IEC 2021

Aprenentatge automàtic que preserva la privadesa

Vicenç Torra

Abril 2021

Departament de Ciències de la Computació, Universitat d'Umeå, Suècia

Índex

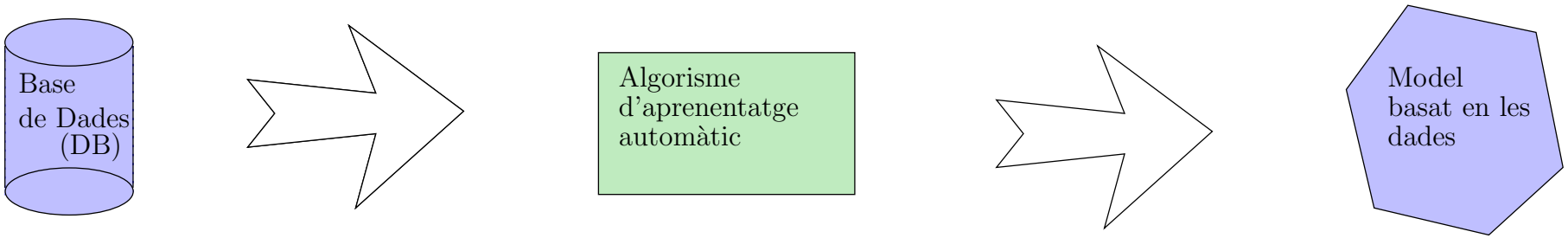
Un context:

Aprenentatge automàtic i models estadístics

Anàlisi de dades i models basats en dades

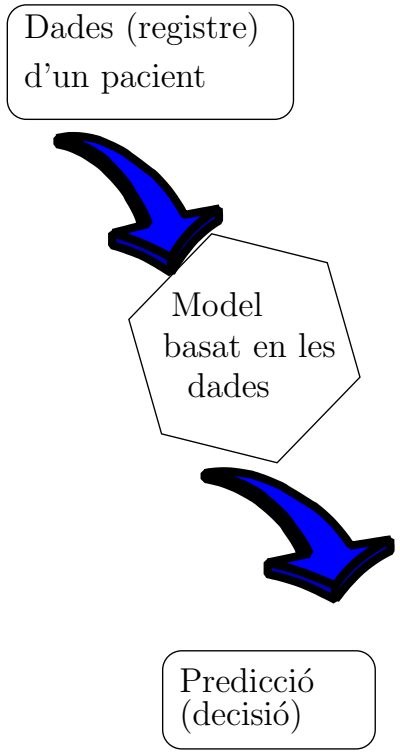
Models basats en les dades

- Construcció d'un model basat en les dades (regressió, regressió logística, xarxes neuronals, regles d'inferència) per predicció, detecció en imatges, suport a la decisió, etc.



Models basats en les dades

- Aplicació d'un model basat en les dades (regressió, regressió logística, xarxes neuronals, regles d'inferència) per predicció, detecció en imatges, suport a la decisió, etc.



Models basats en les dades

- Com apliquem l'aprenentatge automàtic (informal)
  - Accedim a les dades (rellevants, o no)
  - Anàlisi exploratòria de dades.
  - Construïm diversos models (predictors, regressors)
    (considerem diversos tipus de models i diversos parámetres)
  - Seleccionem un model que és bo
    (sigui quina sigui la definició de bo)

- Exemple
  - Predicció de la durada de l'ingrés hospitalari[1]

[1] Domingo A, Trenchs V, Fasheh W, Quintillá J, Caritg J, Luaces C., Bronquiolitis: factors predictius de la durada de l'ingrés hospitalari, Pediatr. Catalana 2005; 65: 77-81. https://dialnet.unirioja.es/servlet/articulo?codigo=5629404
https://www.nature.com/articles/sdata201635

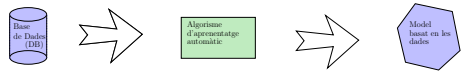
Models basats en les dades

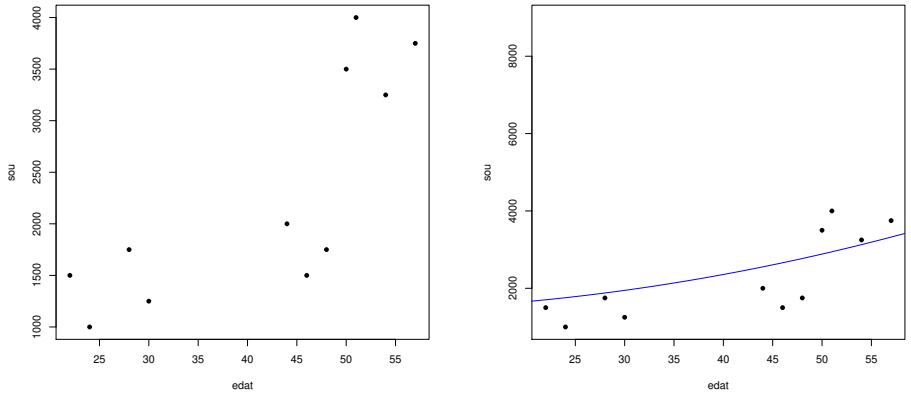
- Com apliquem l'aprenentatge automàtic (informal)
  - Accedim a les dades (rellevants, o no)
  - Anàlisi exploratòria de ades.
  - Construïm diversos models (predictors, regressors)
    (considerem diversos tipus de models i diversos parámetres)
  - Seleccionem un model que és bo
    (sigui quina sigui la definició de bo)

- Exemple
  - Predicció de la durada de l'ingrés hospitalari per bronquiolitis
  - bo: Donat un conjunt de casos amb durada coneguda, que la predicció sigui el més semblant possible a la donada
  - Considerant diversos factors obtindrem diverses prediccions: Prematuritat, presència de virus respiratori sincicial (VRS), edat (mesos), etc.

