AI-hw7

PB21081601 张芷苒

- 13.15
- 13.18
- 13.21
- 13.22
- 14.12
- 14.13

> **13.15** After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease and that the test is 99% accurate (i.e., the probability of testing positive when you do have the disease is 0.99, as is the probability of testing negative when you don't have the disease). The good news is that this is a rare disease, striking only 1 in 10,000 people of your age. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

The good news that the disease is rare means that it is statistically unlikely you actually have it despite the positive test. Given the rarity of the disease (1 in 10,000) and the accuracy of the test (99%), the probability that you actually have the disease after testing positive is relatively low.
Using Bayes' theorem to calculate, the chances you actually have the disease even after testing positive are approximately 0.98%.
This demonstrates that a positive test result in the context of a rare disease does not necessarily indicate you have the disease, largely due to the low base rate (prevalence) of the disease.

> **13.18** Suppose you are given a bag containing n unbiased coins. You are told that $n-1$ of these coins are normal, with heads on one side and tails on the other, whereas one coin is a fake, with heads on both sides.
> **a.** Suppose you reach into the bag, pick out a coin at random, flip it, and get a head. What is the (conditional) probability that the coin you chose is the fake coin?
> **b.** Suppose you continue flipping the coin for a total of k times after picking it and see k heads. Now what is the conditional probability that you picked the fake coin?
> **c.** Suppose you wanted to decide whether the chosen coin was fake by flipping it k times. The decision procedure returns $fake$ if all k flips come up

> heads; otherwise it returns $normal$. What is the (unconditional) probability that this procedure makes an error?

a. denote the event of picking the fake coin as F and getting a head as H. The total probability of getting a head is:

$$P(H) = P(H \mid F)P(F) + P(H \mid F^c)P(F^c) = 1 \cdot \frac{1}{n} + \frac{1}{2} \cdot \frac{n-1}{n} = \frac{1}{n} + \frac{n-1}{2n} = \frac{n+1}{2n}$$

Thus, the probability that the coin is fake given it landed heads is:

$$P(F \mid H) = \frac{P(H \mid F)P(F)}{P(H)} = \frac{\frac{1}{n}}{\frac{n+1}{2n}} = \frac{2}{n+1}$$

b. After flipping the coin k times and seeing k heads, the revised probability becomes:

$$P(F \mid \text{k heads}) = \frac{P(\text{k heads} \mid F)P(F)}{P(\text{k heads})} = \frac{1 \cdot \frac{1}{n}}{\frac{1}{n} + \frac{1}{2^k} \cdot \frac{n-1}{n}} = \frac{1}{1 + \frac{n-1}{2^k}}$$

c. The error probability in the decision procedure depends on two scenarios:

- Choosing a normal coin and getting k heads in a row (false positive).
- Choosing the fake coin and the procedure still misidentifies it (not applicable as it always identifies the fake correctly).

Thus, the error probability is solely from the false positive:

$$P(\text{error}) = P(\text{normal} \mid \text{k heads}) = \frac{\frac{n-1}{2^k}}{1 + \frac{n-1}{2^k}} = \frac{n-1}{n-1+2^k}$$

This represents the probability that a normal coin shows k heads in a row, incorrectly leading to a decision of "fake."

> **13.21** (Adapted from Pearl (1988).) Suppose you are a witness to a nighttime hit-and-run accident involving a taxi in Athens. All taxis in Athens are blue or green. You swear, under oath, that the taxi was blue. Extensive testing shows that, under the dim lighting conditions, discrimination between blue and green is 75% reliable.
> **a.** Is it possible to calculate the most likely color for the taxi? (*Hint:* distinguish carefully between the proposition that the taxi is blue and the proposition that it $appears$ blue.)
> **b.** What if you know that 9 out of 10 Athenian taxis are green?

a. To determine the most likely color of the taxi, it is necessary to consider both the reliability of the witness and the base rate of each taxi color. The witness's testimony has a 75% reliability in distinguishing the correct color under dim light. Therefore, there's a 75% chance the witness correctly identified the taxi as blue when it is indeed blue, and a 25% chance the witness mistakenly identified it as blue when it was actually green.

b. If 9 out of 10 taxis in Athens are green, then only 10% are blue. Using Bayes' theorem, the probability the taxi is blue given that it appeared blue to the witness can be calculated as follows:

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

where (P(B)) is the probability the taxi is blue (0.1), (P(A \mid B)) is the probability of the taxi appearing blue if it is blue (0.75), and (P(A)) is the total probability of the taxi appearing blue. This is calculated as:

$$P(A) = P(A \mid B)P(B) + P(A \mid G)P(G) = 0.75 \cdot 0.1 + 0.25 \cdot 0.9 = 0.075 + 0.225 = 0.3$$

Substituting these values into Bayes' theorem:

$$P(B \mid A) = \frac{0.75 \cdot 0.1}{0.3} = 0.25$$

Thus, despite the witness identifying the taxi as blue, it is still more likely that the taxi was green, given the high base rate of green taxis.

