Privacidad: Definición, Modelos, Propiedades y Aplicaciones

Álvaro J. Riascos Villegas

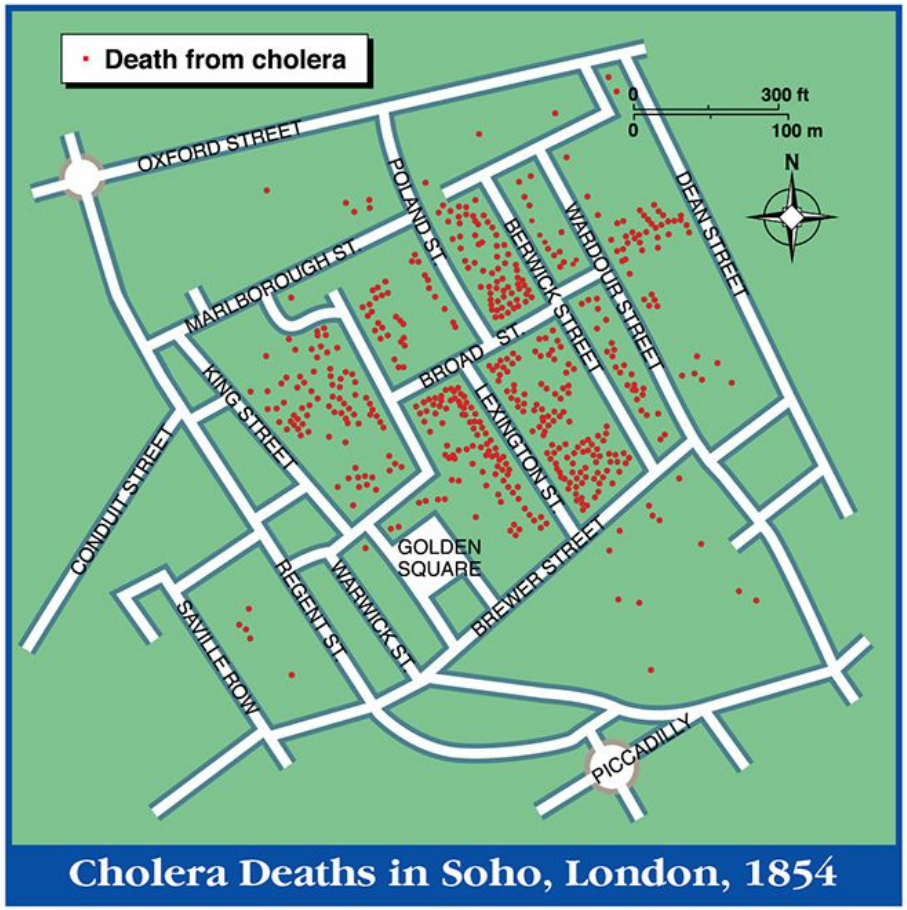
17 de junio de 2022

Introducción

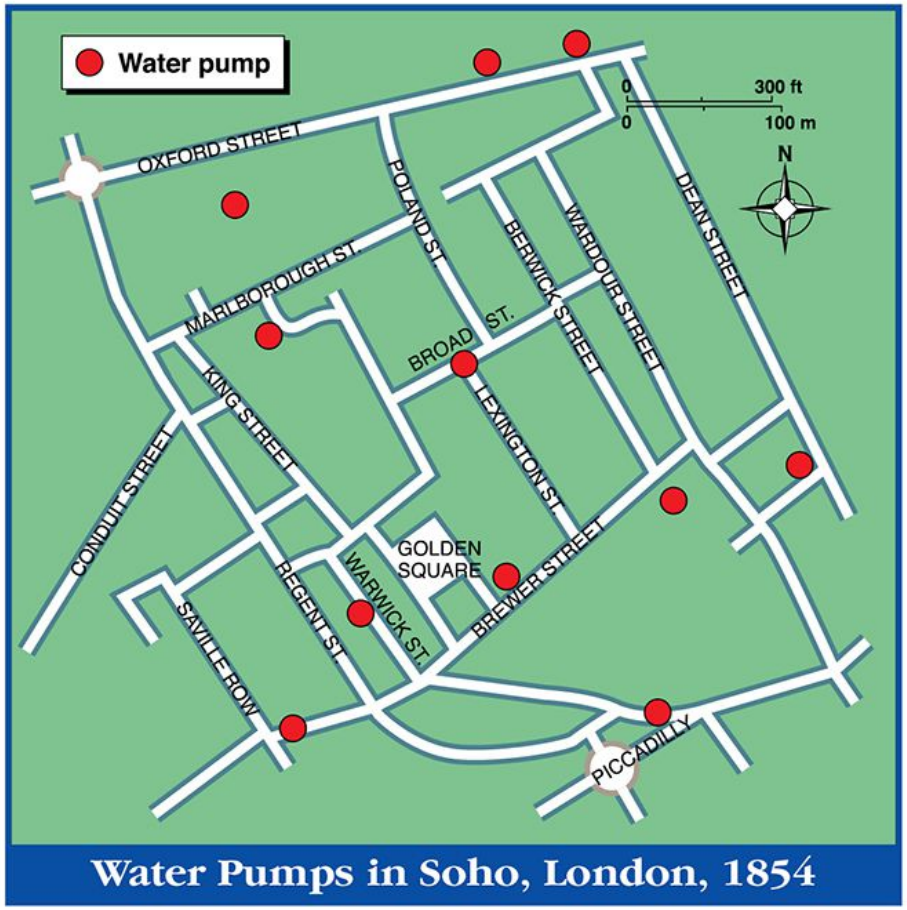
- El derecho a la privacidad ha evolucionado con el tiempo hasta convertirse en un pilar de la sociedad moderna.
- En Estados Unidos han estimado que conociendo el código postal, fecha de nacimiento y sexo, con 87 de probabilidad se puede identificar la persona.
- Veamos algunos ejemplos famosos:
  - Muertes por colera Londres.
  - Gobernador de Massachusetts (1997).
  - Netflix `https://www.wired.com/2009/12/netflix-privacy-lawsuit`

- En el 2006 Netflix lanza un concurso para mejorar su sistema de recomendación.
- Hizo público únicamente la las calificaciones de 500 usuarios de $18,000$ películas. Un total de 100 millones de registros.
- En Estados Unidos revelar la lista de películas que una persona alquila tiene una multa de $USD2,500$.
- Cruzando la base de Netflix con las calificaciones públicamente disponibles (con nombre) en Internet Movie Data Base (IMDB) se puede identificar con probabilidad muy alta el nombre en la base de Netflix.
- Netflix `https://www.wired.com/2009/12/netflix-privacy-lawsuit`

Cólera Londres: Casos



Cólera Londres: Repositorios de Agua



- Este es un ejemplo del uso de datos (privados) para el bienestar social.

Historias Clínicas y Netflix

- Gobernador de Massachusetts (1997): hace públicos todas las historias clínicas de empleados del estado (debidamente anonimizados).
- Muy poco tiempo despúes recibió su historia clínica por correo.
- Al comparar los datos con la información de votantes registrados del estado se encontraban en común las variables: zip code, fecha de naciemiento y sexo.
- Netflix:
  `https://www.wired.com/2009/12/netflix-privacy-lawsuit`

Introducción

- Una base de datos es usalmente una representacion de información poblacional.
- El analista tiene como objetivos aprender de la población a partir de la base de datos.
- El problema formal consiste en diseñar una forma de responder a consultas a una base de datos de tal forma que la respuesta no revele información privada de las personas registradas en la base de datos.

