COS 424 HW 3

1. **EM for Mixtures of Bernoullis (10)** via Kevin Murphy: Consider a set of observations $x_1, \cdots, x_N$ where the observations come from a mixture of Bernoillis:

$$p(x|\boldsymbol{\mu}, \boldsymbol{\pi}) = \sum_{k=1}^{K} \pi_k \text{Bern}(x|\mu_k)$$

(a) Draw the directed graphical model.

(b) Derive the E step.

(c) Derive the M step. Why are your results reasonable?

2. **Initialization for HMMs (5)** via Christopher Bishop: Show that if any elements of the parameters $\boldsymbol{\pi}$ or transition matrix $\mathbf{A}$ for a hidden Markov model are initially set to zero, then those elements will remain zero in all subsequent updates of the EM algorithm.

3. **EM for HMMs with mixture of Gaussian observations (15)** via Kevin Murphy: Consider an HMM where the observation model has the form

$$p(\mathbf{x}_t|z_t = j, \boldsymbol{\theta}) = \sum_{k=1}^{K} w_{jk} \mathcal{N}(\mathbf{x}_t|\mu_{jk}, \sigma_{jk})$$

(a) Draw the directed graphical model.

(b) Derive the E step.

(c) Derive the M step.

4. **Explaining Data (10)** via Serafim Batzoglou: Say I want to find a single sequence of states in an HMM that explain my data. Is it better to find the states that maximize the distribution of the states given the evidence, p(states | observations) or is it better to find the states that are most likely to have produced the observations, p(observations | states)? Why? Can you give an example where the other one fails?

5. **Flippin' transitions (10)** Consider a modified HMM where a coin is flipped and a transition is only chosen if the coin comes up heads. The probability of heads for the coin is p.

(a) What is the distribution of the time spent in a state for this modified HMM with transition matrix $\mathbf{A}$?

(b) How does this model differ from the HMM?

Open Ended (50) Apply a technique from the last two weeks to real world data and discuss your findings.