CSE 250A. Assignment 6

Out: *Mon Nov 19*
Due: *Wed Nov 28*

6.1 Belief updating

In this problem, you will derive recursion relations for real-time updating of beliefs based on incoming evidence. These relations are useful for situated agents that must monitor their environments in real-time.

(a) Consider the discrete hidden Markov model (HMM) with hidden states S_t, observations O_t, transition matrix a_{ij} and emission matrix b_{ik}. Let

$$q_{it} \;=\; P(S_t\!=\!i|o_1, o_2, \ldots, o_t)$$

denote the conditional probability that S_t is in the i^{th} state of the HMM based on the evidence up to and including time t. Derive the recursion relation:

$$q_{jt} = \frac{1}{Z_t} b_j(o_t) \sum_i a_{ij} q_{it-1} \quad \text{where} \quad Z_t = \sum_{ij} b_j(o_t) a_{ij} q_{it-1}.$$

Justify each step in your derivation—for example, by appealing to Bayes rule or properties of conditional independence.

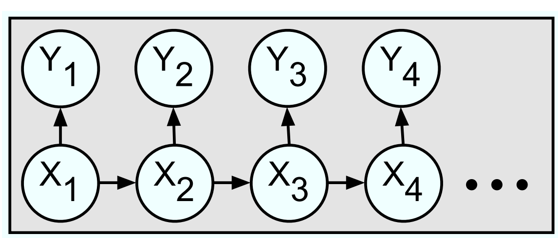
(b) Consider the dynamical system with *continuous, real-valued* hidden states X_t and observations Y_t, represented by the belief network shown below. By analogy to the previous problem (replacing sums by integrals), derive the recursion relation:

$$P(x_t|y_1, y_2, \ldots, y_t) \;=\; \frac{1}{Z_t} P(y_t|x_t) \int dx_{t-1}\, P(x_t|x_{t-1})\, P(x_{t-1}|y_1, y_2, \ldots, y_{t-1}),$$

where Z_t is the appropriate normalization factor,

$$Z_t = \int dx_t\, P(y_t|x_t) \int dx_{t-1}\, P(x_t|x_{t-1})\, P(x_{t-1}|y_1, y_2, \ldots, y_{t-1}).$$

In principle, an agent could use this recursion for real-time updating of beliefs in arbitrarily complicated continuous worlds. In practice, why is this difficult for all but Gaussian random variables?



6.2 Forward-backward algorithm

Consider a discrete HMM with hidden states S_t, observations O_t, transition matrix $a_{ij} = P(S_{t+1}=j|S_t=i)$ and emission matrix $b_{ik} = P(O_t=k|S_t=i)$. In class, we defined the quantities:

$$\begin{aligned} \alpha_{it} &= P(o_1, o_2, \ldots, o_t, S_t=i), \\ \beta_{it} &= P(o_{t+1}, o_{t+2}, \ldots, o_T|S_t=i), \end{aligned}$$

for a particular observation sequence $\{o_1, o_2, \ldots, o_T\}$ of length T. Show that the likelihood can in fact be computed from the $\alpha\beta$-values at any time t, using the formula:

$$P(o_1, o_2, \ldots, o_T) = \sum_{ij} \alpha_{it} a_{ij} b_j(o_{t+1}) \beta_{jt+1}.$$

6.3 Viterbi algorithm

Consider a discrete HMM with hidden states $S_t \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, binary observations $O_t \in \{0, 1\}$, and the following parameters for its initial state distribution, transition matrix, and emission matrix:

$$\begin{aligned} P(S_1=i) &= \frac{1}{9} \text{ for all } i, \\ P(S_{t+1}=j|S_t=i) &= \begin{cases} 0.99 & \text{for } i=j \\ 0.00125 & \text{for } i \neq j \end{cases} \\ P(O_t=1|S_t=i) &= \frac{i}{10} \end{aligned}$$

Download the ASCII file **hmm.dat** from the course web site, which contains a bit sequence $\{o_1, o_2, \ldots, o_T\}$ of $T = 10000$ observations. Use the Viterbi algorithm to compute the most likely sequence of hidden states conditioned on this particular sequence of observations. Turn in a print-out of your *source code*, as well as a *plot* of the most likely sequence of hidden states versus time. (The correct answer will reveal a highly recognizable mathematical sequence.) You may program in the language of your choice.

