Bayesian Reasoning

Adapted from slides by
Tim Finin and
Marie desJardins.
Outline

• Probability theory
• Bayesian inference
  – From the joint distribution
  – Using independence/factoring
  – From sources of evidence
Abduction

• Abduction is a reasoning process that tries to form plausible explanations for abnormal observations
  – Abduction is distinctly different from deduction and induction
  – Abduction is inherently uncertain
• Uncertainty is an important issue in abductive reasoning
• Some major formalisms for representing and reasoning about uncertainty
  – Mycin’s certainty factors (an early representative)
  – Probability theory (esp. Bayesian belief networks)
  – Dempster-Shafer theory
  – Fuzzy logic
  – Truth maintenance systems
  – Nonmonotonic reasoning
Abduction

• **Definition** (Encyclopedia Britannica): reasoning that derives an explanatory hypothesis from a given set of facts
  – The inference result is a *hypothesis* that, if true, could *explain* the occurrence of the given facts

• **Examples**
  – Dendral, an expert system to construct 3D structure of chemical compounds
    • Fact: mass spectrometer data of the compound and its chemical formula
    • KB: chemistry, esp. strength of different types of bounds
    • Reasoning: form a hypothetical 3D structure that satisfies the chemical formula, and that would most likely produce the given mass spectrum
Abduction examples (cont.)

– Medical diagnosis
  • Facts: symptoms, lab test results, and other observed findings (called manifestations)
  • KB: causal associations between diseases and manifestations
  • Reasoning: one or more diseases whose presence would causally explain the occurrence of the given manifestations

– Many other reasoning processes (e.g., word sense disambiguation in natural language process, image understanding, criminal investigation) can also been seen as abductive reasoning
Comparing abduction, deduction, and induction

**Deduction:**
- major premise: All balls in the box are black
- minor premise: These balls are from the box
- conclusion: These balls are black

**Abduction:**
- rule: All balls in the box are black
- observation: These balls are black
- explanation: These balls are from the box

**Induction:**
- case: These balls are from the box
- observation: These balls are black
- hypothesized rule: All balls in the box are black

**Deduction** reasons from causes to effects
**Abduction** reasons from effects to causes
**Induction** reasons from specific cases to general rules
Characteristics of abductive reasoning

• “Conclusions” are hypotheses, not theorems (may be false even if rules and facts are true)
  – E.g., misdiagnosis in medicine

• There may be multiple plausible hypotheses
  – Given rules A => B and C => B, and fact B, both A and C are plausible hypotheses
  – Abduction is inherently uncertain
  – Hypotheses can be ranked by their plausibility (if it can be determined)
Characteristics of abductive reasoning (cont.)

- Reasoning is often a hypothesize-and-test cycle
  - **Hypothesize**: Postulate possible hypotheses, any of which would explain the given facts (or at least most of the important facts)
  - **Test**: Test the plausibility of all or some of these hypotheses
  - One way to test a hypothesis H is to ask whether something that is currently unknown—but can be predicted from H—is actually true
    - If we also know A => D and C => E, then ask if D and E are true
    - If D is true and E is false, then hypothesis A becomes more plausible (**support** for A is increased; **support** for C is decreased)
Characteristics of abductive reasoning (cont.)

- Reasoning is **non-monotonic**
  - That is, the plausibility of hypotheses can increase/drop as new facts are collected
  - In contrast, deductive inference is **monotonic**: it never change a sentence’s truth value, once known
  - In abductive (and inductive) reasoning, some hypotheses may be discarded, and new ones formed, when new observations are made
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**)
Bayesian reasoning

• Probability theory

• Bayesian inference
  – Use probability theory and information about independence
  – Reason diagnostically (from evidence (effects) to conclusions (causes)) or causally (from causes to effects)

• Bayesian networks
  – Compact representation of probability distribution over a set of propositional random variables
  – Take advantage of independence relationships
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:
   • \(P(\text{true}) = 1; P(\text{false}) = 0\)

