Problem Set 4

Issued: Thursday, Oct. 11, 2018 **Due:** Tuesday, Oct. 23, 2018

Problem 4.1

Consider a hidden Markov model (HMM) with states x_i and observations y_i for $i = 1, 2, \ldots, n$ for some n. The states are binary-valued, i.e., $\mathsf{x}_i \in \{0, 1\}$. Moreover, the model is homogeneous, i.e., the potentials are given by

$$\psi_i(x_s, x_{s+1}) = \psi(x_s, x_{s+1}), \quad \psi_i'(x_s, y_s) = \psi'(x_s, y_s), \quad \text{for } i = 1, 2, \ldots, n$$

Given the observations $\mathsf{y}_i = y_i$ for $i = 1, \ldots, n$, we are interested in state estimates $\hat{x}_i(y_1, \ldots, y_n)$ for $i = 1, \ldots, n$ that maximize the probability that at least one of those state estimates $\hat{x}_i$ is correct.

(a) The desired state estimates can be expressed in the form

$$(\hat{x}_1, \ldots, \hat{x}_n) = \arg\min_{\hat{x}_1', \ldots, \hat{x}_n'} p_{\mathsf{x}_1, \ldots, \mathsf{x}_n | \mathsf{y}_1, \ldots, \mathsf{y}_n}(f(\hat{x}_1'), \ldots, f(\hat{x}_n') | y_1, \ldots, y_n).$$

Determine the function $f(\cdot)$.

(b) Show that if only the marginal distributions $p_{\mathsf{x}_i | \mathsf{y}_1, \ldots, \mathsf{y}_n}(x_i | y_1, \ldots, y_n)$, $i = 1, \ldots, n$, for the model are available, the desired state estimates cannot be determined. In particular, construct two HMMs whose marginals coincide, but whose desired state estimates differ. *Hint:* It suffices to consider a model with $n = 2$, and in which the observations are independent of the states and thus can be ignored. Accordingly, express your answer in the form of two distributions $p_{\mathsf{x}_1, \mathsf{x}_2}(\cdot, \cdot)$ and $p'_{\mathsf{x}_1, \mathsf{x}_2}(\cdot, \cdot)$.

(c) Construct an example of an HMM in which our desired estimates are not the same as the MAP estimates obtained from running the Viterbi (i.e., max-product or min-sum) algorithm on our model. The same hint as in part (b) applies, so again give your answer in the form of a distribution $p_{\mathsf{x}_1, \mathsf{x}_2}(\cdot, \cdot)$.

(d) You are given two pieces of code (e.g., matlab scripts).

The first routine implements the forward-backward algorithm, taking as input potential functions that describe a homogeneous HMM, and an associated list of n observations. It produces as output the list of marginal distributions for each of the associated n states conditioned on the full set of n observations, for the specified HMM.

The second routine implements the Viterbi algorithm, taking the same inputs as the forward-backward routine, but producing as output the most probable sequence of states x_i given the full set of n observations.

Describe how to use one or both of these routines to compute the desired estimates $\hat{\mathsf{x}}_i(y_1, \ldots, y_n)$ for $i = 1, \ldots, n$ for our model of interest, assuming the potentials are strictly positive. You are free to use these routines with any input values you like (whether or not related to the model of interest), and you can further process the outputs of these routines to compute the desired state estimates. However, in such further processing, you are not allowed to (re)use the model's potential functions or observations.

Problem 4.2 (practice)

In this problem we will use the Kalman Filter to track the location of an object traveling with constant velocity. We have noisy measurements of the velocity at regular time intervals. Let x_i^1 and x_i^2 denote the true location of the object (in 2 dimensions) at time instance $i, i \geq 0$. Similarly, let x_i^3 and x_i^4 denote the true velocity of the object in both dimensions at time instance $i, i \geq 0$. We assume that velocity does not change between time instances. We shall assume that the absolute time difference between two time instances is Δt.

(a) Let the state vector $\mathbf{x_i} = [x_i^1, x_i^2, x_i^3, x_i^4]$. Argue that

$$\mathbf{x}_i = \mathbf{A}\mathbf{x}_{i-1}$$

for some matrix A.

Due to various exogenous effect, the changes in state and velocity happens uncertainly. We shall model this aspect by modifying linear dynamical system evolution as

$$\mathbf{x}_i = \mathbf{A}\mathbf{x}_{i-1} + \mathbf{w}, \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^4$ is random variable drawn independently at each time instance and assumed to be Gaussian with mean $\mathbf{0}$ and covariance matrix $Q \in \mathbb{R}^{4\times4}$.

Our goal is to estimate the position and velocity at each time. However, we can not observe it. Instead, we can observe only velocity.

(b) Argue that the observation, $\mathbf{y}_i$ at time instance i, obeys

$$\mathbf{y}_i = \mathbf{C}\mathbf{x}_i$$

for appropriate linear matrix $C \in \mathbb{R}^{2\times4}$.