Models basats en les dades

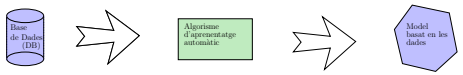
- Construcció d'un model basat en les dades: edat → ingressos

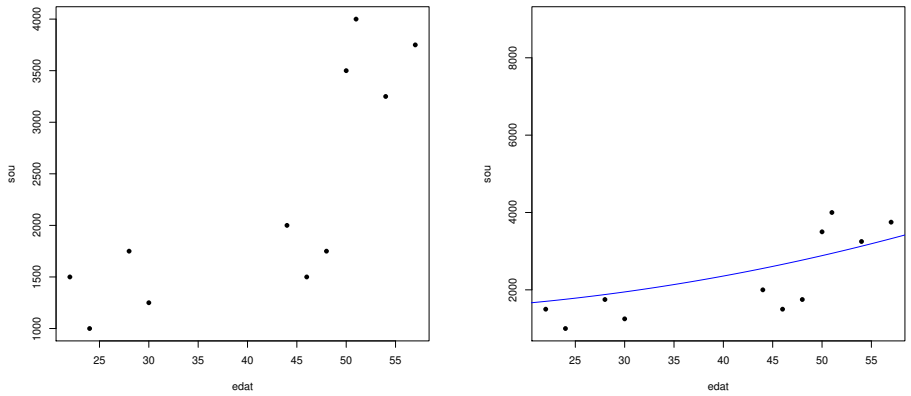


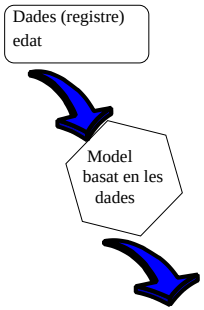


Models basats en les dades

- Construcció d'un model basat en les dades: edat → ingressos







$$\text{sou} = 1418.63 + 0.5864 * \text{edat}^2$$

Quins són els ingressos d'Aina Cohen? (edat=25, ingressos=?)

Anàlisis de dades i models basats en dades

i privadesa

Models basats en les dades i privadesa

- Qüestions rellevants en relació a la privadesa
  - Qui té accés a les dades? $\Rightarrow$ control d'accés
    - ▷ Actors diferents tenen permisos diferents (accés a les dades):
      Admissió hospitalària, personal mèdic, personal d'infermeria, servei de farmàcia (farmacoteràpia individualitzada), laboratori clínic (proves diagnòstiques), etc.
    - ▷ però també
      servei técnic informàtic, analistes/científics de dades

Models basats en les dades i privadesa

- Qüestions rellevants en relació a la privadesa
  - Qui té accés a les dades? $\Rightarrow$ control d'accés
    - ▷ Actors diferents tenen permisos diferents (accés a les dades):
      Admissió hospitalària, personal mèdic, personal d'infermeria, servei de farmàcia (farmacoteràpia individualitzada), laboratori clínic (proves diagnòstiques), etc.
    - ▷ però també
      servei técnic informàtic, analistes/científics de dades

- El control d'accés no és suficient
  - L'accés sembla apropiat, però algunes inferències poden implicar revelació d'informació sensible

Models basats en les dades i privadesa

- Qüestions rellevants en relació a la privadesa
  - De tot allò a què podem accedir, podem inferir-ne res que no hauríem de saber? Per exemple,
    - Podem trobar algú que coneixem a la base de dades (BD) a partir de la informació disponible?
    - Podem descobrir informació confidencial a partir de la nostra visió restringida de la base de dades (o anonimitzada) ?
    - Podem descobrir informació confidencial a partir d'informació agregada ?
    - Podem descobrir informació confidencial a partir d'un model?
  - Si això és possible, que ens cal per evitar-ho? $\Rightarrow$ Privadesa de dades

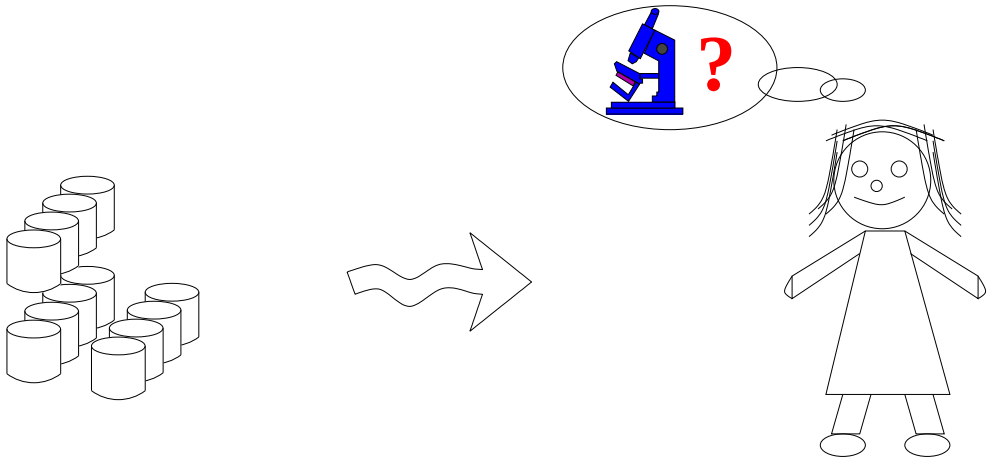
Models de privadesa

Dos exemples esperonadors

L'anonimització és més difícil del que sembla

Dos exemples esperonadors

Exemple #1. La base de dades amb registres (de pacients)



- Base de dades dels pacients: dades d'ingrés (data d'ingrés, durada), variables (antecedent de prematuritat, realització d'analítica).
- Anàlisis: predicció de la durada mitjana, per època de l'any.

Dos exemples esperonadors

Exemple #1. La base de dades amb registres (de pacients)

- Solució. Una anonimització naïf: eliminar noms i targetes sanitàries
- L'anonimització naïf no funciona

Dos exemples esperonadors

Exemple #1. La base de dades amb registres (de pacients)

- Solució. Una anonimització naïf: eliminar noms i targetes sanitàries
- L'anonimització naïf no funciona
  Hi ha informació de les dades *"anonimitzades"* que implica revelació d'informació confidencial

~~Darth Vader~~, Catedral nacional de Washington, Northwest, Washington D.C.

Imatge de la wikipedia

Dos exemples esperonadors

Exemple #1. La base de dades amb registres (de pacients)

- L'anonimització naïf no funciona
- Exemple:
  - Predicció de la durada d'ingrés: base de dades amb

    (any de naixement, població, análisi/malaltia), durada

    2019, Castelló, a, 3 dies
    2020, Castelló, b, 2 dies
    2020, Vila-real, c, 5 dies
    2019, Vila-real, a, 2 dies
    2020, L'Alcora, b, 4 dies
    ...

Dos exemples esperonadors

Exemple #1. La base de dades amb registres (de pacients)

- L'anonimització naïf no funciona
- Exemple:
  - Predicció de la durada d'ingrés: base de dades amb
    (any de naixement, població, análisi/malaltia), durada
    2019, Castelló, a, 3 dies
    2020, Castelló, b, 2 dies
    2020, Vila-real, c, 5 dies
    2019, Vila-real, a, 2 dies
    2020, L'Alcora, b, 4 dies
    ...

    Tanmateix si també hi tenim:
    2020, Figueroles, xxx, 2 dies
    2019, La Foia, xxx, 4 dies [2]
- Divulgació d'informació confidencial
  - Primer, re-identificació (sabem que (2019, Figueroles) és Joana), i
  - després, anàlisi/malaltia = xxx.

