Master M2 MVA 2016/2017 - Graphical models
Homework 3

These exercises are due on or before Wednesday January 4th, 2017, and should be submitted on the Moodle. They can be done in groups of two students. The write-up can be in English or in French. Please submit your answers as a pdf file that you will name `MVA_DM3_<your_name>.pdf` if you worked alone or `MVA_DM3_<name1>_<name2>.pdf` with both of your names if you work as a group of two. Indicate your name(s) as well in the documents. Please submit your code as a separate zipped folder and name it `MVA_DM3_<your_name>.zip` if you worked alone or `MVA_DM3_<name1>_<name2>.zip` with both of your names if you worked as group of two. Note that your files should weight no more than 16Mb.

1 HMM - Implementation

We consider the same training data as in the previous homework, provided as the "EMGaussienne.dat" file (and we will test on the corresponding testing data from the "EMGaussienne.test" file), but this time we use an HMM model to account for the possible temporal structure of the data. The data are of the form $u_t = (x_t, y_t)$ where $u_t = (x_t, y_t) \in \mathbb{R}^2$, for $t = 1, \ldots, T$. The goal of this exercise is to implement the probabilistic inference algorithm and the EM algorithm to learn parameters as well as the Viterbi algorithm. It is recommended to make use of the code of the previous homework.

We consider the following HMM model : the chain (q_t) has $K = 4$ possible states, with an initial probability distribution $\pi \in \mathbb{R}^4$ and a probability transition matrix $A \in \mathbb{R}^{4\times 4}$, and conditionally on the current states we have observations obtained from Gaussian emission probabilities $u_t|q_t = i \sim \mathcal{N}(\mu_i, \Sigma_i)$.

1. Implement the recursions α et β seen in class (and that can be found in the polycopié as well) to compute $p(q_t|u_1, \ldots, u_T)$ and $p(q_t, q_{t+1}|u_1, \ldots, u_T)$.
2. Using the same parameters for the means and covariance matrix of the 4 Gaussians as the ones obtained in the previous homework, taking a uniform initial probability distribution π, and setting A to be the matrix with diagonal coefficients $A_{ii} = \frac{1}{2}$ and off-diagonal coefficients $A_{ij} = \frac{1}{6}$ for all $(i, j) \in \{1, \ldots, 4\}^2$, compute α_t and β_t for all t on the test data ("EMGaussienne.test" file) and compute $p(q_t|u_1, \ldots, u_T)$. Finally, represent $p(q_t|u_1, \ldots, u_T)$ for each of the 4 states as a function of time for the 100 first datapoints in the file. Note that only the 100 first points should be plotted by that filtering should be done with all the data (i.e. $T = 500$). This will be the same for the subsequent questions.

(In Matlab the command subplot might be handy to make long horizontal plots.)

3. Derive the estimation equations of the EM algorithm.

4. Implement the EM algorithm to learn the parameters of the model $(\pi, A, \mu_k, \Sigma_k, k = 1 \ldots, 4)$. The means and covariances could be initialized with the ones obtained in the previous homework. Learn the model from the training data in"EMGaussienne.dat".

5. Plot the log-likelihood on the train data "EMGaussienne.dat" and on the test data "EMGaussienne.test" as a function of the iterations of the algorithm. Comment.

6. Return in a table the values of the log-likelihoods of the Gaussian mixture models and of the HMM on the train and on the test data. Compare these values. Does it make sense to make this comparison ? Conclude. Compare these log-likelihoods as well with the log-likelihoods obtained for the different models in the previous homework.

7. Provide a description and pseudo-code for the Viterbi decoding algorithm (aka MAP inference algorithm or max-product algorithm) that estimates the most likely sequence of states, i.e. $\arg\max_q p(q_1, \ldots, q_T | y_1, \ldots, y_T)$

8. Implement Viterbi decoding. For the set of parameters learned with the EM algorithm, compute the most likely sequence of states with the Viterbi algorithm and represent the data in 2D with the cluster centers and with markers of different colors for the datapoints belonging to different classes.

9. For the datapoints in the test file "EMGaussienne.test", compute the marginal probability $p(q_t | u_1, \ldots, u_T)$ for each point to be in state $\{1, 2, 3, 4\}$ for the parameters learned on the training set. For each state plot the probability of being in that state as a function of time for the 100 first points (i.e., as a function of the datapoint index in the file).

10. For each of these same 100 points, compute their most likely state according to the marginal probability computed in the previous question. Make a plot representing the most likely state in $\{1, 2, 3, 4\}$ as function of time for these 100 points.

11. Run Viterbi on the test data. Compare the most likely sequence of states obtained for the 100 first data points with the sequence of states obtained in the previous question. Make a similar plot. Comment.

12. In this problem the number of states was known. How would you choose the number of states if you did not know it ?