Overview

- Object recognition
- Combining multiple evidence: Bayesian updating
- Belief network

Example: Object Recognition

Image 2D Obj 3D Obj

1

Combining Multiple Evidences

Suppose we have these conditional probabilities $P(\text{Cavity}|\text{Toothache})$ and $P(\text{Cavity}|\text{Catch})$. What if we want to know $P(\text{Cavity} | \text{Toothache} \land \text{Catch})$? These are the alternatives:

- Look up the joint probability table: not practical or even impossible in most cases
- We can calculate

$$P(\text{Cavity} | \text{Ache} \land \text{Catch}) = \frac{P(\text{Ache} \land \text{Catch} | \text{Cav}) P(\text{Cav})}{P(\text{Ache} \land \text{Catch})}$$

but, calculating the new conditional prob is a pain.

Bayesian Updating

An Alternative: gradually work in the multiple evidences – Bayesian Updating

- Reformulate the Bayes’ rule so that conditional probability of events given combined evidences (such as $P(\text{A} | \text{B} \land \text{C})$) are not necessary.
- Use domain knowledge to replace the more complex conditional probabilities with known, simpler ones (utilize conditional independence).

Bayesian updating makes combining evidences efficient.
Bayesian Updating: Example

We want to calculate $P(Cavity|Ache \land Catch)$:

$$P(Cav|Ache \land Catch) = P(Cav|Ache) \frac{P(Catch|Ache \land Cav)}{P(Catch|Ache)}$$

$$= P(Cav) \frac{P(Ache|Cav)}{P(Ache)} \frac{P(Catch|Ache \land Cav)}{P(Catch|Ache)}$$

Problem is that $P(Catch|Ache \land Cav)$ may be equally hard to calculate. However, we can make these assumptions (**Conditional Independence** of $Ache$ and $Catch$ given $Cav$):

$$P(Catch|Cav \land Ache) = P(Catch|Cav)$$

$$P(Ache|Cav \land Catch) = P(Ache|Cav)$$

Only thing that remains is $P(Ache)P(Catch|Ache) = P(Catch \land Ache)$, which can be eliminated by normalization (**Exercise**: try this – see Exercise 14.7).

Bayesian Updating: Example (cont’d)

So, after replacing the factors:

$$P(Cav|Ache \land Catch) = \alpha \frac{P(Cav)P(Ache|Cav)}{P(Catch|Ache)}P(Catch|Cav),$$

where $\alpha$ is the normalization constant needed to make $P(Cav|Ache \land Catch)$ add up to 1.

Bayesian Updating: Summary

$X$ and $Y$ are independent given $Z$ (conditional independence):

$$P(X|Y, Z) = P(X|Z)$$

Simplified Bayes’ rule for multiple evidence is$^a$:

$$P(Z|X, Y) = \alpha P(Z)P(X|Z)P(Y|Z),$$

where $\alpha$ is the normalization constant.

Thus, Bayesian Updating makes combining multiple evidence easy.

---

$^a$**Note**: $Z$ is the cause, and $X$ and $Y$ are the effects.

Probabilistic Reasoning

**Belief Network** represent the dependence between random variables, and give a concise specification of the joint probability distribution. It is represented as a directed acyclic graph (DAG):

1. a set of random variables: nodes of the network
2. a set of directed edges from one node to another
3. each node has a conditional probability table that quantifies the effect the parents have on that node. The parents are the nodes pointing to that node.
4. the graph has no cycles
New burglar alarm was installed.

- The alarm can be triggered by either an actual burglary or an earthquake.
- Neighbors John and Mary agreed to call you at work when they hear the alarm.

Example question: If you got calls from John and Mary, what is the chance of it being a totally false alarm (not a burglary, nor an earthquake)?

You can ask any conjunctive combination.

Each node has a conditional probability table fully describing $P(\text{Current}|\text{Parent}_1, \text{Parent}_2, \ldots, \text{Parent}_n)$:

<table>
<thead>
<tr>
<th>Burglary</th>
<th>Earthquake</th>
<th>$P$</th>
<th>$\neg P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>0.950</td>
<td>0.050</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>0.950</td>
<td>0.050</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>0.290</td>
<td>0.710</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>0.001</td>
<td>0.999</td>
</tr>
</tbody>
</table>

$P = P(\text{Alarm}|\text{Burglary, Earthquake})$

The network can be viewed as

- a representation of the joint probability distribution (this view is helpful when constructing the network), or
- an encoding of a collection of conditional independence statements (this view is helpful when designing effective inference procedures).

Belief Network: Representing Joint Prob. Dist. (cont’d)

Each row in the conditional probability tables are:

\[ P(X_1 = x_1, \ldots, X_n = x_n) = P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | Parent(X_i)) \]

Parent\( (X_j) \) refers to the event when \( X_j = x_j \).

Imagine a case where each node in the example has a \( T \) or \( F \) assignment. \( X_j \) will then be either \( T \) or \( F \) for all \( j \).

The belief network fully defines a joint probability distribution!

Calculating Probability of a Joint Event

Calculate the probability of the event that the alarm (\( A \)) has sounded but neither a burglary (\( \neg B \)) nor an earthquake (\( \neg E \)) occurred, and both John (\( J \)) and Mary (\( M \)) call:

\[
\begin{align*}
P(J \land M \land A \land \neg B \land \neg E) & = P(J | Prnts(J)) P(M | Prnts(M)) P(A | Prnts(A)) P(\neg B) P(\neg E) \\
& = P(J | A) P(M | A) P(A | \neg B \land \neg E) P(\neg B) P(\neg E)
\end{align*}
\]

\[
P(J | A) P(M | A) P(A | \neg B \land \neg E) P(\neg B) P(\neg E) = 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.00062
\]
Key Points

- How is subjective belief utilized in Bayesian analysis?
- Bayesian updating: why does it make probabilistic inference efficient when multiple evidence comes in?
- Belief network: definition, semantics, extracting probabilities of certain conjunction of events.

Next Time

Chapter 15

- Construction of belief networks
- Inference in belief networks

A Brief Summary of the Course

Topics covered so far:

- Search
- Logical inference
- Probabilistic inference

Topics to be covered:

- Learning
- Special topics