HMC CS 158, Fall 2017
Problem Set 9 Exercises: EM, HMMs

Goals:
- To practice applying Expectation-Maximization to Gaussian Mixture Models.
- To practice making inferences using Hidden Markov Models.
- To consider extensions of Hidden Markov Models.

1 Gaussian Mixture Models [5 pts]

Here, we are estimating a mixture of two Gaussians via the EM algorithm. The mixture distribution over x (where $x \in \mathbb{R}$) is given by

$$P(x; \boldsymbol{\theta}) = P(1)N(x; \mu_1, \sigma_1^2) + P(2)N(x; \mu_2, \sigma_2^2).$$

The parameters for this mixture were randomized initially, and figures illustrating the EM updates were generated. However, your absent-minded instructor mixed the figures up; additionally, one figure was slipped in that does not belong. Your task is to extract the figures of successive updates and explain why your ordering makes sense from the point of view of how the EM algorithm works.

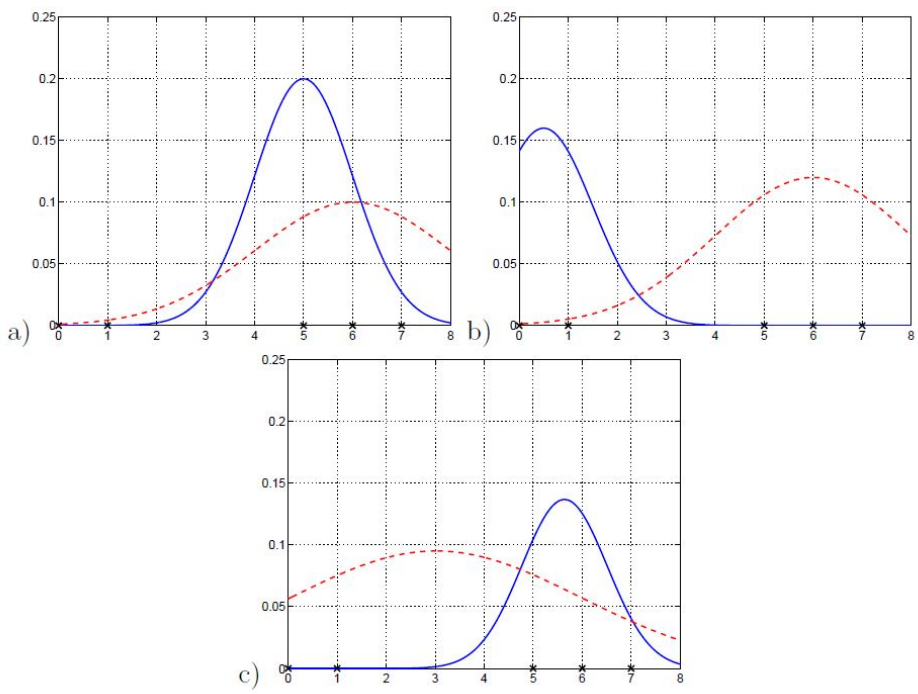

Figure 1: Mixture model with $P(1)N(x; \mu_1, \sigma_1^2)$ in solid blue and $P(2)N(x; \mu_2, \sigma_2^2)$ in dashed red. Data $\{x^{(i)}\}_{i=1}^{5} = \{0, 1, 5, 6, 7\}$ is shown as asterisks along the x-axis.

(a) **(1 pts)** For this mixture model, briefly explain how you could use the above graphs to identify the most likely posterior assignment, i.e. j that maximizes $P(j|x)$.

Parts of this assignment are adapted from course material by Tommi Jaakola (MIT) and Jenna Wiens (UMich).

(b) **(4 pts)** Assign two of figures to the correct steps in the EM algorithm, e.g.

```
Step 0: Figure ? -- Initial Mixture Distribution
Step 1: Figure ? -- Mixture Distribution after one EM Iteration
```

and briefly explain how the mixture you chose for "step 1" follows from the mixture you chose for "step 0". That is, explain both the "E-step" and the "M-step" of the EM algorithm using your chosen figures.

2 Hidden Markov Models [7 pts]

You are keen on monitoring your health over the summer. You start monitoring yourself the day after finals. After a semester's hard work, you often end up with a cold (state $= C$), though you feel fine (state $= F$) other days. Inspired by CS 158, you seek to model your health with an HMM, where there are two possible output symbols, $L = low$ or $H = high$ energy.

For this, you need to specify the model parameters $\theta = \{\pi, a, b\}$, where π is the initial state distribution, a are the state transition probabilities, and b are the output distributions. Based on past observations (and visits to the doctor's office), you determine a few of the parameters of the HMM. The initial state probabilities are

$$\pi_F = P(q_1 = F) = 0.49 \quad \text{and} \quad \pi_C = P(q_1 = C) = 0.51,$$

the state transition probabilities are

$$a_{FF} = P(q_{t+1} = F | q_t = F) = 1 \quad \text{and} \quad a_{CF} = P(q_{t+1} = F | q_t = C) = 1,$$

and the output probabilities are

$$b_F(H) = P(O_t = H | q_t = F) = 0.51 \quad \text{and} \quad b_C(L) = P(O_t = L | q_t = C) = 0.99.$$

Throughout this problem, make sure to show your work to receive full credit.

(a) **(2 pts)** There are two unspecified transition probabilities and two unspecified output probabilities. What are the missing probabilities, and what are their values?

(b) **(2 pts)** What is the most frequent output symbol (L or H) to appear in the first position of sequences generated from this HMM?

(c) **(2 pts)** What is the sequence of three output symbols that has the highest probability of being generated from this HMM model?

(d) **(1 pts)** Based on the above results, what can you infer about your summer? (We are not looking for a serious technical answer here; rather, we want you to interpret your findings above to something understandable in layman's terms.)

3 More HMMs! [8 pts]

We now ask you to learn more about HMMs by reviewing the literature – of course, this is a skill that you will use repeatedly both in college and post-college. Use whatever resources you find helpful in addressing the following questions. Provide the answer in your own words, and please cite any resources used.

Hint: You might start with Section IV of the Rabiner tutorial (on the course website). If you do, remember to cite it.

(a) How could we extend HMMs to handle continuous observations? What additional model parameters need to be specified?

(b) We showed in class that for the basic HMM, state durations follow a geometric distribution. How could we extend HMMs to handle arbitrary state duration distributions? What additional model parameters need to be specified?