Research Highlight: An efficient semismooth Newton based algorithm for convex clustering

Work of YUAN Yancheng, Prof. SUN Defeng and Prof. TOH Kim Chuan

Clustering is a fundamental problem in unsupervised learning because of its importance in many applications. Popular methods like K-means, may suffer from instability as they are prone to get stuck in its local minima. Moreover, prior knowledge on the number of clusters are required for popular clustering model like K-means, which may not be available in many real applications. Recently, the sum-of-norms (SON) model (also known as clustering path), which is a convex relaxation of the hierarchical clustering model, has been proposed and attracted a lot of attention because of its success in empirical experiments. Although numerical algorithms like alternating direction method of multipliers (ADMM) and alternating minimization algorithm (AMA) have been proposed to solve the convex clustering model, it is known to be very challenging to solve large-scale problems. In this paper, we propose a semismooth Newton based augmented Lagrangian method (SSNAL) for large-scale convex clustering problems. Extensive numerical experiments on both simulated and real data demonstrate that our algorithm is highly efficient and robust for solving large-scale problems. Moreover, the numerical results also show the superior performance and scalability of our algorithm comparing to existing first-order methods. Our algorithm can be more than 100 times faster than ADMM and AMA in solving large instances.

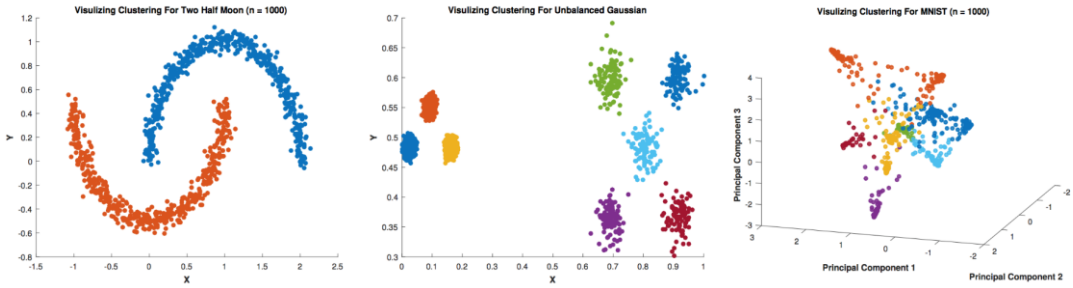

Figure 1: Selected Recovery Results by Convex Clustering Model.

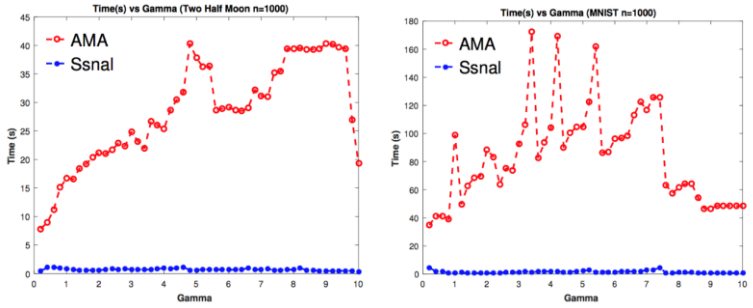

Figure 2: Stability Comparison: Computational time comparison on selected datasets for AMA and SSNAL for different parameters.

Reference:

Yancheng Yuan, Defeng Sun and Kim Chuan Toh, An efficient semismooth Newton based algorithm for convex clustering**,** Proceedings of the 35-th International Conference on Machine Learning (ICML), Stockholm, Sweden,2018.