Introducción

- Por ejemplo, las personas de la base podrían no tener incentivos a hacer parte de la base en la medida que no se proteja su privacidad: Participar de un estudio sobre las consecuencias de fumar puede tener eventualmente consecuencias para el (e.g., prima de seguros).
- Las personas podrían no tener incentivos a decir la verdad en una encuesta: Un encuesta que pregunta sobre algo potencialmente privado (e.g., horientacion sexual, uso de drogas, etc).

Introducción: Ejemplo

Example (Warner 1965)

Suponga que en una encuesta se pregunta por algun atributo (propiedad) para el cual no existen incentivos a decir la verdad. Considere este algoritmo (privado):

- Lanzar moneda: si cae cara responder la verdad.
- Si cae sello volver a lanzar la moneda. Si cae cara decir si, si cae sello decir no.

Este algorimto preserva una forma de privacidad: uno no puede ser juzgado por la respuesta: *plausible deniability*). Sin embargo, se puede estimar con precisión la proporción p de personas que tiene la propiedad. La proporción P de respuestas es: $\frac{1}{2}p + \frac{1}{4}$.

- La aleatorización es escencial.

Privacidad Diferencial Informalmente

- Un proceso de consulta de información satisface privacidad diferencial si analista que hace la consulta no sabe nada adicional sobre un individuo, si este individuo entra o no en la consulta de la base de datos (i.e., que el individuo se incluya en la consulta no lo afecta a el de forma diferencial comparado a no participar de la consulta).

El Problema de Privacidad: Formalmente

- Una base de datos x con dominio D, (i.e., $x \in D^n$), D es el dominio en el que se encuentran los atributos de cada registro. Suponemos que cada registro corresponde a una persona. El número de personas no está fijo.
- Los atributos pueden ser un vector de dimensión d y $D = R^d$ (atributos numéricos) o $D = \{0, 1\}^d$ atributos binarios, etc.
- Una base de datos x (don dominio finito $\mid D \mid$) se puede representar con su histograma: $x = (x_i)_{i \in D}$ donde x_i es el número de elementos de la base que tienen el tipo i (vector de atributos). En este caso x se puede ver como un elemento de $N^{|x|}$.

El Problema de Privacidad: Formalmente

- La norma de una base de datos x es: $\|x\| = \sum_{i=1}^{|D|} \mid x_i \mid$
- La distancia entre dos bases de datos x, y con el mismo dominio es $\|x - y\|$ (i.e., Hamming) es el número de registros que son distintos entre dos bases de datos (i.e., un registro es distinto a otro si difieren en algún atributo).
- Una consulta o algoritmo (probablmente aleatorio) es $M : N^{|x|} \rightarrow \Delta(B)$ donde B es el rango de M.

Definición de Privacidad

Definition (ϵ - indistinguible)

Dado $\epsilon > 0$ decimos que un algoritmo es ϵ indistinguible si para todo x, x' dos bases de datos con dominio D que difieren solamente en un registro, entonces para cualquier algoritmo M y transcripción $S \subseteq B$:

$$P[M(x) \in S] \leq \exp^{\epsilon} P[M(x') \in S] \tag{1}$$

- Si ϵ es pequeño, la definición anterior requiere:

$$\frac{P[M(x) \in S]}{P[M(x') \in S]} \in (1 - \epsilon, 1 + \epsilon) \tag{2}$$

- ϵ es el parámetro de filtración. Si $\epsilon = 0$ no se filtra información por cambiar la base en un registro (más privacidad). La consulta arroja la misma información no importa cual base de datos se use.

Definición de Privacidad

- Cualquier costo de estar en esa base sería igual a no hacerlo.

