Privacidad: Definición, Modelos, Propiedades y Aplicaciones

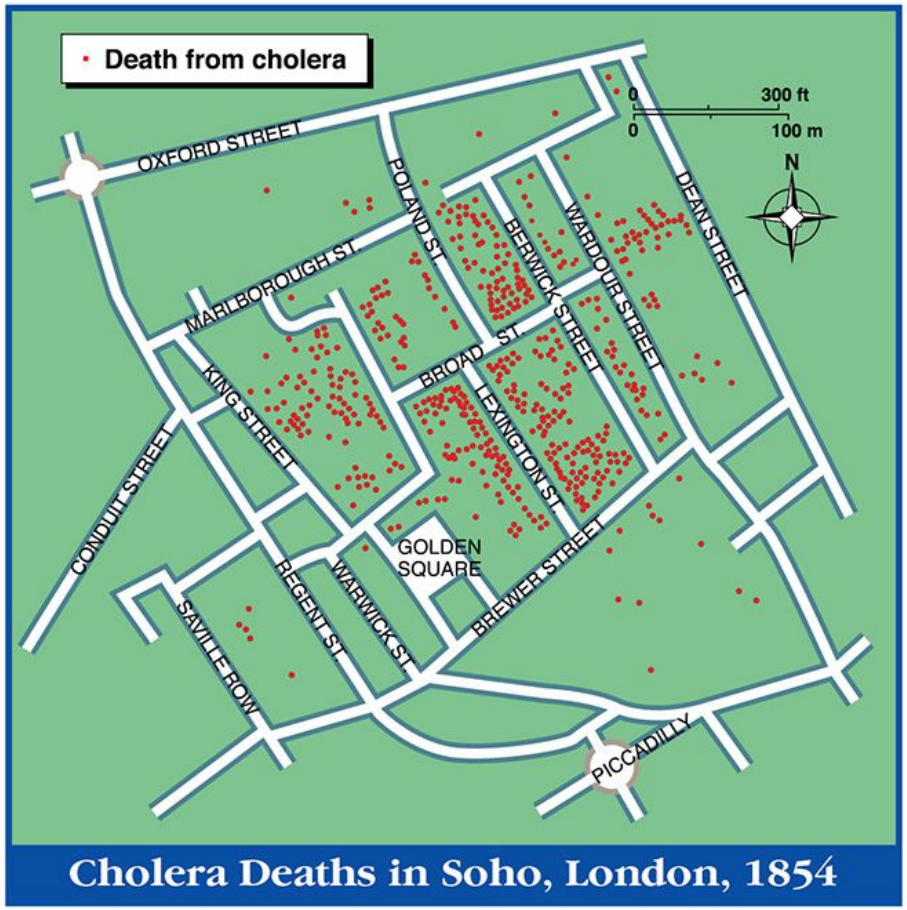
Álvaro J. Riascos Villegas

4 de julio de 2020

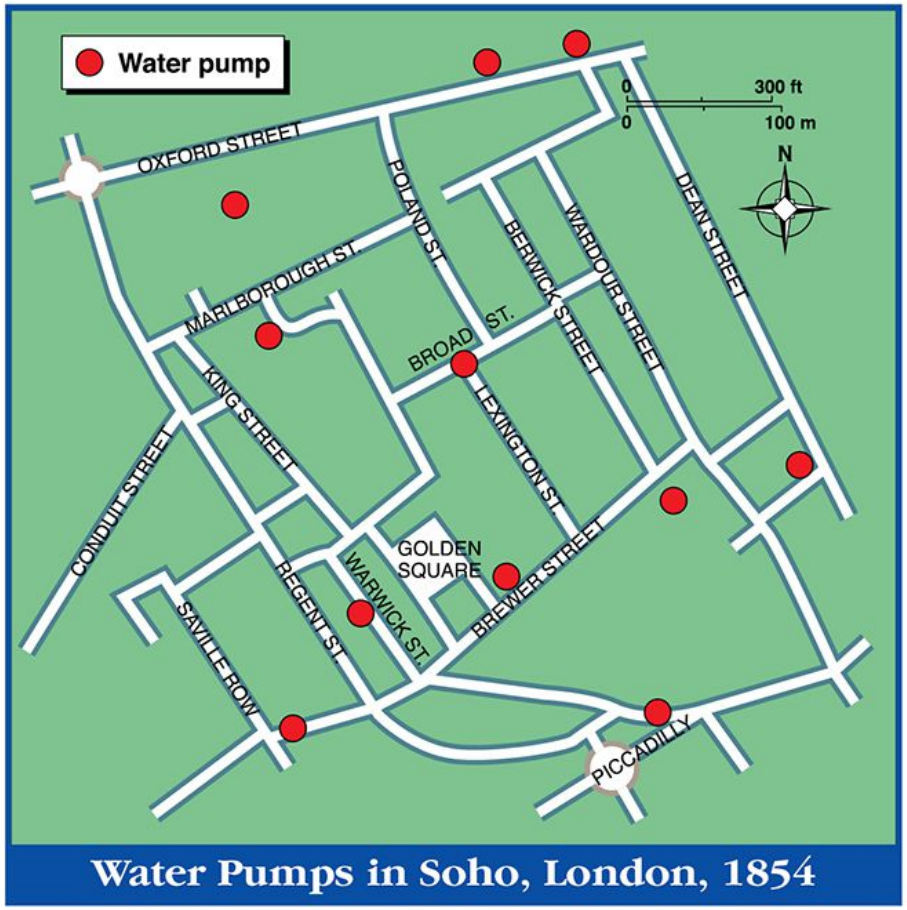
Introducción

- 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 pueden identificar la persona.
- Veamos algunos ejemplos famosos:
  - Muertes por colera Londres.
  - Governador de Massachusetts (1997).
  - 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 bien público.

Historias Clínicas y Netflix

- Governador de Massachusetts (1997): hace públicos todas las historias clínicas de empleados del estado (debidamente anonimizados).
- Muy poco tiempo despuúes recibio su historia clínica por correo.
- Al comparar los datos con la información de votantes registrados del estado se encontraban en común las varibles: 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.
- 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.

El Problema de Privacidad: Formalmente

- Una base de datos es $x \in D^n$ donde D es el dominio en el que se encuentran los atributos de cada registro y existen n registros. Suponemos que cada registro corresponde a una persona.
- Los atributos pueden ser un vector de dimensión d y $D = R^d$ (atributos numéricos) o $D = \{0,1\}^d$ atributos binarios, etc.
- La distancia entre dos bases de datos (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).
- Sea Q una consulta a una base de datos x. El resultado de la consulta (transcripción) es $T_Q(x)$.
- Dada una variable aleatoria X, vamos a denotar por $P[X = x]$ la densidad de probabiliad.

Definición de Privacidad

Definition (ϵ - indistinguible)

