EE798L: Machine Learning in Wireless Communications   **EE Dept. IITK**

MATLAB Assignment-3

EM for Gaussian mixture modelling (10 marks)   *March. 2024*

We will code EM algorithm for Gaussian mixture model. For this, we will first generate synthetic data set, and then calculate the update expressions.

1. *Generation of synthetic data set*: Assume that the two-dimensional dataset consists of two Gaussians with following mean and covariances

   (a) $\boldsymbol{\mu}_1 = [3,3]^T, \boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$.

   (b) $\boldsymbol{\mu}_2 = [1,-3]^T, \boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$.

   We assume that component 1 is more likely than component 2, and will use $\pi_1 = 0.8$ and $\pi_2 = 0.2$. Generate $N = 500$ points for the above dataset.

2. Using the above data set, and by assuming $K = 2$, iteratively perform the following EM algorithm steps.

   - Step-1: update current estimates of model components π_k, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ by fixing assignment probabilities q_{nk}. For the first iteration, randomly choose them.[1]
   - Step-2: update assignments q_{nk} to reflect the new values of the model parameters.

3. Iteratively perform the Step 1 and Step 2, till the lower bound converges. After convergence, assign the dataset points to the respective clusters. Also, satisfy yourself that the values of π_k, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are not very different from the values used while generating dataset.

---

[1]Do not choose the actual values used above in generating the dataset in the first iteration. We want to start with some random values, and then see that they converge to the actual ones at the convergence.

Please follow these Coding instructions:

- Properly comment your code.

- The code should execute and generate the desired output.

- Your submission should be self-contained (should include all the files required for running it).

- Avoid hard-coding the values of the variables for specific configurations. The code should be generic.

Please follow these submission instructions

- Deadline is 8th of April, 11:59 pm.

- All codes should be in one .zip folder. Please do not submit separate files.

- Upload your properly commented in drive link given below. Name your code as rollno.zip.

    – https://tinyurl.com/yhjyvnku

- Please submit one final zip file.

- Please do not mail your file to me.

Please also read this carefully.

- Each one of you have to individually do all the reading and MATLAB assignments. You can discuss with your friends but you will have to completely write your own code.

- Copying also means sharing your code with some else for them to copy. We will not differentiate between the two acts, and both such cases will be awarded zero. Our decision will be final.