CS221 Fall 2018 Homework 7
SUNet ID: 05794739
Name: Luis Perez
Collaborators:

By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

Problem 1

(a) We compute $\mathbb{P}(C_2 = 1 \mid D_2 = 0)$. We note that by the factor graph, we have the following:

$$\mathbb{P}(C_2 = c_2 \mid D_2 = 0) \propto p(D_2 = 0 \mid C_2 = c_2) \sum_{c_1 \in \{0,1\}} p(C_2 = c_2 \mid C_1 = c_1)p(C_1 = c_1)$$

$$\propto p(D_2 = 0 \mid C_2 = c_2) \sum_{c_1 \in \{0,1\}} p(C_2 = c_2 \mid C_1 = c_1)$$

$(p(C_1 = c_1) = 0.5$, which we can drop since it's just a proportionality constant)

$$\propto p(D_2 = 0 \mid C_2 = c_2)$$
$(\forall c_2, \sum_{c_1} p(c_2 \mid c_1) = 1$ is a valid probability distribution)
$$\propto p(D_2 = 0 \mid C_2 = c_2)$$

Note that $p(d_2 \mid c_2)$ is a valid probability distribution, so the proportionality constant is 1. Then we have:

$$\mathbb{P}(C_2 = 0 \mid D_2 = 0) = p(D_2 = 0 \mid C_2 = 0) = 1 - \eta$$
$$\mathbb{P}(C_2 = 1 \mid D_2 = 0) = p(D_2 = 0 \mid C_2 = 1) = \eta$$

(b) We compute $\mathbb{P}(C_2 = 1 \mid D_2 = 0, D_3 = 1)$. We note that by the factor graph, we have the following:

$\mathbb{P}(C_2 = c_2 \mid D_2 = 0, D_3 = 1)$
$\propto \mathbb{P}(D_3 = 1 \mid D_2 = 0, C_2 = c_2)p(D_2 = 0 \mid C_2 = c)p(c_2 = c_2)$ (Bayes' Rule)
$\propto \mathbb{P}(D_3 = 1 \mid D_2 = 0, C_2 = c_2)p(D_2 = 0 \mid C_2 = c)$ $(\mathbb{P}(C_2 = c_2) = \frac{1}{2})$
$\propto \mathbb{P}(D_3 = 1 \mid C_2 = c_2)p(D_2 = 0 \mid C_2 = c)$ $(D_3 \perp D_2 \mid C_2)$

$$\propto \left[\sum_{c_3 \in \{0,1\}} p(D_3 = 1 \mid C_3 = c_3, C_2 = c_2)p(C_3 = c_3 \mid C_2 = c_2) \right] p(D_2 = 0 \mid C_2 = c)$$
$$\text{(LOTP)}$$

$$\propto \left[\sum_{c_3 \in \{0,1\}} p(D_3 = 1 \mid C_3 = c_3)p(C_3 = c_3 \mid C_2 = c_2) \right] p(D_2 = 0 \mid C_2 = c)$$
$$(D_3 \perp C_2 \mid C_3)$$

From the above and given the previous result, we compute directly the requested values. We just plug-in and lookup the corresponding conditional distributions:

$$\mathbb{P}(C_2 = 0 \mid D_2 = 0, D_3 = 1) \propto [\eta(1 - \epsilon) + (1 - \eta)\epsilon](1 - \eta)$$
$$\mathbb{P}(C_2 = 1 \mid D_2 = 0, D_3 = 1) \propto [\eta\epsilon + (1 - \eta)(1 - \epsilon)]\eta$$

From the abive and the fact that we must have a valid distributions, we arrive at the following:

$$\mathbb{P}(C_2 = 1 \mid D_2 = 0, D_3 = 1) = \frac{[\eta\epsilon + (1 - \eta)(1 - \epsilon)]\eta}{[\eta\epsilon + (1 - \eta)(1 - \epsilon)]\eta + [\eta(1 - \epsilon) + (1 - \eta)\epsilon](1 - \eta)}$$

(c) We now compute the probabilities requested where $\epsilon = 0.1$ and $\eta = 0.2$.

   (i) We have:
   $$\mathbb{P}(C_2 = 1 \mid D_2 = 0) = \eta = 0.2$$

   and

   $$\mathbb{P}(C_2 = 1 \mid D_2 = 0, D_3 = 1) = \frac{[0.2(0.1) + (0.8)(0.9)]0.2}{[0.2(0.1) + (0.8)(0.9)]0.2 + [0.2(0.9) + (0.8)(0.1)](0.8)} = 0.4157$$

   (ii) By adding the second sensor reading, the probability that the car was at $C_1 = 1$ increased. This change makes sense, since the second sensor reading $D_3 = 1$ sees that car at position 1. Given that our sensor error rate is low ($\eta = 0.2$), it's likely that the car is at position 1, according to this second reading. Furthermore, since cars change positions with low probability $\epsilon = 0.2$, the probability that $C_2 = 1$ must increase. However, note that the probability is still less than 0.5. This is because our original sensor reading of $D_2 = 0$ has less room for error, so it's still more likely that the positions at $t = 2$ was 0 and that the car simply moved. This is because of our relatively low η value.

