✖

Inference Final Exam

10 试题

1
point

1.

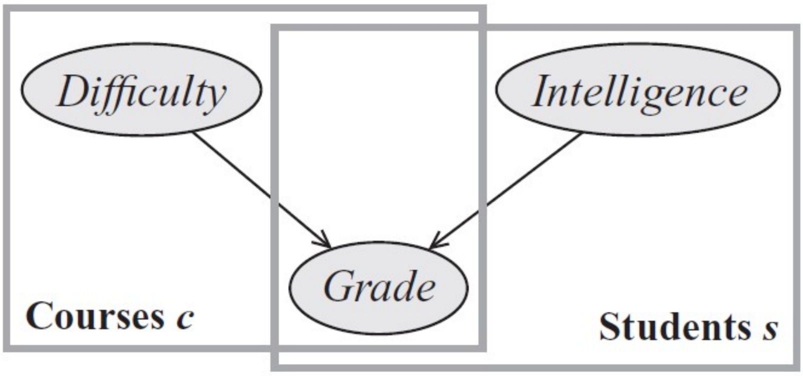
Reparameterization. Suppose we have a calibrated clique tree T and calibrated cluster graph G for the same Markov network, and have thrown away the original factors. Now we wish to reconstruct the joint distribution over all the variables in the network only from the beliefs and sepsets. Is it possible for us to do so from the beliefs and sepsets in T ? Separately, is it possible for us to do so from the beliefs and sepsets in G ?

- ○ It is not possible in T or G .
- ○ It is possible in G but not in T .
- ● It is possible in both T and G
- ○ It is possible in T but not in G .

1
point

2.

***Markov Network Construction.** Consider the unrolled network for the plate model shown below, where we have n students and m courses. Assume that we have observed the grade of all students in all courses. In general, what does a pairwise Markov network that is a minimal I-map for the conditional distribution look like? (Hint: the factors in the network are the CPDs reduced by the observed grades. We are interested in modeling the conditional distribution, so we do not need to explicitly include the Grade variables in this new network. Instead, we model their effect by appropriately choosing the factor values in the new network.)



- ○ A fully connected graph with instantiations of the Difficulty and Intelligence variables.
- ○ Impossible to tell without more information on the exact grades observed.
- ● A fully connected bipartite graph where instantiations of the Difficulty variables are on one side and instantiations of the Intelligence variables are on the other side.
- ○ A graph over instantiations of the Difficulty variables and instantiations of the Intelligence variables, not necessarily bipartite; there could be edges between different Difficulty variables, and there could also be edges between different Intelligence variables.
- ○ A bipartite graph where instantiations of the Difficulty variables are on one side and instantiations of the Intelligence variables are on the other side. In general, this graph will not be fully connected.

1
point

3.

Clique Tree Construction. Consider a pairwise Markov network that consists of a graph with m variables on one side and n on the others. This graph is bipartite but fully connected, in that each of the m variables on the one side is connected to all and only the n variables on the other side. Define the size of a clique to be the number of variables in the clique. There exists a clique tree T^* for the pairwise Markov network such that the size of the largest clique in T^* is the smallest amongst all possible clique trees for this network. What is the size of the largest sepset in T^* ?

Note: if you're wondering why we would ever care about this, remember that the complexity of inference depends on the number of entries in the largest factor produced in the course of message passing, which in turn, is affected by the size of the largest clique in the network, amongst other things.

Hint: Use the relationship between sepsets and conditional independence to derive a lower bound for the size of the largest sepset, then construct a clique tree that achieves this bound.

- ○ $m+n+1$
- ○ $m+n$
- ○ mn
- ○ $mn+1$
- ○ $\max(m,n)$
- ● $\min(m,n)$
- ○ $\min(m,n)+1$
- ○ $\max(m,n)+1$

1
point

4.
Uses of Variable Elimination. Which of the following quantities can be computed using the sum-product variable elimination algorithm? (In the options, let X be a set of query variables, and E be a set of evidence variables in the respective networks.) You may select 1 or more options.

- [ ] The most likely assignment to the variables in a Markov network.
- [ ] $P(X \mid E=e)$ in a Bayesian network
- [x] $P(X \mid E=e)$ in a Markov network
- [x] $P(X)$ in a Bayesian network

1
point

5.
***Time Complexity of Variable Elimination.** Consider a Bayesian network taking the form of a chain of n variables, $X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n$, where each of the X_i can take on k values. Assume we eliminate the X_i starting from X_2 , going to $X_3, \ldots, X_n$ and then back to X_1 . What is the computational cost of running variable elimination with this ordering?

- ( ) $O(nk)$
- ( ) $O(kn^2)$
- ( ) $O(k^n)$
- (•) $O(nk^3)$

1
point

6.

***Numerical Issues in Belief Propagation.** In practice, one of the issues that arises when we propagate messages in a *clique tree* is that when we multiply many small numbers, we quickly run into the precision limits of floating-point numbers, resulting in arithmetic underflow. One possible approach for addressing this problem is to renormalize each message, as it's passed, such that its entries sum to 1. Assume that we do not store the renormalization factor at each step. Which of the following statements describes the consequence of this approach?

- ( ) This renormalization will give rise to incorrect marginals at calibration.
- ( ) This does not change the results of the algorithm: when the clique tree is calibrated, we can obtain from it both the partition function and the correct marginals.
- ( ) Calibration will not even be achieved using this scheme.
- (●) We will be unable to extract the partition function, but the variable marginals that are obtained from renormalizing the beliefs at each clique will still be correct.