Dado $\epsilon > 0$ decimos que un mecanismo es ϵ indistinguible si para todo $x, x' \in D^n$ dos bases de datos que difieren solamente en un registro, entonces para cualquier consulta Q y transcripción t:

$$P[T_Q(x) = t] \leq \exp^{\epsilon} P[T_Q(x') = t] \quad (1)$$

- Obsérvese que si ϵ es pequeño, la definición anterior requiere: $\frac{P[T_Q(x)=t]}{P[T_Q(x')=t]} \in (1-\epsilon, 1+\epsilon)$.
- ϵ es el parámetro de filtración. Si $\epsilon = 0$ no se filtra información por cambiar la base en un registro (más privacidad).
- Cualquier costo de estar en esa base sería igual a no hacerlo.
- Por simplicidad usualmente vamos a escribir $T_Q(x) = f(x)$.

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:

- Obsérvese que $\frac{h(y)}{h(y')} \leq e^{\epsilon|y-y'|}$ para cualquier par de bases de datos que difieran en máximo un registro.
- Para ver esto, si x, x' difierne a lo sumo en un registro: $\frac{h(t-f(x))}{h(t-f(x'))} \leq e^{\epsilon|f(x)-f(x')|} \leq e^{\epsilon}$

Sensibilidad

Definition (Sensibilidad)

La sesibilidad L_1 de $f : D^n \rightarrow R^d$ es el menor número $S(f)$ tal que para todo $x, x' \in D^n$ que difieren en a lo sumo un registro:

$$| f(x) - f(x') | \leq S(f) \tag{2}$$

- Obsérvese que $S(f)$ es una propieda de f y no depende de una base de datos específica.

Example (Sumas)

Si $D = \{0, 1\}$ y $f(x) = \sum_i x_i$ entonces $S(f) = 1$ con la métricas estándar de R.

Sensibilidad

Example (Histogramas)

Sea $B_1, ..., B_d$ una partición del dominio y $f : D^n \to Z^d$ una función que calcula el número de registros que caen en cada uno de los elementos de la partición. La sensibilidad L_1 de f es 2.

