Seguridad, privacidad y aspectos legales

Introducción a la Privacidad Diferencial

Judith Sáinz-Pardo Díaz

Grupo de Computación Avanzada y e-Ciencia
Instituto de Física de Cantabria (IFCA) - CSIC-UC

Máster universitario en ciencia de datos / Master in Data Science

Tabla de contenidos

Resumen: técnicas de anonimización

K-Anonimato

Una base de datos es k-anónima, si para cada fila r hay al menos $k-1$ filas idénticas.

ID	Ciudad	Edad	Causa última hospitalización
1	Santander	20-25	Neumonía
2	Santander	20-25	Apendicitis
3	Santander	20-25	COVID-19
4	Bilbao	25-30	Neumonía
5	Bilbao	25-30	Enfermedad coronaria
6	Oviedo	40-45	COVID-19
7	Oviedo	40-45	COVID-19

K-Anonimato

Una base de datos es k-anónima, si para cada fila r hay al menos $k-1$ filas idénticas.

ID	Ciudad	Edad	Causa última hospitalización
1	Santander	20-25	Neumonía
2	Santander	20-25	Apendicitis
3	Santander	20-25	COVID-19
4	Bilbao	25-30	Neumonía
5	Bilbao	25-30	Enfermedad coronaria
6	Oviedo	40-45	COVID-19
7	Oviedo	40-45	COVID-19

k-anónima con $k = 2$ respecto a los CI ciudad y edad.

L-diversidad

Una base de datos k-anónima es l-diversa con respecto a un AS si para cada clase de equivalencia existen al menos *l* diferentes valores de AS.

ID	Ciudad	Edad	Causa última hospitalización
1	Santander	20-25	Neumonía
2	Santander	20-25	Apendicitis
3	Santander	20-25	COVID-19
4	Bilbao	25-30	Neumonía
5	Bilbao	25-30	Enfermedad coronaria
6	Oviedo	40-45	COVID-19
7	Oviedo	40-45	COVID-19
8	Oviedo	40-45	Apendicitis

L-diversidad

Una base de datos k-anónima es l-diversa con respecto a un AS si para cada clase de equivalencia existen al menos l diferentes valores de AS.

ID	Ciudad	Edad	Causa última hospitalización
1	Santander	20-25	Neumonía
2	Santander	20-25	Apendicitis
3	Santander	20-25	COVID-19
4	Bilbao	25-30	Neumonía
5	Bilbao	25-30	Enfermedad coronaria
6	Oviedo	40-45	COVID-19
7	Oviedo	40-45	COVID-19
8	Oviedo	40-45	Apendicitis

k-anónima con $k = 2$ respecto a los CI ciudad y edad, y l-diversa con $l = 2$.

Input Privacy - Output Privacy

Input Privacy

Input Privacy ("privacidad de las entradas") es una garantía de que una o más personas pueden participar en un proceso de cálculo, de manera que ninguna de las partes sepa nada sobre las entradas del resto.

Son ejemplos de herramientas para garantizar Input Privacy la criptografía de llave pública y la encriptación homomórfica (HE).

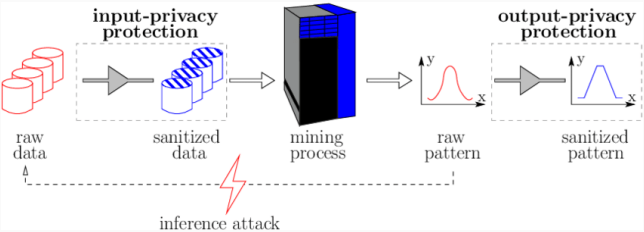
Output Privacy

Output Privacy ("privacidad de la salida") trata de garantizar que la salida de un flujo de información no puede revertirse para aprender atributos específicos sobre la entrada.

Un ejemplo de herramienta para garantizar Output Privacy es la privacidad diferencial (DP).

Idea intuitiva

- Ejemplo input privacy: envío de mensajes usando una clave pública para cifrarlos, y una clave privada para descifrarlos.
- Ejemplo output privacy: en el procesado de imágenes, ocultar el fondo o quitar elementos que puedan revelar información no deseada.



