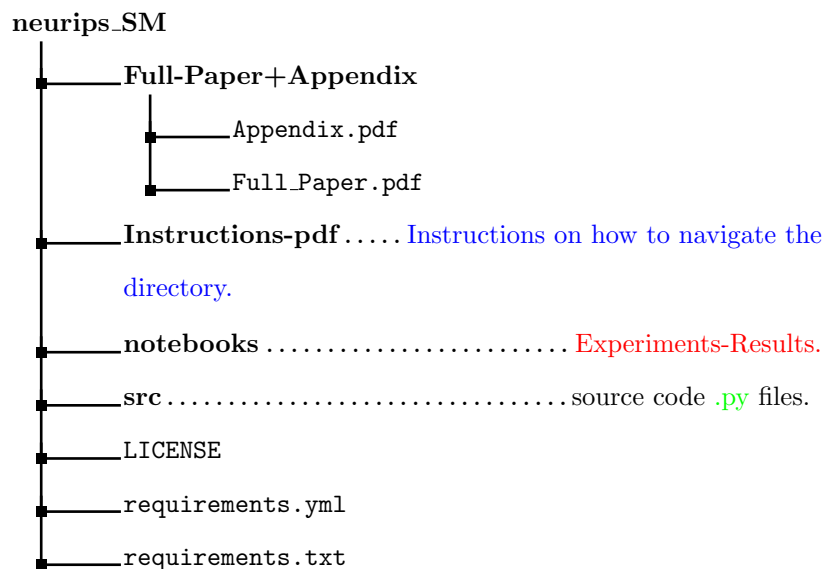


Supplementary Materials Instructions

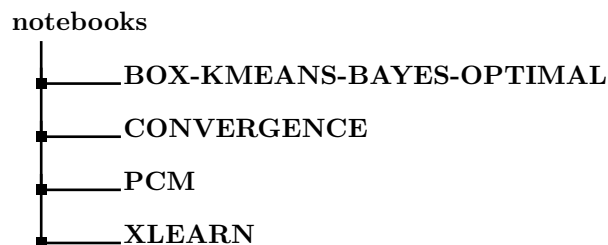
1 Folder Structure

Our supplementary materials **.zip** file has the following folder structure:



We will analyze each important subfolder separately.

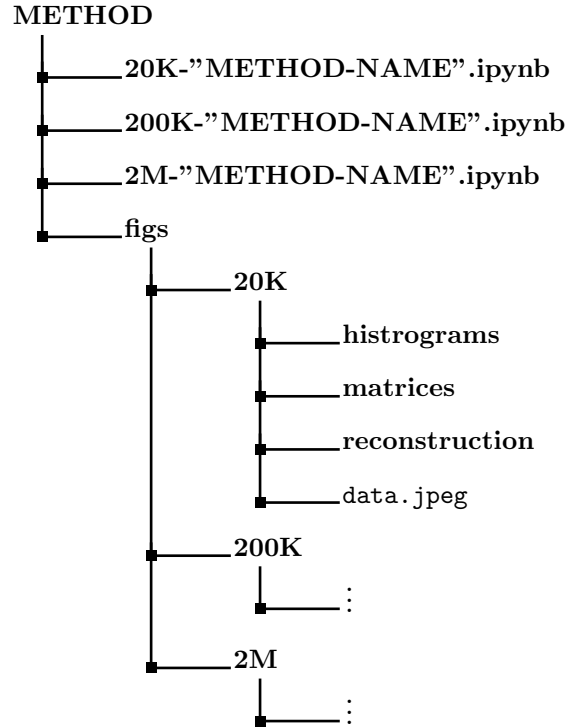
1.1 notebooks



The **notebooks** folder contains all the experiments presented in the main paper **and** the appendix in the form of **python jupyter notebooks**. In the paper we present three competing methods. In the appendix we present the **BOX** method, from which **PCM** is based on as well as an alternative clustering to **BOX**, which is **KMEANS** clustering.

1. PCM (main text, **notebooks/PCM**).
2. X-Learner (main text, **notebooks/XLEARN**).
3. Bayes Optimal (main text, **notebooks/BOX-KMEANS-BAYES-OPTIMAL**).
4. BOX (appendix, **notebooks/BOX-KMEANS-BAYES-OPTIMAL**).
5. KMEANS (appendix, **notebooks/BOX-KMEANS-BAYES-OPTIMAL**).

The main results are under the **Methods** folders, and contain notebooks that produce the results for the 20K, 200K, 2M sizes of the experiments under the **figs** folder. The structure is,



The **CONVERGENCE** folder contains the notebook and the results for the **appendix C Cluster Homogeneity**.

2 Instructions to Reproduce the Experiments

There are two options in order to be able to run the notebooks and reproduce the experiments.

- **With conda (tested in Ubuntu 20:** Create an environment using the **requirements.yml** file, with the command:

```
conda env create -f requirements.yml
```

- This will create an environment with the name **neurips**.
- Activate the environment with:

```
conda activate neurips
```

- From inside the environment run:

```
jupyter notebook
```

- The jupyter server will open up and you can run the notebooks.
- **DO NOT CHANGE THE FOLDER STRUCTURE.** Aside from the external packages, installed from the commands above, notebooks use code from the **src** folder.

3 Final Comments

Inside the **.zip** file all notebooks are pre-runned and all the results are pre-generated. One could just browse the notebooks and see the results naturally from the notebook presentation. Running the notebooks will re-populate the folders containing the results.