Car - Stanford CS 221 Fall 2017-2018

Rohit Apte

2017-11-13

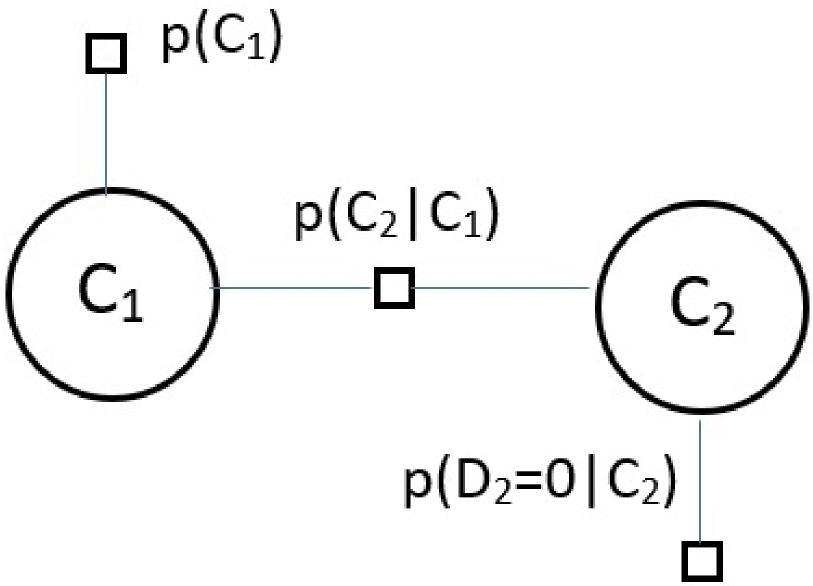

Figure 1: Factor graph for 1a

1 Problem 1: Warmup

a Suppose we have a sensor reading for the second timestep, $D_2 = 0$. Compute the posterior distribution $P(C2 = 1|D2 = 0)$. We encourage you to draw out the (factor) graph.

Lets go through the 5 steps.
Step 1 - Remove variables that are not ancestors. We can remove D_1, C_3 and D_3.
Step 2 - Convert Bayesian network into factor graph
Step 3 - Condition on the evidence ($D_2 = 0$). We can shade and disconnect D_2. We are left with the following graph.

Step 4 - Remove marginalized variables disconnected from A_2. We already removed D_2. There is nothing else to remove.
Step 5 - Do work. We will eliminate vaiables starting with C_1.
We know that

C_1	$p(C_1)$
0	0.5
1	0.5

C_1	C_2	$p(C_2\|C_1)$
0	0	1-ϵ
0	1	ϵ
1	0	ϵ
1	1	1-ϵ

Therefore the joint probability is

C_1	C_2	$p(C_1, C_2)$
0	0	0.5(1-ϵ)
0	1	0.5ϵ
1	0	0.5ϵ
1	1	0.5(1-ϵ)

Eliminating C_1 we get

C_2	$p(C_2)$
0	0.5
1	0.5

We also know that

C_2	D_2	$p(D_2\|C_2)$
0	0	1-η
0	1	η
1	0	η
1	1	1-η

Conditioning on $D_2 = 0$ we have

C_2	D_2	$p(C_2, D_2)$
0	0	0.5(1-η)
1	0	0.5η

Therefore $P(C_2 = 1|D_2 = 0) = \frac{0.5\eta}{0.5\eta+0.5(1-\eta)} = \eta$

b 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 $P(C_2 = 1|D_2 = 0, D_3 = 1)$. The resulting expression might be moderately complex. We encourage you to draw out the (factor) graph.

From 1a we know that conditioning for $D_2 = 0$

C_2	D_2	$p(C_2, D_2 = 0)$
0	0	0.5(1-η)
1	0	0.5η

and

C_2	C_3	$p(C_3\|C_2)$
0	0	(1-ϵ)
0	1	ϵ
1	0	ϵ
1	1	(1-ϵ)

Therefore the joint probability is

C_2	D_2	C_3	$p(C_2, D_2 = 0, C_3)$
0	0	0	$0.5(1-\eta)(1-\epsilon)$
0	0	1	$0.5(1-\eta)\epsilon$
1	0	0	$0.5\eta\epsilon$
1	0	1	$0.5\eta(1-\epsilon)$

Conditioning on $D_3 = 1$

C_2	D_2	C_3	D_3	$p(C_2, D_2 = 0, C_3, D_3 = 1)$
0	0	0	1	$0.5(1-\eta)(1-\epsilon)\eta$
0	0	1	1	$0.5(1-\eta)\epsilon(1-\eta)$
1	0	0	1	$0.5\eta\epsilon\eta$
1	0	1	1	$0.5\eta(1-\epsilon)(1-\eta)$

