



Incremental Clustering: The Case for Extra Clusters

Margareta Ackerman*

Florida State University, Tallahassee, FL, USA - mackerman@fsu.edu

Sanjoy Dasgupta

UC San Diego, La Jolla, CA, USA - dasgupta@eng.ucsd.edu

The explosion in the amount of data available for analysis often necessitates a transition from batch to incremental clustering methods, which process one element at a time and typically store only a small subset of the data. We initiate the formal analysis of incremental clustering methods focusing on the types of cluster structure that they are able to detect. We find that the incremental setting is strictly weaker than the batch model, proving that a fundamental class of cluster structures that can readily be detected in the batch setting is impossible to identify using any incremental method. Furthermore, we show how the limitations of incremental clustering can be overcome by allowing additional clusters.

Keywords: clustering; incremental algorithms; online algorithms.