

# Mixture Models and EM Algorithm

Unnikrishnan Muraleedharan, Master of Science in Computer Science

University of Dublin, Trinity College, 2022

Supervisor: Jason Wyse

The main idea of this work is to understand the importance of clustering. Unsupervised Learning is essential as most of the data available in the real world are not labelled. This dissertation mainly focuses on the functioning of the EM Algorithm on Mixture Models. It also discusses the various clustering techniques like KMeans and KMeans++ algorithm and their main differences. It also establishes how KMeans can be said to be a rudimentary form of EM Algorithm. Further, the importance of initialisation in the EM Algorithm is shown by implementing the algorithm on the same dataset using different initialisations. Performance is measured with the help of the number of iterations for convergence and agreeability using a synthetic data generated. KMeans++ was observed to have better performance. The actual usage of clustering is demonstrated by applying it to a real-world dataset and observing the results. The BIC(Bayesian Information Criteria) value was used to find the ideal clusters in the data.