Introduction to Probabilistic Graphical Models
Practical Session 2

Umut Şimşekli
Télécom ParisTech, Université Paris-Saclay, Paris, France
umut.simsekli@telecom-paristech.fr

Instructions: The codes should be written in Python and no scanned papers will be accepted. Each student should submit an individual report (no group work allowed). Put all your files (code and report – could be a single notebook file) in a zip file and submit it through moodle before November 6 2018, 23:50. Late submissions will not be accepted.

EM for GMMs

Question 1

In Gaussian Mixture Models, we often need to evaluate quantities that are defined as follows:

$$\gamma_i(x) = \frac{\pi_i \mathcal{N}(x; \mu_i, \Sigma_i)}{\sum_{j=1}^K \pi_j \mathcal{N}(x; \mu_j, \Sigma_j)},$$

where $\pi_i \in [0, 1]$ and $\mathcal{N}$ denotes the multivariate Gaussian distribution. A direct computation of these quantities might be problematic in practice since all the terms $(\pi_i \mathcal{N}(x; \mu_i, \Sigma_i))$ might be very small, and we might end up with 0/0. Derive mathematically (don't try to take the derivative!) and implement a function for numerically stable computation of $\{\gamma_i\}_{i=1}^K$.

Hint: first compute $\ell_i = \log \pi_i \mathcal{N}(x; \mu_i, \Sigma_i)$ in a numerically stable way (be careful when computing $\log \det \Sigma_i$). Then use a trick similar to the one we used for 'log_sum_exp'.

Question 2

Let us consider a Gaussian Mixture Model (GMM), given as follows:

$$p(x_n) = \sum_{i=1}^K \pi_i \mathcal{N}(x_n; \mu_i, \Sigma_i). \tag{1}$$

where $\{x_n\}_{n=1}^N$ is a set of observed data points. Derive the M-Step of the Expectation-Maximization algorithm for this model, to find $\pi_{1:K}^{(t+1)}, \mu_{1:K}^{(t+1)}, \Sigma_{1:K}^{(t+1)}$, where t denotes the iteration number.

Question 3

Consider the model given in Equation 1. Set $K = 3$, $\pi_1 = 0.3$, $\pi_2 = 0.2$, $\pi_3 = 0.5$, $\mu_1 = [0; 0]$, $\mu_2 = [1; 2]$, $\mu_3 = [2; 0]$, $\Sigma_1 = [1.00, -0.25; -0.25, 0.50]$, $\Sigma_2 = [0.50, 0.25; 0.25, 0.50]$, $\Sigma_3 = [0.50, -0.25; -0.25, 1]$.

1. Generate a dataset $\{x_n\}_{n=1}^N$ by using the model definition (set $N = 1000$). Visualize the dataset.
2. Implement the EM algorithm for GMMs (be careful about numerical stability!).

(a) Forget about the true parameters $\pi_{1:K}$, $\mu_{1:K}$, and $\Sigma_{1:K}$ for now. By only considering the dataset $\{x_n\}_{n=1}^N$ that is generated in the previous step, run the EM algorithm after randomly initializing the parameter estimates $\pi_{1:K}^{(0)}$, $\mu_{1:K}^{(0)}$, and $\Sigma_{1:K}^{(0)}$. Visualize the intermediate results by plotting the contours of the estimated Gaussians.

(b) While running the EM algorithm, compute the log-likelihood. Plot the log-likelihood vs iterations (be careful about numerical stability!).

(c) Run the EM algorithm with different initializations for $\pi_{1:K}^{(0)}$, $\mu_{1:K}^{(0)}$, and $\Sigma_{1:K}^{(0)}$. How sensitive is the algorithm for different initial values?