Problem Statement and Goals - Gaussian Mixture Model

Wong, Kim Ying

19 Jan 2024

0.1 Problem

We implement a Gaussian mixture model for clustering based on EM algorithm. Given a certain dataset with labeled observations, the software should be able to clusters the data points into the categories that they belongs to in an ideal case.

0.2 Inputs and Outputs

The input of the software will be observations (only predictors without the labels) in the dataset, assumed that it should be arrays or matrices. The output of the software should be an array of the predicted labels and could be verified with the given labels in dataset.

0.3 Stakeholders

The software can serve primarily as a tool for data scientists and researchers in the field of statistical learning. It provides a simple and fast implementation in Gaussian mixture model and could also facilitates

0.4 Environment

We aim at a portable software, which could be operated in any OS and any hardware. No software or extra libraries should be needed in an ideal case, while if the computational or development cost is too high, some external but common libraries in linear algebra or statistical method will be utilized.

0.5 Goal

Our goal is to provide a user-friendly, fast and robust software in implementing Gaussian mixture model. The software would be compared to some current libraries for statistical learning. These software will serve as a benchmark in terms of the accuracy and performance.