

Projection based Clustering

Michael C. Thrun¹, Alfred Ultsch¹

¹Databionics Research Group, University of Marburg, Germany

Many data mining methods rely on some concept of the dissimilarity between pieces of information encoded in the data of interest. These methods can be used for cluster analysis. However, no generally accepted definition of clusters exists in the literature [Hennig et al., 2015]. Here, consistent with Bouveyron et al., it is assumed that a cluster is a group of similar objects [Bouveyron et al., 2012]. The clusters are called *natural* because they do not require a dissection; instead, they are clearly separated in the data [Duda et al., 2001, Theodoridis/Koutroumbas, 2009, pp. 579, 600]. These clusters can be identified by distance or density based high-dimensional structures.

Dimensionality reduction techniques are able to reduce the dimensions of the input space to facilitate the exploration of structures in high-dimensional data. If they are used for visualization, they are called projection methods. The generalized U*-matrix technique is applicable for these and can be used to visualize both distance- and density-based structures [Ultsch/Thrun, 2017]. The idea that the abstract U*-matrix (AU-matrix) can be used for clustering [Ultsch et al., 2016]. The distances required for hierarchical clustering are defined by the AU-matrix [Lötsch/Ultsch, 2014]. Using this distance we propose a clustering approach for every projection method based on the U*-matrix visualization of a topographic map. [Thrun 2017]. The number of clusters and the cluster structure can be estimated by counting the valleys in a topographic map [Thrun et al., 2016]. If the number of clusters and the clustering method are chosen correctly, then the clusters will be well separated by mountains in the visualization. Outliers are represented as volcanoes and can be interactively marked in the visualization after the automated clustering process.

- [Bouveyron et al., 2012] **Bouveyron, C., Hammer, B., & Villmann, T.**: Recent developments in clustering algorithms, Proc. ESANN, Citeseer, **2012**.
- [Duda et al., 2001] **Duda, R. O., Hart, P. E., & Stork, D. G.**: *Pattern classification*, (Second Edition ed.), Ney York, USA, John Wiley & Sons, ISBN: 0-471-05669-3, **2001**.
- [Hennig et al., 2015] **Hennig, C., Meila, M., Murtagh, F., & Rocci, R.**: *Handbook of cluster analysis*, New York, USA, CRC Press, ISBN: 9781466551893, **2015**.
- [Lötsch/Ultsch, 2014] **Lötsch, J., & Ultsch, A.**: Exploiting the Structures of the U-Matrix, in Villmann, T., Schleif, F.-M., Kaden, M. & Lange, M. (eds.), Proc. Advances in Self-Organizing Maps and Learning Vector Quantization, pp. 249-257, Springer International Publishing, Mittweida, Germany, **2014**.
- [Theodoridis/Koutroumbas, 2009] **Theodoridis, S., & Koutroumbas, K.**: *Pattern Recognition*, (Fourth Edition ed.), Canada, Elsevier, ISBN: 978-1-59749-272-0, **2009**.
- [Thrun, 2017] **Thrun, M. C.**: *A System for Projection Based Clustering through Self-Organization and Swarm Intelligence*, (Doctoral dissertation), Philipps-Universität Marburg, Marburg, **2017**.
- [Thrun et al., 2016] **Thrun, M. C., Lerch, F., Lötsch, J., & Ultsch, A.**: *Visualization and 3D Printing of Multivariate Data of Biomarkers*, in Skala, V. (Ed.), International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG), Vol. 24, Plzen, <http://wscg.zcu.cz/wscg2016/short/A43-full.pdf>, **2016**.
- [Ultsch et al., 2016] **Ultsch, A., Behnisch, M., & Lötsch, J.**: ESOM Visualizations for Quality Assessment in Clustering, In Merényi, E., Mendenhall, J. M. & O'Driscoll, P. (Eds.), *Advances in Self-Organizing Maps and Learning Vector Quantization: Proceedings of the 11th International Workshop WSOM 2016, Houston, Texas, USA, January 6-8, 2016*, (10.1007/978-3-319-28518-4_3pp. 39-48), Cham, Springer International Publishing, **2016**.
- [Ultsch/Thrun, 2017] **Ultsch, A., & Thrun, M. C.**: *Credible Visualizations for Planar Projections*, in Cottrell, M. (Ed.), 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM), IEEE Xplore, France, **2017**.