6.4 Continuous density HMM

In class, we studied discrete HMMs with discrete hidden states and observations, as well as linear dynamical systems with continuous hidden states and observations.

This problem considers a *continuous density* HMM, which has discrete hidden states but continuous observations. Let $S_t \in \{1, 2, ..., n\}$ denote the hidden state of the HMM at time t, and let $X_t \in \Re$ denote the real-valued scalar observation of the HMM at time t. The continuous density HMM makes the same Markov assumptions as the discrete HMM in class. In particular, the joint distribution over sequences $S = \{S_t\}_{t=1}^T$ and $X = \{X_t\}_{t=1}^T$ is given by:

$$P(S, X) = P(S_1) \prod_{t=2}^{T} P(S_t|S_{t-1}) \prod_{t=1}^{T} P(X_t|S_t).$$

In a continuous density HMM, however, the distribution $P(X_t|S_t)$ must be parameterized since the random variable X_t is no longer discrete. Suppose that the observations are modeled as Gaussian random variables:

$$P(X_t = x|S_t = i) \;=\; \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right]$$

with state-dependent means and variances. Indicate whether each of the following distributions is Gaussian (univariate or multivariate) or non-Gaussian. Briefly justify your answers.

(a) $P(X_1, X_2, \ldots, X_T|S_1, S_2, \ldots, S_T)$

(b) $P(X_1)$

(c) $P(X_t|X_1, X_2, \ldots, X_{t-1})$

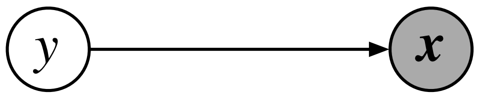
(d) $P(X_1, X_2, \ldots, X_T)$

(e) $P(X_t, X_{t'}|S_t, S_{t'})$

(f) $P(X_t|S_{t-1})$

6.5 Mixture model decision boundary

Consider a multivariate Gaussian mixture model with two mixture components. The model has a hidden binary variable $y \in \{0, 1\}$ and an observed vector variable $\vec{x} \in \mathcal{R}^d$, with graphical model:



$$P(y\!=\!i) \;=\; \pi_i \qquad\qquad P(\vec{x}|y\!=\!i) \;=\; (2\pi)^{-\frac{d}{2}}|\Sigma_i|^{-\frac{1}{2}}e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T\Sigma_i^{-1}(\vec{x}-\vec{\mu}_i)}$$

The parameters of the Gaussian mixture model are thus the prior probabilities π_0 and π_1, the mean vectors $\vec{\mu}_0$ and $\vec{\mu}_1$, and the covariance matrices Σ_0 and Σ_1.

(a) Compute the posterior distribution $P(y\!=\!1|\vec{x})$ as a function of the parameters $(\pi_0, \pi_1, \vec{\mu}_0, \vec{\mu}_1, \Sigma_0, \Sigma_1)$ of the Gaussian mixture model.

(b) Consider the special case of this model where the two mixture components share *the same* covariance matrix: namely, $\Sigma_0 = \Sigma_1 = \Sigma$. In this case, show that your answer from part (a) can be written as:

$$P(y = 1|\vec{x}) = \sigma(\vec{w} \cdot \vec{x} + b) \quad \text{where} \quad \sigma(z) = \frac{1}{1 + e^{-z}}.$$

As part of your answer, you should express the parameters $(\vec{w}, b)$ of the sigmoid function explicitly in terms of the parameters $(\pi_0, \pi_1, \vec{\mu}_0, \vec{\mu}_1, \Sigma)$ of the Gaussian mixture model.