Bayesian Reasoning

Chapter 13

Today's class

- Probability theory
- Bayesian inference
  - From the joint distribution
  - Using independence/factoring
  - From sources of evidence

Sources of uncertainty

- Uncertain **inputs**
  - Missing data
  - Noisy data
- Uncertain **knowledge**
  - Multiple causes lead to multiple effects
  - Incomplete enumeration of conditions or effects
  - Incomplete knowledge of causality in the domain
  - Probabilistic/stochastic effects
- Uncertain **outputs**
  - Abduction and induction are inherently uncertain
  - Default reasoning, even in deductive fashion, is uncertain
  - Incomplete deductive inference may be uncertain

▸ Probabilistic reasoning only gives probabilistic results (summarizes uncertainty from various sources)

Decision making with uncertainty

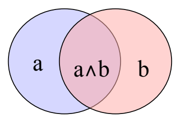
Rational behavior:

- For each possible action, identify the possible outcomes
- Compute the **probability** of each outcome
- Compute the **utility** of each outcome
- Compute the probability-weighted **(expected) utility** over possible outcomes for each action
- Select the action with the highest expected utility (principle of **Maximum Expected Utility**)

Why probabilities anyway?

- Kolmogorov showed that three simple axioms lead to the rules of probability theory
  - De Finetti, Cox, and Carnap have also provided compelling arguments for these axioms

1. All probabilities are between 0 and 1:
   - $0 \leq P(a) \leq 1$
2. Valid propositions (tautologies) have probability 1, and unsatisfiable propositions have probability 0:
   - P(true) = 1 ; P(false) = 0
3. The probability of a disjunction is given by:
   - $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



5

Probability theory

- **Random variables**
  - Domain
- Alarm, Burglary, Earthquake
  - Boolean (like these), discrete, continuous
- **Atomic event**: complete specification of state
- Alarm=True ∧ Burglary=True ∧ Earthquake=False
  alarm ∧ burglary ∧ ¬earthquake
- **Prior probability**: degree of belief without any other evidence
- P(Burglary) = .1
- **Joint probability**: matrix of combined probabilities of a set of variables
- P(Alarm, Burglary) =

	alarm	¬alarm
burglary	.09	.01
¬burglary	.1	.8

6

Probability theory (cont.)

- **Conditional probability**: probability of effect given causes
- **Computing conditional probs**:
  - $P(a \mid b) = P(a \wedge b) / P(b)$
  - P(b): **normalizing** constant
- **Product rule**:
  - $P(a \wedge b) = P(a \mid b) P(b)$
- **Marginalizing**:
  - $P(B) = \Sigma_a P(B, a)$
  - $P(B) = \Sigma_a P(B \mid a) P(a)$ (**conditioning**)

- P(burglary | alarm) = .47
  P(alarm | burglary) = .9
- P(burglary | alarm) =
  P(burglary ∧ alarm) / P(alarm)
  = .09 / .19 = .47
- P(burglary ∧ alarm) =
  P(burglary | alarm) P(alarm) =
  .47 * .19 = .09
- P(alarm) =
  P(alarm ∧ burglary) +
  P(alarm ∧ ¬burglary) =
  .09+.1 = .19

7

Example: Inference from the joint

	alarm		¬alarm	
	earthquake	¬earthquake	earthquake	¬earthquake
burglary	.01	.08	.001	.009
¬burglary	.01	.09	.01	.79

P(Burglary | alarm) = α P(Burglary, alarm)
  = α [P(Burglary, alarm, earthquake) + P(Burglary, alarm, ¬earthquake)
  = α [(.01, .01) + (.08, .09)]
  = α [(.09, .1)]

Since P(burglary | alarm) + P(¬burglary | alarm) = 1, α = 1/(.09+.1) = 5.26
  (i.e., P(alarm) = 1/α = .19 – **quizlet**: how can you verify this?)

P(burglary | alarm) = .09 * 5.26 = .474

P(¬burglary | alarm) = .1 * 5.26 = .526

8

Exercise: Inference from the joint

p(smart ∧ study ∧ prep)	smart		¬smart	
	study	**¬study**	**study**	**¬study**
prepared	.432	.16	.084	.008
¬prepared	.048	.16	.036	.072

- **Queries:**
  - **What is the prior probability of *smart*?**
  - **What is the prior probability of *study*?**
  - **What is the conditional probability of *prepared*, given *study* and *smart*?**
- **Save these answers for next time! ☺**

Independence

- When two sets of propositions do not affect each others' probabilities, we call them **independent**, and can easily compute their joint and conditional probability:
  - Independent (A, B) → P(A ∧ B) = P(A) P(B), P(A | B) = P(A)
- For example, {moon-phase, light-level} might be independent of {burglary, alarm, earthquake}
  - Then again, it might not: Burglars might be more likely to burglarize houses when there's a new moon (and hence little light)
  - But if we know the light level, the moon phase doesn't affect whether we are burglarized
  - Once we're burglarized, light level doesn't affect whether the alarm goes off
- We need a more complex notion of independence, and methods for reasoning about these kinds of relationships

Exercise: Independence

p(smart ∧ study ∧ prep)	smart		¬smart	
	study	**¬study**	**study**	**¬study**
prepared	.432	.16	.084	.008
¬prepared	.048	.16	.036	.072

