EM v.s. DBSCAN

Miguel Alcón Doganoc
Universitat Politècnica de Catalunya
Barcelona, Spain
miguel.alcon@est.fib.upc.edu

ABSTRACT

1 INTRODUCTION

2 ALGORITHMS

2.1 EM

Expectation-maximization algorithm (EM) [1, 2] is an iterative method to find maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables. The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step.

2.2 DBSCAN

Density-based spatial clustering of applications with noise (DBSCAN) is a data clustering algorithm. Specifically, it is a density-based clustering non-parametric algorithm. So, given a set of points in some space, it groups together points that are closely packed together (points with many nearby neighbors), marking as outliers points that lie alone in low-density regions (whose nearest neighbors are too far away).

Given ϵ and $minPts$, the DBSCAN algorithm can be abstracted into the following steps:

(1) Find the points in the ϵ neighborhood of every point, and identify the core points with more than $minPts$ neighbors.
(2) Find the connected components of core points on the neighbor graph, ignoring all non-core points.
(3) Assign each non-core point to a nearby cluster if the cluster is an ϵ neighbor, otherwise assign it to noise.

3 DATA SET

4 EXPERIMENTATION

5 IMPLEMENTATION

6 CONCLUSIONS

REFERENCES

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.
[2] Wikipedia. Expectation–maximization algorithm. https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm, 2019.