CS4786/5786: Machine Learning for Data Science, Spring 2015
4/23/2015: Lecture 23 handout: Learning the parameters of a Bayes Net (or at least an HMM)

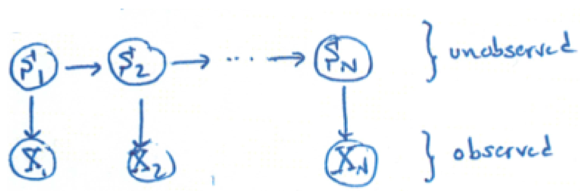
Outline

(0) so far, we've talked about doing *inference* on a Bayes Net given its associated parameters (conditional distributions) θ.

(1) learning parameter values: maximum-likelihood $\longrightarrow$ EM

(2) the case of HMMs: graphical models that are a natural fit for sequences over time, and are nearly chain-like trees

Goals Additional preparation for real-life utilization of EM "from scratch" for parameter learning on a different model
- ... where "real life" will be simulated by the upcoming A3



S_t is the state at time t, takes values from $1..K$. Example: the type of content your professor is delivering: 1 = "announcement", 2 = "important content", 3 = "optional content", 4 = "joke".
- Generic states indicated by variables i and j.

X_t is the observation at time t. Example: relative volume (loudness) of her voice at time t. So possible values $x \in \Re$.

Parameters θ:

- Trans($j \leftarrow i$): prob that given in state i, the next state is j
- Out($x \uparrow i$): prob that given in state i, the output is x. In our case, this is given by state i's associated Gaussian distribution, parametrized by
  - μ_i: mean
  - σ_i^2: variance
- Start(i): prob that the state at time 1 is i

I. Question Let $\vec{x}$ stand for $x_1, x_2, \ldots x_N$. Write down the log likelihood function $\log P(\vec{x}|\theta)$ in a form that would let you compute the partial derivative $\frac{\partial \log P(\vec{x}|\theta)}{\partial \mathrm{Trans}(4 \leftarrow 3)}$