---

1 point

7.

Convergence in Belief Propagation. Suppose we ran belief propagation on a cluster graph G and a clique tree T for the same Markov network that is a perfect map for a distribution P. Assume that both G and T are valid, i.e., they satisfy family preservation and the running intersection property. Which of the following statements regarding the algorithm are true? You may select 1 or more options.

- [ ] If the algorithm converges, the final *clique* beliefs in T, when renormalized to sum to 1, are true marginals of P.
- [ ]

- [ ] Belief propagation always converges on G .
- [x] Assuming the algorithm converges, if a variable X appears in two cliques in T , the marginals $P(X)$ computed from the the two clique beliefs must agree.
- [x] Belief propagation always converges on T .

---

1 point

8.
Metropolis-Hastings Algorithm. Assume we have an $n \times n$ grid-structured MRF over the variables $X_{i,j}$. Let $\mathbf{X_i} = \{\mathbf{X_{i,1}}, \ldots, \mathbf{X_{i,n}}\}$ and $\mathbf{X_{-i}} = \mathcal{X} - \mathbf{X_i}$. Consider the following instance of the Metropolis-Hastings algorithm: at each step, we take our current assignment $\mathbf{x_{-i}}$ and use exact inference to compute the conditional probability $P(\mathbf{X_i} \mid \mathbf{x_{-i}})$. We then sample $\mathbf{x_i}'$ from this posterior distribution, and use that as our proposal. What is the correct acceptance probability for this proposal?

Hint: what is the relationship between this and Gibbs sampling?

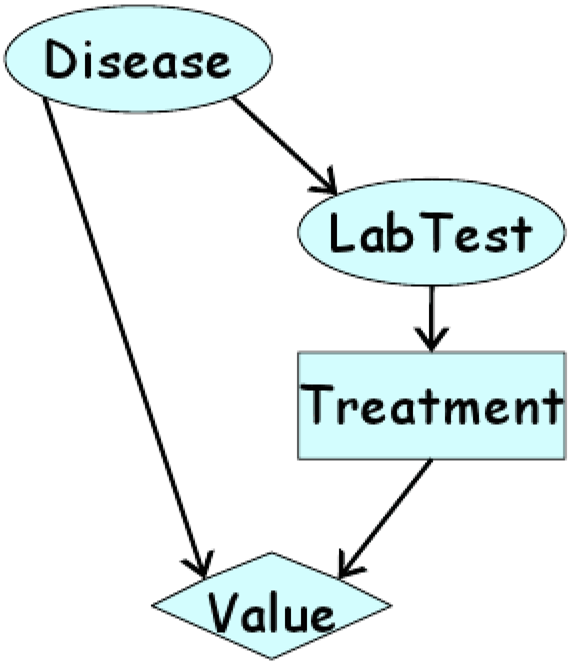
- ( ) $P(\mathbf{x_i} \mid \mathbf{x_{-i}})/P(\mathbf{x_i'} \mid \mathbf{x_{-i}})$
- (●) 1
- ( ) $P(\mathbf{x_i'}, \mathbf{x_{-i}})/P(\mathbf{x_i}, \mathbf{x_{-i}})$
- ( ) $P(\mathbf{x_i'} \mid \mathbf{x_{-i}})/P(\mathbf{x_i} \mid \mathbf{x_{-i}})$
- ( ) $P(\mathbf{x_i}, \mathbf{x_{-i}})/P(\mathbf{x_i'}, \mathbf{x_{-i}})$

---

1 point

9.

***Value of Information.** In the influence diagram on the right, when does performing LabTest have value? That is, when would you want to observe the LabTest variable?



Hint: Think about when information is valuable in making a decision.

- (●) When there is some lab value l such that $argmax_t \sum_d P(d|l)V(d,t) \neq argmax_t \sum_d P(d)V(d,t)$
- ( ) When there is some disease d such that $argmax_t V(d,t) \neq argmax_t \sum_d P(d)V(d,t)$
- ( ) When there is some treatment t such that $V(D,t)$ is different for different diseases D.
- ( ) When $P(D|L)$ is different from $P(D)$.

---

1 point

10.

***Belief Propagation.**

Say you had a probability distribution P_Φ encoded in a set of factors Φ , and that you constructed a loopy cluster graph C to do inference in it. While you were performing loopy belief propagation on this graph, lightning struck and your computer shut down; to your horror, when you booted it back up, the only information you could recover were the graph structure C and the cluster beliefs at the current iteration. (For each cluster, the cluster belief is its initial potential multiplied by all incoming messages. You don't have access to the sepset beliefs, the messages, or the original factors Φ .) Assume the lightning struck before you had finished, i.e., the graph is **not yet calibrated**. Can you still recover the original distribution P_Φ from this? Why?

- [ ] We can reconstruct the original distribution by taking the product of cluster beliefs and normalizing it.
- [ ] We can reconstruct the (unnormalized) original distribution by taking the ratio of the product of cluster beliefs to sepset beliefs, and the sepset beliefs can be obtained by marginalizing the cluster beliefs.
- [x] We can't reconstruct the (unnormalized) original distribution because we don't have the sepset beliefs to compute the ratio of the product of cluster beliefs to sepset beliefs.
- [ ] We can't reconstruct the original distribution because we were preforming loopy belief propagation, and the reparameterization property doesn't hold when it's loopy.

- [x] I, **伟臣 沈**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.
了解荣誉准则的更多信息

提交测试