Homework for Probabilistic Graphical Model

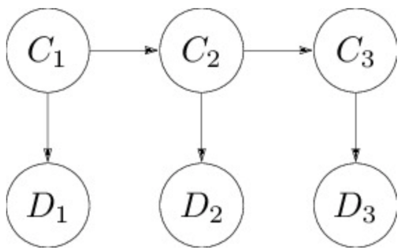
2024 Fall

Overview

Due 2024-12-24(YYYY-MM-DD, ISO 8601), 23:59, UTC+08:00
This homework includes written problems related to Lecture 12.
Please write your answers in a `pdf` file. You can use LaTeX, convert from Word, from notebook apps on tablets, scan your handwritten documents, etc. Your derivation and proof should be completed in the `pdf` file. Besides, your section titles and subsection titles should be the number of questions. i.e., # Q1; ## Q1.1.
Finally, please submit the `report_${student_id}.pdf` file to Web Learner before **2024-12-24, 23:59, UTC+08:00**.

1 Part 1 of Car Tracking (36 pts)

First, let us look at a simplified version of the car tracking problem. For this problem only, let $C_t \in \{0, 1\}$ be the actual location of the car we wish to observe at time step $t \in \{1, 2, 3\}$. Let $D_t \in \{0, 1\}$ be a sensor reading for the location of that car measured at time t. Here's what the Bayesian network (it's an HMM, in fact) looks like:



The distribution over the initial car distribution is uniform; that is, for each value $c_1 \in \{0, 1\}$: $p(c_1) = 0.5$. The following local conditional distribution governs the movement of the car (with probability ϵ, the car moves). For each $t \in \{2, 3\}$:

$$p(c_t|c_{t-1}) = \begin{cases} \epsilon, & \text{if } c_t \neq c_{t-1}, \\ 1 - \epsilon, & \text{if } c_t = c_{t-1}. \end{cases} \tag{1}$$

The following local conditional distribution governs the noise in the sensor reading (with probability η, the sensor reports the wrong position). For each $t \in \{1, 2, 3\}$:

$$p(d_t|c_t) = \begin{cases} \eta, & \text{if } d_t \neq c_t, \\ 1 - \eta, & \text{if } d_t = c_t. \end{cases} \tag{2}$$

Below, you will be asked to find the posterior distribution for the car's position at the second time step (C_2) given different sensor readings.

Important: For the following computations, try to follow the general strategy described in the lecture (marginalize non-ancestral variables, condition, and perform variable elimination). Try to delay normalization until the very end. You'll get more insight than trying to chug through lots of equations.

Q1.1 (10 pts) Suppose we have a sensor reading for the second timestep, $D_2 = 0$. Compute the posterior distribution $\mathbb{P}(C_2 = 1|D_2 = 0)$.

Q1.2 (10 pts) Suppose a time step has elapsed and we got another sensor reading, $D_3 = 1$, but we are still interested in C_2. Compute the posterior distribution $\mathbb{P}(C_2 = 1|D_2 = 0, D_3 = 1)$. Their resulting expression might be moderately complex.

Q1.3 (16 pts) Suppose $\epsilon = 0.1$ and $\eta = 0.2$.

(a) (4 pts) Compute and compare the probabilities $\mathbb{P}(C_2 = 1|D_2 = 0)$ and $\mathbb{P}(C_2 = 1|D_2 = 0, D_3 = 1)$. Give numbers, round your answer to 4 significant digits.

(b) (6 pts) How did adding the second sensor reading $D_3 = 1$ change the result? Explain your intuition for why this change makes sense in terms of the car positions and associated sensor observations.

(c) (6 pts) What would you have to set ϵ while keeping $\eta = 0.2$ so that $\mathbb{P}(C_2 = 1|D_2 = 0) = \mathbb{P}(C_2 = 1|D_2 = 0, D_3 = 1)$? Explain your intuition in terms of the car positions with respect to the observations.

2 Part 2 of Car Tracking (34 pts)

So far, we have assumed that we have a distinct noisy distance reading for each car, but in reality, our microphone would just pick up an undistinguished set of these signals, and we wouldn't know which distance reading corresponds to which car. First, let's extend the notation from before: let $C_{ti} \in \mathbb{R}^2$ be the location of the i-th car at the time step t, for $i = 1, \dots, K$ and $t = 1, \dots, T$. Recall that all the cars move independently according to the transition dynamics as before.

