rate

(C) $\theta_{t+1} \leftarrow \theta_t - \eta_t \mathbf{g}_t$

DP-Stochastic gradient descent

[Abadi et al, 2016]

Given a mini-batch,

(A) $\mathbf{g}_t(x_i) \leftarrow \nabla_{\theta_t} \mathcal{L}(\theta_t, x_i)$

(A.1) $\bar{\mathbf{g}}_t \leftarrow \dfrac{\mathbf{g}_t(x_i)}{\max(1, \|\mathbf{g}_t(x_i)\|_2 / C)}$

Each gradient has a limited sensitivity

(B) $\mathbf{g}_t = \dfrac{1}{L} \sum_i \mathbf{g}_t(x_i)$

Learning rate

(C) $\theta_{t+1} \leftarrow \theta_t - \eta_t \mathbf{g}_t$

DP-Stochastic gradient descent

[Abadi et al, 2016]

Given a mini-batch,

(A) $\mathbf{g}_t(x_i) \leftarrow \nabla_{\theta_t} \mathcal{L}(\theta_t, x_i)$

(A.1) $\bar{\mathbf{g}}_t \leftarrow \dfrac{\mathbf{g}_t(x_i)}{\max(1, \|\mathbf{g}_t(x_i)\|_2 / C)}$

Each gradient has a limited sensitivity

(B) $\mathbf{g}_t = \dfrac{1}{L} \sum_i \mathbf{g}_t(x_i) \longrightarrow \tilde{\mathbf{g}}_t \leftarrow \dfrac{1}{L} \sum_i \left[\bar{\mathbf{g}}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 I) \right]$

Learning rate

(C) $\theta_{t+1} \leftarrow \theta_t - \eta_t \tilde{\mathbf{g}}_t$

DP-Stochastic gradient descent

[Abadi et al, 2016]

Given a mini-batch,

(A) $\mathbf{g}_t(x_i) \leftarrow \nabla_{\theta_t} \mathcal{L}(\theta_t, x_i)$

(A.1) $\bar{\mathbf{g}}_t \leftarrow \dfrac{\mathbf{g}_t(x_i)}{\max(1, \|\mathbf{g}_t(x_i)\|_2 / C)}$

Each gradient has a limited sensitivity

(B) $\mathbf{g}_t = \dfrac{1}{L} \sum_i \mathbf{g}_t(x_i) \longrightarrow \tilde{\mathbf{g}}_t \leftarrow \dfrac{1}{L} \sum_i \left[\bar{\mathbf{g}}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 I) \right]$

Learning rate

(C) $\theta_{t+1} \leftarrow \theta_t - \eta_t \mathbf{g}_t \longrightarrow \theta_{t+1} \leftarrow \theta_t - \eta_t \tilde{\mathbf{g}}_t$

DP parameters

DP-Stochastic gradient descent

[Abadi et al, 2016]

Given a mini-batch,

(A) $\mathbf{g}_t(x_i) \leftarrow \nabla_{\theta_t} \mathcal{L}(\theta_t, x_i)$

(A.1) $\bar{\mathbf{g}}_t \leftarrow \dfrac{\mathbf{g}_t(x_i)}{\max(1, \|\mathbf{g}_t(x_i)\|_2 / C)}$

Each gradient has a limited sensitivity

(B) $\mathbf{g}_t = \dfrac{1}{L} \sum_i \mathbf{g}_t(x_i) \longrightarrow \tilde{\mathbf{g}}_t \leftarrow \dfrac{1}{L} \sum_i \left[\bar{\mathbf{g}}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 I) \right]$

(C) $\theta_{t+1} \leftarrow \theta_t - \eta_t \mathbf{g}_t \longrightarrow \theta_{t+1} \leftarrow \theta_t - \eta_t \tilde{\mathbf{g}}_t$

