Bayesian Reasoning

Chapter 13

Thomas Bayes, 1701-1761
Today’s topics

• Review probability theory
• Bayesian inference
  – From the joint distribution
  – Using independence/factoring
  – From sources of evidence
• Naïve Bayes algorithm for inference and classification tasks
Consider

• Your house has an alarm system
• It should go off if a burglar breaks into the house
• It can go off if there is an earthquake
• How can we predict what’s happened if the alarm goes off?
  – Someone has broken in!
  – It’s a minor earthquake
Probability theory 101

• **Random variables**
  – Domain

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

• **Prior probability**: degree of belief without any other evidence or info

• **Joint probability**: matrix of combined probabilities of set of variables

• Alarm, Burglary, Earthquake
  – Boolean (like these), discrete, continuous

  • Alarm=TRUE and Burglary=TRUE and Earthquake=False

  • P(Burglary) = 0.1
  P(Alarm) = 0.1
  P(Earthquake) = 0.000003

  • P(Alarm, Burglary) =

<table>
<thead>
<tr>
<th></th>
<th>alarm</th>
<th>¬alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td>burglary</td>
<td>.09</td>
<td>.01</td>
</tr>
<tr>
<td>¬burglary</td>
<td>.1</td>
<td>.8</td>
</tr>
</tbody>
</table>
Probability theory 101

- **Conditional probability**: prob. of effect given causes

- **Computing conditional probs**:
  - \( P(a \mid b) = \frac{P(a \land b)}{P(b)} \)
  - \( P(b) \): normalizing constant

- **Product rule**:
  - \( P(a \land b) = P(a \mid b) \times P(b) \)

- **Marginalizing**:
  - \( P(B) = \sum_a P(B, a) \)
  - \( P(B) = \sum_a P(B \mid a) \times P(a) \) (conditioning)

- \( P(\text{burglary} \mid \text{alarm}) = .47 \)
  \( P(\text{alarm} \mid \text{burglary}) = .9 \)

- \( P(\text{burglary} \mid \text{alarm}) = \frac{P(\text{burglary} \land \text{alarm})}{P(\text{alarm})} = \frac{.09}{.19} = .47 \)

- \( P(\text{burglary} \land \text{alarm}) = P(\text{burglary} \mid \text{alarm}) \times P(\text{alarm}) = .47 \times .19 = .09 \)

- \( P(\text{alarm}) = P(\text{alarm} \land \text{burglary}) + P(\text{alarm} \land \neg \text{burglary}) = .09 + .1 = .19 \)

<table>
<thead>
<tr>
<th></th>
<th>alarm</th>
<th>( \neg \text{alarm} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>burglary</td>
<td>.09</td>
<td>.01</td>
</tr>
<tr>
<td>( \neg \text{burglary} )</td>
<td>.1</td>
<td>.8</td>
</tr>
</tbody>
</table>
Example: Inference from the joint

<table>
<thead>
<tr>
<th></th>
<th>alarm</th>
<th></th>
<th>¬-alarm</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>earthquake</td>
<td>¬earthquake</td>
<td>earthquake</td>
<td>¬earthquake</td>
</tr>
<tr>
<td>burglary</td>
<td>.01</td>
<td>.08</td>
<td>.001</td>
<td>.009</td>
</tr>
<tr>
<td>¬burglary</td>
<td>.01</td>
<td>.09</td>
<td>.01</td>
<td>.79</td>
</tr>
</tbody>
</table>

\[
P(\text{burglary} \mid \text{alarm}) = \alpha P(\text{burglary, alarm}) \\
= \alpha \left[ P(\text{burglary, alarm, earthquake}) + P(\text{burglary, alarm, ¬earthquake}) \right] \\
= \alpha \left[ (.01, .01) + (.08, .09) \right] \\
= \alpha \left[ (.09, .1) \right]
\]

Since \( P(\text{burglary} \mid \text{alarm}) + P(\neg\text{burglary} \mid \text{alarm}) = 1 \), \( \alpha = 1/(.09+.1) = 5.26 \) (i.e., \( P(\text{alarm}) = 1/\alpha = .19 \) – quizlet: how can you verify this?)

\[
P(\text{burglary} \mid \text{alarm}) = .09 \times 5.26 = .474
\]

\[
P(\neg\text{burglary} \mid \text{alarm}) = .1 \times 5.26 = .526
\]
Consider

- A student has to take an exam
- She might be smart
- She might have studied
- She may be prepared for the exam
- How are these related?
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>.432</td>
<td>.16</td>
</tr>
<tr>
<td>¬prepared</td>
<td>.048</td>
<td>.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*?
Exercise:
Inference from the joint

