IEOR E4525
M. Haugh

Due Monday May 2^{nd} 2016

Assignment 9

1. *(Barber Exercise 23.3)*
Consider an HMM with three states ($K = 3$) and two output symbols ($M = 2$), with a left-to-right state transition matrix

$$\mathbf{A} = \begin{pmatrix} 0.5 & 0.0 & 0.0 \\ 0.3 & 0.6 & 0.0 \\ 0.2 & 0.4 & 1.0 \end{pmatrix}$$

where $\mathbf{A}_{ij} := \mathrm{P}(h_{t+1} = i \,|\, h_t = j)$, emission matrix

$$\mathbf{B} = \begin{pmatrix} 0.7 & 0.4 & 0.8 \\ 0.3 & 0.6 & 0.2 \end{pmatrix}$$

where $\mathbf{B}_{ij} := \mathrm{P}(v_t = i \,|\, h_t = j)$, and initial state probability vector $\mathbf{a} = (0.9\ 0.1\ 0.0)^\top$. Suppose the observed symbol sequence is $v_{1:3} = (1, 2, 1)$.

   (a) Compute $\mathrm{P}(v_{1:3})$.

   (b) Compute $\mathrm{P}(h_1 \,|\, v_{1:3})$.

   (c) Find the most probable hidden state sequence, i.e. solve $\operatorname{argmax}_{h_{1:3}} \mathrm{P}(h_{1:3} \,|\, v_{1:3})$.

2. *(Correction Smoothing)*
In this question we will see an alternative method to compute the smoothed distribution. This method of calculating the smoothed distribution is called *correction smoothing* and it requires the α-recursion to be completed first. This is in contrast to the *parallel smoothing* algorithm that we described in class.

   (a) Show how $\mathrm{P}(h_t \,|\, h_{t+1}, v_{1:t})$ can be quickly computed after the α-recursions have been completed.

   (b) Let $\gamma(h_t) := \mathrm{P}(h_t \,|\, v_{1:T})$. Write $\gamma(h_t)$ in terms of $\gamma(h_{t+1})$ and hence derive a recursion for calculating $\gamma(h_t)$ efficiently.

3. *(Barber Exercise 23.5)*
Show that if an HMM transition matrix $\mathbf{A}$, emission matrix $\mathbf{B}$ and initial distribution are initialized to uniformly constant values, then the EM algorithm fails to update the parameters meaningfully.

4. *(The EM Algorithm with (Mixture of) Gaussian Emissions)*

   (a) Consider a HMM with Gaussian emissions. In particular, conditional on $h = h_k$, we assume the observation $\mathbf{v} \in \mathbb{R}^d$ has a multivariate normal distribution with mean μ_k and variance-covariance matrix, $\mathbf{\Sigma}_k$. Derive the EM updates for this problem.

   (b) Without doing any explicit calculations, is there any difficulty in extending the EM algorithm to the case where the emission distribution is a mixture of Gaussians? Justify your answer. (A mixture of Gaussians is used in many applications for the emission distribution.)

5. *(Extra-Credit: the N-Max Product Algorithm)*
   Write out the details of an algorithm that determines the N most likely hidden paths of a HMM given the observations $v_{1:T}$.

   Hint: Adapt the Viterbi algorithm which solves the problem for the case where $N = 1$. In the first phase of the algorithm you work backwards from time T and then at each time and state pair, (t, h_t) say, you need to store **two** $N \times 1$ vectors which can be used to determine the N most likely paths starting from (t, h_t).

   In the "back-tracking" or second phase of the algorithm you begin at time 1 and work forwards using your knowledge from the first phase. In particular at any time t in the second phase you should know the N most likely paths from time 1 up to time t. When you move on to the next time period, $t + 1$, you need to be careful!