

dml: Distance Metric Learning in R

Yuan Tang¹, Tao Gao², and Nan Xiao³

¹ Ant Financial Services Group, San Mateo, CA 94402, USA ² Minghong Investment Management, Shanghai 200122, China ³ Seven Bridges Genomics, Inc., Cambridge, MA 02142, USA

DOI: [10.21105/joss.01036](https://doi.org/10.21105/joss.01036)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 18 October 2018

Published: 23 October 2018

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

Distance metric is widely used in the machine learning literature. We used to choose a distance metric according to a priori (e.g. Euclidean Distance, L1 Distance, etc.) or according to the result of cross validation within small class of functions (e.g. choosing order of polynomial for a kernel). Actually, with priori knowledge of the data, we could learn a more suitable distance metric with (semi-)supervised distance metric learning techniques. `dml` (Tang, Gao, & Xiao, 2015) is such an R package aims to implement a collection of algorithms for (semi-)supervised distance metric learning.

The `dml` package provides native R implementations for a collection of *Distance Metric Learning* algorithms, including both global and local methods such as *Relevant Component Analysis* (Shental, Hertz, Weinshall, & Pavel, 2002), *Discriminative Component Analysis* (Peltonen, Goldberger, & Kaski, 2007), and *Local Fisher Discriminant Analysis* (Sugiyama, 2007). A list of all the implemented algorithms can be found in the `dml` [package reference manual](#). These methods are widely applied in feature extraction, dimensionality reduction, clustering, information retrieval, and computer vision problems.

Additionally, implementations for the variants of the methods are also available in `dml` package. For example, since it was built on top of the `lfda` (Tang, 2017; Tang & Li, 2016) package, users also have access to the family of *Local Fisher Discriminant Analysis* methods, which includes *Local Fisher Discriminant Analysis*, *Kernel Local Fisher Discriminant Analysis*, and *Semi-supervised Local Fisher Discriminant Analysis* (Sugiyama, Idé, Nakajima, & Sese, 2010). To make the results of these methods easy for users to interpret and analyze, both static and interactive visualizations for the results are available through `ggfortify` (Horikoshi & Tang, 2018; Tang, Horikoshi, & Li, 2016) and `autoplotly` (Tang, 2018a, 2018b) packages respectively.

References

- Horikoshi, M., & Tang, Y. (2018). *Ggfortify: Data visualization tools for statistical analysis results*. Retrieved from <https://CRAN.R-project.org/package=ggfortify>
- Peltonen, J., Goldberger, J., & Kaski, S. (2007). Fast semi-supervised discriminative component analysis. In *Machine learning for signal processing, 2007 IEEE workshop on* (pp. 312–317). IEEE.
- Shental, N., Hertz, T., Weinshall, D., & Pavel, M. (2002). Adjustment learning and relevant component analysis. In *European conference on computer vision* (pp. 776–790). Springer.
- Sugiyama, M. (2007). Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of machine learning research*, 8(May), 1027–1061.

Sugiyama, M., Idé, T., Nakajima, S., & Sese, J. (2010). Semi-supervised local fisher discriminant analysis for dimensionality reduction. *Machine learning*, 78(1-2), 35.

Tang, Y. (2017). *Local fisher discriminant analysis*. Retrieved from <https://CRAN.R-project.org/package=lfda>

Tang, Y. (2018a). *Autoplotly: An R package for automatic generation of interactive visualizations for statistical results*. Retrieved from <https://CRAN.R-project.org/package=autoplotly>

Tang, Y. (2018b). Autoplotly: An R package for automatic generation of interactive visualizations for statistical results. *Journal of Open Source Software*, 3(24), 657. Retrieved from <https://doi.org/10.21105/joss.00657>

Tang, Y., & Li, W. (2016). Lfda: An R package for local fisher discriminant analysis and visualization. *arXiv preprint arXiv:1612.09219*. Retrieved from <https://arxiv.org/abs/1612.09219>

Tang, Y., Gao, T., & Xiao, N. (2015). *Dml: Distance metric learning in R*. Retrieved from <https://CRAN.R-project.org/package=dml>

Tang, Y., Horikoshi, M., & Li, W. (2016). Ggfortify: Unified interface to visualize statistical result of popular R packages. *The R Journal*, 8(2), 478–489. Retrieved from <https://journal.r-project.org/>