<table>
<thead>
<tr>
<th>p(smart \land study \land prep)</th>
<th>smart</th>
<th>\neg\text{smart}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>study</td>
<td>\neg\text{study}</td>
</tr>
<tr>
<td>prepared</td>
<td>.432</td>
<td>.16</td>
</tr>
<tr>
<td>\neg\text{prepared}</td>
<td>.048</td>
<td>.16</td>
</tr>
</tbody>
</table>

Queries:
- What is the prior probability of \textit{smart}?
- What is the prior probability of \textit{study}?
- What is the conditional probability of \textit{prepared}, given \textit{study} and \textit{smart}?

\[ p(\text{smart}) = .432 + .16 + .048 + .16 = 0.8 \]
## Exercise:
### Inference from the joint

<table>
<thead>
<tr>
<th>p(smart  ^  study  ^  prep)</th>
<th>smart</th>
<th>\neg smart</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>study</td>
<td>\neg study</td>
</tr>
<tr>
<td>prepared</td>
<td>.432</td>
<td>.16</td>
</tr>
<tr>
<td>\neg prepared</td>
<td>.048</td>
<td>.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*?
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>.432</td>
<td>.16</td>
</tr>
<tr>
<td>¬prepared</td>
<td>.048</td>
<td>.16</td>
</tr>
</tbody>
</table>

Queries:
- What is the prior probability of \textit{smart}?
- **What is the prior probability of \textit{study}?**
- What is the conditional probability of \textit{prepared}, given \textit{study} and \textit{smart}?

\[
p(\text{study}) = .432 + .048 + .084 + .036 = 0.6
\]
Exercise: Inference from the joint

<table>
<thead>
<tr>
<th>p(smart &amp; study &amp; 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>.432</td>
<td>.16</td>
</tr>
<tr>
<td>¬prepared</td>
<td>.048</td>
<td>.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?
**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>.432</td>
<td>.16</td>
</tr>
<tr>
<td>¬prepared</td>
<td>.048</td>
<td>.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*?

\[
p(\text{prepared}|\text{smart, study}) = \frac{p(\text{prepared, smart, study})}{p(\text{smart, study})} = \frac{.432}{(.432 + .048)} = 0.9
\]
Independence

• When variables don’t affect each others’ probabilities, they are independent; we can easily compute their joint & conditional probability:
  Independent(A, B) → P(A\&B) = P(A) * P(B) or P(A | B) = P(A)
• \{moonPhase, lightLevel\} might be independent of \{burglary, alarm, earthquake\}
  – Maybe not: burglars may be more active during a new moon because darkness hides their activity
  – But if we know light level, moon phase doesn’t affect whether we are burglarized
  – If burglarized, light level doesn’t affect if alarm goes off
• Need a more complex notion of independence and methods for reasoning about the relationships
Exercise: Independence

<table>
<thead>
<tr>
<th>p(smart &amp; study &amp; 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>.432</td>
<td>.16</td>
</tr>
<tr>
<td>¬preparation</td>
<td>.048</td>
<td>.16</td>
</tr>
</tbody>
</table>

Queries:

– Q1: Is *smart* independent of *study*?
– Q2: Is *prepared* independent of *study*?

How can we tell?
Exercise: Independence

<table>
<thead>
<tr>
<th>p(smart $\land$ study $\land$ prep)</th>
<th>smart</th>
<th>$\neg$smart</th>
</tr>
</thead>
<tbody>
<tr>
<td>study $\land$ $\neg$study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prepared</td>
<td>.432</td>
<td>.16</td>
</tr>
<tr>
<td>$\neg$prepared</td>
<td>.048</td>
<td>.16</td>
</tr>
</tbody>
</table>

Q1: Is *smart* independent of *study*?

- You might have some intuitive beliefs based on your experience
- You can also check the data

Which way to answer this is better?
Exercise: Independence

Q1: Is *smart* independent of *study*?

Q1 true iff $p(\text{smart} | \text{study}) = p(\text{smart})$

$p(\text{smart} | \text{study}) = p(\text{smart}, \text{study}) / p(\text{study})$

$= (0.432 + 0.048) / 0.6 = 0.8$

$0.8 = 0.8$, so *smart* is independent of *study*
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>.432</td>
<td>.16</td>
</tr>
<tr>
<td>¬prepared</td>
<td>.048</td>
<td>.16</td>
</tr>
</tbody>
</table>

Q2: Is prepared independent of study?
• What is prepared?
• Q2 true iff
Exercise: Independence

Q2: Is prepared independent of study?