[2] Figueroles (l'Alcalatén), població 2020: 523 (font: INE))

Dos exemples esperonadors

L'anonimització naïf no funciona: demostrat en diversos treballs

- Diversos casos de publicació de dades
  - AOL, Netflix (històric de cerques, avaluacions de pel·lícules)
  - 3.7% (9.1/248 milions) de la població dels Estats Units és única donat el codi postal, génere i més i any de naixement.
- De forma similar
  - Posicions de mòbils (on dues posicions et poden identificar: la casa i la feina)
  - targetes de fidelitat, compres amb targetes de crèdits, etc.
  - i més casos[3]
  - En la BD de la botiga de l'Antoni hi ha totes les compres dels clients.

⋆ Dades de grans dimensions + dades fortament identificables

[3] K. El Emam, E. Jonker, L. Arbuckle, B. Malin (2011) A Systematic Review of Re-Identification Attacks on Health Data, PLOS one 6:12 e28071.

Dos exemples esperonadors

Exemple #1. La base de dades amb registres (de pacients)

- Divulgació d'identitat: Trobar-hi algú
  - Trobem la Joana en la BD de l'hospital
  - Trobem la Colometa en la BD de la botiga de l'Antoni

  (pot no tenir cap rellevància en el nostre context)
- Divulgació d'atribut: Aprenem alguna cosa d'algú
  - com ara la malaltía de la Joana
  - o si ha comprat salfumant la Colometa

  (pot no tenir cap rellevància en el nostre context)

Dos exemples esperonadors

Exemple #1. La base de dades amb registres (de pacients)

- Divulgació d'identitat: Trobar-hi algú
  - Trobem la Joana en la BD de l'hospital
  - Trobem la Colometa en la BD de la botiga de l'Antoni

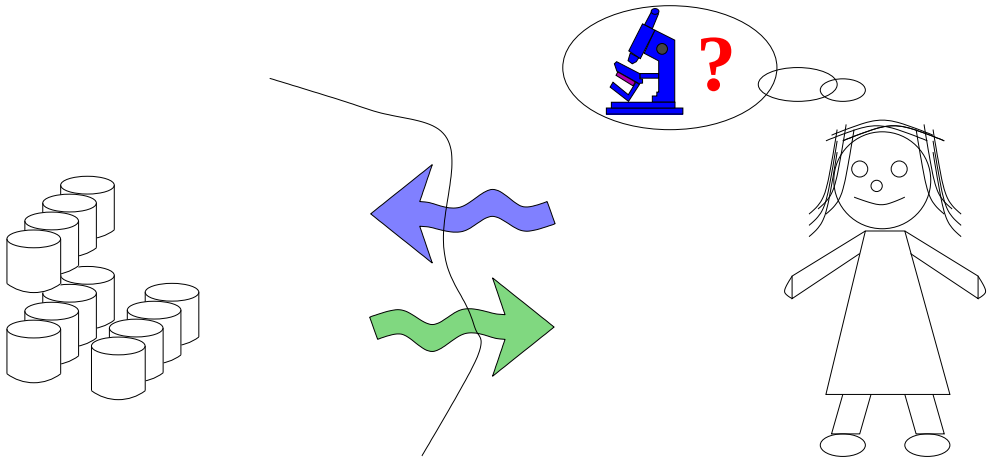
  (pot no tenir cap rellevància en el nostre context)
- Divulgació d'atribut: Aprenem alguna cosa d'algú
  - com ara la malaltía de la Joana
  - o si ha comprat salfumant la Colometa

  (pot no tenir cap rellevància en el nostre context)

- Trobar algú en una BD acostuma a implicar descobrir-ne alguna cosa
  - Divulgació d'identitat implica divulgació d'atribut

Dos exemples esperonadors

Exemple #2. Sou mitjà (o bé, en general, un altre càlcul)



Dos exemples esperonadors

Exemple #2. Sou mitjà

- Els ingressos mitjans són un valor agregat, així doncs, no són dades personals
  Calculem $\sum_{i=1}^{n} x_i/n$

Dos exemples esperonadors

Exemple #2. Sou mitjà

- Els ingressos mitjans són un valor agregat, així doncs, no són dades personals
  Calculem $\sum_{i=1}^{n} x_i/n$
- Però això no funciona .....!!

  'I sense something. A presence I have not felt since . . . '

  (Darth Vader, La guerra de les galàxies IV: Una nova esperança)

- El resultat pot donar-nos informació de qui hi ha a la base de dades
  - El salari mitjà de la unitat de psiquiatria

    $\Rightarrow$ descobrim que hi han atès l'única persona rica del poble

Dos exemples esperonadors

Exemple #2. Sou mitjà

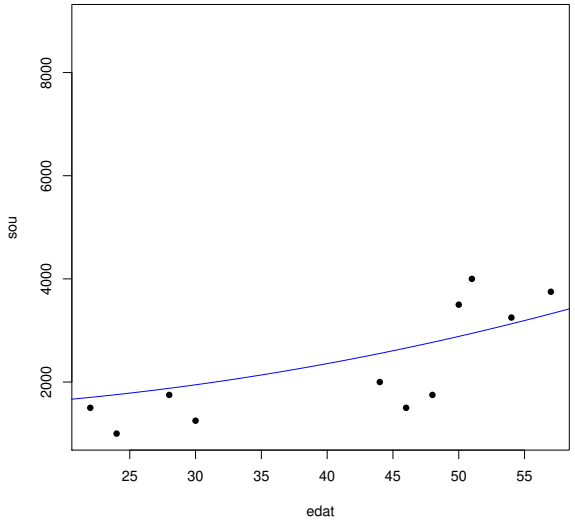
- Q: Ingressos mitjans dels admesos (unitat, població) – psiquiatria
  (similar, salari mitjà per professions en una població)
- Els ingressos mitjans és un valor agregat, així doncs, tot correcte?
  - Exemple:
    1000 2000 3000 2000 1000 6000 2000 10000 2000 4000
    $\Rightarrow$ mitjana = 3300
  - Si afegim els ingressos de Dona Obdúlia de Montcada 100,000
    $\Rightarrow$ mitjana = 12090,90 !
    (un valor molt alt canvia significativament la mitjana)
    $\Rightarrow$ podem inferir que Dona Obdúlia és entre els pacients de la unitat