> **13.22** Text categorization is the task of assigning a given document to one of a fixed set of categories on the basis of the text it contains. Naive Bayes models are often used for this task. In these models, the query variable is the document category, and the “effect” variables are the presence or absence of each word in the language; the assumption is that words occur independently in documents, with frequencies determined by the document category.
> **a.** Explain precisely how such a model can be constructed, given as “training data” a set of documents that have been assigned to categories.
> **b.** Explain precisely how to categorize a new document.
> **c.** Is the conditional independence assumption reasonable? Discuss.

a. Constructing a Naive Bayes Text Categorization Model:

1. **Gather Data**: Collect a set of categorized documents.
2. **Prepare Text**: Tokenize text into words, possibly removing common words.
3. **Calculate Probabilities**:
    - **Prior Probability (P(C))**: Calculate the proportion of documents in each category.

- **Conditional Probability (P(w_i | C))**: Compute the frequency of each word in each category, using smoothing to handle zero occurrences.

b. Categorizing a New Document:

1. **Process the Document**: Tokenize the document into words.
2. **Compute Posterior for Each Category**:

$$P(C|\text{document}) \propto P(C) \prod_{i=1}^{n} P(w_i|C)$$

   Calculate this for each category and select the one with the highest probability.

c. Reasonableness of Conditional Independence Assumption:

- **Simplification**: Assumes the presence of one word doesn't affect the presence of another given the category. This simplifies calculations significantly.
- **Effectiveness**: Despite potential inaccuracies due to word dependencies in real language, Naive Bayes can perform well in practice, especially where the dataset is large or the text is straightforward.
- **Limitations**: In complex texts where word interactions matter, models that account for word dependencies might perform better.

> **14.12** Two astronomers in different parts of the world make measurements M_1 and M_2 of the number of stars N in some small region of the sky, using their telescopes. Normally, there is a small possibility e of error by up to one star in each direction. Each telescope can also (with a much smaller probability f) be badly out of focus (events F_1 and F_2), in which case the scientist will undercount by three or more stars (or if N is less than 3 , fail to detect any stars at all). Consider the three networks shown in Figure 14.22.
> **a.** Which of these Bayesian networks are correct (but not necessarily efficient) representations of the preceding information?
> **b.** Which is the best network? Explain.
> **c.** Write out a conditional distribution for $\mathbf{P}(M_1 \mid N)$, for the case where $N \in 1,2,3$ and $M_1 \in 0,1,2,3,4$. Each entry in the conditional distribution should be expressed as a function of the parameters e and/or f.
> **d.** Suppose $M_1 = 1$ and $M_2 = 3$. What are the possible numbers of stars if you assume no prior constraint on the values of N ?
> **e.** What is the most likely number of stars, given these observations? Explain how to compute this, or if it is not possible to compute, explain what additional information is needed and how it would affect the result.

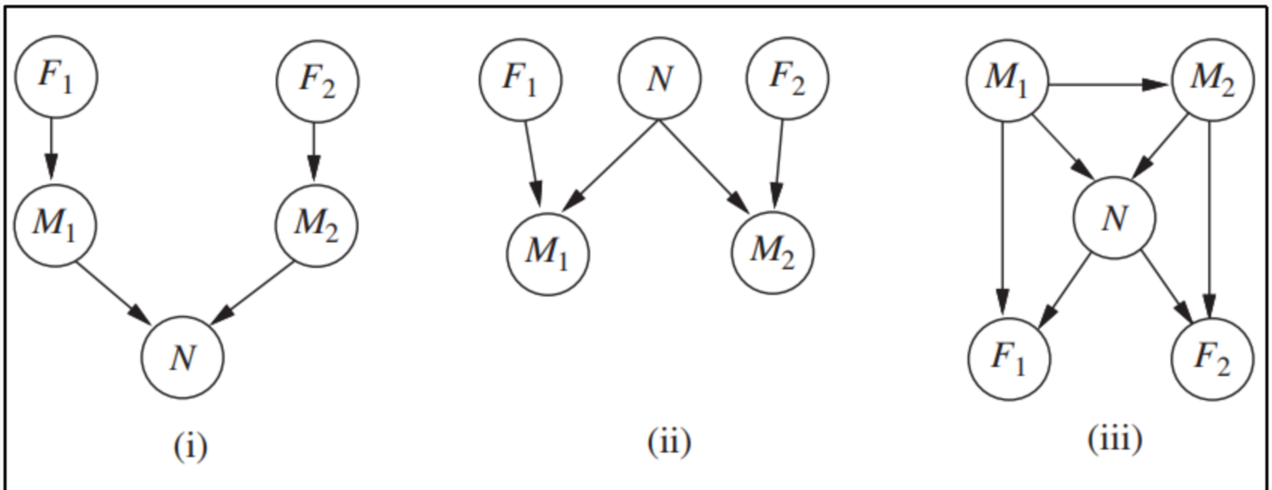

Figure 14.22 Three possible networks for the telescope problem.

a. Correct Bayesian Networks:

- Networks (i) and (iii) are correct representations. In both, the measurements M_1 and M_2 depend on the actual number of stars N and the focus states F_1 and F_2 of the respective telescopes. Network (ii) is incorrect as it suggests that N depends on M_1 and M_2, which is backwards causality.