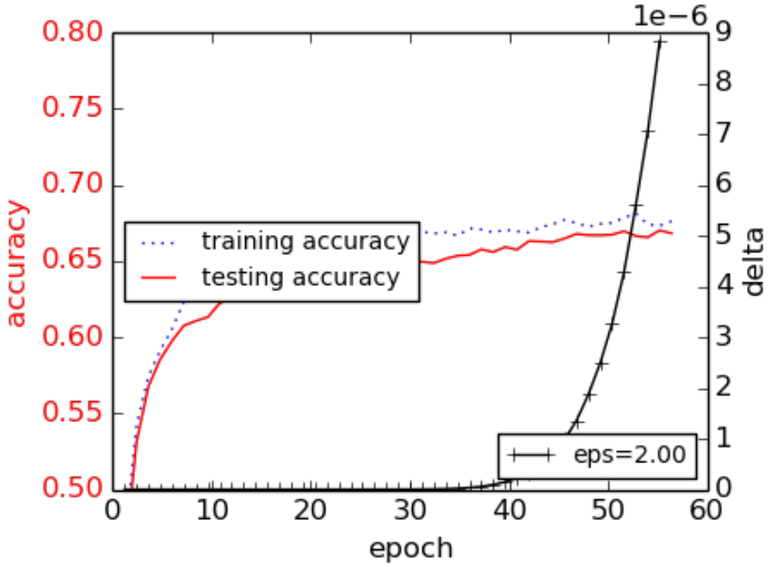
Learning rate

DP parameters

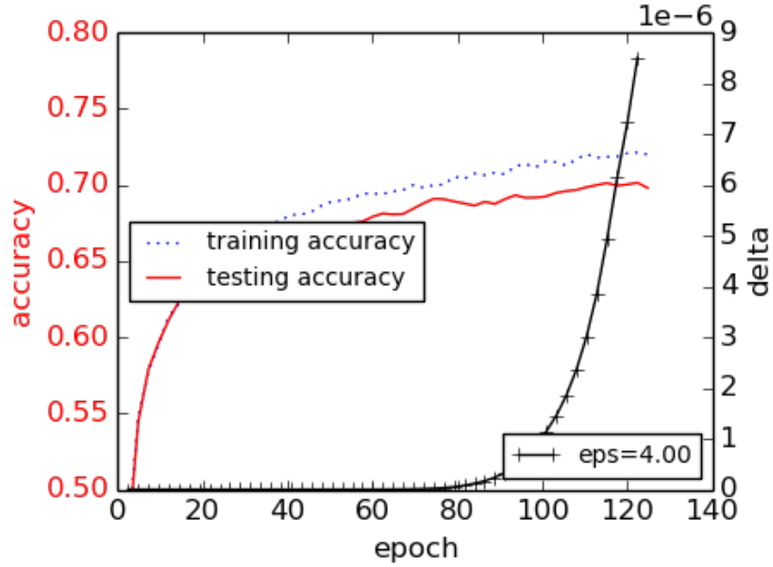
- Moments accountant composition to compute the total privacy loss incurred during training

DP-SGD on CIFAR10 data

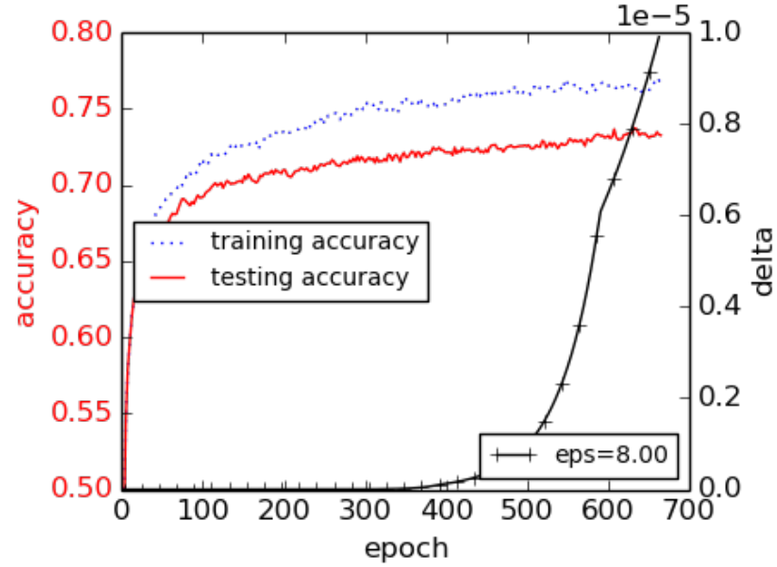
[Abadi et al, 2016]



(1) $\varepsilon = 2$



(2) $\varepsilon = 4$



(3) $\varepsilon = 8$

- DP-SGD with a high level of noise (epsilon=2) provides a less accurate classifier than DP-SGD with a lowe level of noise (epsilon=8)
- The gap between training and test accuracies gets smaller with more additive noise. Noise added for privacy seems helpful for better generalization.

