Bayesian networks (1)

Lirong Xia

Rensselaer

Spring 2017
Project 2

» alpha-beta
  • v>beta
  • v<alpha
  • different from book/slides

» depth
  • each depth means a full round of Pacman-ghost
  • not Pacman or ghost
  • may have multiple ghosts

» My OH moved to Thursday at 3 pm
Random variables and joint distributions

- **A random variable** is a variable with a domain
  - Random variables: capital letters, e.g. W, D, L
  - Values: small letters, e.g. w, d, l

- **A joint distribution** over a set of random variables: $X_1, X_2, \ldots, X_n$ specifies a real number for each assignment (or outcome)
  - $p(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$
  - $p(x_1, x_2, \ldots, x_n)$

- This is a **special (structured) probability space**
  - Sample space $\Omega$: all combinations of values
  - Probability mass function is $p$

- **A probabilistic model** is a joint distribution over a set of random variables
  - Will be our focus of this course

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): combine collapsed rows by adding

\[ p(X_1 = x_1) = \sum_{x_2} p(X_1 = x_1, X_2 = x_2) \]

<table>
<thead>
<tr>
<th>T</th>
<th>W</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

\[
W: \begin{array}{cc}
\text{sun} & 0.6 \\
\text{rain} & 0.4
\end{array}
\]
\[
T: \begin{array}{cc}
\text{hot} & 0.5 \\
\text{cold} & 0.5
\end{array}
\]
\[
\begin{array}{cc}
0.4 & 0.1 \\
0.2 & 0.3
\end{array}
\]
Conditional Distributions

- Conditional distributions are probability distributions over some variables given fixed values of others.

|       | $p(W|T = \text{hot})$ | $p(W|T = \text{cold})$ |
|-------|----------------------|------------------------|
| $W$   | $p$                  | $p$                    |
| sun   | 0.8                  | 0.4                    |
| rain  | 0.2                  | 0.6                    |

<table>
<thead>
<tr>
<th>$T$</th>
<th>$W$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>hot</td>
<td>sun</td>
<td>0.4</td>
</tr>
<tr>
<td>hot</td>
<td>rain</td>
<td>0.1</td>
</tr>
<tr>
<td>cold</td>
<td>sun</td>
<td>0.2</td>
</tr>
<tr>
<td>cold</td>
<td>rain</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Joint Distributions
Independence

- Two variables are independent in a joint distribution if for all \( x, y \), the events \( X=x \) and \( Y=y \) are independent:

\[
p(X, Y) = p(X)p(Y)
\]

\[
\forall x, y \quad p(x, y) = p(x)p(y)
\]

- The joint distribution factors into a product of two simple ones
- Usually variables aren’t independent!
The Chain Rule

- Write any joint distribution as an incremental product of conditional distributions

\[ p(x_1, x_2, x_3) = p(x_1) p(x_2 | x_1) p(x_3 | x_1, x_2) \]
\[ p(x_1, x_2, \ldots x_n) = \prod_{i} p(x_i | x_1 \ldots x_{i-1}) \]

- Why is this always true?
  - Key: \( p(A|B) = p(A,B)/p(B) \)
Today’s schedule

- Conditional independence
- Bayesian networks
  - definitions
  - independence
Conditional Independence among random variables

- $p(\text{Toothache}, \text{Cavity}, \text{Catch})$

- If I don’t have a cavity, the probability that the probe catches in it doesn’t depend on whether I have a toothache:
  - $p(+\text{Catch}|+\text{Toothache},-\text{Cavity}) = p(+\text{Catch}|-\text{Cavity})$

- The same independence holds if I have a cavity:
  - $p(+\text{Catch}|+\text{Toothache},+\text{Cavity}) = p(+\text{Catch}|+\text{Cavity})$

- Catch is *conditionally independent* of toothache given cavity:
  - $p(\text{Catch}|\text{Toothache},\text{Cavity}) = p(\text{Catch}|\text{Cavity})$

- Equivalent statements:
  - $p(\text{Toothache}|\text{Catch},\text{Cavity}) = p(\text{Toothache}|\text{Cavity})$
  - $p(\text{Toothache},\text{Catch}|\text{Cavity}) = p(\text{Toothache}|\text{Cavity}) \times p(\text{Catch}|\text{Cavity})$
  - One can be derived from the other easily (part of Homework 1)
Conditional Independence