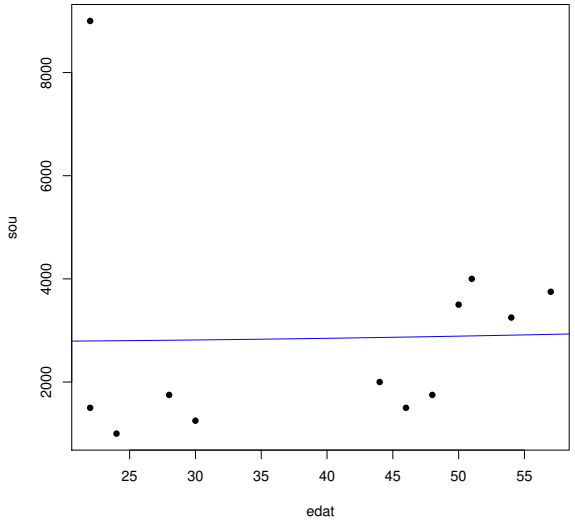
  Obi-Wan Kenobi és a l'Estrella de la Mort

Dos exemples esperonadors

Exemple #2. on un altre càlcul

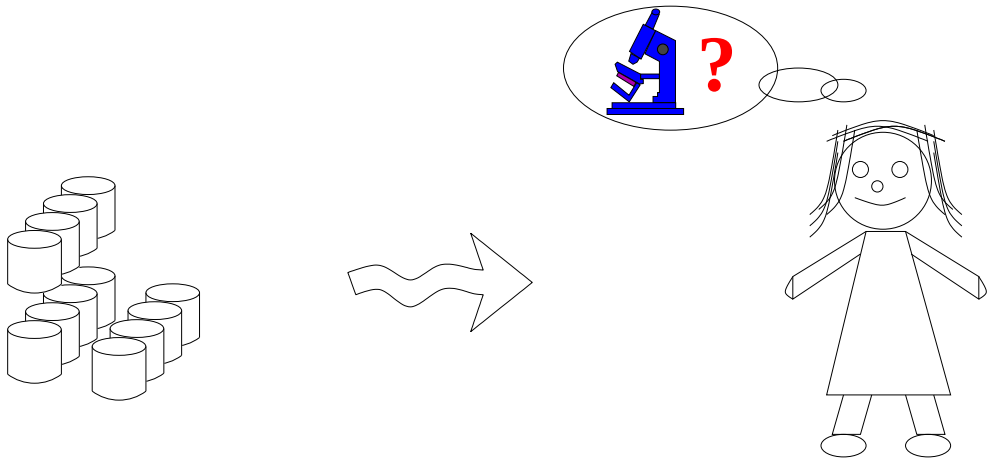
- Q: Les regresions (i altres models) poden patir atacs de pertinença (hi ha Dona Obdúlia a la BD?)





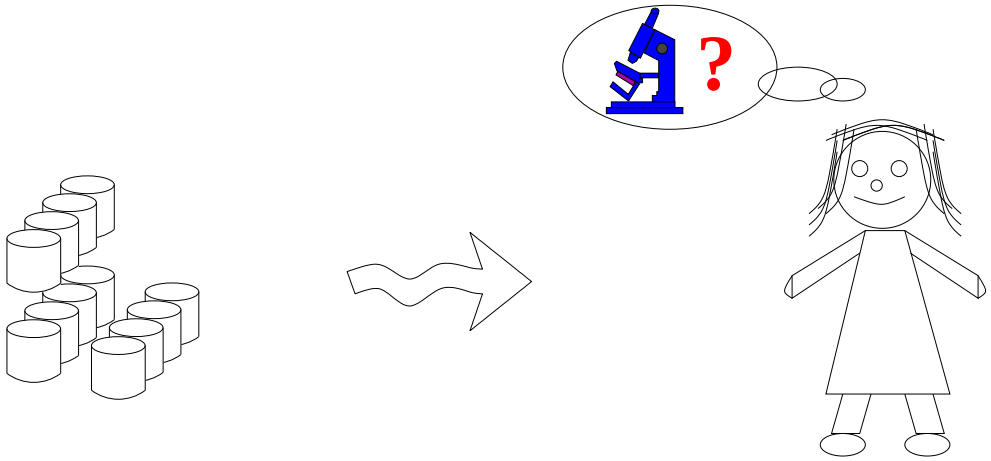
$\text{sou} = 1418.63 + 0.5864 * \text{edat}^2$ vs. $\text{sou} = 2774 + 0.04639 * \text{edat}^2$

Què són els models de privadesa?



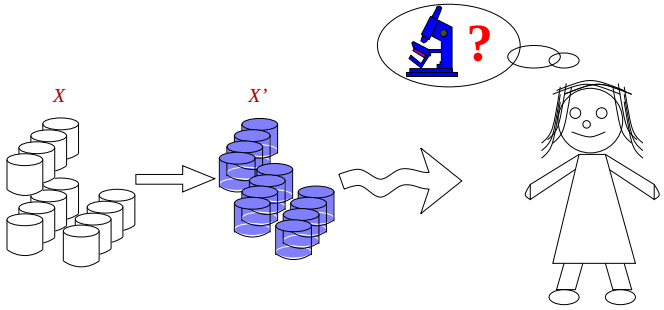
Models de privadesa

Models de privadesa. Definicions computacionals de privadesa



Models de privadesa:

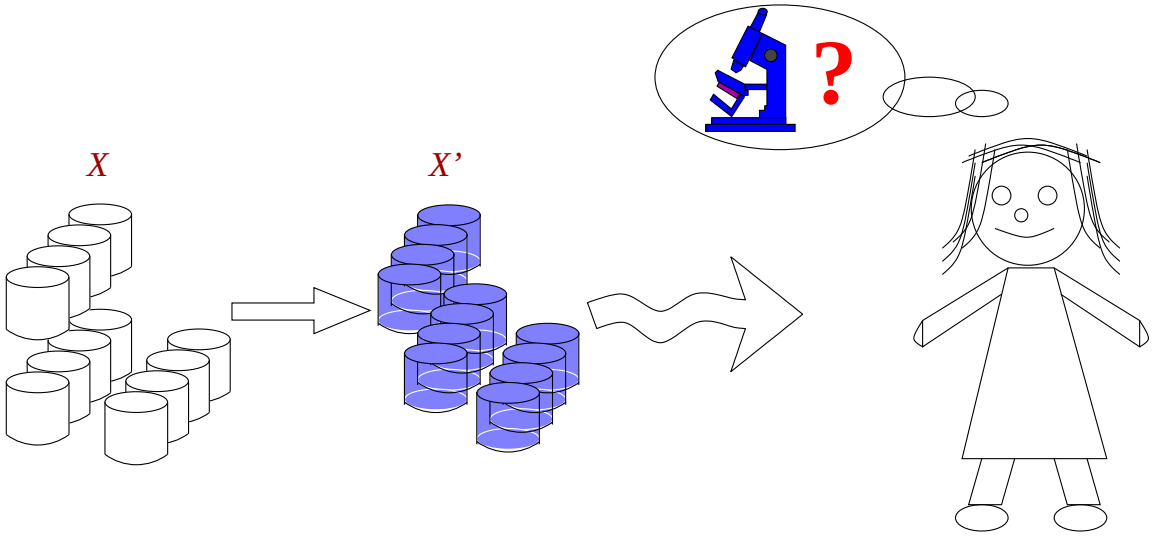
Evitem la reidentificació



Models de privadesa

Models de privadesa. Objectiu: volem evitar la reidentificació

- **Privadesa per la reidentificació.** Evitem trobar un registre
- **k-anonimitat.** Fem cada registre de la BD indistingible d'altres $k-1$



Podem trobar-lo  ? volem evitar-ne la possibilitat ...

