Homework 2

February 2023

Exercise 4: Find at least 3 examples k-mean doesn't work

- In cases of high dimensional data (M ¿ ¿ N) neither K-means, nor MAP-DP are likely to be appropriate clustering choices. Methods have been proposed that specifically handle such problems, such as a family of Gaussian mixture models that can efficiently handle high dimensional data. Since MAP-DP is derived from the nonparametric mixture model, by incorporating subspace methods into the MAP-DP mechanism, an efficient high-dimensional clustering approach can be derived using MAP-DP as a building block. We leave the detailed exposition of such extensions to MAP-DP for future work.

- The data cannot be described by an exponential family distribution. Clustering such data would involve some additional approximations and steps to extend the MAP approach. Fortunately, the exponential family is a rather rich set of distributions and is often flexible enough to achieve reasonable performance even where the data cannot be exactly described by an exponential family distribution.