統計的機械学習
プライバシー保護

教授：中川裕志

助教：佐藤一誠、荒井ひろみ

大学院生（数理情報学専攻）：博士1名、修士4名

その他研究員を含め合計12名のメンバーがいます。

中川研究室
情報基盤センター　5階　504室

中川教員室
情報基盤センター　3階　314室

工学部6号館

授業　など

- 工学部
  - 数理手法IV「統計的機械学習入門」
  - 3,4年生夏学期　水曜日5限 今日開講！

- 大学院　情報理工学系研究科　数理情報学専攻
  - 数理情報学輪講　冬学期担当
  - 言語情報科学　隔年　冬学期　火曜日5限（今年は休講）

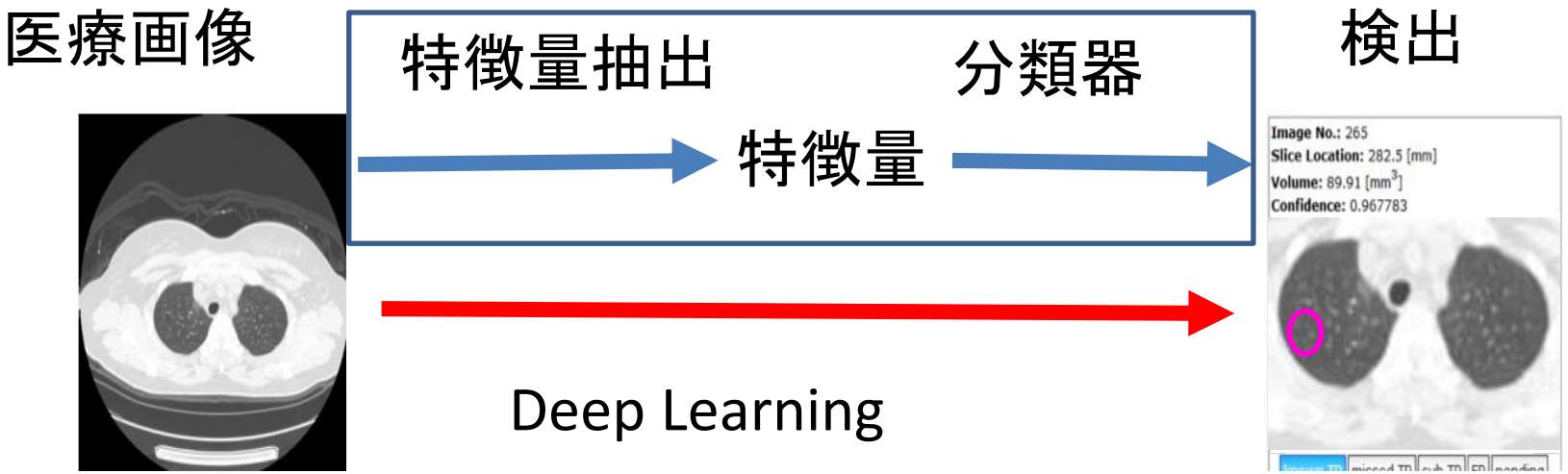
- 中川研に興味のある方はいつでも見学に来てください
  - 研究室HPのURLは「中川裕志」でググるとトップに出てきます！
  - nakagawa@dl.itc.u-tokyo.acにメールをしていただけると助かります。

本年度の研究： 数理モデル

- 最適化問題を利用する機械学習
  - オンライン学習
    - 1データ毎に学習し分類器を更新
  - Multi-Armed Bandit
    - 複数の選択肢のうち選んだ1個だけの成功失敗がわかる。
    - 広告の最適化に応用

Deep Learningによる医療画像識別のための特徴量検出

- 現在はハンドメイドで”有用そうな”特徴量が作られている
- この作業を自動化or高精度化したい。
- 病変データから分類に適した特徴量を自動で学習したい。



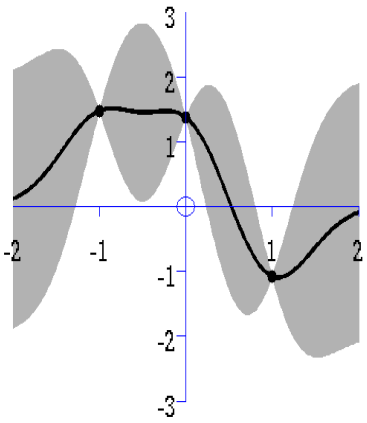
BO(Baysian Optimization)