Models de privadesa

Models de privadesa. Objectiu: volem evitar la reidentificació

- **Privadesa per la reidentificació.** Evitem trobar un registre
- **k-anonimitat.** Fem cada registre de la BD indistingible d'altres $k-1$

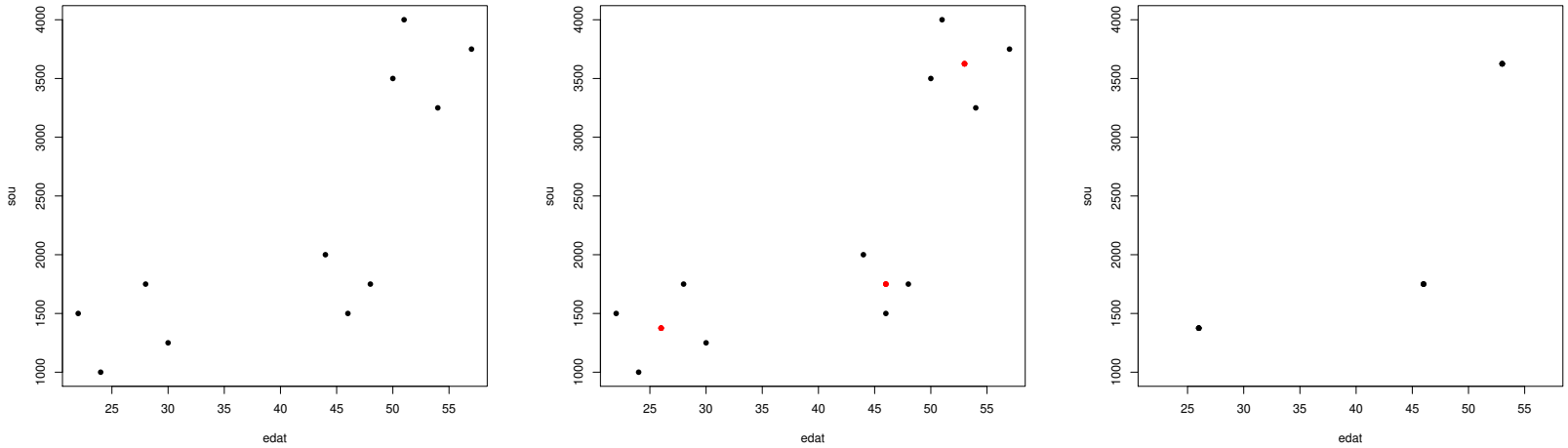
Com ho fem? Canviem el nivell de detall de les dades o hi afegim soroll

- Soroll additiu:
  $x' = x + r$ amb $r \sim N(0, b)$: 2019 $\to$ 2018
- Generalització: $x' = comarca(població(x))$:
  Figueroles $\to$ l'Alcalatén
- Microagregació:
  Fem uns grups de mida mínima i en publiquem la mitjana
- ...

Microagregació

Mecanisme de protecció. Microagregació. grups: mínim k registres

- **Model de privadesa.** k-Anonimitat ($k = 3$)



Base de dades: (edat, sou)

- Agrupació original: {(22,1500), (24,1000), (28, 1750), (30, 1250)}
- Dades protegides: {(26, 1375),(26, 1375),(26, 1375),(26, 1375)}

- **Formalització.** $u_{ij} = 1$ sii x_j en el i-èssim grup; v_i centroide

  Minimitzar $SSE = \sum_{i=1}^{g} \sum_{j=1}^{n} u_{ij} (d(x_j, v_i))^2$

  Subjecte a $\sum_{i=1}^{g} u_{ij} = 1$ for all $j = 1, \ldots, n$

  $2k \geq \sum_{j=1}^{n} u_{ij} \geq k$ for all $i = 1, \ldots, g$

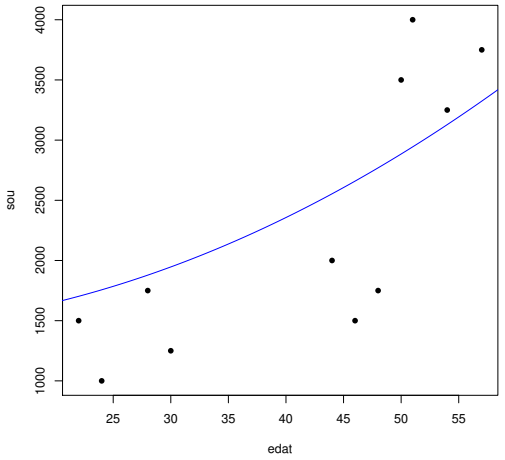
  $u_{ij} \in \{0, 1\}$

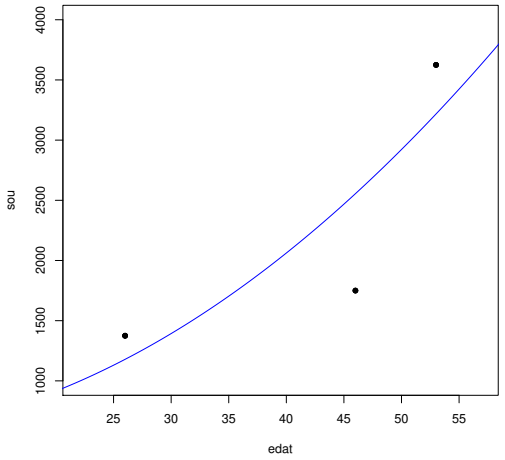
Microagregació

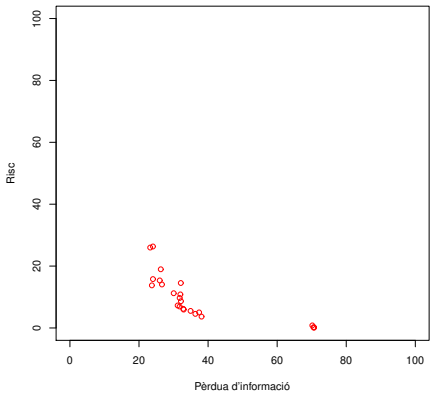
Mecanisme de protecció. Microagregació. grups: mínim k registres

- Les agrupacions protegeixen l'anonimat de les dades, però volem preservar-ne també la utilitat

Podem seguir inferint el sou d'Aina Cohen? (edat=25, ingressos=?)



