The Journal of Open Source Software

lfda: Local Fisher Discriminant Analysis in R

Yuan Tang[1] **and Wenxuan Li**[2]

1 Ant Financial Services Group, San Mateo, CA 94402, USA **2** Purdue University, West Lafayette, IN 47907, USA

DOI: 10.21105/joss.01572

Software

- Review
- Repository
- Archive

Submitted: 13 July 2019
Published: 30 July 2019

Summary

Fisher discriminant analysis (Scholkopft & Mullert, 1999) is a popular choice to reduce the dimensionality of the original dataset. It maximizes between-class scatter and minimizes within-class scatter. It works really well in practice but lacks some considerations for multimodality. Multimodality exists within many applications, such as disease diagnosis, where there may be multiple causes for a particular disease. In this situation, *Fisher discriminant analysis* cannot capture the multimodal characteristics of the clusters. To deal with multimodality, *local-preserving projection* (Niyogi, 2004) preserves the local structure of the data in that it keeps nearby data pairs in the original data space close in the embedding space. As a result, multimodal data could be embedded and its local structure will not be lost.

Later on, a new dimensionality reduction method called *local Fisher discriminant analysis (LFDA)* (Sugiyama, 2006) was proposed to combine both advantages of Fisher discriminant analysis and those of local-preserving projection in a way that between-class separability is maximized while within-class separability is minimized and its local structure is preserved. Furthermore, with the help of kernel trick, local Fisher discriminant analysis can also be extended to deal with non-linear dimensionality reduction situations.

Principal components analysis (PCA) (Jolliffe, 2002) is another popular choice for performing dimension reduction. However, in practice sometimes principal components analysis generates bad principal components that cannot explain a great amount of variance in the original dataset. For example, if the original dataset has six dimensions and we reduce the dimension into three using *PCA*, the three principal components might not capture some important characteristics and variance in the original dataset. Using this result from *PCA* sometimes misleads the analysis due to the poor pre-processing that losses a lot of the essential information in the original dataset. On the other hand, metric learning methods such as *LFDA* in particular can, surprisingly, enhance the distinctive characteristics of the original dataset and pull data points that have similar characteristics close to each other. However, both *PCA* and *LFDA* have their own shortcomings and a combined approach called *semi-supervised local Fisher discriminant analysis* (Sugiyama, Idé, Nakajima, & Sese, 2010) mix the supervised and unsupervised approaches to provide a more stable result.

The `lfda` (Tang, 2017; Tang & Li, 2016) package is an R package that provides the implementation for the abovely mentioned methods *local Fisher discriminant analysis*, *kernel local Fisher discriminant analysis*, and *semi-supervised local Fisher discriminant analysis* so researchers could quickly experiment different variations of *local Fisher discriminant analysis* methods in different applications. As of the time of writing, it's the first package with those methods implemented in native R language. It also provides visualization functions to easily visualize the dimension reduction results by using `rgl` (Adler, Murdoch, & others, 2016) for 3D visualization, `ggfortify` (Horikoshi & Tang, 2018; Tang, Horikoshi, & Li, 2016) for 2D visualization in `ggplot2` (Wickham, 2009) style, or `autoplotly` (Tang, 2018a, 2018b) for interactive visualizations in `plotly` (Sievert et al., 2018) style. `lfda` is also included in `dml` (Tang, Gao, & Xiao, 2015, 2018) package together with implementations of other *distance metric learning* algorithms.

References

Adler, D., Murdoch, D., & others. (2016). *Rgl: 3D visualization using opengl*. Retrieved from https://CRAN.R-project.org/package=rgl

Horikoshi, M., & Tang, Y. (2018). *Ggfortify: Data visualization tools for statistical analysis results*. Retrieved from https://CRAN.R-project.org/package=ggfortify

Jolliffe, I. (2002). *Principal component analysis*. Wiley Online Library.

Niyogi, X. (2004). Locality preserving projections. In *Neural information processing systems* (Vol. 16, p. 153). MIT.

Scholkopft, B., & Mullert, K.-R. (1999). Fisher discriminant analysis with kernels. *Neural networks for signal processing IX*, *1*(1), 1. doi:10.1109/NNSP.1999.788121

Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., & Despouy, P. (2018). *Plotly: Create interactive web graphics via 'plotly.js'*.

Sugiyama, M. (2006). Local fisher discriminant analysis for supervised dimensionality reduction. In *Proceedings of the 23rd international conference on machine learning* (pp. 905–912). ACM. doi:10.1145/1143844.1143958

Sugiyama, M., Idé, T., Nakajima, S., & Sese, J. (2010). Semi-supervised local fisher discriminant analysis for dimensionality reduction. *Machine learning*, *78*(1-2), 35. doi:10.1007/s10994-009-5125-7

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. doi: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., Gao, Tao, & Xiao, N. (2018). Dml: Distance metric learning in R. *Journal of Open Source Software*, *3*(30), 1036. doi:10.21105/joss.01036

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. doi:10.32614/RJ-2016-060

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer Science & Business Media. doi:10.1007/978-0-387-98141-3