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

**Team name**

|  |
| --- |
| GenSynDS |

**Submission filenames(s)**

|  |  |
| --- | --- |
| Hider | hider.py |
| Seeker | seeker.py |

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

|  |  |
| --- | --- |
| Hider | Barycenters |
| Seeker | Nearest neighbor |

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

|  |  |
| --- | --- |
| Hider | Return barycenters of pairs of time series that are within the same cluster (2 clusters in total) |
| Seeker | Find samples that are closest to generated data (1-NN) according to various features |

**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 | Data is normalized then clustered according to variability (standard deviation of difference between consecutive observations). Within each cluster, we compute barycenters of pairs of consecutive time series according to the SoftDTW metric. Then we scale it back and return results. |
| Seeker | The seeker is almost identical to the 1-KNN baseline, but we added to the existing features (mean and standard deviation) some features about the mask if available (sum of mask). |

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

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

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

|  |  |
| --- | --- |
| Hider | As a threshold for making the clusters we pick variability = 0.5 (after observation on the training set). |
| Seeker | No hyperparameter used. |

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

|  |  |
| --- | --- |
| Hider | None. |
| 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 | **Inputs:** Dataset, D, random seed  **Hyperparameters:** variability threshold (default 0.5)   1. Rescale all time series to mean 0 and std 1 2. Compute variability aka std of difference between consecutive observations 3. Cluster 1 is defined as variability > 0.5, cluster 2 as variability <= 0.5 4. For each cluster    1. For each pair T1, T2 of consecutive time series within this cluster       1. Compute the SoftDTW barycenter B of T1 and T2       2. Append B \* T1.std + T1.mean to the generated data 5. Return the generated data (without mask) |
| Seeker | **Inputs:** Dataset, D, random seed   1. generated\_features ← means and stds of generated data 2. enlarged\_features ← means and stds of enlarged data 3. If there are generated and enlarged masks    1. Append to generated\_features the sums of masks of generated data    2. Append to enlarged\_features the sums of masks of enlarged data 4. Use the KNN seeker baseline on generated\_features and enlarged\_features |

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)).