Lecture Graphical Models
`https://ml01.zrz.tu-berlin.de/wiki/Main/SS09_GraphicalModels`
Machine Learning Group, TU Berlin

Instructors: Dr. Ulf Brefeld, Dr. Marc Toussaint
Tutor: Tobias Lang, `lang@cs.tu-berlin.de`

Sheet 3

Due: 12 May 2009

Exercises 1-3 contain programming tasks. Please send your programs to `lang@cs.tu-berlin.de` and bring some printouts of your results to the tutorial so that you can explain your observations.

1. Hidden Markov Model inference (1/3)

We have $T+1$ (latent) random variables $X_0, .., X_T$ with domain $\{0, .., K-1\}$, and $T+1$ (observed) random variables $Y_0, .., Y_T$ with domain $\{0, .., K-1\}$. (To start with, choose $T = 100$ and $K = 256$.) We have the Bayesian Network

$$P(X_0, .., X_T, Y_0, .., Y_T) = P(X_0) \left[\prod_{t=1}^{T} P(X_t|X_{t-1})\right] \left[\prod_{t=0}^{T} P(Y_t|X_t)\right] . \qquad (1)$$

a) Draw this Bayes Net graphically.

We make the following specific assumptions about the CPTs:

$$P(X_0 = i) = 1/K \qquad (2)$$

$$P(X_t = i \,|\, X_{t-1} = j) = A_{ij} \quad \text{with} \quad A_{ij} \propto \exp(\frac{-|i-j|^2}{2w^2}) \qquad (3)$$

$$P(Y_t = j \,|\, X_t = i) = B_{ij} \quad \text{with} \quad B_{ij} \propto \begin{cases} 1 & \text{if } i = j \\ \epsilon & \text{if } i \neq j \end{cases} \qquad (4)$$

Note that $\propto$ means that each column needs to be normalized.

b) Write a routine that computes these matrices for parameters w and ϵ. Test your program with $w = 10$ and $\epsilon = 0.01$.

2. Hidden Markov Model inference (2/3)

Implement the forward-backward inference algorithm in the HMM of the last exercise.

Assume that we observed only $Y_0 = 0$ and $Y_T = K - 1$.

Compute the posterior marginal $P(X_t)$ for each t. To do this, first realize that $Y_1, .., Y_{T-1}$ can be eliminated and

$$P(x_0, .., x_T \,|\, y_0, y_T) \propto 1/K \Big[\prod_{t=1}^{T} A_{x_t, x_{t-1}} \Big] \; B_{y_0, x_0} \; B_{y_T, x_T} \; . \tag{5}$$

This is a pairwise factor graph over variables $X_0, .., X_T$. (In other words, before implementing things, realize that we do not need to actually represent the Y's in the code because they are either observed (induce a factor over x_t) or are eliminated.) Since this is a pair-wise factor graph, use the special case BP equations for pair-wise factor graphs. Note that, to resolve the resursive definition of messages, you need to compute forward messages $\mu_{x_t, x_{t-1}}$ iterating forward $t = 1, .., T$, and compute backward messages μ_{x_{t-1}, x_t} iterating backward $t = T, .., 1$.

Implementation: No fancy things! Write a single simple C or Java routine! Allocate a block of $T \times K$ double memory for the fwd messages, and a block of $T \times K$ double memory for the bwd messages; compute them; compute the posterior beliefs and output them in an ascii file with $(T+1) \times K$ numbers. Use gnuplot `splot 'filename' matrix` to plot this posterior. That's it.

3. Hidden Markov Model inference (3/3)

Assume we also observe $Y_{\frac{T}{2}} = 0$. How does this change our belief about the X_t, i.e., the marginals $P(X_t)$?

What happens if we choose very small w and ϵ?

4. Bonus geek question (unmarked): Relation to Gaussian Processes

Explain how the above HMM is related to a Gaussian process with a Matérn kernel.

5. Elimination ordering

Take a look at the ASIA network (see slide 3 of lecture 2). Represent this Bayesian network as a factor graph.

We are interested in the probability $P(A \,|\, D)$ and want to infer it by means of the elimination algorithm. What is a good elimination ordering for the variables? Justify your choice.