Q2 true iff \( p(\text{prepared} | \text{study}) = p(\text{prepared}) \)

\[
p(\text{prepared} | \text{study}) = \frac{p(\text{prepared}, \text{study})}{p(\text{study})} = \frac{(.432 + .084)}{.6} = .86
\]

\[0.86 \neq 0.8, \text{ so prepared not independent of study}\]
Bayes’ rule

Derived from the product rule:

- \( P(A, B) = P(A | B) \times P(B) \)  
  \# from definition of conditional probability

- \( P(B, A) = P(B | A) \times P(A) \)  
  \# from definition of conditional probability

- \( P(A, B) = P(B, A) \)  
  \# since order is not important

So...

\[
P(A | B) = \frac{P(B | A) \times P(A)}{P(B)}
\]
Useful for diagnosis!

• *C is a cause, E is an effect:*  
  – \( P(C|E) = P(E|C) \times P(C) / P(E) \)

• **Useful for diagnosis:**  
  – E are (observed) effects and C are (hidden) causes,  
  – Often have model for how causes lead to effects \( P(E|C) \)  
  – May also have info (based on experience) on frequency of causes \( P(C) \)  
  – Which allows us to reason abductively from effects to causes \( P(C|E) \)
Ex: meningitis and stiff neck

- Meningitis (M) can cause stiff neck (S), though there are other causes too
- Use S as a diagnostic symptom and estimate \( p(M|S) \)
- Studies can estimate \( p(M) \), \( p(S) \) & \( p(S|M) \), e.g. \( p(M)=0.7, p(S)=0.01, p(M)=0.00002 \)
- Harder to directly gather data on \( p(M|S) \)
- Applying Bayes’ Rule:
  \[
p(M|S) = \frac{p(S|M) \times p(M)}{p(S)} = 0.0014
  \]
Reasoning from evidence to a cause

• In the setting of diagnostic/evidential reasoning

\[
P(H_i \mid E_j) = P(H_i) \times P(E_j \mid H_i) / P(E_j)
\]

– Know prior probability of hypothesis \( P(H_i) \)
– Want to compute the posterior probability \( P(H_i \mid E_j) \)

Bayes’ s theorem:
Simple Bayesian diagnostic reasoning

• **Naive Bayes classifier**

• Knowledge base:
  – Evidence / manifestations: \( E_1, \ldots, E_m \)
  – Hypotheses / disorders: \( H_1, \ldots, H_n \)
    
    Note: \( E_j \) 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): \( E_1, \ldots, E_l \)

• Goal: Find the hypothesis \( H_i \) with highest posterior
  – \( \text{Max}_i P(H_i \mid E_1, \ldots, E_l) \)
Simple Bayesian diagnostic reasoning

• Bayes’ rule:

\[ P(H_i \mid E_1 \ldots E_m) = \frac{P(E_1 \ldots E_m \mid H_i) P(H_i)}{P(E_1 \ldots E_m)} \]

• Assume each 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 only care about relative probabilities for \( H_i \), then:

\[ P(H_i \mid E_1 \ldots E_m) = \alpha P(H_i) \prod_{j=1}^{m} P(E_j \mid H_i) \]
Limitations

• Can’t easily handle multi-fault situations or cases where intermediate (hidden) causes exist:
  – Disease D causes syndrome S, which causes correlated manifestations M₁ and M₂

• Consider composite hypothesis H₁ ∩ H₂, where H₁ & H₂ independent. What’s relative posterior?

\[
P(H₁ \cap H₂ \mid E₁, ..., Eₙ) = \alpha \ P(E₁, ..., Eₙ \mid H₁ \cap H₂) \ P(H₁ \cap H₂)
\]

\[
= \alpha \ P(E₁, ..., Eₙ \mid H₁ \cap H₂) \ P(H₁) \ P(H₂)
\]

\[
= \alpha \ \prod_{j=1}^{n} P(E_j \mid H₁ \cap H₂) \ P(H₁) \ P(H₂)
\]

• How do we compute \( P(E_j \mid H₁ \cap H₂) \)?
Summary

• Probability a rigorous formalism for uncertain knowledge
• **Joint probability distribution** specifies probability of every atomic event
• Answer queries by summing over atomic events
• Must reduce joint size for non-trivial domains
• **Bayes rule:** compute from known conditional probabilities, usually in causal direction
• **Independence & conditional independence** provide tools
• Next: Bayesian belief networks