b. Best Network:

- Network (iii) is the best representation because it correctly shows the dependency of M_1 and M_2 not only on N but also allows for the influence of each telescope's focus directly on its measurement. It also captures the interaction between the measurements that can be influenced by N.

c. Conditional Distribution for $\mathbf{P}(M_1 \mid N)$:

For $N \in \{1, 2, 3\}$ and $M_1 \in \{0, 1, 2, 3, 4\}$, the conditional distribution $P(M_1 \mid N)$ can be expressed as:

$$P(M_1 = N + 1|N) = e \quad \text{(small error adding one star)}$$

$$P(M_1 = N - 1|N) = e \quad \text{(small error subtracting one star)}$$

$$P(M_1 = N|N) = 1 - 2e - f \quad \text{(no error, not out of focus)}$$

$$P(M_1 = 0|N) = f \quad \text{(telescope out of focus, undercount by three or more, or detect none)}$$

For values of M_1 outside these ranges, the probabilities are 0.

d. Possible Numbers of Stars (N) for $M_1 = 1$ and $M_2 = 3$:

- From $M_1 = 1$: N could be 0 (if M_1 underestimated by 1 due to error), 1, or 2 (if M_1 overestimated by 1 due to error).

- From $M_2 = 3$: N could be 2 (if M_2 overestimated by 1 due to error), 3, or 4 (if M_2 underestimated by 1 due to error).
- Combining these, the possible values for N considering both measurements are 2.

e. Most Likely Number of Stars:

- To compute the most likely number of stars, Bayesian inference using the combined probabilities $P(N \mid M_1 = 1, M_2 = 3)$ based on the joint probabilities of M_1 and M_2 given N is needed. The calculation requires knowing the joint distribution:

$$P(N \mid M_1 = 1, M_2 = 3) \propto P(M_1 = 1 \mid N) \times P(M_2 = 3 \mid N) \times P(N)$$

- Additional information required includes the prior probability distribution $P(N)$ (the likelihood of different N values without any measurement). This distribution affects the result by updating the probabilities based on observed data (measurements). Without $P(N)$, exact computation isn't possible.

> **14.13** Consider the network shown in Figure 14.22 (ii), and assume that the two telescopes work identically. $N \in 1, 2, 3$ and $M_1, M_2 \in 0, 1, 2, 3, 4$, with the symbolic CPTs as described in Exercise 14.12. Using the enumeration algorithm (Figure 14.9 on page 525), calculate the probability distribution $\mathbf{P}(N \mid M_1 = 2, M_2 = 2)$.

function ENUMERATION-ASK(X, **e**, bn) **returns** a distribution over X
  inputs: X, the query variable
        e, observed values for variables **E**
        bn, a Bayes net with variables $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$ /* **Y** = *hidden variables* */

  $\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty
  for each value x_i of X **do**
    $\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL(bn.VARS, $\mathbf{e}_{x_i}$)
      where $\mathbf{e}_{x_i}$ is **e** extended with $X = x_i$
  return NORMALIZE($\mathbf{Q}(X)$)

---

function ENUMERATE-ALL($vars$, **e**) **returns** a real number
  if EMPTY?($vars$) **then return** 1.0
  $Y \leftarrow$ FIRST($vars$)
  if Y has value y in **e**
    then return $P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), **e**)
    else return $\sum_y P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}_y$)
      where $\mathbf{e}_y$ is **e** extended with $Y = y$

Figure 14.9 The enumeration algorithm for answering queries on Bayesian networks.

To calculate the probability distribution $\mathbf{P}(N \mid M_1 = 2, M_2 = 2)$ using the enumeration algorithm for the network shown in Figure 14.22 (ii), follow these steps:

Step-by-Step Calculation

1. **Define the Query and Evidence:**
   - Query Variable: N
   - Evidence: $M_1 = 2$ and $M_2 = 2$
2. **Enumeration Ask Function:**
   - Initialize the distribution $Q(N)$ for each value of $N \in \{1, 2, 3\}$.
3. **Compute $Q(N)$:**
   - Calculate for each N:

$$Q(N) = \sum_{N} P(M_1 = 2 \mid N) P(M_2 = 2 \mid N) P(N)$$

   - Utilize the Conditional Probability Tables (CPTs) for M_1 and M_2, assuming $P(N) = 1/3$ for uniform prior distribution.

Simplified Error Model (Assumed)

- Error probability e for small error and f for focus error, calculate:

$$Q(N = n) = P(M_1 = 2 | N = n) P(M_2 = 2 | N = n) P(N = n)$$

  - This is derived from the telescopes' accuracy and focus probability.

Normalizing $Q(N)$:

- Normalize $Q(N)$ so that the probabilities sum to 1:

$$P(N = n | M_1 = 2, M_2 = 2) = \frac{Q(N = n)}{\sum_{n=1}^{3} Q(N = n)}$$

  - Compute for $n = 1, 2, 3$ and normalize.

Specific values for e and f are required to complete these calculations. Adjust the computation based on the assumed values for these error probabilities.