Statistica Sinica: Preprint
doi:10.5705/ss.202015.0261

One of existing sparse clustering approaches, $\ell_1$-$k$-means, maximizes the weighted between-cluster sum of squares subject to the $\ell_1$ penalty. Although many recent efforts have been devoted to its efficient implementation, its application to high-dimensional data still poses potential challenges. One challenge is that it still unavoidably keeps a portion of noise features. In this paper, we propose a new sparse clustering method based on an $\ell_{\infty}/\ell_0$ penalty, which we call $\ell_0$-$k$-means, to reduce the number of noise features. Another challenge is the incompletion of theoretical analysis for sparse clustering, especially the absence of some basic concepts. In this paper, we introduce new definitions of an optimal partition and a noise feature for facilitating the theoretical analysis. With these definitions, we prove both $\ell_1$-$k$-means and $\ell_0$-$k$-means have a screening consistent property under proper conditions. These conditions indicate that $\ell_0$-$k$-means has better feature selection performance than $\ell_1$-$k$-means. Experiments on synthetic as well as real data justify the outperforming results of our proposed method.