- Unconditional (absolute) independence very rare
- Conditional independence is our most basic and robust form of knowledge about uncertain environments:
- Definition: X and Y are conditionally independent given Z, if
  - \( \forall x, y, z: p(x, y|z) = p(x|z) \times p(y|z) \)
  - or equivalently, \( \forall x, y, z: p(x|z, y) = p(x|z) \)
  - X, Y, Z are random variables
  - written as \( X \perp Y|Z \)
- Brain teaser: in a probabilistic model with three random variables XYZ
  - If X and Y are independent, can we say X and Y are conditionally independent given Z
  - If X and Y are conditionally independent given Z, can we say X and Y are independent?
  - Bonus questions in Homework 1
The Chain Rule

- \( p(X_1, \ldots, X_n) = p(X_1) \ p(X_2|X_1) \ p(X_3|X_1, X_2) \ldots \)

- Trivial decomposition:

\[
p(\text{Catch}, \text{Cavity}, \text{Toothache}) = p(\text{Cavity}) \ p(\text{Catch} | \text{Cavity}) \ p(\text{Toothache} | \text{Catch}, \text{Cavity})
\]

- With assumption of conditional independence:
  - \( \text{Toothache} \perp \text{Catch} | \text{Cavity} \)

\[
p(\text{Toothache}, \text{Catch}, \text{Cavity}) = p(\text{Cavity}) \ p(\text{Catch} | \text{Cavity}) \ p(\text{Toothache} | \text{Cavity})
\]

- Bayesian networks/ graphical models help us express conditional independence assumptions
Bayesian networks: Big Picture

Using full joint distribution tables
- Representation: n random variables, at least $2^n$ entries
- Computation: hard to learn (estimate) anything empirically about more than a few variables at a time

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>W</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>summer</td>
<td>hot</td>
<td>sun</td>
<td>0.30</td>
</tr>
<tr>
<td>summer</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>summer</td>
<td>cold</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>sun</td>
<td>0.10</td>
</tr>
<tr>
<td>winter</td>
<td>hot</td>
<td>rain</td>
<td>0.05</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>sun</td>
<td>0.15</td>
</tr>
<tr>
<td>winter</td>
<td>cold</td>
<td>rain</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Bayesian networks: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions
Example Bayesian networks: Car

- **Initial observation:** car won’t start
- **Orange:** “broken” nodes
- **Green:** testable evidence
- **Gray:** “hidden variables” to ensure sparse structure, reduce parameters
Graphical Model Notation

- Nodes: variables (with domains)
- Arcs: interactions
  - Indicate “direct influence” between variables
  - Formally: encode conditional independence (more later)
- For now: imagine that arrows mean direct causation (in general, they don’t!)
Example: Coin Flips

- \( n \) independent coin flips (different coins)

- No interactions between variables: independence
  
  - Really? How about independent flips of the same coin?
  
  - How about a skillful coin flipper?
  
  - Bottom line: build an application-oriented model
Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic

- Model 1: independence

- Model 2: rain causes traffic

- Which model is better?
Example: Burglar Alarm Network

- **Variables:**
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!
Bayesian network

- Definition of Bayesian network (Bayes’ net or BN)
- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents’ values
  - CPT: conditional probability table
  - Description of a noisy “causal” process

A Bayesian network = Topology (graph) + Local Conditional Probabilities
Probabilities in BNs

- Bayesian networks implicitly encode joint distributions
  - As a product of local conditional distributions
    \[ p(x_1, x_2, \cdots x_n) = \prod_{i=1}^{n} p(x_i \mid \text{parents}(X_i)) \]
  - Example:
    \[ p(+\text{Cavity}, +\text{Catch}, -\text{Toothache}) \]
- This lets us reconstruct any entry of the full joint
- Not every BN can represent every joint distribution
  - The topology enforces certain conditional independencies
Example: Coin Flips

\begin{align*}
p(X_1) & \quad p(X_2) & \quad p(X_n) \\
\begin{array}{|c|c|} \hline h & 0.5 \\
\hline t & 0.5 \\
\hline \end{array} & \quad \begin{array}{|c|c|} \hline h & 0.5 \\
\hline t & 0.5 \\
\hline \end{array} & \quad \begin{array}{|c|c|} \hline h & 0.5 \\
\hline t & 0.5 \\
\hline \end{array} \end{align*}

\[ p(h, h, t, h) = \]