   (iii) We would have to set $\epsilon = 0.5$. This is because with $\epsilon = 0.5$, at each time-step, the car has equal probabiilty of staying at the current position or alternating to a new one. As such, the sensor readings then become indendent. An additional reading of $D_3 = 1$ gives no further information about the position at $t = 2$, since the car was equally likely to have stayed or swapped positions.

Problem 2

(a) In "submission.py"

Problem 3

(a) In "submission.py"

Problem 4

(a) In "submission.py"

Problem 5

(a) We can calculate the expression directly as follows.

$$\mathbb{P}(C_{11} = c_{11}, C_{12} = c_{12} \mid E_1 = e_1)$$

$$\propto \mathbb{P}(E_1 = e_1 \mid C_{11} = c_{11}, C_{12} = c_{12})p(c_{11})p(c_{12)} \hspace{3cm} (C_{11} \perp C_{12})$$

$$\propto p(c_{11})p(c_{12}) \sum_{d_{11},d_{12}} p(E_1 = e_1 \mid D_{11} = d_{11}, D_{12} = d_{12})\mathbb{P}(D_{11} = d_{11}, D_{12} = d_{12} \mid C_{11} = c_{11}, C_{12} = c_{12})$$

$$\text{(Law of Total Probability)}$$

$$\propto \frac{1}{2!}p(c_{11})p(c_{12}) \sum_{(i,j)\in\{(1,2),(2,1)\}} \mathbb{P}(D_{11} = e_{1i}, D_{12} = e_{1j} \mid C_{11} = c_{11}, C_{12} = c_{12})$$

(Expanding out permutation distribution, noting that support exists only over permutations of e_1)

$$\propto p(c_{11})p(c_{12}) \sum_{(i,j)\in\{(1,2),(2,1)\}} p(e_{1i} \mid C_{11} = c_{11})p(e_{1j} \mid C_{12} = c_{12})$$

$$(D_{11} \perp D_{12} \mid C_{11}, C_{12})$$

$$\propto p(c_{11})p(c_{12}) \sum_{(i,j)\in\{(1,2),(2,1)\}} p_{\mathcal{N}}(e_{1i}; ||a_1 - c_{11}||, \sigma^2)p_{\mathcal{N}}(e_{1j}; ||a_1 - c_{12}||, \sigma^2)$$

$$\text{(Definition of } p(d_{1i} \mid c_{1i}))$$

(b) Suppose we have an assignment $c_1 = (c_{11}, \cdots, c_{1K})$ which maximizes the value of $\mathbb{P}(C_{11} = c_{11}, C_{1k} = c_{1k} \mid E_1 = e_1)$. We make two claims.

First, we claim that any permutation of this assignment (ie, a new assignment $c_1' = (c_{11}', \cdots, c_{1K}'')$ where c_1' is a permutation of c_1) will necessarily have the same probability value. The intuitive explanation is that this new assignment is indistinguisable to our sensors from the previous assignment. Intuitively, this permuted assignment will generate a new set of distinguishable distances $d_1' = (d_{11}', \cdots, d_{1K}')$. Given that our priors $p(c_{1i})$ are all the same, these distances will be generated from a distribution which is a permutation of the orignal distance dynamics.

However, from these distances, a set of final observations $e_1' = (e_{11}', \cdots, e_{1K}')$ will be generated which consist of a uniform re-shuffling. Now given the set-up in our problem, this set of observations e_1' is indistinguishable from the observations e_1 generated by our original assignment, since the observations are both uniformly selected random permutations of the same set of distances (since our priors are all equivalent). As such, it must be the case that all permutations of an assignment have the same probability.

We now claim that the number of assignments that will achieve the maximum value is at least $K!$. This follow naturally from the following facts:

(a) There must be at least one assignment which achieves the maximual value, let's call it c^m

(b) Any permutation of this assignment will achieve the same value (as discussed above)

(c) This assignment is such that $c^m_{1i} \neq c^m_{1j}, \forall i \neq j$.

Since the number of distinct permutations of K distinct values is $K!$, we have that the number of assignments that achieve the maximum value is at least $K!$.

(c) The treewidth will be $2K - 1$ since the no matter the elimination order, we'll need to produce a factor involving the K observations and the K distance variables (minus the elimated variable).