Examples of DP-ML algorithms

- DP-EM for clustering & density estimation

- DP-VI for topic modeling

- DP-SGD for deep learning

DP in industry

DP in machine learning systems

Google's Randomized Aggregatable Privacy-Preserving Ordinal Response (RAPPOR)

- A DP technique for crowdsourcing statistics, e.g., for collecting data about the Google Chrome users.
- Based on Randomised response technique!

DP in machine learning systems

Google's Randomized Aggregatable Privacy-Preserving Ordinal Response (RAPPOR)

- A DP technique for crowdsourcing statistics, e.g., for collecting data about the Google Chrome users.
- Based on Randomised response technique!

Introduce Randomness

1. flip a coin
2. if tails, respond truthfully
3. if heads, flip another coin
4. if heads, "yes". Say "no" otherwise.

DP in machine learning systems

Google's Randomized Aggregatable Privacy-Preserving Ordinal Response (RAPPOR)

- A DP technique for crowdsourcing statistics, e.g., for collecting data about the Google Chrome users.
- Based on Randomised response technique!

Introduce Randomness

1. flip a coin
2. if tails, respond truthfully
3. if heads, flip another coin
4. if heads, "yes". Say "no" otherwise.

True answer 75% of time

DP in machine learning systems

Google's Randomized Aggregatable Privacy-Preserving Ordinal Response (RAPPOR)

- A DP technique for crowdsourcing statistics, e.g., for collecting data about the Google Chrome users.
- Based on Randomised response technique!

Introduce Randomness

1. flip a coin
2. if tails, respond truthfully
3. if heads, flip another coin
4. if heads, "yes". Say "no" otherwise.

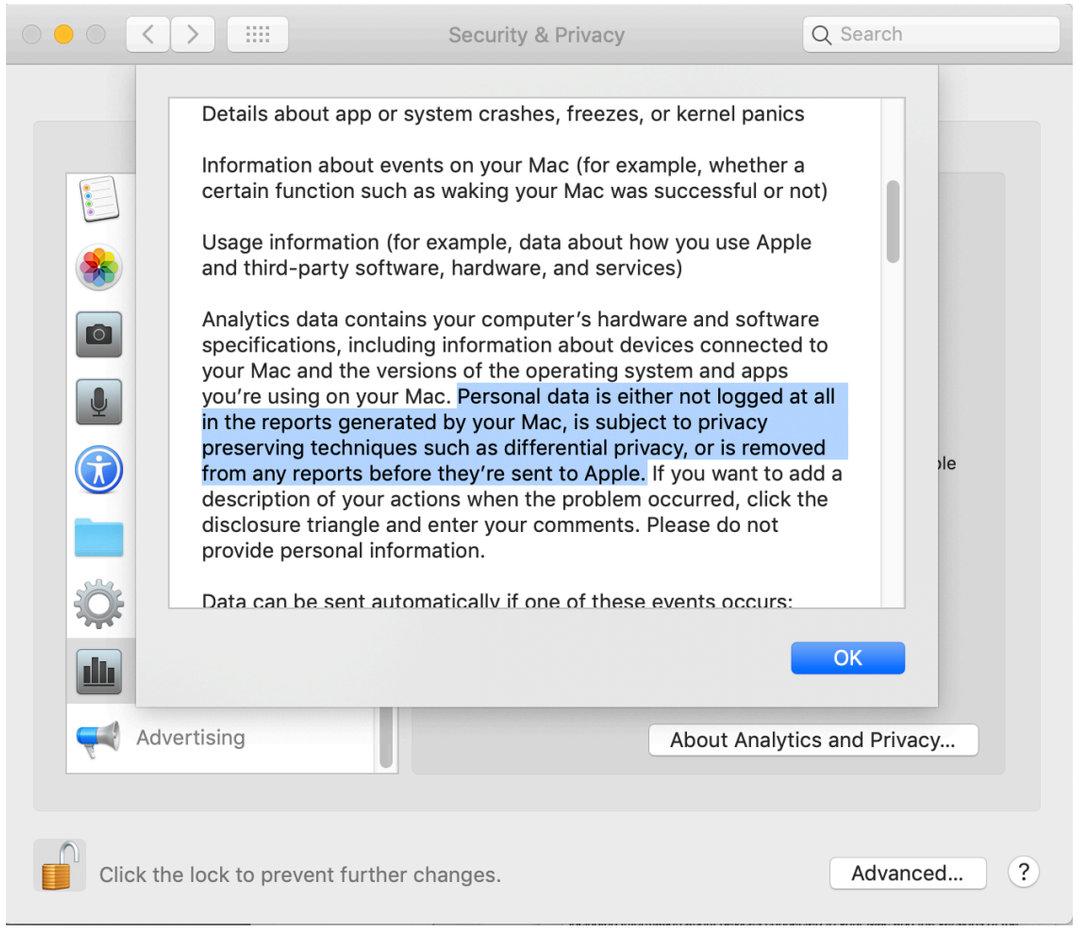
True answer 75% of time

accurate population statistics
while preserving the privacy
of the individual answers!