Definición de Privacidad

Example (Warner 1965)

n individuos responden $x_i \in \{0, 1\}$. Cada individuo responde sinceramente, x_i con probabilidad $\frac{e^\epsilon}{1+e^\epsilon}$ o miente con probabilidad complementaria. Consideremos dos bases que difieren en la respuesta de un solo individuo
$x = (x_1, ..., x_{n-1}, x_n), x' = (x_1, ..., x_{n-1}, x'_n)$. El resultado de una consulta sería un elemento de $\{0, 1\}^n$. Sin embargo, si reportamos el resultado de la consulta con el mecanismo anterior este preserva ϵ privacidad diferencial (suponiendo que todos responde de forma independiente). Para ver esto hay que comparar: $\frac{P[M(x)=t]}{P[M(x')=t]}$

Distribución de Laplace

- La distribución de Laplace con densidad $h(y) = \frac{1}{2\lambda} \exp(- \mid y \mid /\lambda)$, va ser fundamental en lo que resta.
- Esta distribución con parámetro λ tiene media cero y varianza $2\lambda^2$.

Definición de Privacidad

Example (Sumas)

Sea $x \in \{0,1\}^n$ y una consulta $f(x) = \sum_i x_i$. Considere el mecanismo $M(x) = f(x) + Y$, donde $Y \sim Laplace(\frac{1}{\epsilon})$. Este mecanismo es ϵ - indistinguible:

Introducción

- Local differential privacy: tiene como propósito permitir la interaccion entre un agregador de informacion y usuario de tal forma que los datos nunca sean extraidos de un repositorio local privado (con información potencialmente sensible).
- Por ejemplo: los navegadores de internet (e.g., Chrome) o los sistemas operativos que recolectan información sobre el uso del sistema (e.g., MS Windows).
- La posibilidad de extraer información que no comprometa la privacidad es crucial para la construcción de modelos de aprendizaje de máquinas.

Formalmente

- Vamos a tener un agregador de información y un número n de usuarios.
- Para proteger su privacidad cada usuaruo i perturba su información usando un mecanismo (estocástico) M (o función de perturbación) para reportarle al agregador $M(x_i)$ en vez del verdadero registro x_i.

Estimando Promedios

- Por simplicidad nos vamos a concentrar en un solo atributo numérico x_i.
- El problema es estimar el promedio: $\frac{1}{n}\sum_{i=1}^{n} x_i$.
- El problema de entrenar modelo nos concentramos en regresión lineal, logístico y SVM.

Estimando Promedios

- Sea $M(x_i) = x_i + Y$, donde Y se distrbute Laplace con parámetro $\lambda = \frac{2}{\epsilon}$.
- Luego cada agente responde $x_i^* = x_i + Y$, que es un estimador no sesgado del verdadero valor x_i con varianza $\frac{8}{\epsilon^2}$.
- Agregando se obetiene un estimador no sesgado del promedio con $\frac{1}{n}\sum_{i=1}^{n} x_i^*$ con varianza $O(\frac{1}{\epsilon\sqrt{n}})$

Lineal, Logísitico y SVM

- Estos modelos de ML se pueden estimar usando la siguientes funciones de pérdidad.
  - Regresión lineal: $L(\beta, x_i, y_i) = (x_i^T \beta - y_i)^2$
  - Logistica: $L(\beta, x_i, y_i) = log(1 + \exp(-y_i x_i^T \beta))$
  - SVM: $L(\beta, x_i, y_i) = max\{0, 1 - y_i x_i^T \beta\}$
- La estimación con regularizacion sería minimizar:

$$L(\beta, \hat{x}_i, y_i) = L(\beta, x_i, y_i) + \frac{\lambda}{2} \mid\mid \beta \mid\mid_2^2 \tag{3}$$

