Feedback — Tree Learning and Hill Climbing

You achieved a score of 6.00 out of 6.00

(The title of this quiz would fit right into a wilderness course.)

Question 1

Detective! You are a detective tracking down a serial murderer who has already claimed 1,000 victims, thereby giving you a large enough training set. No one else has been able to catch him (or her), but you are certain that there is a method (specifically, a Bayesian network) in this madness. You decide to model the murderer's activity with a **tree-structured network** (meaning that each node has **at most** one parent): this network has (observed) variables such as the location of the previous crime, the gender and occupation of the previous victim, the current day of the week, etc., and a single unobserved variable, which is the location of the next murder. The aim is to predict the location of the next murder.

Unfortunately, you have forgotten all of your classical graph algorithms, but fortunately, you have a copy of The Art of Computer Programming (volume 4B) next to you. Which graph algorithm do you look up to help you find the optimal tree-structured network? Assume that the structure score we are using satisfies score decomposability and score equivalence.

Your Answer		Score	Explanation
Finding the maximum-weight	✔	1.00	The tree-structured Bayesian network that we

undirected spanning forest (i.e., a set of undirected edges such that there is at most one path between any pair of nodes).		eventually want to construct is directed, However, if we have score equivalence, finding the maximum-weight undirected spanning forest is equivalent to finding the maximum-weight directed spanning forest and is easier to implement.
Total	1.00 / 1.00	

Question 2

***Recovering Directionality.** Once again, assume that our structure score satisfies score decomposability and score equivalence. After we find the optimal undirected spanning forest (containing n nodes), how can we recover the optimal directed spanning forest (and catch the murderer)?

If more than one option is correct, pick the faster option; if the options take the same amount of time, pick the more general option.

Your Answer		Score	Explanation
Pick any arbitrary root, and direct all edges away from it. This takes $O(n)$ time.	✔	1.00	No matter which root we pick, the resulting trees are in the same I-equivalence class; in fact, there are no valid directed trees that cannot be obtained with this procedure. Because of score equivalence, it does not matter which root we pick.
Total		1.00 / 1.00	

Question 3

***Augmenting Trees.** It turns out that the tree-structured network we learnt in the preceding questions was not sufficient to apprehend the murderer, allowing him to claim his 1001th victim. Not one to be discouraged, you decide to increase the expressiveness of your network.

Assume that we now want to learn a hybrid naive-Bayes/tree-structured network, where we have a single class variable C as well as the variables $X_1, \ldots, X_n$. In this model, each X_i has C as a parent, and there is also a tree connecting the X_i's; that is, each X_i, in addition to C, may also have up to one other parent X_j. For our baseline network G_0, we are going to use the naive Bayes network, in which each X_i has only C as a parent. We are thus aiming to optimize the difference in likelihood scores $\mathrm{Score}_L(G : \mathcal{D}) - \mathrm{Score}_L(G_0 : \mathcal{D})$, where D is the training dataset.

If we use the appropriate spanning tree algorithm to find the optimal forest structure, what is the correct edge weight to use for $w_{j \to i}$? In these options, $M = 1001$ is the size of our training dataset, and $I_{\hat{P}}(\mathbf{A}, \mathbf{B})$ is the mutual information in the empirical distribution of the variables in set $\mathbf{A}$ with the variables in set $\mathbf{B}$.

Your Answer		Score	Explanation
$M \cdot (I_{\hat{P}}(X_i; X_j, C) - I_{\hat{P}}(X_i; C))$	✔	1.00	$w_{j \to i} = \mathrm{FamScore}_L(X_i \| X_j, C : D) - \mathrm{Fa}$ In the case of the likelihood score, this gives us $M(I_{\hat{P}}(X_i; X_j, C) - I_{\hat{P}}(X_i; C))$, since the depend on X_i and cancel each other out.
Total		1.00 / 1.00	

Question 4

Trees vs. Forests. Congratulations! Your hybrid naive-Bayes/tree-structured network managed to correctly predict where the criminal would be next, allowing the police to catch him (or her) before the 1002th victim got attacked. The grateful populace beg you to return to studying probabilistic graphical models.

While re-watching the video lectures, you begin to wonder if the algorithm we have been using to learn tree-structured networks can produce a forest, rather than a single tree. Assume that we use the likelihood score, and also assume that the maximum spanning forest algorithm breaks ties (between equal-scoring trees) arbitrarily. Which of the following is true? In this question, interpret "forest" to mean a set of two or more disconnected trees.