Again, due to exogenous reasons, there will be uncertainty in the observation which we model as follows:

$$\mathbf{y}_i = \mathbf{C}\mathbf{x}_i + \mathbf{v},$$

where $\mathbf{v} \in \mathbb{R}^2$ is noise random variable that is drawn independently each time instance and has Gaussian distribution with mean 0 and covariance matrix R.

The initial state, and the noise vectors at each step $\{\mathbf{x}_0, \mathbf{w}_1, , \mathbf{w}_i, \mathbf{v}_1\mathbf{v}_i\}$ are all assumed to be mutually independent. As described in the lecture notes, $\mathbf{x}_0 \sim \mathtt{N}(\mu_0, \mathbf{\Lambda_0})$, where $\mu_0 = 0$, and the Kalman Filter can then be implemented using Belief Propagation. However, note that μ_0 need not be zero and in such a scenario the development in the lecture notes will need appropriate modifications.

We are interested in **causal** estimates of the state vector. To that end,

(c) Argue that $p_{\mathbf{x}_0^N;\mathbf{y}_0^N}(\mathbf{x}_0^N;\mathbf{y}_0^N)$ follows Gaussian Hidden Markov Model. Identify the "factors" associated with the Gaussian Hidden Markov Model. Use the standard representation (i.e. means and covariances) to describe the model.

(d) Transform the description in (c) into the "information" form so as to be able to apply Belief Propagation algorithm.

(e) Write down the forward pass equations of the Belief Propagation algorithm and the equations you need to estimate the mean and covariance of the state x_i at time instance i using all the information available till time instance i, i.e. causal estimates.

(f) Now let us put the algorithm to practice. Specifically, let us assume

- Time step, $\Delta t = 0.1$
- Number of steps, $N = 100$
- Initial uncertainty, $\Lambda_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
- Initial state mean, $\mu_0 = \begin{bmatrix} 0 & 0 & 20 & 10 \end{bmatrix}^T$
- System uncertainty, $Q = \begin{bmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix}$
- Measurement uncertainties, $R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Compute the estimates of the state (i.e. position and velocity) using the forward pass equation at each time instance using information till that time (causal estimation). Compare these estimates with actual data (based on the constant velocity assumption) and plot them using your favorite plotter.

(g) Now do the same computation as in (f) using ρR in place of R for $\rho \in \{0.1, 10\}$. Based the observations of these results along with those in (f), draw the conclusion that "as noise increases the estimation becomes poorer".

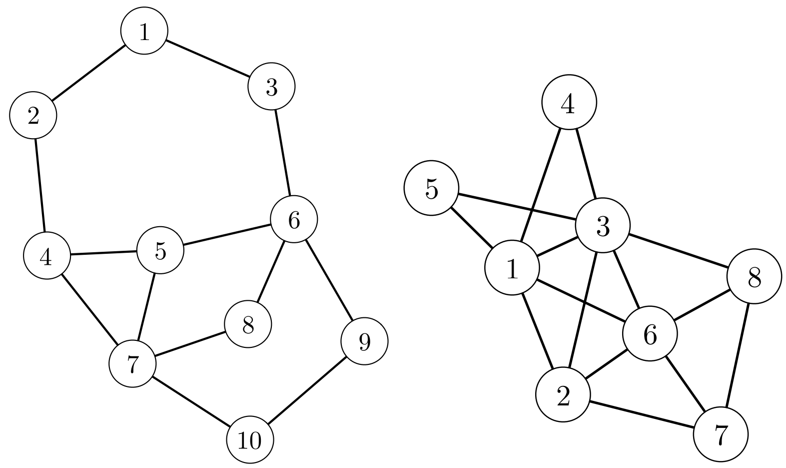
Problem 4.3

For each of the graphs in Figure 4.3, construct a junction tree by first chordalizing and then applying a greedy algorithm. You should write each step of the greedy algorithm.

Problem 4.4

Consider the undirected graphical model depicted in Figure 4.4.

(a) Treating node 2 as the root node, draw the computation tree corresponding to the first 4 iterations of loopy belief propagation.