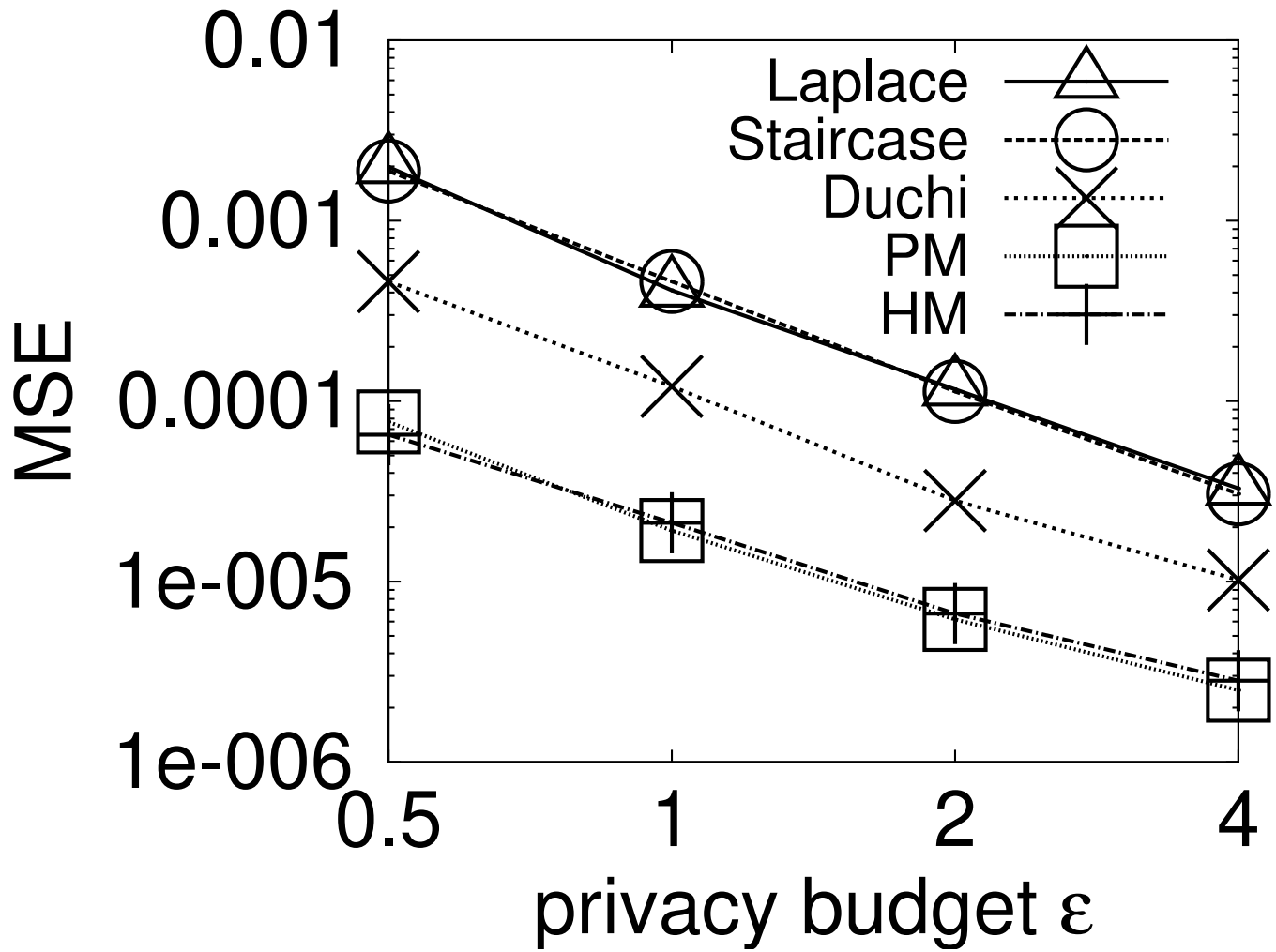
Lineal, Logísitico y SVM

- El método del gradiente desendiente itera.