Let $D_{ti} \in \mathbb{R}$ be the noisy distance measurement of the i-th car, which is now not directly observed. I.e., $D_{ti} \sim \mathcal{N}(\|a_t - C_{ti}\|, \sigma^2)$, where a_t is the position of the sensor at time t . Instead, we observe the **set** of distances $D_t = \{D_{t1}, \dots, D_{tK}\}$. (For simplicity, we'll assume that all distances are distinct values.) Alternatively, you can think of $E_t = (E_{t1}, \dots, E_{tK})$ as a list which is a uniformly random permutation of the noisy distances $(D_{t1}, \dots, D_{tK})$. For example, suppose $K = 2$ and $T = 2$. Before, we might have gotten distance readings of 1 and 2 for the first car and 3 and 4 for the second car at time steps 1 and 2, respectively. Now, our sensor readings would be permutations of $\{1, 3\}$ (at time step 1) and $\{2, 4\}$ (at time step 2). Thus, even if we knew the second car was distance 3 away at time $t = 1$, we wouldn't know if it moved further away (to distance 4) or closer (to distance 2) at time $t = 2$.

Q2.1 (18 pts) Suppose we have $K = 2$ cars and one time step $T = 1$. Write an expression for the conditional 2 distribution $\mathbb{P}(C_{11}, C_{12}|E_1 = e_1)$ as a function of the PDF of a Gaussian $p_{\mathcal{N}}(\nu; \mu, \sigma^2)$ and the prior probability $p(c_{11})$ and $p(c_{12})$ over car locations. Your final answer should not contain variables d_{11}, d_{12}. Remember that $p_{\mathcal{N}}(\nu; \mu, \sigma^2)$ is the probability of a random variable, ν, in a Gaussian distribution with mean μ and standard deviation σ.

Hint: for $K = 1$, the answer would be

$$\mathbb{P}(C_{11} = c_{11}|E_1 = e_1) \propto p(c_{11}) p_{\mathcal{N}}(e_{11}; \|a_1 - c_{11}\|, \sigma^2),$$

where a_t is the position of the sensor at time t ($D_{ti} \sim \mathcal{N}(\|a_t - C_{ti}\|, \sigma^2)$). You might find it useful to draw the Bayesian network and think about the distribution of E_t given $D_{t1}, \ldots, D_{tK}$.
Q2.2 (16 pts) Assuming the prior $p(c_{1i})$ is the same for all i, please write an expression of $\mathbb{P}(C_{11} = c_{11}, \ldots, C_{1K} = c_{1K}|E_1 = e_1)$. Then show that the number of assignments for all K cars $(c_{11}, \ldots, c_{1K})$ that obtain the maximum value of $\mathbb{P}(C_{11} = c_{11}, \ldots, C_{1K} = c_{1K}|E_1 = e_1)$ is at least $K!$.
You can also assume that the car locations that maximize the probability above are unique ($C_{1i} \neq c_{1j}$ for all $i \neq j$).
Note: you don't need to produce a complicated proof for this question. It is acceptable to provide a clear explanation based on your intuitive understanding of the scenario.

3 Two definitions for Bayes Net (30 pts)

There are two definitions for Bayes Net: (there are others, e.g. d-separation.)
Factorization The joint distribution can be written as the product of the distribution of each variable, conditional on its parents.
Local Markov Each variable is conditionally independent of its non-descendants given its parent variables.
We want to show their equivalence for any given directed acyclic graphs $G = (V, E)$, namely:

$$\begin{aligned} & \forall X \in V, p(X, NonDe(X)|Pa(X)) = p(X|Pa(X))p(NonDe(X)|Pa(X)) \\ \Leftrightarrow \quad & p(V) = \prod_{X \in V} p(X|Pa(X)) \end{aligned}$$

where $NonDe(X)$ and $Pa(X)$ is the set of non-descendents and parents of X, respectively.
Note: If $p(B) = 0$, we define $p(A|B) = 0$.
Hint $\Rightarrow$: Consider a topological ordering.
Hint $\Leftarrow$: For each $X \in V$, consider a proper topological ordering.

Grading

This assignment counts for a final grade of 10 points. Following is the conversion of your homework score (implied by points after each exercise) and final grade:

$$\text{Final Grade} = \left\lceil \frac{\text{This homework score}}{100} \times 10 \right\rceil$$

We will actively be checking for plagiarism. If plagiarism is found, your final grade will be **F**.