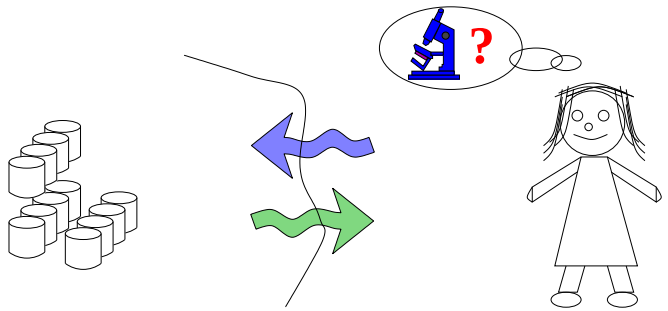




Microagregació difusa. Els límits de les agrupacions no són nítids, podem assignar un registre a diverses agrupacions,
i podem reduir la influencia de les dades atípiques
(els ingressos de Dona Obdúlia de Montcada)

Models de privadesa:

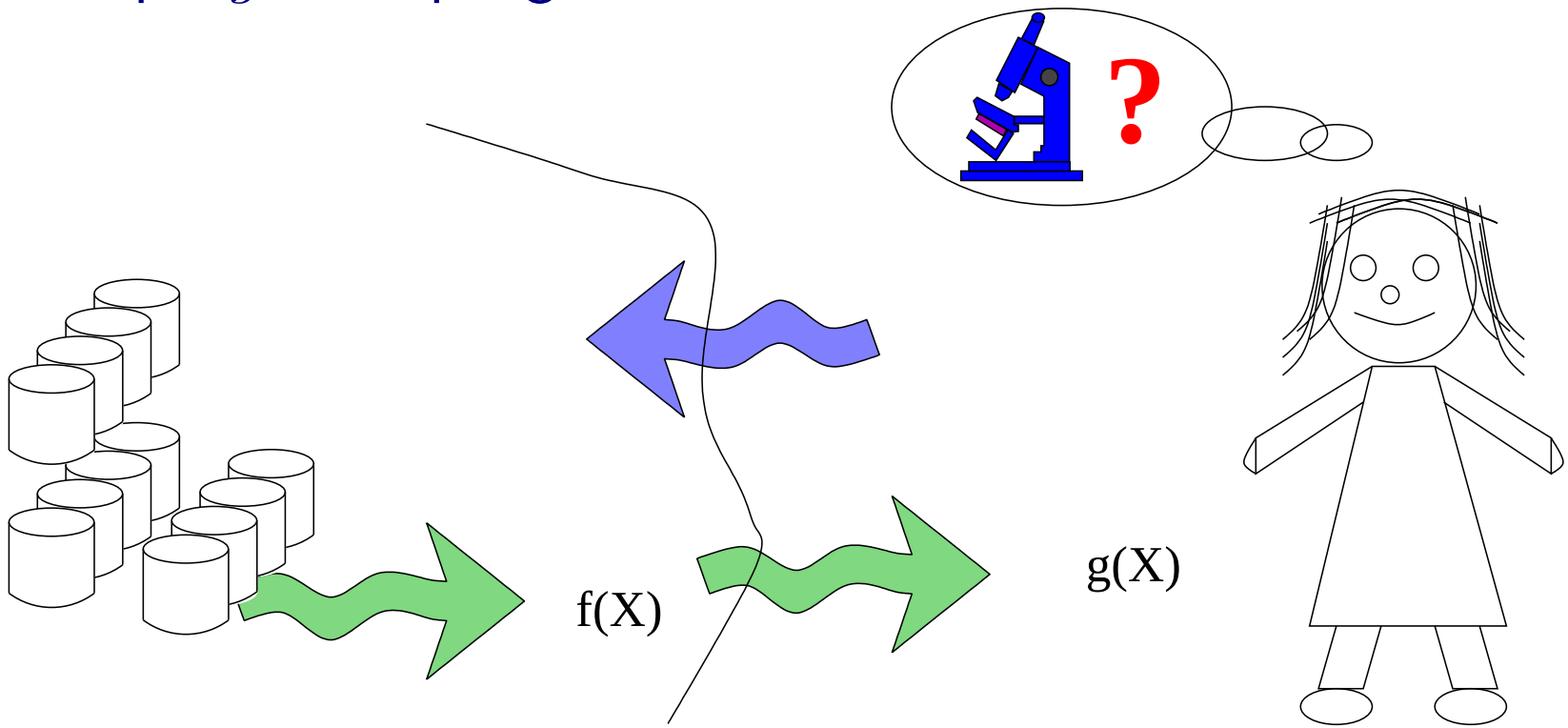
Evitem inferències a partir dels càlculs



Models de privadesa

Models de privadesa. Volem evitar les inferències a partir d'un càlcul

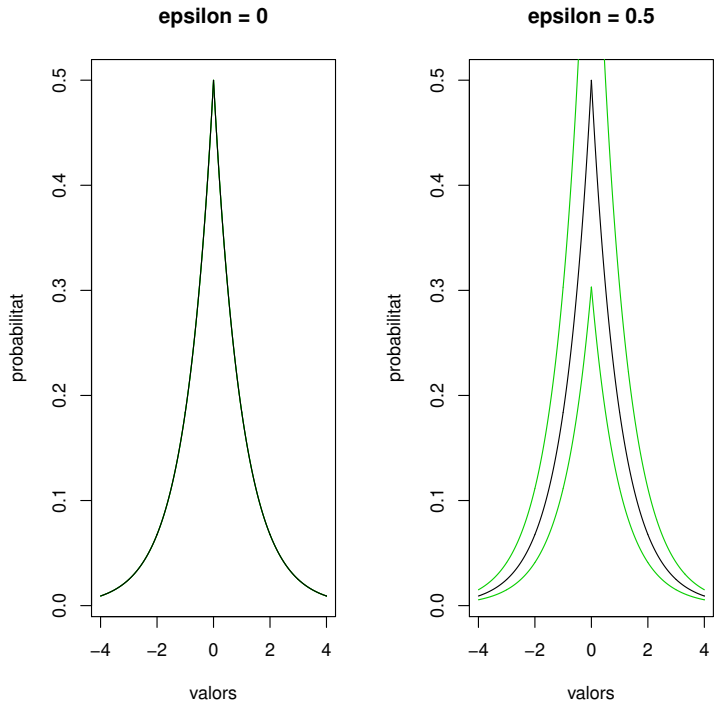
- **Privadesa diferencial.**
  El resultat no depèn (massa) de la presència (o no) d'un registre
- Implementació: en lloc de calcular $f(X)$ calculem $g(X)$,
  i per tal que g no depengui massa de l'entrada hi introduïm soroll