Theorem (Caso no interactivo)

Si $f : D^n \to R^d$ *entonces el mecanismo:*
$M(x) = f(x) + (Y_1, ..., Y_d)$ *es* ϵ *- indistinguible, donde* Y_i *son i.i.d Laplace:* $(\frac{S(f)}{\epsilon})$.

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.
- La información de cada usuario es $x_i \in D^d$.
- Sin pérdida de generalida suponemos que los atributos numéricos están en $[-1, 1]$ y los categoricos en $\{1, ..., k\}$.
- 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.

Formalmente

Definition (ϵ - privacidad diferencial local)

Dado $\epsilon > 0$, decimos que M satisface ϵ - privacidad diferencial local si:

$$P[M(x) = x^*] \leq \exp^{\epsilon} P[M(x') = x^*] \tag{3}$$

- Privacidad local es un caso particular de Privacidad, en la que el agregador solicita el registro individual x_i de cada agente i.
- La definición sugiere que cuando hay privacidad local, cuando el agente reporta x^*, el agregador no puede diferenciar si el verdadero regsitro es x_i o x_i' con una confianza alta (i.e., superior a ϵ).
- El objetivo es (1). Mostrar como se puede lograr privacidad local cuando las consultas son valores esperados e histogramas y (2). Se quiere estimar modelos de aprendizaje de máquinas que se calibran minimizando el riesgo empírico.

Formalmente

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

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:

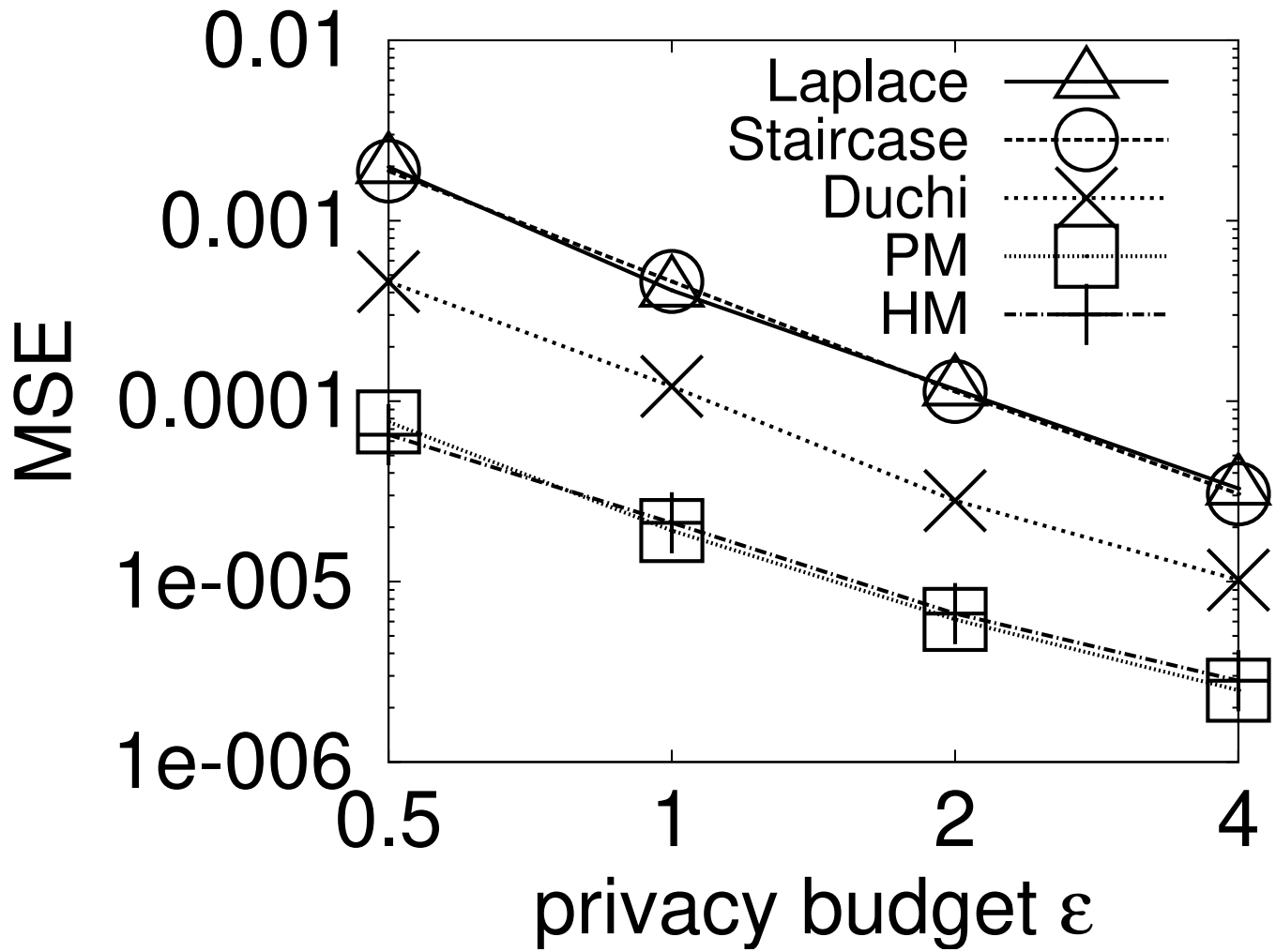
$$L(\beta, \hat{x}_i, y_i) = L(\beta, x_i, y_i) + \frac{\lambda}{2} \mid\mid \beta \mid\mid_2^2 \tag{4}$$

