README: Computational Time

The folder contains 2 subfolders: `Table3` and `Figure6`. These two folders include code and functions for generating Table 3 and Figure 6 in Section 4.

Subfolder `Table3`

File name	Description
`soybean.R`	Code for obtaining imputation time for the `soybean` dataset
`nwts.R`	Code for obtaining imputation time for the `nwts` dataset
`nhanes.R`	Code for obtaining imputation time for the `nhanes` dataset
`credit.R`	Code for obtaining imputation time for the `credit` dataset
`adult.R`	Code for obtaining imputation time for the `adult` dataset

After running the above files, outputs would be stored in the folder `result`. This folder contains computational time results for different datasets.

File name	Description
`soybean.RData`	Computational time results for the `soybean` dataset
`nwts.RData`	Computational time results for the `nwts` dataset
`nhanes.RData`	Computational time results for the `nhanes` dataset
`credit.RData`	Computational time results for the `credit` dataset
`adult.RData`	Computational time results for the `adult` dataset

We can then obtain the summary statistics for these results.

File name	Description
`summary.R`	Summary of computational time for obtaining Table 3 in Section 4.

Datasets used in this section

- Dataset `soybean` is originally from the UCI Machine Learning Repository (Dua and Graff 2019). The dataset used in Section 4 is downloaded from the `R` package `mlbench` (Leisch and Dimitriadou 2021).
- Dataset `nwts` (D'Angio et al. 1989) can be downloaded from the `R` package `addhazard` (Hu et al. 2017).
- Dataset `nhanes` used in Section 4 can be downloaded from the `R` package `hexbin` (Carr et al. 2020).
- Dataset `credit` (Yeh and Lien 2009) is from the UCI Machine Learning Repository (Dua and Graff 2019). Users can also directly use the datafile `credit.rda` in the `Table3` folder.

- Dataset `adult` (Becker and Kohavi 1996) is from the UCI Machine Learning Repository . Users can also load this dataset from the R package `misle` (Deng 2021).

Subfolder `Figure6`

File name	Description
`time1.R`	Code for obtaining running time for simulated data with 100 observations and different numbers of continuous features (11, 21 and 31).
`time1b.R`	Code for obtaining running time for simulated data with 100 observations and different numbers of binary features (11, 21 and 31).
`time1c.R`	Code for obtaining running time for simulated data with 100 observations and different numbers of a mix of binary and 3-class categorical features (11, 21 and 31).
`time2.R`	Code for obtaining running time for simulated data with 1000 observations and different numbers of continuous features (11, 21 and 31).
`time2b.R`	Code for obtaining running time for simulated data with 1000 observations and different numbers of binary features (11, 21 and 31).
`time2c.R`	Code for obtaining running time for simulated data with 1000 observations and different numbers of a mix of binary and 3-class categorical features (11, 21 and 31).
`time3.R`	Code for obtaining running time for simulated data with 10000 observations and different numbers of continuous features (11, 21 and 31).
`time3b.R`	Code for obtaining running time for simulated data with 10000 observations and different numbers of binary features (11, 21 and 31).
`time3c.R`	Code for obtaining running time for simulated data with 10000 observations and different numbers of a mix of binary and 3-class categorical features (11, 21 and 31).
`time4.R`	Code for obtaining running time for simulated data with 100000 observations and different numbers of continuous features (11, 21 and 31).
`time4b.R`	Code for obtaining running time for simulated data with 100000 observations and different numbers of binary features (11, 21 and 31).
`time4c.R`	Code for obtaining running time for simulated data with 100000 observations and different numbers of a mix of binary and 3-class categorical features (11, 21 and 31).
`timepfm.R`	Code for generating Figure 6 in Section 4.

References

Becker, B., and Kohavi, R. (1996), "adult Dataset," https://archive.ics.uci.edu/ml/datasets/adult.

Carr, D., ported by Nicholas Lewin-Koh, Maechler, M., and contains copies of lattice functions written by Deepayan Sarkar (2020), *hexbin: Hexagonal Binning Routines*, R package version 1.28.1. Available at https://CRAN.R-project.org/package=hexbin.

D'Angio, G. J., Breslow, N., Beckwith, J. B., Evans, A., Baum, E., Delorimier, A., Fernbach, D., Hrabovsky, E., Jones, B., Kelalis, P. et al. (1989), "Treatment of Wilms' Tumor. Results of the Third National Wilms' Tumor Study," *Cancer*, 64, 349–360.

Deng, Y. (2021), *misle: Multiple Imputation Through Statistical Learning*, R package. Available at https://github.com/agnesdeng/misle.

Dua, D., and Graff, C. (2019), "UCI Machine Learning Repository," University of California, Irvine, School of Information and Computer Sciences. Available at http://archive.ics.uci.edu/ml.

Hu, J. K., Breslow, N., and Chan, G. (2017), *addhazard: Fit Additive Hazards Models for Survival Analysis*, R package version 1.1.0. Available at https://CRAN.R-project.org/package=addhazard.

Leisch, F., and Dimitriadou, E. (2021), *mlbench: Machine Learning Benchmark Problems*, R package version 2.1-3. Available at https://cran.r-project.org/web/packages/mlbench.

Yeh, I.-C., and Lien, C.-h. (2009), "The Comparisons of Data Mining Techniques for the Predictive Accuracy of Probability of Default of Credit Card Clients," *Expert Systems with Applications*, 36, 2473–2480.