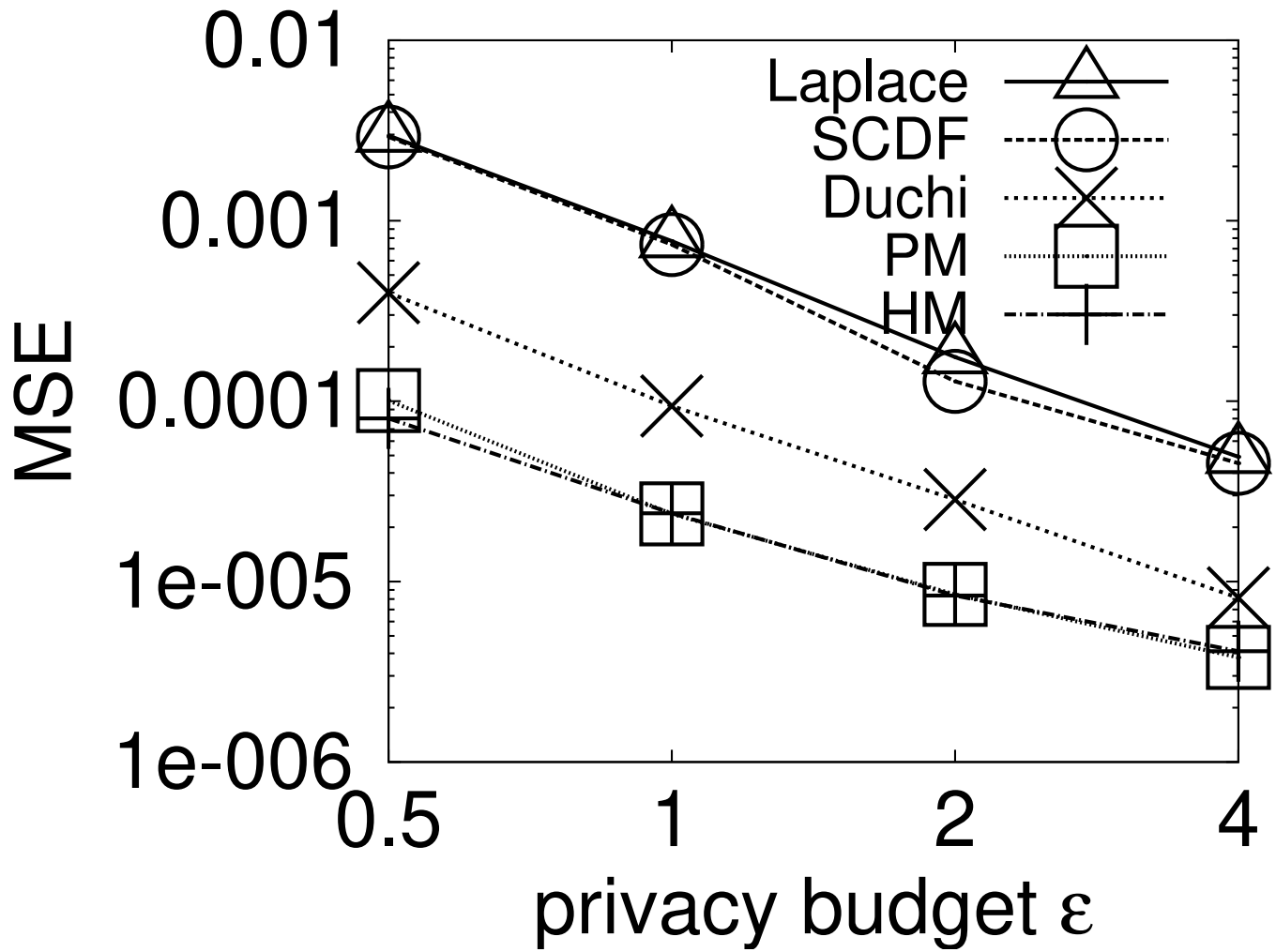
$$\beta_{t+1} = \beta_t - \gamma \nabla \frac{1}{n} \sum_i \hat{L}(\beta, x_i, y_i) \tag{4}$$

- La versión estocástica aproxima la suma con muestreos aleatorios (minibatches).
- La estimación con privacidad local consulta $\nabla \hat{L}(\beta, x_i, y_i)$ que se reporta con un mecanismo que satisfaga privacidad local.

