Representation Final Exam

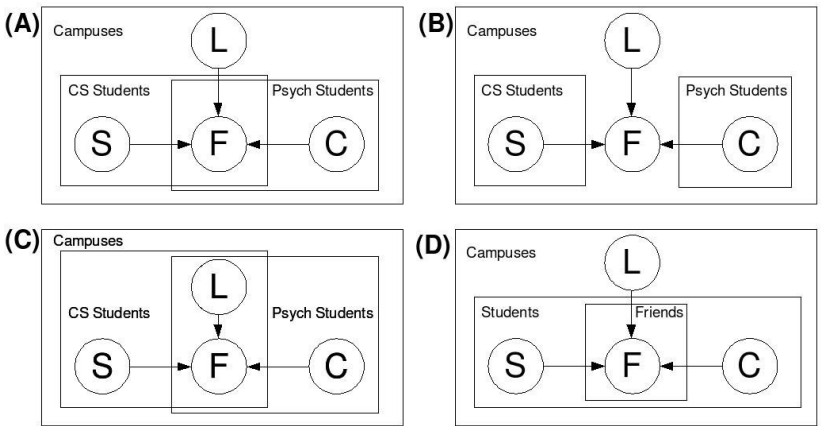
8 试题

1 point

1.

Template Model Representation. Consider the following scenario:

On each campus there are several Computer Science students and several Psychology students (each student belongs to one xor the other group). We have a binary variable L for whether the campus is large, a binary variable S for whether the CS student is shy, a binary variable C for whether the Psychology student likes computers, and a binary variable F for whether the Computer Science student is friends with the Psychology student. Which of the following plate models can represent this scenario?



- (A)
- None of these plate models can represent this

scenario

- (B)
- (C)

---

1 point

2.

Partition Function. Which of the following is a use of the partition function?

- One can divide factor products by the partition function in order to convert them into probabilities.
- One can subtract the partition function from factor products in order to convert them into probabilities.
- The partition function is useless and should be ignored
- The partition function is used only in the context of Bayesian networks, not Markov networks.

---

1 point

3.

***I-Equivalence.** Let T be any directed tree (not a polytree) over n nodes, where $n \geq 1$. A directed tree is a traditional tree, where each node has at most one parent and there is only one root, i.e., all but one node has exactly one parent. (In a polytree, nodes may have multiple parents.) How many networks (including itself) are I-equivalent to T ?

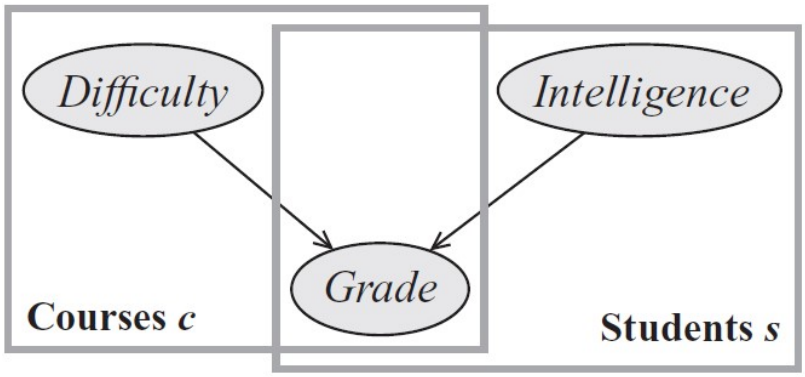
- n
- $n+1$
- $n!$

Depends on the specific structure of T .

1 point

4.