ブラックボックス関数の最大化を解く方法

- ハイパーパラメータのチューニングが最適なパラメータ探索の計算時間がかかりとても大変

- BOはこの探索を組織的に効率よく行う方法

- Gaussian Proccess
- 関数からサンプルした点が同時ガウス分布に属すると仮定
- 関数が取りうる上界が大きいものを選んでいく



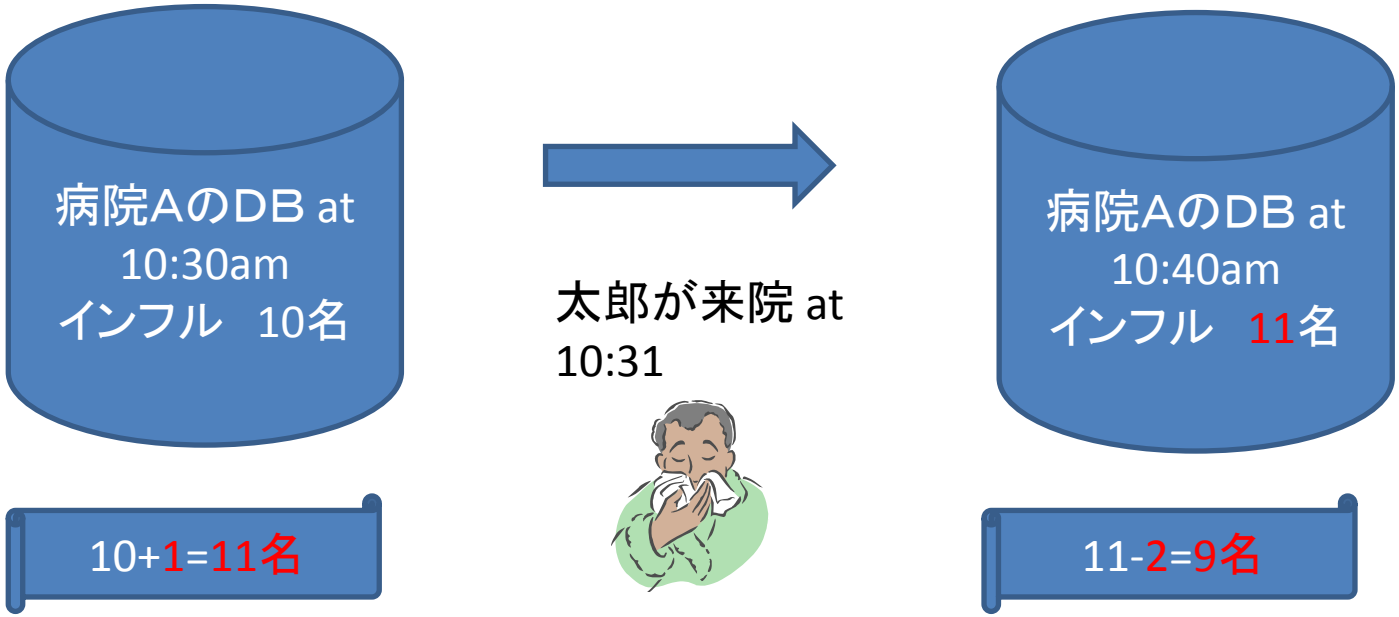
本年度の研究： プライバシー保護技術

- ビッグデータを利活用
  - そのうちでも個人データ（購買履歴など）は宝の山？
  - でも、個人データなので、プライバシーを保護しないと利用できません
- 3種類の方法があります。
  - データベースを改変する方法（k-匿名化）
  - データベースへの質問への応答に雑音加算する方法（差分プライバシー）
  - 暗号を使う方法（準同型公開鍵暗号：暗号化したまま加算ならできる）