Models de privadesa

Models de privadesa. Volem evitar les inferències a partir d'un càlcul

- **Privadesa diferencial.**
  El resultat no depèn (massa) de la presència (o no) d'un registre.
- Implementació: en lloc de calcular $f(X)$ calculem $g(X)$,
  típicament $g(X) = f(X) + r$ amb $r \sim L(0, b)$ (distribució Laplace)



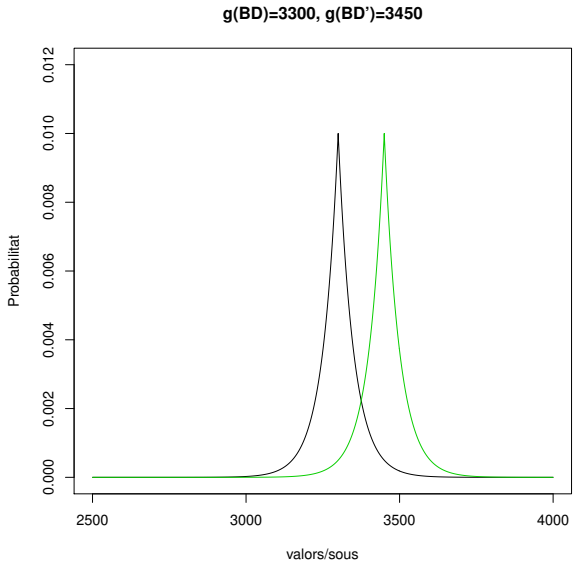
Definició. El resultat $g(D)$ satisfà privadesa diferencial en grau ϵ si per a qualsevol BD_1 i BD_2 tenim que per tot $S \subseteq Range(K_q)$,
$Pr[K_q(BD_1) \in S] \leq e^{\epsilon} Pr[K_q(BD_2) \in S]$

- Com més petit és ϵ més s'han d'assemblar les dues distribucions

Models de privadesa

Models de privadesa. Volem evitar les inferències a partir d'un càlcul

- **Privadesa diferencial.** Implementació
  - Definim $g(X) = f(X) + r$ amb $r \sim L(0, b)$ (distribució de Laplace)
  - Exemple amb $f(BD) = 3300$ i $f(BD') = 3450$, amb distribució de Laplace $L(0, 50)$



- El valor de b a $L(0, b)$ depèn de ϵ (el nivell de privadesa) i la sensitivitat de la funció f a les base de dades possibles

Models de privadesa

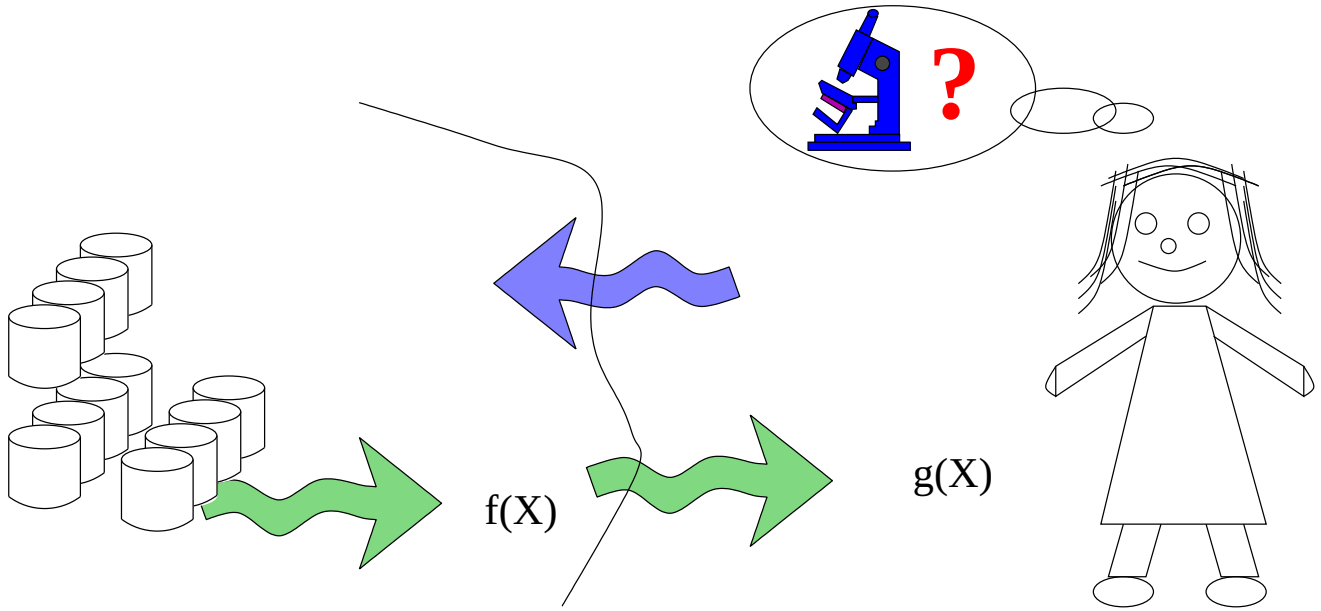
Models de privadesa. Volem evitar les inferències a partir d'un càlcul

- **Privadesa diferencial.**
  Existeixen altres mecanismes per les funcions no numèriques
  i, per exemple, per xarxes neuronals / deep learning, arbres de decisió
- Les solucions són robustes a atacs de pertinença
  (recordem Dona Obdúlia de Montcada!)

