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

**Team name**

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

**Submission filenames(s)**

|  |  |
| --- | --- |
| Hider |  |
| Seeker | final\_seeker.zip |

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

|  |  |
| --- | --- |
| Hider |  |
| Seeker | Nearsest Neighbour |

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

|  |  |
| --- | --- |
| Hider |  |
| Seeker | Nearsest Neighbour using timestamps only. |

**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 |  |
| Seeker | This is a modified version of the baseline KNN seeker. Our insight is that, the sampling length and interval is discrepant across different users. We found that doing KNN classification on the time dimension only is actually better than doing it over all the dimensions. |

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

|  |  |
| --- | --- |
| Hider |  |
| Seeker | No loss functions used, except il-2 distance is used to calculate the distance between data pairs. |

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

|  |  |
| --- | --- |
| Hider |  |
| Seeker | No hyperparameters |

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

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

**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 |  |
| Seeker | **Inputs:** Dataset D   1. Take out the time domain from the datasets and discard the rest dimensions, squeeze the enlarged dataset to [no1 x 100] and the genegrated dataset to [now x 100] 2. For each data entry in enlarge dataset   For each generated data entry  Check the distance between timestamps in enlarge dataset and  generated dataset  Find the minimum distance from 1-NN generated data   1. Check the threshold distance for top gen\_no for 1-NN distance 2. Return the decision for reidentified data |

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)).