

Programming in Python: Final Project

Jakub Rydzewski jr@fizyka.umk.pl

I. PROJECT

Write a Python package for reducing data dimensionality. Given the high-dimensional representation of data \mathbf{X} , implement a Python class (or function) that preprocesses the data, projects the data onto two- or three-dimensional space, and plots the data in the low-dimensional embedding \mathbf{Y} .

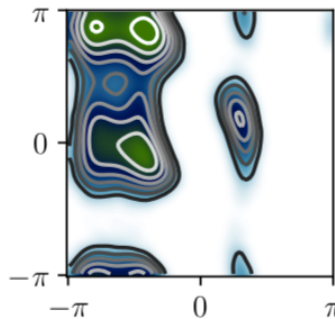
II. WORKFLOW

1. Download the high-dimensional data from a public FTP server at Freie Universität Berlin (<https://markovmodel.github.io/mdshare/>) using:

```
import numpy as np
import mdshare

dataset = mdshare.fetch('alanine-dipeptide-3x250ns-heavy-atom-distances.npz')
with np.load(dataset) as f:
    X = np.vstack([f[key] for key in sorted(f.keys())])
```

2. Write a Python package (using classes or functions) that, among all methods, implements a method called $\mathbf{Y} = \text{fit}(\mathbf{X})$ that takes a high-dimensional tensor and returns its projection.
3. Compare your embedding with the following figure:



4. Write a short report or present your findings during classes.

III. REQUIREMENTS

1. **Tools.** Python 2.7 or 3.*, Jupyter notebook, and Python packages: `numpy`, `matplotlib`, `sklearn`, `keras` / `tensorflow` / `torch`.
2. **Command-line Interface.** The package should be able to work in the command-line mode. Use `argparse` to process important flags. For instance:
`./dimred.py -data /path/to/data -parameter_a 1e-5 -max_iter=1000 ...`
3. **Documentation.** The code should be documented.
4. **Repository.** Make a git repository `XXX_PPSeminar/` on GitHub; it can be a private or public repository. Each project member must be able to access the repository. It should have the following high-level directory structure:

```

-- doc/
-- 2019-pp-report/
--   report.ipynb
--   report.pdf
-- etc/
-- 2017-03-25-whitewwq.jpg
-- 2017-04-03-whiteboard.jpg
-- 2017-04-06-cow-comments.md
-- 2017-04-08-jake-comments.pdf
-- src/
-- checkpoints/
-- codebase/
-- log/
-- out/
-- script1.py
-- script2.py
-- README.md

```

5. **Presentation or Report.** Present your project during the classes (in Polish) or write a report (in English) that explain what is done, how to use the package, and what is the difference between the figure and your projection.