Frame-based Matrix Factorization

Sebastian Mair[1], Ahcène Boubekki[1,2], and Ulf Brefeld[1]

[1] Leuphana University, Lüneburg, Germany
{mair,boubekki,brefeld}@leuphana.de
[2] German Institute for Educational Research, Frankfurt am Main, Germany

Abstract. Archetypal Analysis is the method of choice to compute interpretable matrix factorizations. Every data point is represented as a convex combination of factors, i.e., points on the boundary of the convex hull of the data. This renders computation inefficient. In this paper, we make two contributions. First, we show that the set of vertices of a convex hull, the so-called *frame*, can be efficiently computed by a quadratic program. We provide theoretical and empirical results for our proposed approach and provide further strategies for scaling up the computation. Second, we make use of the frame to accelerate Archetypal Analysis. This is done by restricting Archetypal Analysis to the frame and just to the frame. Empirical results show, that the novel method often yields similar reconstruction errors as baseline competitors in practice but is much faster to compute. This is especially beneficial when several executions are needed for determining the amount of latent factors. Considering the cumulative runtime in such a scenario outperforms all baselines. In addition, we demonstrate the application of an autoencoder by comparing the quality of the embedding to a neural network.

Keywords: convex hull, frame, matrix factorization, fast approximation

Remark. This work [1] was originally presented at the International Conference on Machine Learning 2017.

Acknowledgments. This work has been funded in parts by the German Federal Ministry of Education and Science BMBF under grant QQM/01LSA1503C.

References

1. Mair, S., Boubekki, A., Brefeld, U.: Frame-based Data Factorizations. International Conference on Machine Learning (2017)