Exploiting the Frame for Active Learning in Multi-class Classification

Sebastian Mair [1] **Ulf Brefeld** [1]

Many learning tasks can be efficiently approximated by leveraging the geometry of data. In particular, we restrict learning task to the frame of the data. The frame is the subset of minimum cardinality that yields the same convex hull as the data itself. As seen in Figure 1, we can interpret the frame as the vertices of the polygon that is induced by the data points. The frame can be trivially computed with any convex hull algorithm that supports higher dimensionalities such as Quickhull (Barber et al., 1996). In practice, however, dispensable triangulations render convex hull algorithms infeasible. More efficient approaches which only compute the frame are based on linear (Dulá & Helgason, 1996; Ottmann et al., 2001; Dulá & López, 2012) or quadratic programming (Mair et al., 2017).

Particularly the approach in Mair et al. (2017) allows to efficiently compute the frame for applications like matrix factorization. Restricting the problem to the frame maintains the performance at a much lower computational cost.

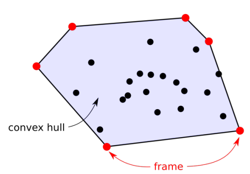

Figure 1. Illustration of the frame.

We now propose to exploit the frame in semi-supervised classification tasks. Consider a labeled data set $\mathcal{X}_l = \{(\mathbf{x}_i, y_i)\}_{i=1}^{l} \subset \mathbb{R}^d \times \{1, \ldots, C\}$ consisting of l points, each of them belonging to one of C classes. Furthermore, we have access to an unlabeled set of instances $\mathcal{X}_u = \{\mathbf{x}_1, \ldots, \mathbf{x}_u\}$ with $u \gg l$.

Given a budget k, the task is to train an accurate classifier by querying at most k labels for points of the pool $\mathcal{X}_u$. This setting is known as active learning (Cohn et al., 1996; Tong & Koller, 2001; Settles, 2012). Common strategies are based on uncertainty criteria or heuristics to sequentially query the next data point. After obtaining the label, the newly labeled pair $(\mathbf{x}, y)$ is included in $\mathcal{X}_l$, the model is retrained, and the next instance to be labeled is chosen. Such active learning strategies are usually tailored to binary problems. Extensions to multiple classes are constructed by costly one-vs-one or one-vs-all approaches.

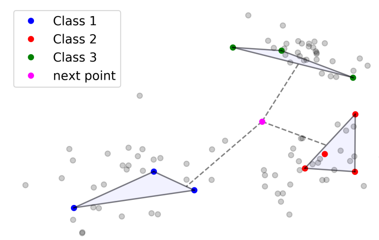

Figure 2. Active Classification.

By contrast, classifiers based on convex hulls (Nalbantov et al., 2006) naturally extend to multi-class problems. We propose to construct one frame per class and to successively dismiss the points lying in the interior of the convex hulls. This information is already given as a side-product of the frame computation. Then, from all candidate points lying outside of the class frames, we choose to label the instance for which the class projections are most similar. In other words, we pick the point which is most likely to be misclassified. This approach is akin to a strategy outlined by Tong & Koller (2001), where instances are successively selected according to their distance to an SVM decision hyperplane in binary classification tasks.

Figure 2 shows an example. There are three classes (blue, red, green) with their respective frames. The chosen instance (shown in magenta) is most likely to be misclassified and is thus selected for labeling by the user. Note that convex hull classifiers naturally yield non-linear decision functions even without using kernels. For cases of overlapping classes, we show how to lift the approach to kernel-induced feature spaces. Furthermore, we extend the approach to archetypal hulls (Thurau, 2010; Chen et al., 2014) in which Archetypal Analysis (AA) (Cutler & Breiman, 1994) is used to obtain an approximation of a convex hull with a priorly specified number of vertices. In summary, we derive an efficient active learning strategy for semi-supervised classification tasks using geometrically inspired convex hull-based approaches.

[1] Leuphana University of Lüneburg, Germany. Correspondence to: Sebastian Mair <mair@leuphana.de>.

References

Barber, C Bradford, Dobkin, David P, and Huhdanpaa, Hannu. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)*, 22(4): 469–483, 1996.

Chen, Yuansi, Mairal, Julien, and Harchaoui, Zaid. Fast and robust archetypal analysis for representation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1478–1485, 2014.

Cohn, David A, Ghahramani, Zoubin, and Jordan, Michael I. Active learning with statistical models. *Journal of artificial intelligence research*, 4:129–145, 1996.

Cutler, Adele and Breiman, Leo. Archetypal analysis. *Technometrics*, 36(4):338–347, 1994.

Dulá, José H and Helgason, Richard V. A new procedure for identifying the frame of the convex hull of a finite collection of points in multidimensional space. *European Journal of Operational Research*, 92(2):352–367, 1996.

Dulá, José H and López, Francisco J. Competing output-sensitive frame algorithms. *Computational Geometry*, 45 (4):186–197, 2012.

Mair, Sebastian, Boubekki, Ahcène, and Brefeld, Ulf. Frame-based Data Factorizations. In *International Conference on Machine Learning*, pp. 2305–2313, 2017.

Nalbantov, Georgi, Groenen, Patrick, and Bioch, Cor. Nearest convex hull classification. Technical report, 2006.

Ottmann, Thomas, Schuierer, Sven, and Soundaralakshmi, Subbiah. Enumerating extreme points in higher dimensions. *Nordic Journal of Computing*, 8(2):179–192, 2001.

Settles, Burr. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1):1–114, 2012.

Thurau, Christian. Nearest archetype hull methods for large-scale data classification. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pp. 4040–4043. IEEE, 2010.

Tong, Simon and Koller, Daphne. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 2(Nov):45–66, 2001.