Guided Projects Artificial Intelligence & Machine Learning

Guided Projects: Unsupervised Learning

Gaussian Mixture Models: Bag of Words Representation

A **Gaussian mixture model is a probabilistic model** that assumes all the data points are generated from a mixture of a **finite number of Gaussian distributions** with unknown parameters. It attempts to find a mixture of **multi-dimensional Gaussian probability distributions** that best model any input dataset allowing the model to learn automatically, i.e. in an unsupervised manner. **The bag-of-words model** is a way of representing text data when **modelling text** with machine learning algorithms which can be combined with **GMM** to get a useful model representation.

Question:

Using a **gaussian mixture model**, perform a simple clustering on the given **2D Dataset**. Try to find the optimal number of clusters using python (you may use any module to implement this). Now implement the same from scratch using python and a dummy dataset generated using **scikit learn dataset** generating functions such as **make blob.**

Dataset Link: Clustering_GMM

https://cdn.analyticsvidhya.com/wp-content/uploads/2019/10/Clustering_gmm.csv