High-Dimensional Clustering

via Features Segmentation and Fusion

Yu-Han He (何郁涵)
國立高雄大學統計學研究所

摘要

This paper proposes a clustering method based on feature segmentation and fusion, denoted by SFC. We divide the features into several groups and apply the clustering method with the features in each feature group separately for the subjects. A fusion method is used to fuse the clustering results. Under mild conditions, we prove that SFC has the advantage of improving the clustering effect compared to using all features for clustering at one time. In addition, the SFC can employ parallel computing when conducting clustering in each feature group to save computational time. We design several simulation scenarios and apply the SFC to four scRNA-seq gene datasets to investigate the performance of the proposed method. The numerical results reveal that the SFC obtains better clustering performances than competitors.

Keywords：feature segmentation, fusion, clustering