**NeurIPS Hide-and-seek Privacy Challenge documentation questionnaire**

**Team name**

|  |
| --- |
| Golden\_Fleece |

**Submission filenames(s)**

|  |  |
| --- | --- |
| Hider | akash\_hider.zip |
| Seeker |  |

**What class of algorithms does your solution belong to?** (e.g. GANs, VAEs, noise-injection, nearest neighbor, etc.)

|  |  |
| --- | --- |
| Hider | Data point random switch |
| Seeker |  |

**Describe your algorithm in one sentence** (e.g. “Noise is added to the original data and then this data is returned.”)

|  |  |
| --- | --- |
| Hider | Group the datapoints in each dimension based on their cumulative distribution, and randomly switch the datapoints that falls into the same bin. |
| Seeker |  |

**Describe your algorithm in words** (e.g. “Noise is drawn from a Gaussian distribution, with mean 0 and variance s, where the dimension is determined by the size of the dataset. This noise is added to the original data to produce a noisy version of the dataset and this noisy dataset is then returned as the synthetic data.”)

|  |  |
| --- | --- |
| Hider | For each feature dimension, aggregate the data from all the users. Then divide the data into a number of bins based on cumulative distribution percentage. For each data point (here by data point I mean one numerical number, I will refer to each user as “data entry”), it is randomly replaced with another value within the same bin. |
| Seeker |  |

**Specify any loss functions used** (e.g. “No loss functions used.”)

|  |  |
| --- | --- |
| Hider | No loss functions used. |
| Seeker |  |

**Specify any hyperparameters and how they are optimized (or preset values)** (e.g. “The noise size, s, is set to 0.1.”)

|  |  |
| --- | --- |
| Hider | * number\_of\_bins: This determines how many bins we aggregate the data in each dimension into. We currently use 33. For example, we have 1000 data in a specific feature dimension, then we calculate the cumulative distribution of the data. Then the minimum 3% of the data falls into bin 1, 3%-6% falls into bin 2... The number is selected based on how many data points are there in a bin on average. If too large, then we only have very limited number of data in each bin, which lead to insuffcient noise added after switching, If too high, the variance inside each bin is too high and will introduce too much noise after switching. |
| Seeker |  |

**Specify any pre-trained models used by your algorithm** (e.g. “None.”)

|  |  |
| --- | --- |
| Hider | None. |
| Seeker |  |

**Pseudo-code for your algorithm**

e.g. **Inputs:** Dataset, D, random seed

**Hyperparameters:** s (default 0.1)

1. Determine dataset dimension: n x d x T

2. Draw N ~ N(0, s), an n x d x T dimensional Gaussian

3. Return D + N

|  |  |
| --- | --- |
| Hider | **Inputs:** Dataset D  **Hyperparameters:**  n\_bins: number of bins   1. Identify all the NaN’s and replace them with 0 2. MinMax Scale each feature dimension to [0,1] 3. Add a very small amout of noise to each data point 4. Loop feature k:   (a) aggregate datapoints of feature k from all the dataentries and times  (b) split the data into n\_bins different bins based on percentage  (c) use an array to store which bin does each data belong to  for datapoint d\_i ([any user, feature k, any time] in D)  switch d\_i with another value in the same bin   1. Reverse the MinMax Scaling 2. Reverse the NaN’s back |
| Seeker |  |

Finally, alongside this document **please also submit a commented version of your code**. Please include:

- Docstrings for each new class/function defined

- Inline comments for your main function/class

The goal of these comments is to tie the code to the description you have provided here. Please do not alter the actual content of your code - only add comments/docstrings.

**Submitting your documentation and commented code**

Please submit your commented code within a .zip or equivalent file type (1 file per solution), and share it with us as an attachment alongside this Word doc.

You can send these via email (to [nm736@cam.ac.uk](mailto:nm736@cam.ac.uk); [james.jordon@wolfson.ox.ac.uk](mailto:james.jordon@wolfson.ox.ac.uk); [es583@cam.ac.uk](mailto:es583@cam.ac.uk)) or DM James Jordon/Evgeny Saveliev on Slack (you can join the workspace [with this URL](https://join.slack.com/t/hideandseekpr-fbc8582/shared_invite/zt-k2h9xye8-RQNen128uXIG2TRsLa_ppA)).