- El método del gradiente desendiente itera.

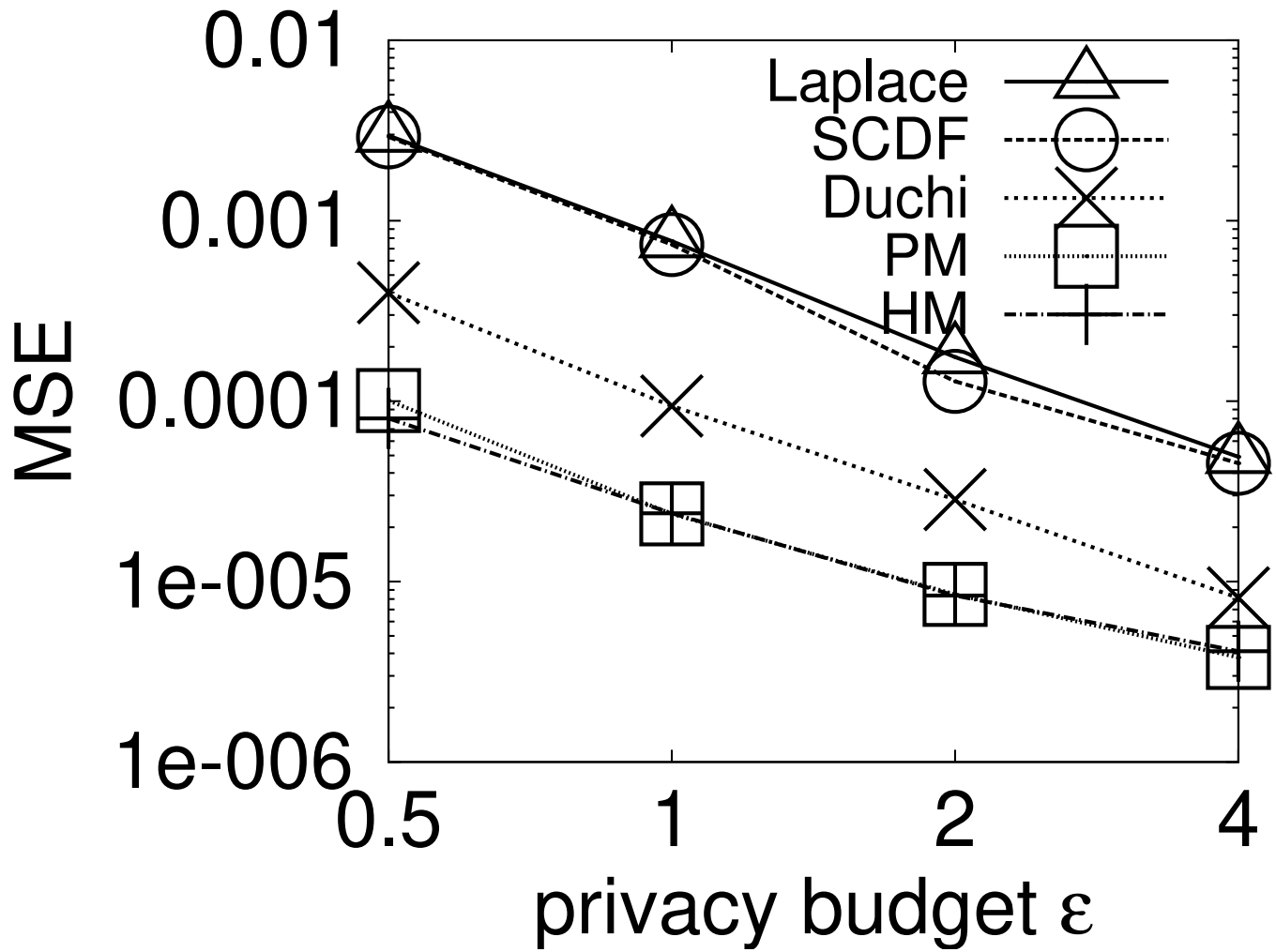
$$\beta_{t+1} = \beta_t - \gamma \nabla \frac{1}{n} \sum_i \hat{L}(\beta, x_i, y_i) \tag{5}$$

- 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.

