✖

Structure Scores

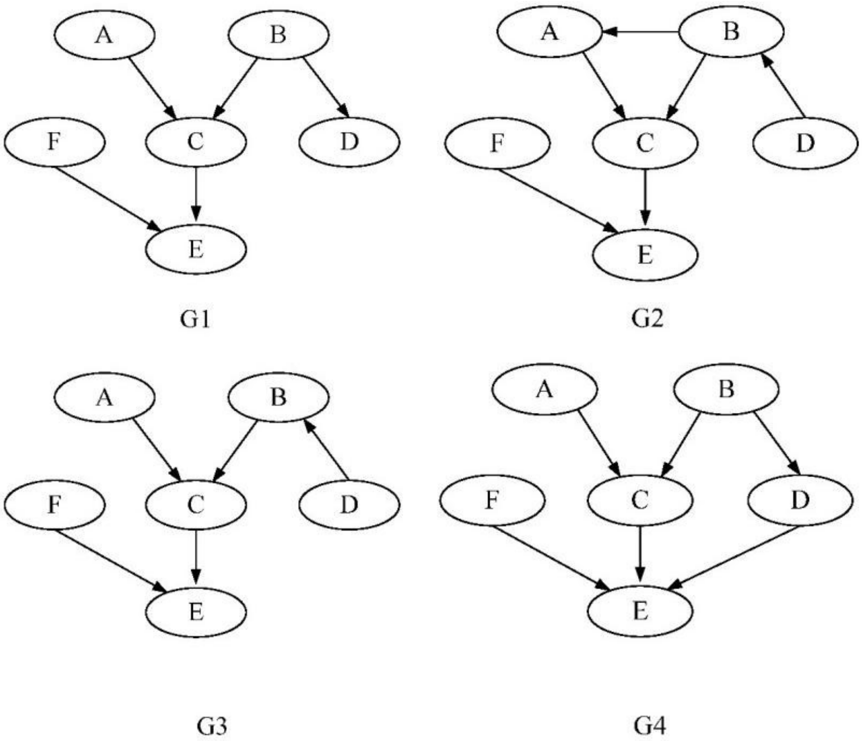
返回第 #3 周课程

5/5 得分（100%)

测验通过！

1 / 1 分

1。
Likelihood Scores. Consider the following 4 graphs:



Which of the following statements about the likelihood scores of the different graphs is/are true? You may choose more than 1 option.

- [ ] $\mathrm{Score}_L(G1 : D) \geq \mathrm{Score}_L(G4 : D)$ for every dataset D

未选择的是正确的

- [x] $\text{Score}_L(G1 : D) = \text{Score}_L(G3 : D)$ for every dataset D

正确
I-equivalent graphs will have the same likelihood score, as the ranges of distributions that they can express are the same.

- [x] $\text{Score}_L(G2 : D) \geq \text{Score}_L(G3 : D)$ for every dataset D

正确
$G2$ is an I-map of $G3$, that is, every independence relation that is in $G2$ is also in $G3$.

Hence, $G2$ can represent all distributions that $G3$ can, and so its likelihood score will not be lower than that of $G3$.

- [ ] $\text{Score}_L(G4 : D) \geq \text{Score}_L(G2 : D)$ for every dataset D

未选择的是正确的

---

1 / 1 分

2。
BIC Scores. Consider the same 4 graphs as in the previous question, but now think about the BIC score. Which of the following statements is/are true?

- [ ] $\text{Score}_{BIC}(G1 : D) \geq \text{Score}_{BIC}(G2 : D)$ for every dataset D

未选择的是正确的

- [ ] $\text{Score}_{BIC}(G1 : D) \geq \text{Score}_{BIC}(G4 : D)$ for every dataset D

未选择的是正确的

- [ ] $\text{Score}_{BIC}(G2 : D) \neq \text{Score}_{BIC}(G3 : D)$ for every dataset D

未选择的是正确的

- [x] $\text{Score}_{BIC}(G1 : D) = \text{Score}_{BIC}(G3 : D)$ for every dataset D

正确
I-equivalent graphs have the same likelihood score, and have the same complexity (in terms of the number of independent parameters). Hence, they have the same BIC score.

---

1 / 1 分

3。
Likelihood Guarantees. Consider graphs $G2$ and $G3$.

We have a dataset D generated from some probability distribution P , and the likelihood scores for $G2$ and $G3$ are $\text{Score}_L(G2 : D)$ and $\text{Score}_L(G3 : D)$, respectively.

Let $\theta^*_{D,2}$ and $\theta^*_{D,3}$ be the maximum likelihood parameters for each network, taken with respect to the dataset D .

Now let $L(X : G, \theta)$ represent the likelihood of dataset X given the graph G and parameters θ , so
$\text{Score}_L(G2 : D) = L(D : G2, \theta^*_{D,2})$ and
$\text{Score}_L(G3 : D) = L(D : G3, \theta^*_{D,3})$.

Suppose that $L(D : G2, \theta^*_{D,2}) > L(D : G3, \theta^*_{D,3})$. If we draw a new dataset E from the distribution P , which of the following statements can we guarantee?

If more than one statement holds, choose the more general statement.

None of the others

正确

$\theta^*_{D,2}$ and $\theta^*_{D,3}$ correspond to the ML estimation from dataset D .

Given the new dataset E , they might not be the ML estimation parameters any longer.

Since the dataset D and E might not be sufficiently large enough to accurately characterize P , there is no guarantee on the relation of likelihood scores of the new dataset.

$L(E:G2,\theta^*_{D,2}) > L(E:G3,\theta^*_{D,3})$

$L(E:G2,\theta^*_{D,2}) < L(E:G3,\theta^*_{D,3})$

$L(E:G2,\theta^*_{D,2}) \neq L(E:G3,\theta^*_{D,3})$

$L(E:G2,\theta^*_{D,2}) \leq L(E:G3,\theta^*_{D,3})$

---

1 / 1 分

4。

Hidden Variables. Consider the case where the generating distribution has a naive Bayes structure, with an unobserved class variable C and its binary-valued children $X_1, \ldots, X_{100}$. Assume that C is strongly correlated with each of its children (that is, distinct classes are associated with fairly different distributions over each X_i). Now suppose we try to learn a network structure directly on $X_1, \ldots, X_{100}$, **without including C in the network**. What network structure are we likely to learn if we have 10,000 data instances, and we are using table CPDs with the **likelihood** score as the structure learning criterion?

Some connected network over $X_1, \ldots, X_{100}$ that is not fully connected nor empty.

The empty network, i.e., a network consisting of only the variables but no edges between them.

A fully connected network, i.e., one with an edge between every pair of nodes.

正确
In the generating distribution, for any pair of variables X_i, X_j , the trail $X_i \leftarrow C \rightarrow X_j$ is active.

Thus, there are no independence relations of the form $X_i \perp X_j$.

This means that when we try to use the likelihood score to learn a network structure over only the X_i 's, we will end up with a fully connected network.

---

1 / 1 分

5。
Hidden Variables. Now suppose that we use the BIC score instead of the likelihood score in the previous question. What network structure are we likely to learn with the same 10,000 data instances?

- ( ) The empty network, i.e., a network consisting of only the variables but no edges between them.
- ( ) A fully connected network, i.e., one with an edge between every pair of nodes.
- (•) Some connected network over $X_1, \ldots, X_{100}$ that is not fully connected nor empty.

正确
Even though a fully connected network may be

the best representation for the true underlying distribution, we don't have enough data to learn it, and the BIC structure penalty will not allow the learning of a network with such high

complexity, given only 10,000 instances.