本年度の研究： プライバシー保護技術

- ビッグデータを利活用
  - そのうちでも個人データ（購買履歴など）は宝の山？
  - でも、個人データなので、プライバシーを保護しないと利用できません
- 3種類の方法があります。
  - データベースを改変する方法（k-匿名化）
  - **データベースへの質問への応答に雑音加算する方法（差分プライバシー）**
  - 暗号を使う方法（準同型公開鍵暗号：暗号化したまま加算ならできる）

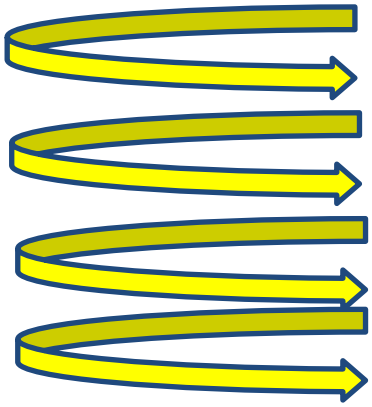
差分プライバシーの有効なケース



DBにインフル患者数を質問し、10:30am で10，10:40am で11と分かると、太郎の来院を知っていた人は太郎がインフルだと分かってしまう。
→ 質問への結果に雑音加算すると回避できる。

差分プライバシーの弱点

攻撃者がDBへ同じ質問
（インフル患者数）を繰り返すと



9

12

8

11

平均すると　(9+12+8+11)/4=10

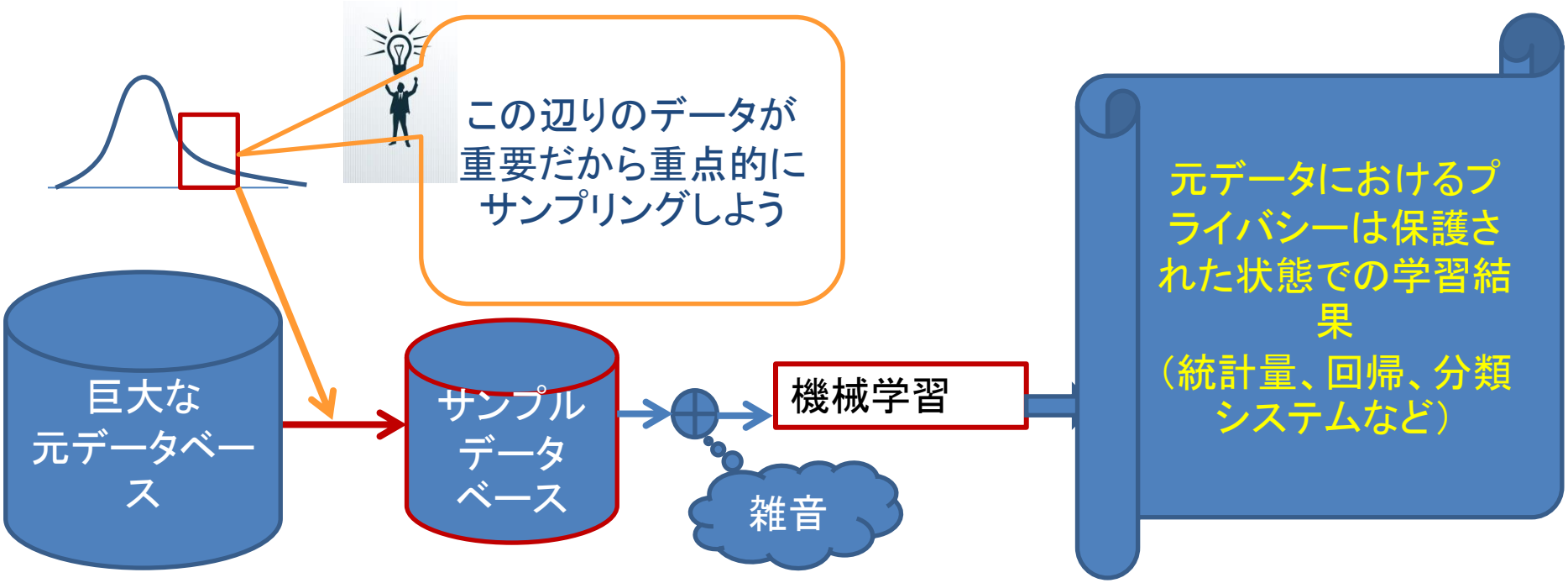
対策1．質問監査
同じ質問が来たら回答を拒否
ただし、同じ質問が過去に来たかどうかのチェックの計算量が大きい

