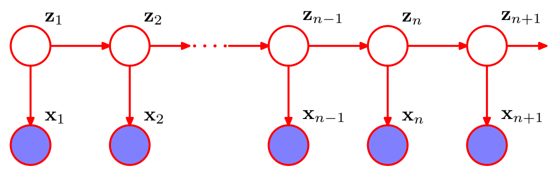
CS 613: Machine Learning	Due May 27th 2008, 6:00 pm (Instructor's mail box)

Homework #3

Lecturer: Prof. Ko Nishino — *Counts for 15% of final grade*

You must solve the following problems solely by yourself. You are expected to support all your answers with clear and detailed arguments, derivations, sketches and/or proofs. Please clearly number your solution with the corresponding problem number and staple all the solution sheets together. No late submission will be accepted without prior permission by the instructor.



Problem 1 By using d-separation, show that the distribution $p(\boldsymbol{x}_a, \ldots, \boldsymbol{x}_N)$ of the observed data for the graphical model shown on the right does not satisfy any conditional independence properties and hence does not exhibit the Markov property at any finite order.

Problem 2 Show that if any elements of the parameters $\boldsymbol{\pi}$ or $\boldsymbol{Z}$ for a hidden Markov model are initially set to zero, then those elements will remain zero in all subsequent updates of the EM algorithm.

Problem 3 Suppose we wish to train a hidden Markov model by maximum likelihood using data that comprises R independent sequences of observations, which we denote by $\boldsymbol{X}^{(r)}$ where $r = 1, \ldots, R$. Show that in the E step of the EM algorithm, we simply evaluate posterior probabilities for the latent variables by running the α and β recursions independently for each of the sequences.

Problem 4 Use the results (2.115) and (2.116), together with the matrix identities (C.5) and (C.7), to derive the results (13.89), (13.90), and (13.91), where the Kalman gain matrix $\boldsymbol{K}_n$ is defined by (13.92). (Equation numbers refer to those in the textbook.)

Problem 5 Consider a special case of the linear dynamical system in which the state variable $\boldsymbol{z}_n$ is constrained to be equal to the previous state variable, which corresponds to $\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{\Lambda} = \boldsymbol{0}$ in (13.75). For simplicity, assume also that $\boldsymbol{V}_0 \rightarrow \infty$ in (13.77) so that the initial conditions for $\boldsymbol{z}$ are unimportant, and the predictions are determined purely by the data. Use proof by induction to show that the posterior mean for state $\boldsymbol{z}_n$ is determined by the average of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. This corresponds to the intuitive results that if the state variable is constant, our best estimate is obtained by averaging the observations. (Equation numbers refer to those in the textbook.)