***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.

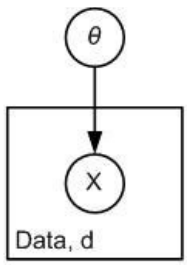
---

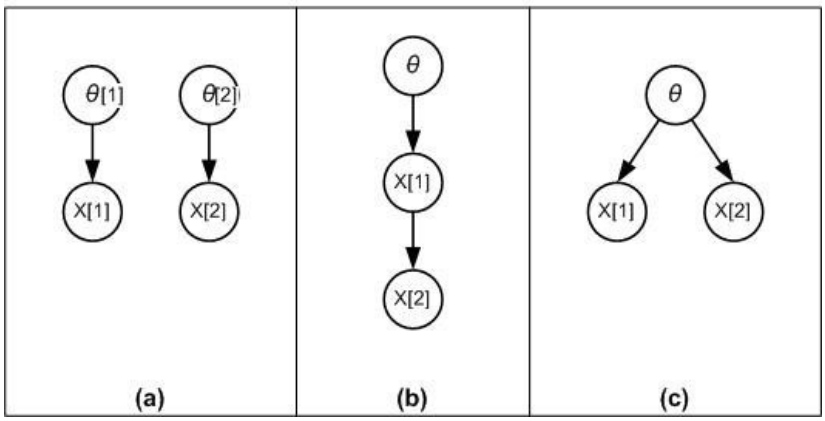
1 point

5.

Grounded Plates.

Which of the following is a valid grounded model for the plate shown? You may select 1 or more options.





- [ ] (b) -- watch out, options are not in order
- [x] (c) -- watch out, options are not in order
- [ ] (a) -- watch out, options are not in order

---

1 point

6.

Independencies in Markov Networks.

Consider the following set of factors: $\Phi = \{\Phi_1(A, B), \Phi_2(B, C, D), \Phi_3(D), \Phi_4(C, E, F)\}$. Now, consider a Markov Network G such that P_Φ factorizes over G. Which of the following is an independence statement that holds in the network? You may select 1 or more options.

- [x] $(A \perp E \mid B)$
- [x] $(A \perp F \mid C)$
- [x] $(B \perp E \mid C)$
- [ ] $(C \perp E \mid B)$
- [ ] $(B \perp E \mid A)$
- [ ] $(C \perp D \mid A)$

---

1 point

7.

Factorization of Probability Distributions.

Consider a directed graph G. We construct a new graph G' by removing one edge from G. Which of the following is always true? You may select 1 or more options.

- [x] If G and G' were undirected graphs, the answers to the other options would not change.
- [ ]

- [ ] Any probability distribution P that factorizes over G also factorizes over G' .
- [x] Any probability distribution P that factorizes over G' also factorizes over G .
- [ ] No probability distribution P that factorizes over G also factorizes over G' .

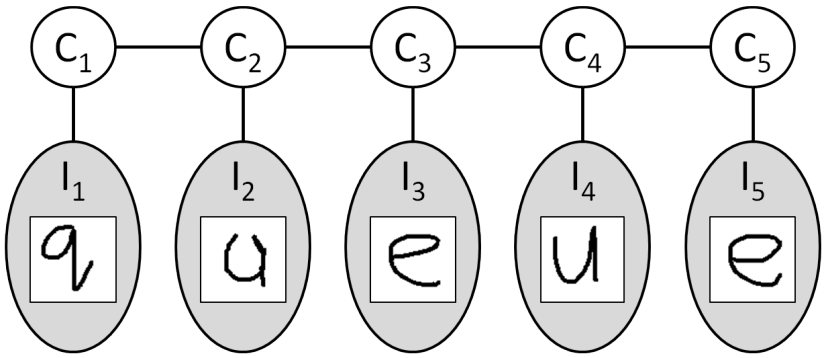
---

1 point

8.

Template Model in CRF.

The CRF model for OCR with only singleton and pairwise potentials that you played around with in PA3 and PA7 is an instance of a template model, with variables $C_1, \ldots, C_n$ over the characters and observed images $I_1, \ldots, I_n$. The model we used is a template model in that the singleton potentials are replicated across different C_i variables, and the pairwise potentials are replicated across character pairs. The structure of the model is shown below:



Now consider the advantages of this particular template model for the OCR task, as compared to a non-template model that has the same structure, but where there are distinct singleton potentials for each C_i variable, and distinct potentials for each pair of characters. Which of the following about the advantage of using a template model is true? You may select 1 or more options.

- [ ] The inference is significantly faster with the template model.
- [x]

- [ ] Parameter sharing could make the model less susceptible to over-fitting when there is less training data.
- [x] The same template model can be used for words of different lengths.
- [ ] The template model can incorporate position-specific features, e.g. q-u occurs more frequently at the beginning of a word, while a non-template model cannot.

- [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