対策2．バイアスのある雑音
期待値＝0ではない雑音の加算
→ データマイニングの精度とのトレードオフ

同じ質問を繰り返してもε-privacyが保てるようにするには、非常に大きな雑音を加算しないとならない→データマイニング精度の低下

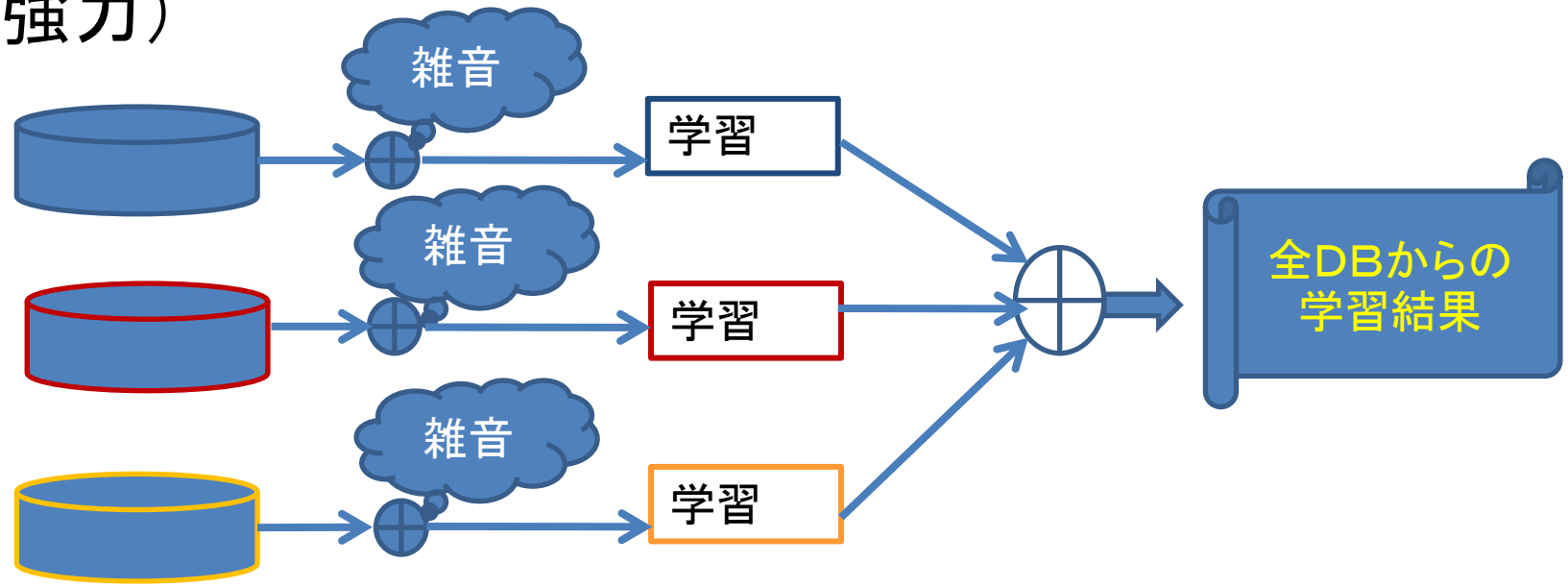
機械学習と差分プライバシーの組み合わせ
重点サンプリング

- 元のデータベースから目的に役立つデータが存在しそうな部分を狙って重点的にサンプルを収集する方法。そのサンプル結果に差分プライバシーの方法で雑音を加えて機械学習する方法を研究します。



機械学習と差分プライバシーの組み合わせ 分散データベースへの応用

- データベースの各レコードに差分プライバシーの方法で雑音を加算してから機械学習する方法
  - 分散したデータベースに適用（個人情報保護が強力）



見学歓迎します

- メールで連絡いただければ日程調整します。
- 研究内容は中川のホームページ
  - http://www.r.dl.itc.u-tokyo.ac.jp/
- あるいはスライドシェアの一覧表(上から辿どれます）
  - http://www.r.dl.itc.u-tokyo.ac.jp/node/57

- 一度、のぞいてみてくださいね。