



COMPARISON OF CLUSTERING ALGORITHMS: AN EXAMPLE WITH PROTEOMIC DATA

**Nairanjana Dasgupta¹, Yibing Chen¹, Ananth Kalyanaraman² and
Sayed Daoud³**

¹Department of Mathematics

Washington State University

U. S. A.

²Department of Computer Sciences

Washington State University

U. S. A.

³Department of Pharmaceutical Sciences

Washington State University

U. S. A.

Abstract

Clustering has become the norm in analyzing genomic and proteomic data. There are various algorithms for which graphs and software are easily available. However, often there is little agreement among these methods; prompting the question of which algorithm should one choose? In this manuscript, we consider eleven different methods of clustering, including partitioning, hierarchical and model based methods, to look at proteomic data from a colon cancer study. We use a pairwise index to evaluate these eleven clustering methods by comparing the clustering results. In an effort to comprehensively

© 2013 Pushpa Publishing House

2010 Mathematics Subject Classification: 62P10.

Keywords and phrases: model based clustering, k -means, hierarchical clustering, colon cancer.

Communicated by Yixin Fang

Received November 7, 2012; Revised December 13, 2012