3. The probability of a disjunction is given by:
   • \(P(a \lor b) = P(a) + P(b) - P(a \land b)\)
Probability theory

- **Random variables**
  - Domain

- **Atomic event**: complete specification of state

- **Prior probability**: degree of belief without any other evidence

- **Joint probability**: matrix of combined probabilities of a set of variables

- Alarm, Burglary, Earthquake
  - Boolean (like these), discrete, continuous
  - \((\text{Alarm}=\text{True} \land \text{Burglary}=\text{True} \land \text{Earthquake}=\text{False})\) or equivalently \((\text{alarm} \land \text{burglary} \land \neg\text{earthquake})\)

- \(\text{P(Burglary)} = 0.1\)

- \(\text{P(Alarm, Burglary)} = \)

<table>
<thead>
<tr>
<th></th>
<th>alarm</th>
<th>\neg\text{alarm}</th>
</tr>
</thead>
<tbody>
<tr>
<td>burglary</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>\neg\text{burglary}</td>
<td>0.1</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Probability theory (cont.)

• Conditional probability: probability of effect given causes
• Computing conditional probs:
  – $P(a | b) = P(a \land b) / P(b)$
  – $P(b)$: normalizing constant
• Product rule:
  – $P(a \land b) = P(a | b) \times P(b)$
• Marginalizing:
  – $P(B) = \sum_a P(B, a)$
  – $P(B) = \sum_a P(B | a) \times P(a)$ (conditioning)

• $P(burglary \mid alarm) = 0.47$
  $P(alarm \mid burglary) = 0.9$
• $P(burglary \mid alarm) =$
  $P(burglary \land alarm) / P(alarm)$
  $= 0.09 / 0.19 = 0.47$
• $P(burglary \land alarm) =$
  $P(burglary \mid alarm) \times P(alarm) =$
  $0.47 \times 0.19 = 0.09$
• $P(alarm) =$
  $P(alarm \land burglary) +$
  $P(alarm \land \neg burglary) =$
  $0.09 + 0.1 = 0.19$
Example: Inference from the joint

<table>
<thead>
<tr>
<th></th>
<th>alarm</th>
<th>¬alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>earthquake</td>
<td>¬earthquake</td>
</tr>
<tr>
<td>burglary</td>
<td>0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>¬burglary</td>
<td>0.01</td>
<td>0.09</td>
</tr>
</tbody>
</table>

\[ P(\text{Burglary} \mid \text{alarm}) = \alpha \ P(\text{Burglary, alarm}) = \alpha \ [P(\text{Burglary, alarm, earthquake}) + P(\text{Burglary, alarm, ¬earthquake})] = \alpha \ [ (0.01, 0.01) + (0.08, 0.09) ] = \alpha \ [ (0.09, 0.1) ] \]

Since \( P(\text{burglary} \mid \text{alarm}) + P(\neg\text{burglary} \mid \text{alarm}) = 1\), \( \alpha = 1/(0.09+0.1) = 5.26 \) (i.e., \( P(\text{alarm}) = 1/\alpha = 0.109 \)  

Quizlet: how can you verify this?)

\[ P(\text{burglary} \mid \text{alarm}) = 0.09 \times 5.26 = 0.474 \]

\[ P(\neg\text{burglary} \mid \text{alarm}) = 0.1 \times 5.26 = 0.526 \]
Exercise: Inference from the joint

<table>
<thead>
<tr>
<th>p(smart ∧ study ∧ prep)</th>
<th>smart</th>
<th>¬smart</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>study</td>
<td>¬study</td>
</tr>
<tr>
<td>prepared</td>
<td>0.432</td>
<td>0.16</td>
</tr>
<tr>
<td>¬prepared</td>
<td>0.048</td>
<td>0.16</td>
</tr>
</tbody>
</table>

• 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 \((A, B) \iff P(A \land B) = P(A) \cdot P(B), \ P(A \mid B) = P(A)\)
• 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

<table>
<thead>
<tr>
<th>p(smart ∧ study ∧ prep)</th>
<th>smart</th>
<th>¬smart</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>study</td>
<td>¬study</td>
</tr>
<tr>
<td>prepared</td>
<td>0.432</td>
<td>0.16</td>
</tr>
<tr>
<td>¬prepared</td>
<td>0.048</td>
<td>0.16</td>
</tr>
</tbody>
</table>

- Queries:
  - Is smart independent of study?
  - Is prepared independent of study?
Conditional independence