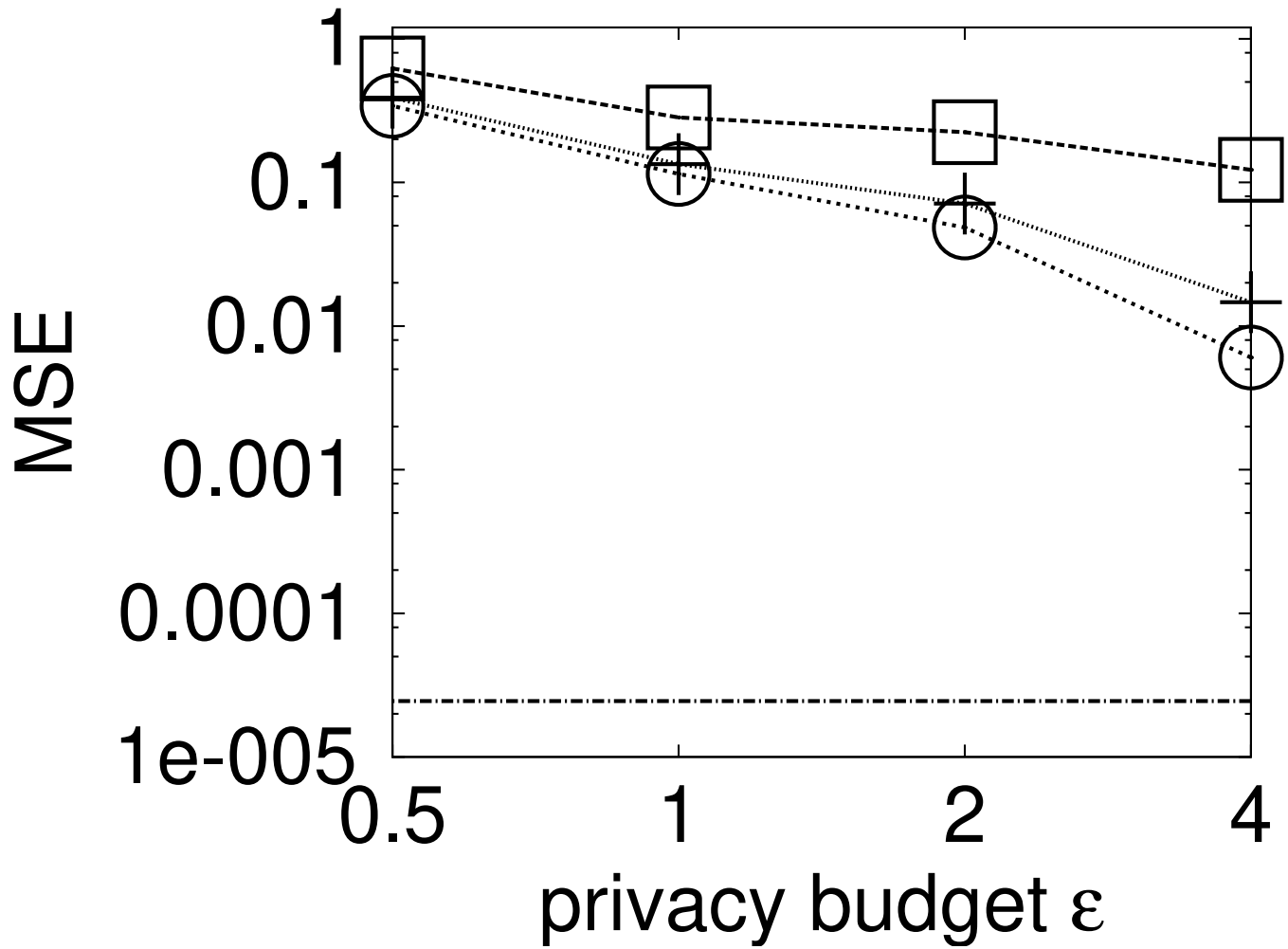
Promedios: Brasil



Promedios: México



Modelos: Brasíl Regresión Lineal



Modelos: Brasíl Logístico