Home (../index.html)

eHarmony matchings

This data set was provided by eHarmony, Inc. (http://www.eharmony.com/) The data consists of pairs of individuals, which either matched (positive example) or did not (negative example). The data is partitioned into two subsets corresponding to two equal-length segments of time. The data is stored in CSV files, organized as follows.

EH-*-data.csv.gz
Each row describes an individual. The first column is an identification number for that individual, and all subsequent columns contain the (numeric) feature values.

EH-*-labels.csv.gz
Each row describes a pairwise interaction. The first column indicates whether the interaction is positive (1) or negative (0). The second and third columns contain identification numbers for the corresponding individuals.

Please refer to the paper below for more details about this data set.

Note

To protect the privacy of users, all features have been obfuscated and normalized. I cannot provide names for the features.

Download

- Slice 1: data (EH-slice1-data.csv.gz) (9.7MB, md5 (EH-slice1-data.csv.gz.md5)), labels (EH-slice1-labels.csv.gz) (2.2MB, md5 (EH-slice1-labels.csv.gz.md5))
- Slice 2: data (EH-slice2-data.csv.gz) (7.4MB, md5 (EH-slice2-data.csv.gz.md5)), labels (EH-slice2-labels.csv.gz) (1.8MB, md5 (EH-slice2-labels.csv.gz.md5))

References

If you use this data, please cite the following paper:

Metric learning to rank
2010
McFee, B. (../) and Lanckriet, G.R.G. (http://cosmal.ucsd.edu/~gert/)   ⇄ (../papers/mlr.bib) | 📄 (../papers/mlr.pdf)
Twenty-seventh International Conference on Machine Learning (ICML).

Source code

The source code for MLR is now hosted on GitHub (http://github.com/bmcfee/mlr).