Removing $C3$

C_2	D_2	D_3	$p(C_2, D_2 = 0, C_3, D_3 = 1)$
0	0	1	$0.5(1-\eta)(1-\epsilon)\eta + 0.5(1-\eta)\epsilon(1-\eta)$
1	0	1	$0.5\eta\epsilon\eta + 0.5\eta(1-\epsilon)(1-\eta)$

Therefore $P(C_2 = 1|D_2 = 0, D_3 = 1) =$

$$\frac{0.5\eta\epsilon\eta+0.5\eta(1-\epsilon)(1-\eta)}{0.5\eta\epsilon\eta+0.5\eta(1-\epsilon)(1-\eta)+0.5(1-\eta)(1-\epsilon)\eta+0.5(1-\eta)\epsilon(1-\eta)}$$

$$= \frac{\eta[\epsilon\eta+(1-\epsilon)(1-\eta)]}{\eta[\epsilon\eta+(1-\epsilon)(1-\eta)]+(1-\eta)[(1-\epsilon)\eta+(1-\eta)\epsilon]}$$

c Suppose $\epsilon = 0.1$ and $\eta = 0.2$.

c.1 Compute and compare the probabilities $P(C_2 = 1D_2 = 0)$ and $P(C_2 = 1D_2 = 0, D3 = 1)$. Give numbers, round your answer to 4 significant digits.

$P(C_2 = 1D_2 = 0) = \eta = 0.2$
$P(C_2 = 1D_2 = 0, D3 = 1) = 0.4157$
$P(C_2 = 1D_2 = 0, D3 = 1) > P(C_2 = 1D_2 = 0)$

C_2 has a 0.5 probability of being at 1 since C_1 is either 0 or 1 with equal probability. The intuition here is that the positions are sticky ($\epsilon = 0.1$) and sensors are reliable ($\eta = 0.2$) so additional observations improve our information of the location of the object.

c.2 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.

The additional sensor reading improved our probability.
The intuition here is that the positions are sticky ($\epsilon = 0.1$) and sensors are reliable ($\eta = 0.2$) so additional observations improve our information of the location of the object.

c.3 What would you have to set ϵ while keeping $\eta = 0.2$ so that $P(C_2 = 1 D_2 = 0) = P(C_2 = 1 D_2 = 0, D_3 = 1)$.? Explain your intuition in terms of the car positions with respect to the observations.

We need to set ϵ to 0.5 while keeping $\eta = 0.2$ so that $P(C_2 = 1|D_2 = 0) = P(C_2 = 1|D_2 = 0, D_3 = 1)$. The intuition here is that the position is no longer sticky (has equal odds of being in position 0 or 1 but the sensors are still reliable with failure rate η).

2 Problem 2

3 Problem 3

4 Problem 4

5 Problem 5:Which car is it?

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 $C_{ti} \in R^2$ be the location of the i-th car at the time step t, for $i = 1, \ldots, K$ and $t = 1, \ldots, T$. Recall that all the cars move independently according to the transition dynamics as before. Let $D_{ti} \in R$ be the noisy distance measurement of the i-th car, which is now not directly observed. Instead, we observe the set of distances $D_t = \{D_{t1}, \ldots, D_{tK}\}$. (For simplicity, we'll assume that all distances are distinct values.) Alternatively, you can think of $E_t = (E_{t1} \ldots, E_{tK})$ as a list which is a uniformly random permutation of the noisy distances $(D_{t1}, \ldots, D_{tK})$. 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.

a Suppose we have $K = 2$ cars and one time step $T = 1$. Write an expression for the conditional distribution $P(C_{11}, C_{12}|E_1 = e_1)$ as a function of the PDF of a Gaussian $\rho_N(v; \mu, \sigma^2)$ 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 $\rho_N(v, \mu, \sigma^2)$ is the probability of a random variable, v, in a Gaussian distribution with mean μ and standard deviation σ.
Hint: for $K = 1$, the answer would be $P(C_{11} = c_{11}|E_1 = e1) \propto p(c_{11})\rho N(e_1; ||a_1 c_{11}||, \sigma^2)$. where a_t is the position of the car at time t. You might find it useful to draw the Bayesian network and think about the distribution of E_t given $D_{t1}, \ldots, D_{tK}$.