Fuente: Ting Wang and Ling Liu. 2011. Output privacy in data mining. ACM Trans. Database Syst. 36, 1, Article 1 (March 2011), 34 pages. DOI: https://doi.org/10.1145/1929934.1929935

Privacidad Diferencial

Ejemplo: Censo EEUU

- El 81 % de las veces bastan tres datos (CP, sexo y fecha de nacimiento) para identificar a alguien en una base de datos. Evaluador de la probabilidad de localizar a un individuo de EEUU o UK en función de estos tres atributos: `https://cpg.doc.ic.ac.uk/individual-risk/`.
- Este porcentaje llega a un 99,98 % si se tienen 15 datos demográficos de alguien que vive en Massachusetts.

Ejemplo: Censo EEUU

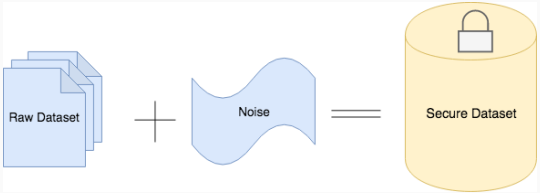
- El 81 % de las veces bastan tres datos (CP, sexo y fecha de nacimiento) para identificar a alguien en una base de datos. Evaluador de la probabilidad de localizar a un individuo de EEUU o UK en función de estos tres atributos: `https://cpg.doc.ic.ac.uk/individual-risk/`.
- Este porcentaje llega a un 99,98 % si se tienen 15 datos demográficos de alguien que vive en Massachusetts.

Ejemplo: Netflix Prize

Objetivo: crear un algoritmo para recomendar a los usuarios las series y películas que mejor se ajustaran a sus preferencias. Investigadores de diferentes universidades demostraron que podían identificar a los usuarios por sus valoraciones de películas usando IMDb.

Motivación - Idea

- La identidad de usuarios de bases de datos que parecen anonimizadas, se puede revelar en algunos casos mediante linkage atacks.
- IDEA: introducir ruido en los datos. Cuanto más ruido se introduzca, más dificil será romper el anonimato. OJO: añadir un excesivo nivel de ruido puede hacer que los datos sean inservibles.
- EJEMPLO: censo de EEUU $\rightarrow$ modificar la edad de ciertos individuos.



Fuente: https://medium.com/secure-and-private-ai-writing-challenge/differential-privacy-e5c7b933ef9e

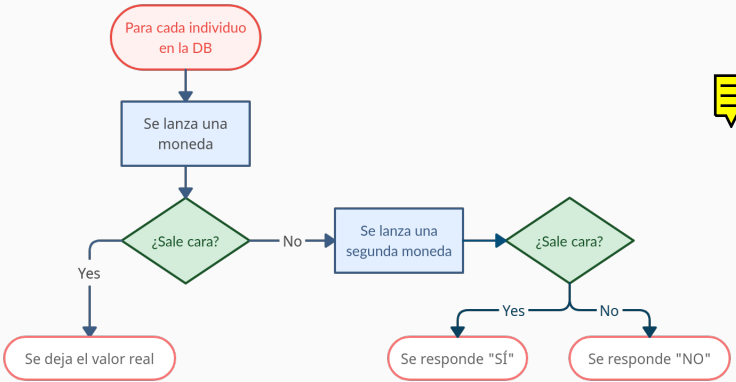
- **EJEMPLO NOTAS:** `https://github.com/judithspd/security_MDS/blob/master/ExampleNotes.ipynb`
- Proteger la privacidad del usuario pero permitiendo que los datos sean significativos para su análisis.
- Hay que seguir estrategias que permitan que el ruido añadido no altere significativamente el análisis, además de contar con un número elevado de datos.

Privacidad Diferencial

La **Privacidad Diferencial (DP)** establece que un algoritmo es diferencialmente privado si viendo su resultado un observador no puede saber si los datos de un individuo concreto están incluidos en la base de datos usada para llegar a dicho resultado.

¿Cómo añadimos ruido a los datos?

Idea intuitiva: lanzamiento de una moneda.
Ante una variable con dos posibles valores (p.e. **SI** o **NO**), el proceso a seguir para asignar un valor para cada individuo que participa en el estudio es el siguiente:



La Privacidad Diferencial no es un algoritmo, es una *definición*.

Definición Formal de Privacidad Diferencial

Un algoritmo probabilista (randomised algorithm) $\mathcal{M}$, con dominio $\mathcal{A}$ y rango $\mathcal{B}$, satisface **ϵ-differential privacy** si para cualesquiera dos entradas adyacentes (difieren en un solo elemento) $a, a' \in \mathcal{A}$ y para cualquier $S \subseteq \mathcal{B}$ se verifica:


$$\mathbb{P}[\mathcal{M}(a) \in S] \leq e^{\epsilon}\mathbb{P}[\mathcal{M}(a') \in S]$$

En la definición anterior el valor de ϵ es el **privacy budget** (presupuesto de privacidad), es decir, permite controlar el nivel de privacidad (la cantidad de pérdida de privacidad permitida).