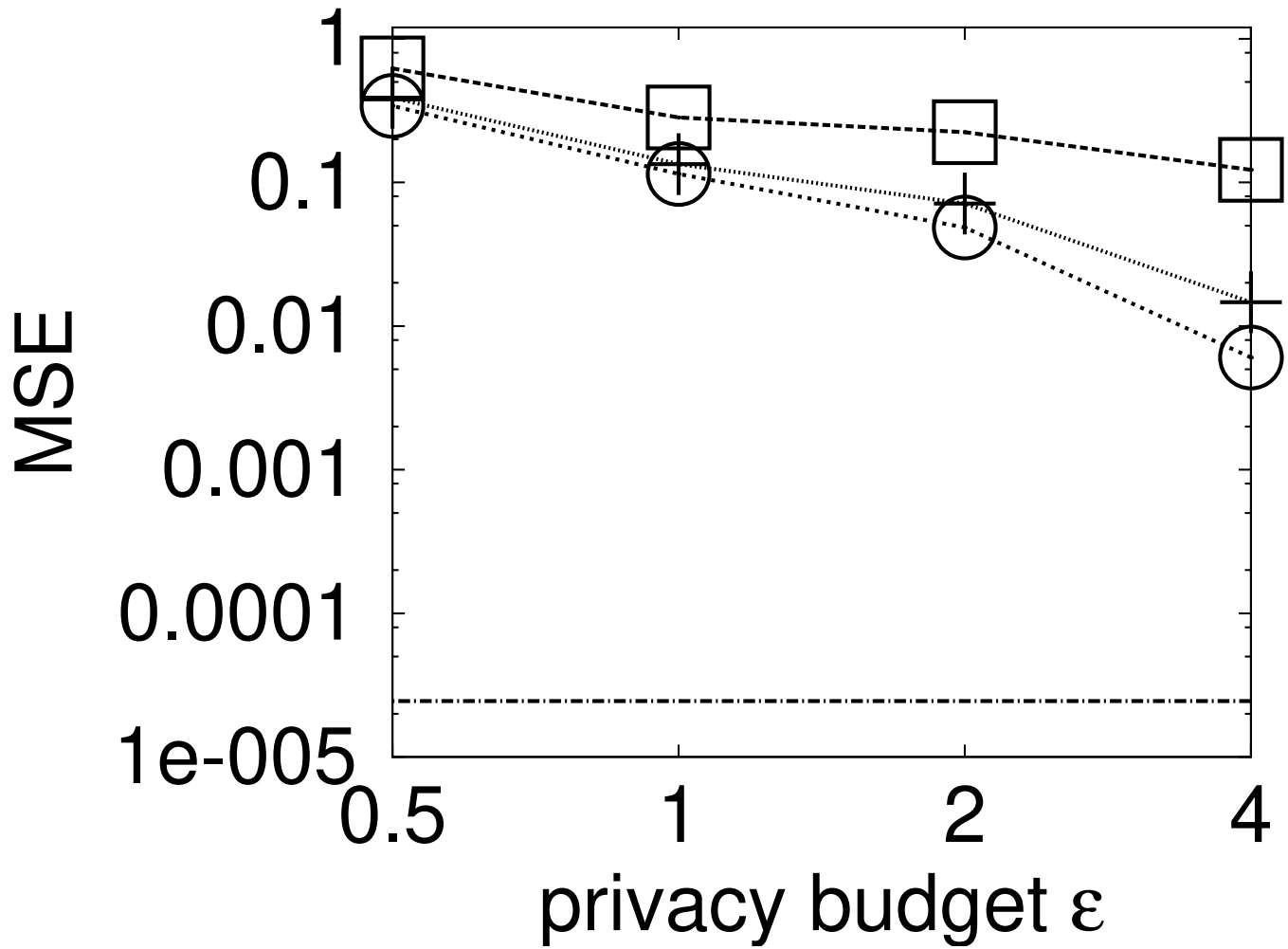
Promedios: Brasil



Promedios: México



Modelos: Brasíl Regresión Lineal



Modelos: Brasíl Logístico