Your Answer		Score	Explanation
It's theoretically possible for the algorithm to produce a forest. However, this will only occur in very contrived and unrealistic circumstances, not in practice.	✔	1.00	A forest will be produced only if we can partition the variables into two disjoint sets A and B, such that all edges $X_j \rightarrow X_i$ with either $X_i \in A, X_j \in B$ or $X_i \in B, X_j \in A$ have weight 0. This will be the case only if all variables in A are independent of the variables in B in the empirical distribution. While this is not impossible, it is very unlikely to happen in practice.
Total		1.00 / 1.00	

Question 5

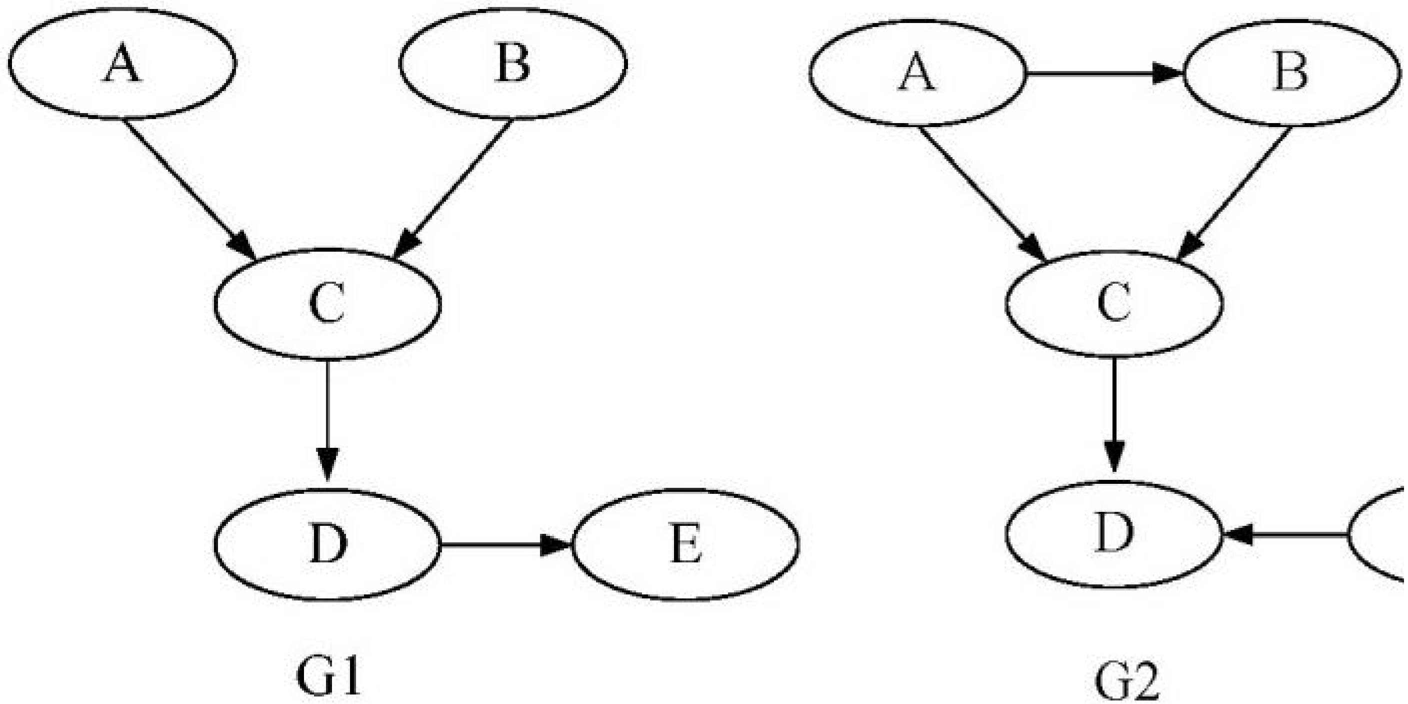
Optimality of Hill Climbing. Jack and Jill come up to you one day with a worried look on their face. "All this while we've been climbing hills, trying to improve upon our graph structure," they say. "We've been considering edge deletions, reversals, and additions at each step. Today, we found that no single edge deletion, reversal, or addition could give us a higher-scoring structure. Are we guaranteed that our current graph is the best graph structure?" What should you tell them? You may assume that their dataset is sufficiently large, and that your answer should hold for a general graph.

Your Answer	Score	Explanation

No - greedy hill-climbing will find only local maxima of the scoring function with respect to our available moves. While it might find the true graph structure on occasion, we cannot guarantee this.	✔	1.00	During greedy hill-climbing, we only make moves that will improve our current structure score. This gets us to a local maximum of the scoring function, but need not necessarily get us to the global optimum.
Total		1.00 / 1.00	

Question 6

Calculating Likelihood Differences. While doing a hill-climbing search, you run into the following two graphs, and need to choose between them using the likelihood score.



What is the difference in likelihood scores, $\mathrm{score}_L(G_1 : D) - \mathrm{score}_L(G_2 : D)$, given a dataset D of size M? Give your answer in terms of the entropy H and mutual information I. The subscripts below denote empirical values according to D: for example, $H_D(X)$ is the empirical entropy of the variable X in the dataset D.

Your Answer		Score	Explanation
$M \times [I_D(D;C) + I_D(E;D) - I_D(B;A) - I_D(D;C,E)]$	✔	1.00	
Total		1.00 / 1.00	