Mathematical Statistics
Stockholm University
Master Thesis **2021:1**
http://www.math.su.se

Statistical survey of clustering using message passing

Hiam Shaba[*]

February 2021

Abstract

Clustering analysis is an important part of machine learning due to the need of grouping data into segments. By handling the abundant amount of data we have today, these analyses have created new opportunities. Different fields such as social science and biology have benefited from cluster analysis and machine learning in general. However, clustering methods usually limit us to certain types of similarity measures and require to make assumptions on the structure of the data. *Affinity propagation* (AP) is a method that addresses these inconveniences. The algorithm can take nonmetric similarity graphs as input and do not require the number of clusters prespecified, which creates great opportunities in a variety of fields. This study aims to scrutinize AP, using the negative squared Euclidean distance as similarity measure, and its inputs. We will also compare it to one of the most common clustering methods, *k-means.* By investigating the method's statistical properties with different test examples, we conclude that the results from AP are similar to k-means. The results show similar clustering of imbalanced, noisy and arbitrarily shaped data. Both methods try to cluster imbalanced and arbitrarily shaped data into balanced spherically shaped clusters, and may find structure in noise when clustering noisy data. Moreover, AP is computationally expensive in computer time when dealing with large datasets, since it needs to be run with multiple self-similarities to find a suitable value. To find the right self-similarity the original authors of AP applied a root-finding method called the bisection method, which is slow. For further studies, we therefore suggest using a faster root-finding method than the bisection method, to increase the efficiency when searching for the right self-similarity.

[*]Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: hiam.shaba@hotmail.com. Supervisor: Chun-Biu Li.