UNIVERSITY COMPUTER SCIENCE • EXERCISE SHEET

13.3 Probability models

Name: _____ Class: _____ Date: _____ **Total: 24 marks**

KEY WORDS Bayesian network • inference

Objective

Compute posteriors by enumeration and respect base rates.

1 Key knowledge and worked examples

■ **The net**

Storm → Delay → {Late, Complaint}, with P(S) = 0.3, P(D given S) = 0.8, P(D given not S) = 0.1, P(C given D) = 0.6, P(C given not D) = 0.05 ✓.

■ **Worked example: the posterior**

$P(C) = 0.2205$; $P(S \mid C) = 0.667$; with $P(S) = 0.03$ the posterior falls to 0.126 ✓.

2 Practice

2.1 What does the network factorise? [3]

2.2 Compute $P(D)$. [3]

2.3 Name the two fallacies this unit warns about. [3]

3 Exam-style questions

3.1 Using the course net, compute $P(D \mid C)$.

Enumerate: $P(C \mid D)P(D)$ and $P(C \mid \neg D)P(\neg D)$; give $P(C)$ and then $P(D \mid C)$. Give $P(S \mid C)$ from the same normaliser and state which cause the complaint implicates more. Explain in one line why the two posteriors sum to more than 1. [8]

3.2 The storm prior drops to 0.03 (rare storms).

Recompute $P(C)$ and $P(S \mid C)$. State which changed most and why. Give the screening-test reading of the same arithmetic (a rare disease, a positive test) and the design lesson for tests used at scale. [7]

Solutions

2.1

- The joint distribution, into per-node CPTs given parents ✓.

2.2

- $P(D) = 0.8 \times 0.3 + 0.1 \times 0.7 = 0.24 + 0.07 = 0.31$ ✓.

2.3

- Dropping the normaliser; swapping $P(C \mid S)$ with $P(S \mid C)$ (the prosecutor's fallacy) ✓.

3.1

- $P(C \mid D)P(D) = 0.6 \times 0.31 = 0.186$ ✓.
- $P(C \mid \neg D)P(\neg D) = 0.05 \times 0.69 = 0.0345$; $P(C) = 0.2205$ ✓.
- $P(D \mid C) = 0.186/0.2205 = \mathbf{0.844}$ ✓.
- $P(S \mid C) = 0.49 \times 0.3/0.2205 = 0.667$ – the complaint implicates Delay more than Storm ✓.
- They do not sum to 1: S and D are not mutually exclusive exhaustive answers – D hides inside both S-branches ✓.

3.2

- $P(C) = 0.49 \times 0.03 + 0.105 \times 0.97 = 0.11655$ ✓.
- $P(S \mid C) = 0.0147/0.11655 = \mathbf{0.126}$ (from 0.667) ✓.
- Changed most: $P(S \mid C)$ – the rare cause loses the most belief, collapsing to one-eighth ✓.
- Screening reading: a rare disease with a decent test still yields most positives being false at low prevalence – design for high specificity or test twice ✓.

exercises.lol exercises.com.cn