Trabalho 3
Privacidade Diferencial – Mecanismo de Laplace

Javam Machado

Outubro 2019

1 Objetivo

- Implementar o mecanismo de Laplace, segundo a privacidade diferencial. Realizar consultas sobre uma base de dados através do mecanismo de Laplace implementado e fornecer os resultados de forma correta e privada.

2 Especificação

- Carregue o conjunto de dados "adult.csv", contendo dados de indivíduos capturados pelo censo estadunidense de 1994.
- A lista de atributos e mais informações dos dados estão na seguinte URL: https://archive.ics.uci.edu/ml/datasets/Adult
- As consultas deverão respeitar os seguintes $\varepsilon = \{0.1, 1, 10\}$.
- A totalidade do orçamento de privacidade *(privacy budget – ε)* deverá ser igualmente distribuída para todas as consultas. Por exemplo: para $\varepsilon = 0.1$, esse valor deverá ser distribuído igualmente entre as consultas.
- Consultas a serem realizadas sobre a totalidade dos dados:
  1. Q1: Média do atributo *age*;
  2. Q2: Somatório do atributo *capital-gain*;
  3. Q3: Média do atributo *hours-per-week*;
  4. Q4: Quantidade de pessoas com $income > 50K$.
- Assume-se que o conjunto de dados utilizado é o universo de dados possíveis, sendo assim possível calcular a sensibilidade global de forma fácil.

3 Requisitos

- Linguagem: Python
- Meio de entrega: criar um repositório chamado "disciplina_privacidade_2019" no Github e compartilhar com os seguintes e-mails: {andre.luis, iago.chaves, israel.vidal, javam.machado}@lsbd.ufc.br. **Todos os trabalhos da disciplina serão entregues através desse repositório**.
- Criar uma pasta "dp_laplace" no repositório "disciplina_privacidade_2019".
- **As respostas devem estar disponíveis através do "result.csv"**, onde esse arquivo apresenta as seguintes colunas: budget, result_q1, result_q2, result_q3, result_q4, sens_q1, sens_q2, sens_q3, sens_q4. As respostas das consultas para cada budget exigido e suas respectivas sensibilidades deverão estar presentes nesse arquivo.
- Equipes de até 2 pessoas.
- Somente um repositório deve ser criado por equipe.
- O arquivo "README.md" no repositório deve conter os componentes da equipe.

4 Avaliação

- Na avaliação será considerada a:
  1. Privacidade;
  2. Corretude do algoritmo;
  3. Corretude das sensibilidades;
  4. Apresentação da equipe;
  5. Qualidade e legibilidade do código.
  6. ***Bônus:*** Plotar a distribuição da Laplace de cada questão.

5 Entrega

- 05 de Novembro de 2019. *Commit* do código até 13:59.