National Institute of Technology

Warangal

Lab 7 Assignment

Name: Ayush Rana

Roll Number: 24CSM2R05

Course Title: Data Privacy

Department: Computer Science and Engineering

Program: M. Tech in Computer Science and Information Security

Semester: 2

Assumptions:

- **Sensitivity:** We assume that the sensitivity Δf for the query (here, the mean of the 'income' attribute) can be approximated by the range of income values.

- **Epsilon (ε):** We set the privacy budget to 1.0. A lower ϵ would yield stronger privacy (more noise) but lower accuracy.

- **No Built-in Laplace Function**: Instead of using np.random.laplace, we implement our own noise generation using the inverse transform sampling method.

Methodology:

- **Laplace Noise Generation:**

  The function `my_laplace_noise` generates noise by drawing a uniform random value u in the interval $[-0.5, 0.5]$ and then applying the inverse transform:

  $$\text{noise} = \mu - b \cdot \text{sign}(u) \cdot \ln(1 - 2|u|)$$

  Here, μ is the mean (set to 0) and b is the scale parameter calculated as $\text{sensitivity}/\epsilon$.

- **Application of DP Mechanism:**
  We add the generated Laplace noise to each income value to create a differentially private version of the income attribute.
- **Empirical Simulation:**
  The function simulate_dp_income_custom runs multiple trials (1000) to compute the DP mean for the full dataset and for a neighboring dataset (one record removed). The near-identical distributions of the DP means demonstrate that the mechanism is robust to small changes, fulfilling the differential privacy guarantee.

Proof of Protection:

- The theoretical guarantee of differential privacy ensures that for any two neighboring datasets, the probability distributions of the outputs are nearly identical.
- The simulation shows that the DP means from the full and neighboring datasets are almost the same, which empirically confirms that the influence of any single record is effectively masked by the added noise.

Initial dataset state:

colums:

['age', 'workclass', 'fnlwgt', 'education', 'education_num',

'marital_status', 'occupation', 'relationship', 'race', 'sex',

'capital_gain', 'capital_loss', 'hours_per_week', 'native_country',

'income', 'age_generalized', 'education_generalized',

'occupation_generalized', 'workclass_generalized',

'marital_generalized', 'native_generalized', 'suppressed'],

dtype='object')

	age	workclass	fnlwgt	education	education_num	marital_status	occupation	relationship	race	sex	capital_gain	capital_loss	hours_per_week	native_country	income
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
32556	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female	0	0	38	United-States	<=50K
32557	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	0	0	40	United-States	>50K
32558	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female	0	0	40	United-States	<=50K
32559	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	Male	0	0	20	United-States	<=50K
32560	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	Female	15024	0	40	United-States	>50K

32561 rows × 15 columns

Modified data:

	age	workclass	fnlwgt	education	education_num	marital_status	occupation	relationship	race	sex	capital_gain	capital_loss	hours_per_week	native_country	income
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	49379
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	45020
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	89020
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	83888
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	79334
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
32556	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female	0	0	38	United-States	98541
32557	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	0	0	40	United-States	78930
32558	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female	0	0	40	United-States	49818
32559	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	Male	0	0	20	United-States	104715
32560	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	Female	15024	0	40	United-States	72182

32561 rows × 15 columns

statistical analysis:

```
Full dataset DP mean:
  Mean: 79957.4620, Std: 634.2368
Neighboring dataset DP mean:
  Mean: 79990.5513, Std: 626.9773
```