Trabalho 5
Privacidade Diferencial Local

Javam Machado

Novembro 2019

1 Objetivo

- Implementar Mecanismos Locais Diferencialmente Privados (protocolos) mais adequados para as consultas, com o objetivo de obter uma melhor utilidade. **Atenção:** este trabalho não é de caráter obrigatório.

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 respostas deverão respeitar os seguintes $\varepsilon = \{1, 2, 5, 10\}$.
- Utilize o budget completo para cada uma das respostas, não sendo necessário dividi-lo entre elas.
- Respostas a serem enviadas por cada usuário de maneira diferencialmente privada:
  1. R_1: valor do atributo *age*. Por conta da alta dimensionalidade deste atributo, utilize uma estratégia de *buckets* para redução de dimensionalidade. Utilize um total de 10 *buckets*;
  2. R_2: Valor do atributo *education*;
  3. R_3: Valor do atributo *sex*;
  4. R_4: Valor do atributo *native-country*.
- Para extrair o domínio dos atributos, analise a documentação no link acima e, se necessário, analise os dados.

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_local" no repositório "disciplina_privacidade_2019".
- Para cada resposta R_i, gerar um arquivo de saída *respostas_r_i.csv*, contendo a resposta diferencialmente privada de todos os usuários no dataset para a resposta R_i.
- Para cada conjunto de respostas, gerar um arquivo *estimativa_r_i.txt*, contendo a estimativa dos valores do domínio do atributo. Por exemplo, para R_3, criaria-se o arquivo *estivativa_r_3.txt* contendo as frequências estimadas de *Female* e de *Male*.
- 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 dos protocolos implementados;
  3. Escolha dos protocolos para cada resposta;
  4. Apresentação da equipe;
  5. Qualidade e legibilidade do código.

5 Entrega

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