Code and data for reproducing the empirical results in the paper "Regularization for Shuffled Data Problems via Exponential Family Priors on the Permutation Group"

This directory is divided into five main subdirectories.

1. Subdirectory `data` contains data files reproducing synthetic and real data analysis results in the paper.

2. Subdirectory `functions` contains `MATLAB` code for specific algorithms and approaches in the paper.

3. Subdirectory `real` contains `MATLAB` code and result files (`.mat`) that reproduce the results in figure 4.

4. Subdirectory `simluation` contains `MATLAB` code and result files (`.mat`) reproducing the simulation results in figure 3.

5. Subdirectory `appendix` contains `MATLAB` code and result files (`.mat`) reproducing the simulation results in appendix.

Guide for the subdirectory `data`

1. The path for `MATLAB` code is already been set up in each code with prefix `run` and `plot`, no need to add data into each folder before running the code

Guide for the subdirectory `functions`

1. Prefix `EM` stands for EM algorithm

2. Prefix `DA` stands for Data Augmentation

3. Prefix `EB` stands for Empirical Bayes

4. Prefix `HB` stands for Hierarchical Bayes

5. Suffix `mixture` stands for the mixture approach in Slawski, Diao and Ben-David (2020)

6. There are mex files in the folder which is created by mex `MATLAB 2019a` .

Guide for the subdirectory `real` and `simulation`

1. Subdirectory contains folders with `.mat` files that regenerate the results in figure 3 and 4.

2. To regenerate the plot, call the `.m` files with prefix `plot`

Guide for the subdirectory `appendix`

1. Subdirectory contains folders with `.mat` files that regenerate the results in figure S.1.

2. To regenerate the plot, call the `.m` files with prefix `plot`

To regenerate the results in the these subdirectories, call the files with prefix run, which in turn automatically call the corresponding files with prefix `experiment`. Note that execution of the latter will over-ride all existing .mat files.