

Homework 9, due November 4th, 11:59pm

October 29, 2020

In this homework, you are required to include in your report the code that you implemented. If you use some code from the web or package, also mention in your report the origin of the code.

1. Implement the EM algorithm for clustering with multivariate Gaussian models discussed in the LearningGM section. Initialize the algorithm with the K-means result (for example Matlab provides a `kmeans` function). Assume there are only two clusters. Report the parameters $\pi_k, \mu_k, \Sigma_k, k = 1, 2$ for the following datasets containing 500-600 observations from \mathbb{R}^2 :

- a) The dataset `xeasy`. (1 point)
- b) The dataset `x1`. (1 point)
- c) The dataset `x2`. (2 points)

Pay attention to the dimensions of the mean centers $\mu_k, k = 1, 2$ and the covariance matrices $\Sigma_k, k = 1, 2$.

2. Now implement the two-step EM algorithm (also known as provable EM) and repeat the questions from problem 1. (3 points)