A menor valor de ϵ, mayor privacidad, pero menor utilidad de los datos para el análisis.

(ϵ, δ)-Privacidad Diferencial

Un algoritmo probabilista (randomised algorithm) $\mathcal{M}$, con dominio $\mathcal{A}$ y rango $\mathcal{B}$, satisface **(ϵ, δ)-differential privacy** si para cualesquiera dos entradas adyacentes (difieren en un solo elemento) $a, a' \in \mathcal{A}$ y para cualquier $S \subseteq \mathcal{B}$ se verifica:

$$\mathbb{P}[\mathcal{M}(a) \in S] \leq e^{\epsilon}\mathbb{P}[\mathcal{M}(a') \in S] + \delta$$

El parámetro δ es la probabilidad de superar el presupuesto de privacidad, es decir, con probabilidad $1 - \delta$ la pérdida de privacidad no será mayor que ϵ.

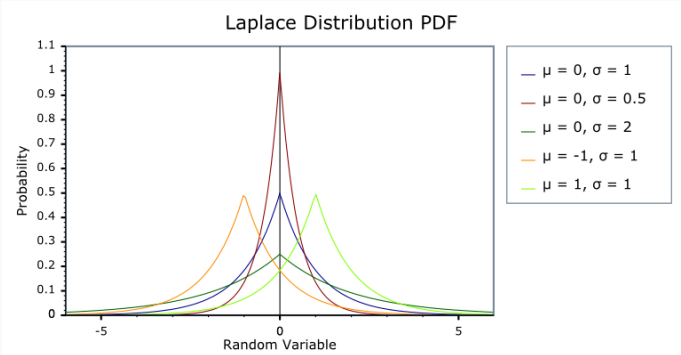
Objetivo:

- **Utilidad:** permitir el "análisis estadístico" de los conjuntos de datos. Es decir, permitir inferir conocimiento sobre los datos, aplicar modelos de ML...
- **Privacidad:** proteger los datos a nivel individual, especialmente ante ataques basados en el uso de información auxiliar.

Interpretación de la definición:

- Todo lo que un adversario aprende sobre mí, también puede aprenderlo a partir de los datos de todos los demás (es decir, la probabilidad de cualquier evento es comparable en los casos en que se incluye o no la información de un cierto individuo en el conjunto de datos).
- Un adversario con capacidad ilimitada de cómputo e información auxiliar no puede romper el nivel de privacidad establecido.

- **Mecanismos para obtener DP:** Mecanismo de Laplace, Mecanismo Exponencial, y Mecanismo Gaussiano entre otros. Información sobre el mecanismo de Laplace: `http://www.gautamkamath.com/CS860notes/lec4.pdf`.
  - Idea intuitiva: en el caso del mecanismo de Laplace, se trata de añadir ruido a partir de la Distribución de Laplace.



Fuente: `https://valelab4.ucsf.edu/svn/3rdpartypublic/boost/libs/math/doc/sf_and_dist/html/math_toolkit/dist/dist_ref/dists/laplace_dist.html`

Ejercicio Privacidad Diferencial

1. Para este ejercicio, vamos a usar la librería de Python *PyDP* (`https://github.com/OpenMined/PyDP`).
2. Descarga los datos diarios de Covid-19 de `https://cnecovid.isciii.es/covid19/` y quédate con los de 2021.
3. Obtén, para cada mes de 2021, la media de los casos diarios, considerando todas las CCAAs, y los casos totales en el mes.
4. Obtén los mismos valores que en el apartado anterior, pero aplicando DP (usa *BoundedMean* y *BoundedSum* de *PyDP*)
5. Compara los resultados obtenidos al usar DP y al usar los datos en bruto. Observa como varían los resultados al cambiar el privacy budget. Ilustra los resultados mediante barplots.

Algunos recursos interesantes

- Friedman, A., Schuster, A. (2010, July). Data mining with differential privacy. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 493-502). https://dl.acm.org/doi/abs/10.1145/1835804.1835868.
- Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., Zhang, L. (2016, October). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security (pp. 308-318). https://dl.acm.org/doi/abs/10.1145/2976749.2978318.
- Harvard University Privacy Tools Project. Differential Privacy. https://privacytools.seas.harvard.edu/differential-privacy.