(a) (b)

Figure 4.3

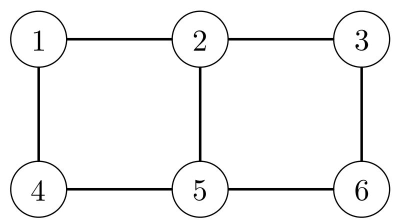


Figure 4.4

Associated with each node $i \in \{1, \ldots, 6\}$ in the graphical model is a binary random variable $\mathsf{x}_i \in \{0, 1\}$, and the joint distribution for these variables is of the form

$$p_{\mathsf{x}_1,\mathsf{x}_2,\mathsf{x}_3,\mathsf{x}_4,\mathsf{x}_5,\mathsf{x}_6}(x_1, x_2, x_3, x_4, x_5, x_6) \propto \psi_2(x_2)\, \psi_5(x_5) \prod_{(i,j)\in\mathcal{E}} \mathbb{1}_{x_i = x_j},$$

where $\mathcal{E}$ is the set of edges in the graph, and, for some $0 < \gamma < 1$,

$$\psi_2(x_2) = \begin{cases} 1-\gamma & x_2 = 0 \\ \gamma & x_2 = 1 \end{cases}, \qquad \psi_5(x_5) = \begin{cases} \gamma & x_5 = 0 \\ 1-\gamma & x_5 = 1 \end{cases}.$$

(b) Determine the sequence of marginal probabilities of x_2 obtained from loopy belief propagation, starting with its initialization, then after each of the first four iterations. Assume parallel message updates, i.e., that a new message is computed along each edge at each iteration of the algorithm.

(c) Consider the messages generated by loopy belief propagation for this graph, with the messages normalized so that the message from i to j satisfies $\sum_{x_j} m_{ij}(x_j) = 1$ for all $(i,j) \in \mathcal{E}$. Suppose

$$m_{23}(x_3) = \begin{cases} \alpha & x_3 = 0 \\ 1-\alpha & x_3 = 1 \end{cases} \qquad \text{and} \qquad m_{21}(x_1) = \begin{cases} \beta & x_1 = 0 \\ 1-\beta & x_1 = 1 \end{cases}.$$

Determine all values of α and β such that there are messages in the graph that both (i) satisfy the belief propagation message update equations everywhere and (ii) give the correct marginal probability at each node.

For messages satisfying (i) and (ii), with $m_{23}(x_3)$ and $m_{21}(x_1)$ as above, specify $m_{12}(x_2)$, $m_{52}(x_2)$, and $m_{32}(x_2)$, expressing these messages in terms of α and β.

Problem 4.5

Consider a ferromagnetic Ising model on the two dimensional grid with periodic boundary conditions (i.e. wrapped on a torus), as in Figure 4.5. Ising spins $x_i \in \{1, -1\}, i \in V$ are associated to the vertices of the grid, and interact along the edges:

$$p(x_V) \propto \exp\left(\beta \sum_{(i,j)\in E} x_i x_j\right)$$

(a) Write down the expression for loopy BP messages $m_{j\to i}(x_i)$.

(b) Look for a solution that is invariant under translation, i.e., $m_{j\to i}(x_i) = m(x_i)$ for all $(i,j) \in E$. You only need to write the equations satisfied by the function $m(\cdot)$.

(c) Let $h = \frac{1}{2\beta}\log\frac{m(+1)}{m(-1)}$ and $Z = \sum_x m(x)$. Parameterize $m(x)$ in terms of the log-likelihood h and normalizer Z. Show that h satisfies the equation $\tanh(\beta h) = \tanh(\beta)\tanh(3\beta h)$. (Hint: use the message equations you derived in (b) to cancel out Z and verify.)

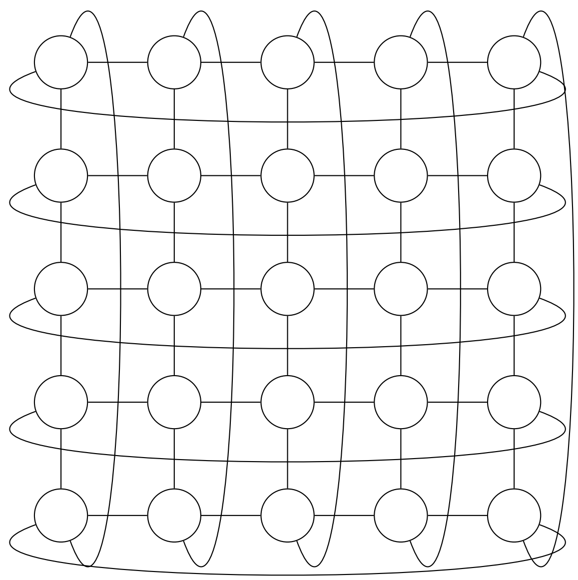


Figure 4.5

(d) Study this equation and show that, for $3\tanh(\beta) > 1$, it has three distinct solutions corresponding to three LBP fixed points.

(e) Consider iterating the LBP updates starting from a translation invariant initial condition. Does the iteration converge to a fixed point? Which one(s)?

(f) Discuss how the fixed points change as the inverse temperature β changes, and its relationship to the paramagnetic-ferromagnetic transition (macroscopically, the overall spin changes from neutral to polarized). What is the approximate value of the critical inverse temperature β_c obtained from LBP? (Remark: look up Peierls argument for a more rigorous analysis of Ising model phase transitions.)