- Absolute independence:
  - A and B are independent if and only if \( P(A \land B) = P(A) \cdot P(B) \); equivalently, \( P(A) = P(A \mid B) \) and \( P(B) = P(B \mid A) \)

- A and B are conditionally independent given C if and only if
  - \( P(A \land B \mid C) = P(A \mid C) \cdot P(B \mid C) \)

- This lets us decompose the joint distribution:
  - \( P(A \land B \land C) = P(A \mid C) \cdot P(B \mid C) \cdot P(C) \)

- 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

<table>
<thead>
<tr>
<th>p(smart ∧ study ∧ prep)</th>
<th>smart</th>
<th>¬smart</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>study</td>
<td>¬study</td>
</tr>
<tr>
<td>prepared</td>
<td>0.432</td>
<td>0.16</td>
</tr>
<tr>
<td>¬prepared</td>
<td>0.048</td>
<td>0.16</td>
</tr>
</tbody>
</table>

• 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:
  – \( P(Y \mid X) = P(X \mid Y) \, P(Y) / P(X) \)

• 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 (\( P(X \mid Y) \))
  – 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 (\( P(Y \mid X) \)).
Bayesian inference

- In the setting of diagnostic/evidential reasoning

- Know prior probability of hypothesis
  - conditional probability
- Want to compute the posterior probability

- Bayes’ theorem (formula 1):

\[
P(H_i | E_j) = \frac{P(H_i)P(E_j | H_i)}{P(E_j)}
\]
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: \( P(E_j \mid H_i), i = 1, \ldots, n; j = 1, \ldots, m \)
  – Cases (evidence for a particular instance): \( E_1, \ldots, E_m \)
  • Goal: Find the hypothesis \( H_i \) with the highest posterior
    – \( \text{Max}_i P(H_i \mid E_1, \ldots, E_m) \)
Bayesian diagnostic reasoning II

• Bayes’ rule says that
  \[ P(H_i \mid E_1, \ldots, E_m) = P(E_1, \ldots, E_m \mid H_i) \frac{P(H_i)}{P(E_1, \ldots, E_m)} \]

• Assume each piece of evidence \( E_i \) is conditionally independent of the others, given a hypothesis \( H_i \), then:
  \[ P(E_1, \ldots, E_m \mid H_i) = \prod_{j=1}^{m} P(E_j \mid H_i) \]

• If we only care about relative probabilities for the \( H_i \), then we have:
  \[ P(H_i \mid E_1, \ldots, E_m) = \alpha P(H_i) \prod_{j=1}^{m} P(E_j \mid H_i) \]
Limitations of simple Bayesian inference

• Cannot easily handle multi-fault situation, 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?
  – $P(H_1 \wedge H_2 \mid E_1, \ldots, E_m) = \alpha \ P(E_1, \ldots, E_m \mid H_1 \wedge H_2) \ P(H_1 \wedge H_2)$
    = $\alpha \ P(E_1, \ldots, E_m \mid H_1 \wedge H_2) \ P(H_1) \ P(H_2)$
    = $\alpha \ \prod_{j=1}^{m} P(E_j \mid H_1 \wedge H_2) \ P(H_1) \ P(H_2)$

• How do we compute $P(E_j \mid H_1 \wedge H_2)$ ??
Limitations of simple Bayesian inference II

• Assume $H_1$ and $H_2$ are independent, given $E_1, \ldots, E_m$?
  $\quad P(H_1 \land H_2 \mid E_1, \ldots, E_m) = P(H_1 \mid E_1, \ldots, E_m) \cdot P(H_2 \mid E_1, \ldots, E_m)$

• This is a very unreasonable assumption
  $\quad$ Earthquake and Burglar are independent, but not given Alarm:
  $\quad$ $P(\text{burglar} \mid \text{alarm, earthquake}) \ll P(\text{burglar} \mid \text{alarm})$

• Another limitation is that simple application of Bayes’ s rule doesn’t allow us to handle causal chaining:
  $\quad$ A: this year’s weather; B: cotton production; C: next year’s cotton price
  $\quad$ A influences C indirectly: $A \rightarrow B \rightarrow C$
  $\quad$ $P(C \mid B, A) = P(C \mid B)$

• Need a richer representation to model interacting hypotheses, conditional independence, and causal chaining

• Next time: conditional independence and Bayesian networks!