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

**Team name**

|  |
| --- |
| Crouching Tiger Hidden DraGAN |

**Submission filenames(s)**

|  |  |
| --- | --- |
| Hider | hider\_hiddendragan.zip |
| Seeker |  |

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

|  |  |
| --- | --- |
| Hider | Wasserstein GAN |
| Seeker |  |

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

|  |  |
| --- | --- |
| Hider | A single timepoint from each sequence synthesized and the output is reshaped into expected form by repeating the single timepoint 100 times. |
| 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 | Preprocessing: The columns at indices 39 & 69 , and 37 & 38 were pairs of variables with the only difference appearing to be related to the date of admission. Patients in the dataset had non-missing observations for, at most, one variable in each of the pairs of columns. |
| Seeker |  |

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

|  |  |
| --- | --- |
| Hider |  |
| Seeker |  |

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

|  |  |
| --- | --- |
| Hider | iterations = 50000  Number of critic iterations per generator iteration = 100  batch\_size = 256  Regulurisation coefficient for gradient penalty (lambda) = 10  Number of neurons per hidden layer = 1024 |
| Seeker |  |

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

|  |  |
| --- | --- |
| Hider | None |
| Seeker | (Improved Training of Wasserstein GANs by Gulrajani et al. (2017) |

**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 (n x d x T)  **Hyperparameters:** iterations, n\_critic, batch\_size,  Preprocess data by combining duplicate variable columns, MAP columns, time column  Discard all but final timepoints, (data now has dimensions n x D)  **for** idx in 1 to iterations:  **for** c in 1 to critic:  Draw an (batch\_size x d) dimensional Gaussian  Draw  Select from Dataset (batch\_size x D)          **End for**  Draw an (batch\_size x d) dimensional Gaussian    **End for**  MAP  Duplicate the variable columns that were combined to match the original format |
| 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)).