We know from Bayes theorem that
$P(A|B) = \frac{p(B|A)p(A)}{p(B)} \propto p(B|A)p(A)$
Therefore in our example when $K = 1$, for time $T = 1$,
$P(C_{11} = c_{11}|E_1 = e_1) \propto p(c_{11})p(E_1 = e_1|C_{11} = c11)$
Since there is only 1 car there is no ambiguity about who the signal belongs to.
$\therefore P(C_{11} = c_{11}|E_1 = e_1) \propto p(c_{11})p(E_1 = e_1|C_{11} = c11)$
$\Rightarrow P(C_{11} = c_{11}|E_1 = e_1) \propto p(c_{11})\rho_N(d_1, ||a_1 - c_{11}||, \sigma^2)$

In the case of $K = 2$ we have $D_t = \{d_{t1}, d_{t2}\}$. For $t = 1$ $D_1 = \{d_{11}, d_{12}\}$. Since we dont know which car the reading is for, $E_{1k} = \{e_{11}, e_{12}\} = \{d_{11}, d_{12}\}$ or $\{d_{12}, d_{11}\}$.
Lets say $E_{1k} = \{d_{11}, d_{12}\}$. Then we have
$P(C_{11} = c_{11}, C_{12} = c_{12}|E_{1k} = \{d_{11}, d_{12}\}) \propto p(c_{11})p(c_{12})P(E_{1k} = \{d_{11}, d_{12}\} |C_{11} = c_{11}, C_{12} = c_{12})$
Since the position of the 2 cars is independant,
$P(D_1, D_2|C_1, C_2) = P(D_1|C_1) \times P(D_2|C_2)$. Therefore we can rewrite the above expression as
$P(C_{11} = c_{11}, C_{12} = c_{12}|E_{1k} = \{d_{11}, d_{12}\}) \propto p(c_{11})p(c_{12})\rho_N(d_{11}, ||a_1 - c_{11}||, \sigma^2)\rho_N(d_{12}, ||a_2 - c_{12}||, \sigma^2)$
Similarly if $E_{1k} = \{d_{12}, d_{11}\}$ then
$P(C_{11} = c_{11}, C_{12} = c_{12}|E_1 = \{d_{12}, d_{11}\}) \propto p(c_{11})p(c_{12})\rho_N(d_{12}, ||a_1 - c_{11}||, \sigma^2)\rho_N(d_{11}, ||a_2 - c_{12}||, \sigma^2)$

Therefore since $E_1 = \{E_{11}, E_{12}\}$, for $K = 2$ we need to average over 2 instances E_{ik}. i.e.
$P(C_{11} = c_{11}, C_{12} = c_{12}|E_1) \propto \frac{1}{2}\sum_{k=1}^{2} p(c_{11})p(c_{12})\rho_N(e_{11}, ||a_1-c_{11}||, \sigma^2)\rho_N(e_{12}, ||a_2-c_{12}||, \sigma^2)$

b Assuming the prior $p(c_{1i})$ is the same for all i, show that the number of assignments for all K cars $(c_{11}, \dots, c_{1K})$ that obtain the maximum value of $P(C_{11} = c_{11}, \dots, C_{1K} = c_{1K}|E_1 = e_1)$ 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 $i \neq j$.

We know that
$P(C_{11} = c_{11}, C_{12} = c_{12}|E_1 = e_1) \propto p(c_{11})p(c_{12})\rho_N(e_{11}, ||a_1-c_{11}||, \sigma^2)\rho_N(e_{12}, ||a_2-c_{12}||, \sigma^2)$
To find the maximum assignment we need to go through the entire list E_1 to find the best value. Since there are K cars, there are $K!$ possible ways to order the distance readings $d_1, \dots, d_k$. Therefore to obtain the maximum probability we would have to go over each of these scenarios and thus there are $K!$ assignments to obtain the maximum value of $P(C_{11} = c_{11}, \dots, C_{1K} = c_{1K}|E_1 = e_1)$.

c For general K, what is the treewidth corresponding to the posterior distribution over all K car locations at all T time steps conditioned on all the sensor readings: $P(C_{11} = c_{11}, \dots, C_{1K} = c_{1K}, \dots, C_{T1} = c_{T1}, \dots, C_{TK} = c_{TK}|E_1 = e_1, \dots, E_T = e_T)$?
Briefly justify your answer.

For each time step we are computing K car locations. We have T timesteps. Therefore we have $T \times K$ factor graph. Therefore the treewidth is $min(T, K)$.

d Now suppose you change your sensors so that at each time step t, they return the list of exact positions of the K cars, but shifted (with wrap around) by a random amount. For example, if the true car positions at time step 1 are $c_{11} = 1, c_{12} = 3, c_{13} = 8, c_{14} = 5$, then e_1 would be $[1, 3, 8, 5], [3, 8, 5, 1], [8, 5, 1, 3]$, or $[5, 1, 3, 8]$, each with probability $1/4$. Describe an efficient algorithm for computing $p(c_{ti}|e_1, \ldots, e_T)$ for any time step t and car i. Your algorithm should not be exponential in K or T.

Since we have K cars that are shifted by a random amount with wrap around, we have K states each with probability $1/K$. This is a discrete uniform distribution.

We basically need to calculate the maximum likelihood estimate for each car using the points.