B-MASTER Reproducibility Instructions

We describe the reproducibility instructions in three sections. Please refer to the following list to see which section describes the reproducibility of the tables and figures below.

Figures

- Figure 1: NA (concept diagram).

- Figure 2: refer to Simulation study 2.

- Figure 3: refer to Simulation study 1.

- Figure 4: refer to Simulation study 2.

- Figure 5,6,7,8,9,10: refer to Real data analysis.

Tables

- Table 1: refer to Simulation study 1.

- Table 2: refer to Simulation study 2.

- Table 3: refer to Real data analysis.

1. Simulation study 1

1. **Create Dummy data:** Go to Simulation study 1 / Dummy Real Data / Generate_Dummy_Real_via_BMASTER.

2. Run BMASTER_MockRealDataGenerate.m to generate estimated coefficient matrix, which is considered as the True coefficient matrix for the rest of this simulation study.

3. Outputs are generated within Generate_Dummy_Real_via_BMASTER / Data.

4. The generated data is copied and placed in Simulation study 1 / Dummy Real Data (already copied, no need to copy again).

5. In Dummy Real Data, run GENERATE_Y_DATA.m to generate 10 realizations of Y based on X and earlier estiamted B ("True beta"). Now data generation is completed.

6. Copy the generated datasets to the sub-folders within Simulation study 1, namely, BMASTER, SSLasso, mSSL, Remmap (already copied, no need to copy again).

7. **B-MASTER results:** Run BMASTER / BMASTER_on_simReal_v3.m.

8. **SSLasso results:** Run SSLasso / SSLASSO_on_simReal.R.

9. **mSSL results:** Run mSSL / mSSL_dpe_on_simReal.R (for dpe method), and mSSL / mSSL_dCpe_on_simReal.R (for dcpe method).

10. **Remmap results:** Run Remmap / remMAP_on_simReal.R (for original method), and Remmap / remMAPBic_on_simReal.R (for BIC-based approximation method).

11. Copy all the outputs to Simulation study 1 / Summary Table and Plots (already copied, no need to copy again).

12. Run Summary Table and Plots / Summary_tables_plots.R to generate **Table 1** outputs, and **Figure 3 (bottom right)**.

13. Run Summary Table and Plots / Post_analysis_plot.R to generate **Figure 3 (up; bottom left)**.

2. Simulation study 2

1. Go to Simulation study 2 and open BMASTER_scalability.m.

2. Setting `IsRhoNonZero` $= 0$ run it for $P = (20, 50, 100, 200, 500, 1000, 2000)$.

3. Setting `IsRhoNonZero` $= 1$ run it for $P = (20, 50, 100, 200, 500, 1000, 2000)$.

4. Run Simulation study 2 / Summary_comp_time.R, that generates **Figure 4**, and two csv files corresponding to the upper and lower halves of **Table 2**.

5. Go to Simulation study 2 / remMap scalability.

6. Run BMASTER_v_remmap.m thrice setting `SampleMultFactor` $= 1, 5, 10$.

7. Run remMAP_scalability.R thrice setting `SampleMultFactor` $= 1, 5, 10$.

8. Run Comparison_plot.R. That generates **Figure 2** (both left and right).

3. Real data analysis

1. Go to Real Data Analysis. Run BMASTER_Real_Data.m.

2. Run Extraction_subset_for_analysis.R.

3. Run Overall_plots.R. That generates **Figure 5,6,7**.

4. Run Post_analysis_subset_1_NEW.R. That generates **Figure 8**.

5. Run Post_analysis_subset_2_NEW.R. That generates **Figure 9**.

6. Run CCA_subset_plots.R. That generates **Figure 10**.

7. Go to Real Data Analysis / Validation on real data.

8. Run BMASTER_Real_Data_validation.m. Also run SSLASSO_RealData_validation.R, mSSL_dCpe_RealData_validation.R, remMAPBic_RealData_validation.R. This sequence of executions generate 4 outputs csv files corresponding to corresponding methods; which results are assembled and presented in **Table 3**.