Homework 2 of CS 165A (Spring 2023)

University of California, Santa Barbara

To be discussed on Apr 26 and May 3, 2023 (Wednesdays)

Notes:

- The homework is optional. You do not need to submit your solutions anywhere and you will not be evaluated by these.
- To maximize your learning, you should try understanding the problems and try solving them as much as you can before the discussion class.
- Feel free to discuss with your peers / form small groups to solve these problems.
- Feel free to discuss any questions with the instructor and the TA in office hours or on Piazza.

---

1 Why should I do this homework?

This homework is given for you to practice what you learned in BayesNet (Problem 1 - 3) and in Search (Problem 4-5). In Problem 1, you will practice modeling with BayesNet.. In Problem 2, you will practice reading conditional independences from the graph. Problem 3 teaches you something about the notorious Hidden Markov Models (HMM). While Problem 3 is a challenge question, part (a) - (c) are short and highly doable.

Problem 4 asks you to write down what we have brainstormed in the lecture on the Missionary and Cannibal example. Problem 5 is a good chance to understand and practice different search algorithms by hands (something that you should perhaps expect one question in the midterm).

2 Homework problems

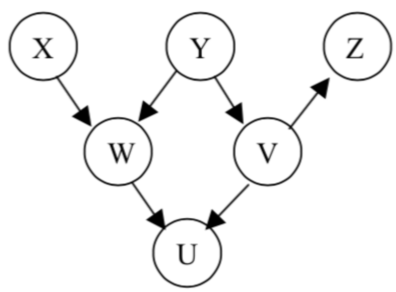
Problem 1 A patient has a probability to recover from a disease that depends on whether s/he receives the drug, how old s/he is and which gender the patient has. A doctor gives a patient a drug dependent on their age and gender. Additionally it is known that age and gender are independent.

(a) Draw the Bayesian network which describes this situation.

(b) How does the factorized probability distribution look like?

(c) Write down the formula to compute the probability that a patient recovers given that you know if s/he gets the drug. Write down the formula using only probabilities which are part of the factorized probability distribution.

Problem 2 Consider the Bayes Net below:



(a) Is it true that $P(X|Y,W) = P(X|W)$? Explain.

(b) Write down the expression for computing $P(X|Y)$ using the above Bayes Net.

(c) Are variables X,W conditionally independent of variables V,Z, given Y? Explain.

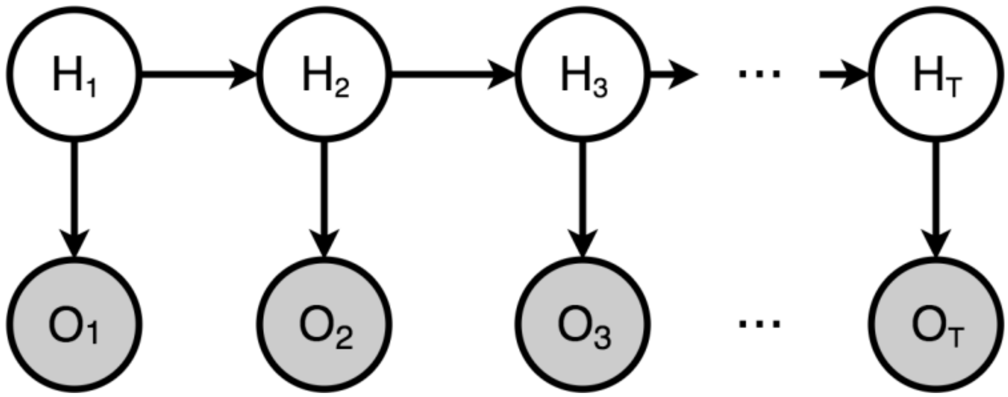
(d) Are variables X,W conditionally independent of variables V,Z, given U? Explain.

(e) Are variables W and Z independent? Explain.

(f) Write down the Markov Blanket of variable W and variable Y.

(g) Assume all the variables are binary, either take value 0 or 1. Write down the expression to compute $P(U = 1, V = 1, W = 1, X = 0, Y = 0, Z = 1)$ using notation like $P(X = 1|W = 0)$.

Problem 3 Hidden Markov Models (Challenge Problem) Let all variables be discrete. In particular, let O_i be a discrete random variable that could take d possible values, and H_i be a discrete random variable that could take k possible values.