Models de privadesa

Models de privadesa. Volem evitar les inferències a partir d'un càlcul

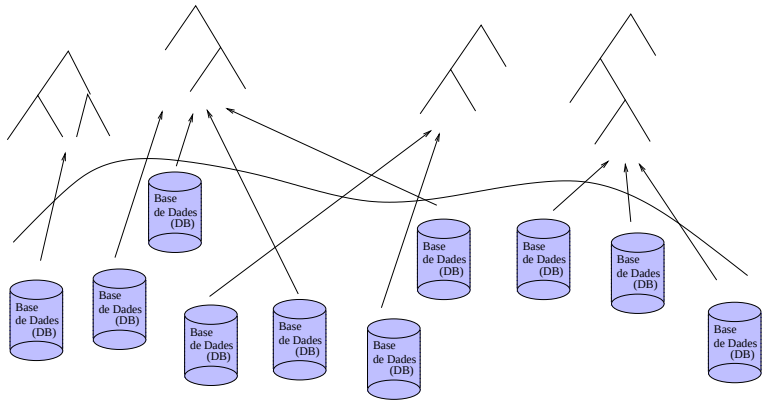
- **Privadesa integral.**
  El resultat és un resultat recurrent
  - Diverses base de dades ens poden proporcionar el mateix resultat
- Privadesa:
  - k bases de dades que generen el mateix resultat (k-anonimitat)
  - negació plausible: jo no hi era allà – Diu Dona Obdúlia



Models de privadesa

Models de privadesa. Volem evitar les inferències a partir d'un càlcul

- **Privadesa integral.**
  El resultat és un resultat recurrent



Definició. El resultat $G = g(D)$ satisfà privadesa integral donat un coneixement pre si $Gen(G, S^*)$ és gran (k BDs) i $\cap_{g \in Gen^*(G,S^*)} g = \emptyset$.
on $Gen(G, S^*) = \{S' | S^* \subseteq S' \subseteq P, A(S') = G\}$
$Gen^*(G, S^*) = \{S' \setminus S^* | S^* \subseteq S' \subseteq P, A(S') = G\}$

- k bases de dades diferents,
  no compartint registres (i suficientment diferents)
  per evitar atacs de pertinença

Models de privadesa

Models de privadesa. Volem evitar les inferències a partir d'un càlcul

- **Privadesa integral.**

  El resultat és un resultat recurrent
  - En l'exemple

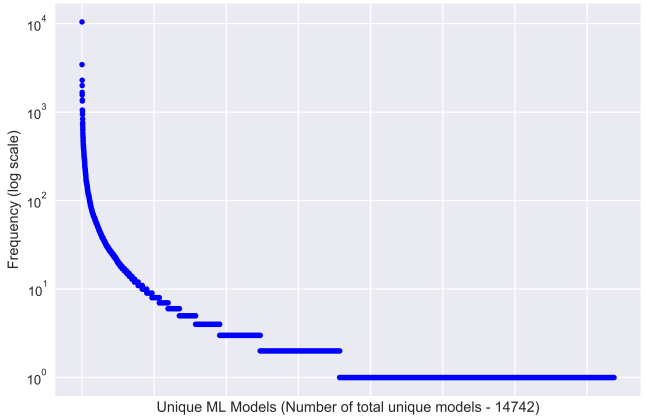
$$\{1000, 2000, 3000, 2000, 1000, 6000, 2000, 10000, 2000, 4000\} \cup \{100000\}$$

  - Diversos subconjunts donen el mateix resultat: mitjana de 3000
    - $\{3000\}$
    - $\{2000, 4000\}$
    - $\{6000, 2000, 1000\}$
    - $\{10000, 1000, 1000, 3000\}$
    - $\{6000, 4000, 1000, 2000, 3000, 2000\}$

Models de privadesa

Models de privadesa. Volem evitar les inferències a partir d'un càlcul

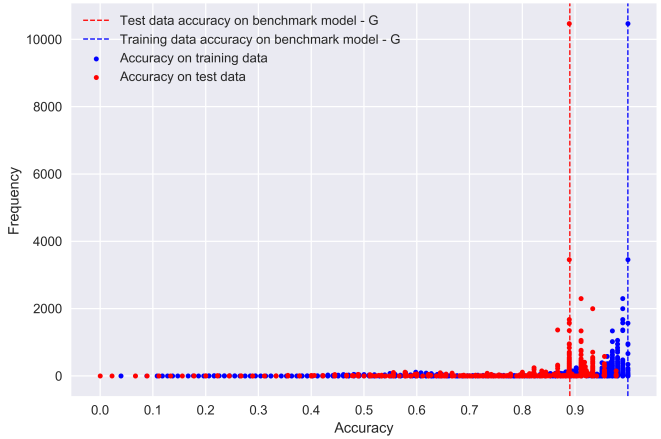
- **Privadesa integral.**
  El resultat és un resultat recurrent
- Els models recurrents també apareixen en aprenentatge automàtic
- Arbres de decisió construïts a partir d'un conjunt de dades (Iris dataset). Models/freq.



Models de privadesa

Models de privadesa. Volem evitar les inferències a partir d'un càlcul

- **Privadesa integral.**
  El resultat és un resultat recurrent
- Els models recurrents també poden tenir bona exactitud
- Arbres de decisió construïts a partir d'un conjunt de dades (Iris dataset). Exactitud/freq.



Models de privadesa

Models de privadesa. Volem evitar les inferències a partir d'un càlcul

- Privadesa diferencial, funció suau
$$f(D) \sim f(D \oplus x)$$
on $D \oplus x$ representa afegir un registre x a la base de dades D
- Privadesa integral, funció recurrent
Si $f^{-1}(G)$ és el conjunt de totes les bases de dades (reals) que poden generar G, exigim que $A^{-1}(G)$ sigui un conjunt gran i divers.

Models de privadesa

Models de privadesa. Volem evitar les inferències a partir d'un càlcul

- Privadesa diferencial, funció suau
$$f(D) \sim f(D \oplus x)$$
on $D \oplus x$ representa afegir un registre x a la base de dades D
- Privadesa integral, funció recurrent
Si $f^{-1}(G)$ és el conjunt de totes les bases de dades (reals) que poden generar G, exigim que $A^{-1}(G)$ sigui un conjunt gran i divers.

- Un exemple de funció simple que satisfà privadesa diferencial:
A és un algorisme que és 1 si el nombre de registres a D és parell i 0 si és senar.
Això és, $f(D) = 1$ si i només si $|D|$ és parell.

Resum

Resum

- Aconseguir una bona anonimització és un desafiament (si volem que les dades siguin útils, és clar)
- És possible aconseguir dades i models protegits que siguin útils i que mantinguin un cert nivell de privadesa.

Gràcies

* Agraïments a Guillermo Navarro-Arribas i Navoda Senavirathne

Referències

Referències relacionades (meves).

- V. Torra (2017) Data Privacy: Foundations, New Developments and the Big Data Challenge, Springer.
- N. Senavirathne, V. Torra (2019) Integrally private model selection for decision trees. Comput. Secur. 83: 167-181
- V. Torra (2020) Fuzzy clustering-based microaggregation to achieve probabilistic k-anonymity for data with constraints, J. Intell. Fuzzy Syst. 39(5): 5999-6008.
- V. Torra, G. Navarro-Arribas, E. Galván (2020) Explaining Recurrent Machine Learning Models: Integral Privacy Revisited. Proc. PSD 2020: 62-73
- http://ppdm.cat/dp/