DP in data collection

Apple Mac OS

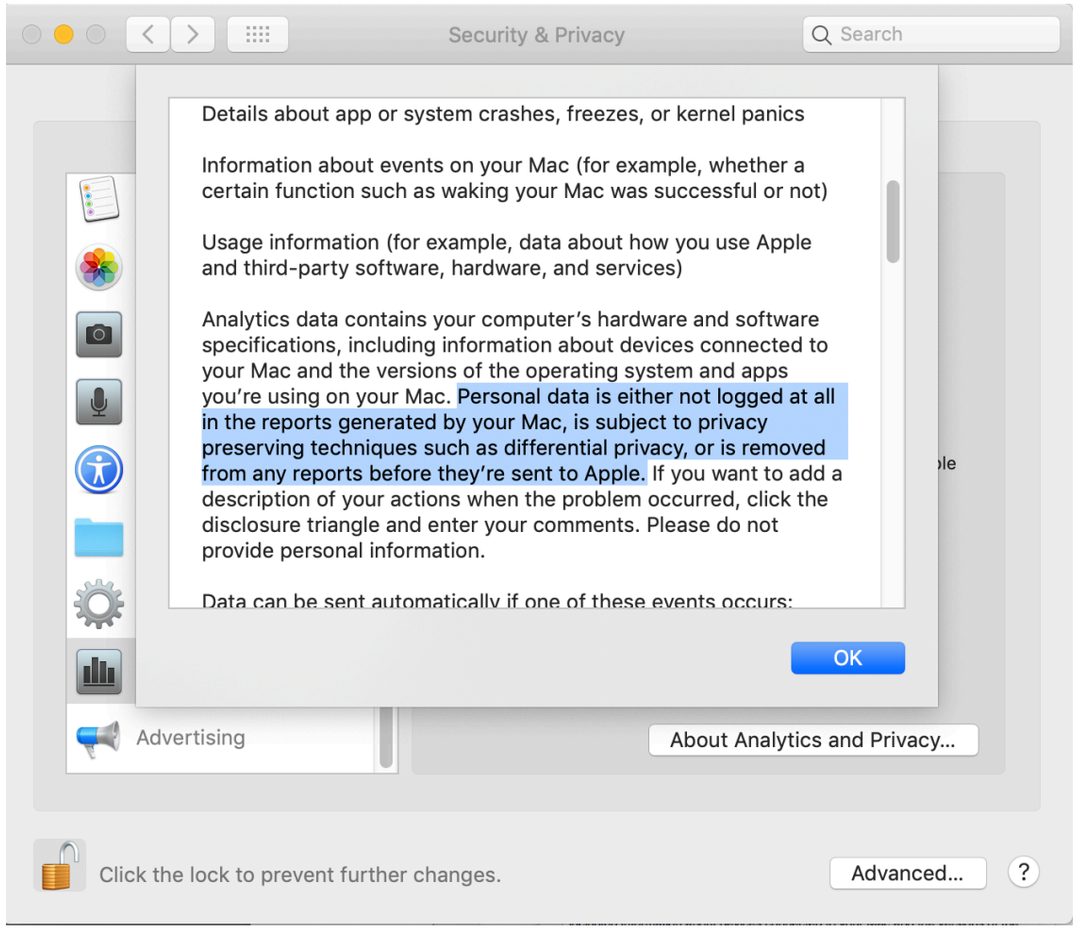
- DP techniques are used to send my own personal data collected by iMAC (analytics).



DP in data collection

Apple Mac OS

- DP techniques are used to send my own personal data collected by iMAC (analytics).



- Specific techniques on a new word and Emoji: OneBitHistogram, CountMedianSketch -> adding noise to data before sending to server!
- Issue: non-transparent & questionable "epsilon" selection [Tang et al. 2017]

Summary

- Privacy is necessary to consider not only for the sake of protection against threats, but also for philanthropic reasons in data sharing.

- Differential privacy seems to dominate the field due to the mathematically provable guarantees.

- Although many progresses have been made in privacy-preserving machine learning, there are still A LOT of work that needs to be done to apply DP in many machine learning/data sharing applications.