The parameters of the HMM model are simply the CPTs of the graphical model, i.e.,

$$\begin{aligned}
&P(H_1) = \theta \in \mathbb{R}^k, \\
&P(H_{i+1}|H_i) = A \in \mathbb{R}^{k\times k} \text{ for all } i = 1, 2, 3, ..., T-1, \\
&P(O_i|H_i) = B \in \mathbb{R}^{d\times k} \text{ for all } i = 1, 2, 3, ..., T.
\end{aligned}$$

Canonically, parameter θ, A, B are called the “initial state distribution”, “transition probabilities” and “emission probabilities” in standard HMM jargon.

Convince yourself the dimensionality of these CPTs are correct.

Note that the transition and emission probabilities are *the same* for all $i = 1, ..., T$.

(a) Write down the joint probability of $P(H_1, ..., H_T, O_1, ..., O_T)$ in factorized form as function of the CPTs θ, A, B.

(b) Write down the probability distribution of the observed variables $P(O_1, ..., O_T)$ as a function of the CPTs θ, A, B.

(hint: This is identical to expressing $P(O_1, ..., O_T)$ using CPTs, but the parameters are shared. The final expression (if you use a matrix form, will be quite clean))

Remark: The above probability distribution $P(O_1, ..., O_T)$ is jointly parametrized by values of $O_1, ..., O_T$, and the values of θ, A, B. When we view it as a function of θ, A, B, while keeping $O_1, ..., O_T$ fixed, Then this function is known as the likelihood function: $L(O_1, ..., O_T; \theta, A, B)$. This measures the likelihood of observing $O_1, ..., O_T$ when the data generating distribution is specified by θ, A, B.

Given a sequence of observation $[O_1, ..., O_T] = [o_1, ..., o_T]$, the parameters A, B, θ that maximizes the likelihood, i.e.

$$[\hat{\theta}, \hat{A}, \hat{B}] = \underset{A,B,\theta}{\operatorname{argmax}}\, L(O_1 = o_1, ..., O_T = o_t; \theta, A, B)$$

is called the maximum likelihood estimator.

Solving the optimization for this MLE is not easy. It is not a convex optimization problem and we will have to use the EM algorithm to find a local optimal solution. The E-step alone requires using dynamic programming — a Forward-Backward algorithm (closely related to the more famous Viterbi algorithm). The EM solution itself is known as the Baum-Welch algorithm. Rest assured. You are *not* going to derive that in this homework.

We will take an alternative route using only things that we have learned from the class.

(c) Show (using the rules of d-separation or otherwise) that for $2 \leq i \leq T-1$, O_{i-1}, O_i, O_{i+1} are conditionally independent given H_i.

(d) Use the conditional independence in (c) to show that:

$$P(O_1, O_2, O_3) = \sum_{i=1}^{k} P(H_2 = i)P(O_1|H_2 = i)P(O_2|H_2 = i)P(O_3|H_2 = i). \qquad (1)$$

(e) Let O_1, O_2, O_3 be discrete random variables with d possible values and H_2 be a discrete random variable with k possible values.

- What is the total number of independent numbers to describe $P(H_2)$, $P(O_2|H_2)$, $P(O_1|H_2), P(O_3|H_2)$ in terms of k and d?

- Let us enumerate all combinations of O_1, O_2, O_3 in (1), how many equations do we get in total?

- Note that the LHS of (1) can be estimated from the data directly and the RHS are all unknown parameters. By solving the system of (nonlinear) equations, we can potentially identify the unknowns: $P(H_2), P(O_2|H_2), P(O_1|H_2), P(O_3|H_2)$. What is a condition on k, d such that we have enough equations to identify all unknowns variables? (Assume that we need one equation for one unknown.)

  (Hint: the number of unknown variables are the same as the number of independent parameters)

(f) If we can solve the nonlinear equations about, we can then identify

$$P(H_2), P(O_2|H_2), P(O_1|H_2), P(O_3|H_2).$$

But these are not the CPTs. If the CPTs are ultimately what we want to learn, then we need an set of equations to convert these quantities back to CPTs.

Write $P(O_2|H_2)$, $P(O_3|H_2)$ and $P(O_1|H_2)$ in terms of the model parameters (the CPTs): θ, A, B.

Problem 4 The missionaries and cannibals problem is usually stated as follows. Three missionaries and three cannibals are on one side of a river, along with a boat that can hold one or two people. Find a way to get everyone to the other side without ever leaving a group of missionaries in one place outnumbered by the cannibals in that place.

(a) Define a state representation.