- **Queries:**
  - **Is *smart* independent of *study*?**
  - **Is *prepared* independent of *study*?**

Conditional independence

- Absolute independence:
  - A and B are **independent** if P(A ∧ B) = P(A) P(B); equivalently, P(A) = P(A | B) and P(B) = P(B | A)
- A and B are **conditionally independent** given C if
  - P(A ∧ B | C) = P(A | C) P(B | C)
- This lets us decompose the joint distribution:
  - P(A ∧ B ∧ C) = P(A | C) P(B | C) P(C)
- Moon-Phase and Burglary are ***conditionally independent given*** Light-Level
- Conditional independence is weaker than absolute independence, but still useful in decomposing the full joint probability distribution

Exercise: Conditional independence

p(smart ∧ study ∧ prep)	smart		¬smart	
	study	**¬study**	**study**	**¬study**
prepared	.432	.16	.084	.008
¬prepared	.048	.16	.036	.072

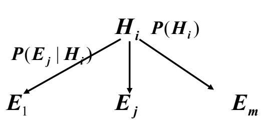
- **Queries:**
  - **Is *smart* conditionally independent of *prepared*, given *study*?**
  - **Is *study* conditionally independent of *prepared*, given *smart*?**

Bayes's rule

- Bayes's rule is derived from the product rule:
  - P(Y | X) = P(X | Y) P(Y) / P(X)
- Often useful for diagnosis:
  - If X are (observed) effects and Y are (hidden) causes,
  - We may have a model for how causes lead to effects (P(X | Y))
  - We may also have prior beliefs (based on experience) about the frequency of occurrence of effects (P(Y))
  - Which allows us to reason abductively from effects to causes (P(Y | X)).

Bayesian inference

- In the setting of diagnostic/evidential reasoning



hypotheses

evidence/manifestations

  - Know prior probability of hypothesis $P(H_i)$
    conditional probability $P(E_j \mid H_i)$
  - Want to compute the *posterior probability* $P(H_i \mid E_j)$
- Bayes's theorem (formula 1):

$$P(H_i \mid E_j) = P(H_i)P(E_j \mid H_i) / P(E_j)$$

Simple Bayesian diagnostic reasoning

- Knowledge base:
  - Evidence / manifestations: $E_1, \ldots E_m$
  - Hypotheses / disorders: $H_1, \ldots H_n$
    - E_j and H_i are **binary**; hypotheses are **mutually exclusive** (non-overlapping) and **exhaustive** (cover all possible cases)
  - Conditional probabilities: $P(E_j \mid H_i)$, i = 1, … n; j = 1, … m
- Cases (evidence for a particular instance): $E_1, \ldots, E_l$
- Goal: Find the hypothesis H_i with the highest posterior
  - $\text{Max}_i\ P(H_i \mid E_1, \ldots, E_l)$

Bayesian diagnostic reasoning II

- Bayes' rule says that
  - $P(H_i \mid E_1, \ldots, E_l) = P(E_1, \ldots, E_l \mid H_i)\, P(H_i) / P(E_1, \ldots, E_l)$
- Assume each piece of evidence E_i is conditionally independent of the others, *given* a hypothesis H_i, then:
  - $P(E_1, \ldots, E_l \mid H_i) = \prod_{j=1}^{l} P(E_j \mid H_i)$
- If we only care about relative probabilities for the H_i, then we have:
  - $P(H_i \mid E_1, \ldots, E_l) = \alpha\, P(H_i) \prod_{j=1}^{l} P(E_j \mid H_i)$

17

Limitations of simple Bayesian inference

- Cannot easily handle multi-fault situations, nor cases where intermediate (hidden) causes exist:
  - Disease D causes syndrome S, which causes correlated manifestations M_1 and M_2
- Consider a composite hypothesis $H_1 \wedge H_2$, where H_1 and H_2 are independent. What is the relative posterior?
  - $P(H_1 \wedge H_2 \mid E_1, \ldots, E_l) = \alpha\, P(E_1, \ldots, E_l \mid H_1 \wedge H_2)\, P(H_1 \wedge H_2)$
    $= \alpha\, P(E_1, \ldots, E_l \mid H_1 \wedge H_2)\, P(H_1)\, P(H_2)$
    $= \alpha \prod_{j=1}^{l} P(E_j \mid H_1 \wedge H_2)\, P(H_1)\, P(H_2)$
- How do we compute $P(E_j \mid H_1 \wedge H_2)$??

18

Limitations of simple Bayesian inference II

- Assume H1 and H2 are independent, given E1, …, El?
  - $P(H_1 \wedge H_2 \mid E_1, \ldots, E_l) = P(H_1 \mid E_1, \ldots, E_l)\, P(H_2 \mid E_1, \ldots, E_l)$
- This is a very unreasonable assumption
  - Earthquake and Burglar are independent, but *not* given Alarm:
    - P(burglar | alarm, earthquake) << P(burglar | alarm)
- Another limitation is that simple application of Bayes's rule doesn't allow us to handle causal chaining:
  - A: this year's weather; B: cotton production; C: next year's cotton price
  - A influences C indirectly: A→ B → C
  - P(C | B, A) = P(C | B)
- Need a richer representation to model interacting hypotheses, conditional independence, and causal chaining
- Next time: conditional independence and Bayesian networks!

19