Probabilistic Unsupervised Learning
Exercise 4: EM for Gaussian Mixture Models

Submission deadline: Monday, December 20, 2021 at 10:00 a.m.

In the programming exercises, support is only
provided for Matlab or Python source code.

1. Consider again the following generative model of the previous exercise sheet:

$$p(c|\Theta) = \pi_c \tag{1}$$

$$p(x|c,\Theta) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\Big(-\frac{1}{2}\frac{(x-\mu_c)^2}{\sigma_c^2}\Big) \tag{2}$$

where $c \in \{1,2\}$, and the parameters Θ are given by $\Theta = (\pi_1, \mu_1, \sigma_1^2, \pi_2, \mu_2, \sigma_2^2)$, with $\pi_1, \pi_2 \in [0,1]$, $\mu_1, \mu_2 \in \mathbb{R}$ and $\sigma_1^2, \sigma_2^2 \in (0,\infty)$. Note that π_1 and π_2 are not independent as we demand that $\pi_1 + \pi_2 = 1$.

[1 *point*] Task A:
Using the generative parameters Θ^{gen} given by $\pi_1 = 0.4$, $\mu_1 = -2$, $\sigma_1^2 = 1$, $\pi_2 = 0.6$, $\mu_2 = 4$, $\sigma_2^2 = 10$, generate $N = 200$ data points. (Hint: you can reuse your code from the last exercise sheet.)

[6 *points*] Task B:
Choose the initial parameters Θ of the generative model to be: $\pi_1 = 0.5$, $\mu_1 = -5$, $\sigma_1^2 = 1$, $\pi_2 = 0.5$, $\mu_2 = 5$, $\sigma_2^2 = 10$. Starting with these parameters, maximize the log-likelihood of the $N = 200$ data points generated in (A) under the generative model above (eqs. 1 and 2). Use the EM algorithm for Gaussian mixture models to do so, i.e., update the parameters Θ using the E- and the M-steps as given in the lecture. After each iteration:

- Compute the likelihood and store it in an array in order to plot the likelihood value vs number of iterations graph after convergence.
- Plot the pdf $p(x|\Theta)$ and compare it to the plot of the pdf and the histogram from the last exercise sheet (tasks 1B and 1C). (Hint: you can reuse your code from the last exercise sheet and plot the pdf into the same window.)

After convergence, compare the learned parameters Θ with the generative parameters Θ^{gen}. Also plot a graph showing the evolution of the likelihood vs. the number of iterations. Remember: the likelihood cannot decrease from one iteration to the next.

[2 *points*] Task C:
Use much less data points (down to $N = 5$) and maximise the data likelihood as in B. Repeat the maximisation for ten different sets of $N = 5$ data points. What do you observe? Do you experience numerical problems? If yes, why?

[2 *points*] Task D:
For another $N = 200$ data points generated as in A use different initial parameters and repeat the likelihood maximisation as described in B. What do you observe? Experiment with different sets of data points and initial conditions, especially with initial conditions very different from Θ^{gen}.

[3 *points*] Task E:
Use the solution you obtain with your EM algorithm to compute the decision boundaries that minimize the classification error. Plot the decision boundaries on the $p(x|\Theta)$ plot from Exercise 3.

2. [15 *bonus points*]
A word of warning: This may be quite a lot of programming work, so only do it if you find it entertaining. The deadline for this bonus task is Monday, January 10, 2022, at 10 a.m.

Generalize your implementation of the previous task for a D dimensional observed space (data space) and C classes. Run the algorithm for $C = 5$, $D = 2$, $N = 1000$, and for generative parameters of your choice. Run the algorithm ten times using different (possibly random) initial values for your parameters in each run (but keep the same set of N data points). Do you observe the same result for each run? If not, why not, and how would you pick the "best" run?