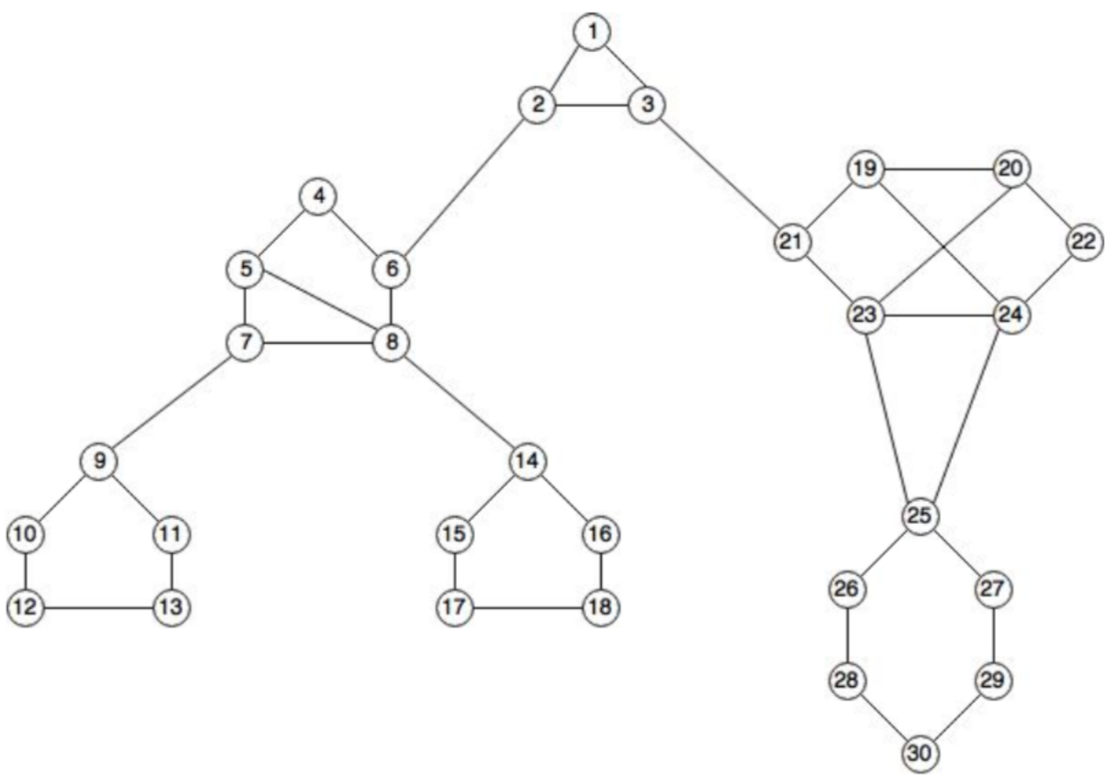
(b) Give the initial and goal states in this representation.

(c) Define the successor function (output available states that are safe) in this representation.

(d) What is the cost function in your successor function?

(e) What is the total number of safe states? Give an example of a state that is safe but unreachable?

Problem 5



Consider the state space diagram shown above. Assume state 12 is the start state and state 30 is the goal state.

1. Assuming a uniform cost of 1 on each edge, simulate the execution of BFS, DFS, IDS (assuming that the depth increases by 1 beginning from 3 to 5) and show the order of states visited. Assume that lower number children are visited first.

2. Now, simulate the execution of bidirectional search (assuming uniform cost of 1 on each edge and BFS as the basic search from each end). At which state do the two searches meet? (3')

3. Now, we consider non-uniform weights on edges. Assume that edges between even-even and odd-odd numbered edges have a cost of 1 and those between even-odd numbered edges have a cost of 2. Repeat the goal search using uniform-cost search.

4. Now, we add a heuristic h to the search. Denote states 1-3 as cluster A, 4-8 as cluster B, 9-13 as cluster C, 14-18 as cluster D, 19-24 as cluster E, and 25-30 as cluster F. Heuristic h estimates costs to the goal state 30 as follows:

   (a) h(30) = 0
   (b) h(all nodes in cluster F except 30) = 1
   (c) h(all nodes in cluster E)=2
   (d) h(all nodes in cluster A)=3
   (e) h(all nodes in cluster B) = 4
   (f) h(all nodes in cluster C)=5
   (g) h(all nodes in cluster D) = 5

   (a) Is this heuristic admissible? Prove or disprove.
   (b) Is it consistent? Prove or disprove.
   (c) If the heuristic is consistent, repeat the search for the goal state using A* (GRAPH-SEARCH).