

CS 328 HW 2

Due: 18/03/2020

The problem numbers below refer to [this linked version of BHK book](#). The usual agreement with respect to collaboration being okay (mention names of collaborators) but not copying. Other aspects of honor code are also to be followed. All code needs to be written by self. Code to be submitted using colab as usual.

1. In this experiment, all code needs to be implemented from scratch.

Generate a mixture of two Gaussians in \mathbb{R}^{32} one with center $\mu_1 = (0, \dots, 0)$ and the other with center $\mu_2 = c \times N(0, 1) \times (1, \dots, 1)$ (we will vary the value of c . Note that μ_2 should be generated only once for each value of c . Each distribution has weights 0.5 and variance 1. First write code to generate 10k points from this mixture distribution. Repeat this experiment for c in $[0.5, 1, 1.5, 2, 3, 4, 8]$.

- (a) For each of the above datasets, use EM to fit the data using 2 Gaussians. Let $\hat{\mu}_1$ and $\hat{\mu}_2$ be the two centers that EM returned after 50 iterations. Report $\|\mu_1 - \hat{\mu}_1\|_2 + \|\mu_2 - \hat{\mu}_2\|_2$.
- (b) For each of the above datasets, also run kmeans++, implemented from scratch with $k = 2$. Run 5 iterations of Lloyd's algorithm after center initializations.

Report two tables, one for each algorithm, for the following error measure. Let $\hat{\mu}_1$ and $\hat{\mu}_2$ be the two centers that your algorithm (EM or kmeans++) returned. Report $\|\mu_1 - \hat{\mu}_1\|_2 + \|\mu_2 - \hat{\mu}_2\|_2$ for each of the datasets. Each table should have a row for each c value.

2. Problem 3.11.
3. Problem 3.12.
4. Problem 3.26.
5. Problem 3.28. Use [this picture](#). Should be done in python, feel free to use OpenCV or matplotlib library.