Bachelor Thesis: Affine Subspace Clustering with SENet

Winter Semester 2025/2026

Background Sparse or self-expressive subspace clustering [4, 2] is a popular approach for clustering high-dimensional data. The approaches center around the idea that a sparse representation of a data point by a combination of other data points from the same dataset would consist only of data points that come from the same subspace. Based on this, the subspace these points lie in can be determined, as groups of data points expressing each other would denote a shared subspace. In turn, these groups also form subspace clusters.
In practice, this means that coefficients are determined that allow each point to be expressed as a linear combination of other points, while applying constraints that minimize a subspace-preserving metric over them. Once the coefficients are determined, they are used to perform clustering with Spectral Clustering [5] or k-Means [3].
Subspaces found with regular sparse subspace clustering always include the origin, limiting the flexibility of detecting subspaces. The setting of affine subspaces, however, allows for more flexible subspace handling as its subspaces do not need to include the origin. This variant can be introduced into regular sparse subspace clustering by extending the data points by an additional 1 coordinate and adapting the constraint so that all the coefficients sum up to 1, rather than minimizing (this is required to maintain the 1 coordinate in the expressed data points)
A major problem of sparse subspace clustering is that storing the self-expressive coefficients scales quadratically with the number of data points, making it unfeasible for very large datasets. SENet [8] is an approach that attempts to address this. Instead of learning a full coefficient matrix, it instead learns a Neural network that approximates the self-expression based on a subset of the data.

Description By default, SENet only handles non-affine subspaces. The goal of this thesis is to introduce the concepts of affine subspaces into the formalization of SENet. A way to approach this would be to add the 1-coordinate to the data used for SENet and to adapt the adjusted constraint into the loss function used to train the neural network (this would work by using $|1 - \sum_{i \neq j} r(f(x_i, x_j; \Theta))|$ instead of $\sum_{i \neq j} r(f(x_i, x_j; \Theta))$ for optimization.

For examination, a mix of synthetic data (using a dedicated data generator for subspace data [1]) and real-world data should be used. For comparison, the affine constraints should be added to a sparse subspace clustering method other than SENet, such as ENSC [6] or SSC-OMP [7].

Tasks

1. Add the affine handling to an implementation of SSC
2. Integrate the affine handling to SENet
3. Compare the performance of affine SENet and affine SSC on a small synthetic dataset
4. Compare the performance of affine and regular SENet on a real dataset

Relevant Resources: `https://github.com/zhangsz1998/Self-Expressive-Network`, `https://github.com/ChongYou/subspace-clustering`, `https://github.com/NanniSchueler/SubCluGen`

Requirements

- Study in the field of computer science
- Understanding of machine learning
- Programming skills in Python
- Beneficial: Data Mining Algorithms II (in or after summer semester 2025)

Additional Information

- Start: in or after August 2025
- Duration of the thesis: 14 weeks

Contact information If you are interested, please send your CV and transcripts to jahn@dbs.ifi.lmu.de

References

[1] Anna Beer, Nadine Sarah Schüler, and Thomas Seidl. A generator for subspace clusters. In Robert Jäschke and Matthias Weidlich, editors, *Proceedings of the Conference on "Lernen, Wissen, Daten, Analysen", Berlin, Germany, September 30 - October 2, 2019*, volume 2454 of *CEUR Workshop Proceedings*, pages 69–73. CEUR-WS.org, 2019.

[2] Ehsan Elhamifar and René Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(11):2765–2781, 2013.

[3] Stuart P. Lloyd. Least squares quantization in PCM. *IEEE Trans. Inf. Theory*, 28(2):129–136, 1982.

[4] René Vidal. Subspace clustering. *IEEE Signal Process. Mag.*, 28(2):52–68, 2011.

[5] Ulrike von Luxburg. A tutorial on spectral clustering. *Stat. Comput.*, 17(4):395–416, 2007.

[6] Chong You, Chun-Guang Li, Daniel P. Robinson, and René Vidal. Oracle based active set algorithm for scalable elastic net subspace clustering. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 3928–3937. IEEE Computer Society, 2016.

[7] Chong You, Daniel P. Robinson, and René Vidal. Scalable sparse subspace clustering by orthogonal matching pursuit. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 3918–3927. IEEE Computer Society, 2016.

[8] Shangzhi Zhang, Chong You, René Vidal, and Chun-Guang Li. Learning a self-expressive network for subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pages 12393–12403. Computer Vision Foundation / IEEE, 2021.