Only distributions whose variables are absolutely independent can be represented by a Bayesian network with no arcs.
Example: Traffic

\[ p(R) = \begin{array}{cc}
+r & 0.25 \\
-r & 0.75
\end{array} \]

\[ p(+r, -t) = \]

\[ p(T|R) = \begin{array}{ccc}
+r & +t & 0.75 \\
+r & -t & 0.25 \\
-r & +t & 0.50 \\
-r & -t & 0.50
\end{array} \]
Example: Alarm Network

<table>
<thead>
<tr>
<th>B</th>
<th>$p(B)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E</th>
<th>$p(E)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| A   | J     | $p(J|A)$ |
|-----|-------|----------|
| +a  | +j    | 0.9      |
| +a  | -j    | 0.1      |
| -a  | +j    | 0.05     |
| -a  | -j    | 0.95     |

| A   | M     | $p(M|A)$ |
|-----|-------|----------|
| +a  | +m    | 0.7      |
| +a  | -m    | 0.3      |
| -a  | +m    | 0.01     |
| -a  | -m    | 0.99     |

| B   | E   | A   | $p(A|B, E)$ |
|-----|-----|-----|------------|
| +b  | +e  | +a  | 0.95       |
| +b  | +e  | -a  | 0.05       |
| +b  | -e  | +a  | 0.94       |
| +b  | -e  | -a  | 0.06       |
| -b  | +e  | +a  | 0.29       |
| -b  | +e  | -a  | 0.71       |
| -b  | -e  | +a  | 0.001      |
| -b  | -e  | -a  | 0.999      |
Size of a Bayesian network

- How big is a joint distribution over \( N \) Boolean variables?
  - \( 2^N \)

- How big is an \( N \)-node net if nodes have up to \( k \) parents?
  - \( O(N \times 2^{k+1}) \)

- Both give you the power to calculate \( p(X_1, \ldots, X_n) \)

- BNs: Huge space savings!

- Also easier to elicit local CPTs

- Also turns out to be faster to answer queries
Bayesian networks

➢ So far: how a Bayesian network encodes a joint distribution

➢ Next: how to answer queries about that distribution
  • Key idea: conditional independence
  • Main goal: answer queries about conditional independence and influence from the graph

➢ After that: how to answer numerical queries (inference)
Conditional Independence in a BN

Important question about a BN:

- Are two nodes independent given certain evidence?
- If yes, can prove using algebra (tedious in general)
- If no, can prove with a counter example

Example: X: pressure, Y: rain, Z: traffic

- Question: are X and Z necessarily independent?
  - Answer: no. Example: low pressure causes rain, which causes traffic
  - X can influence Z, Z can influence X (via Y)
This configuration is a “causal chain”

\[
p(x, y, z) = p(x)p(y|x)p(z|y)
\]

- Is \(X\) independent of \(Z\) given \(Y\)?

\[
p(z|x, y) = \frac{p(x, y, z)}{p(x, y)} = \frac{p(x)p(y|x)p(z|y)}{p(x)p(y|x)} = p(z|y)
\]

Yes!

- Evidence along the chain “blocks” the influence

X: Low pressure
Y: Rain
Z: Traffic
Common Cause

Another basic configuration: two effects of the same cause

- Are X and Z independent?
- Are X and Z independent given Y?

\[
p(z|x, y) = \frac{p(x, y, z)}{p(x, y)} = \frac{p(y) p(x|y) p(z|y)}{p(y) p(x|y)} = p(z|y) \quad \text{Yes!}
\]

- Observing the cause blocks influence between effects.

Y: Project due
X: Many Piazza posts
Z: No one play games
Common Effect

- Last configuration: two causes of one effect (v-structure)
  - Are X and Z independent?
    - Yes: the ballgame and the rain cause traffic, but they are not correlated
    - Still need to prove they must be (try it!)
  - Are X and Z independent given Y?
    - No: seeing traffic puts the rain and the ballgame in competition as explanation?
    - This is backwards from the other cases
      - Observing an effect activates influence between possible causes.

X: Raining
Z: Ballgame
Y: Traffic
The General Case

- Any complex example can be analyzed using these three canonical cases

- General question: in a given BN, are two variables independent (given evidence)?

- Solution: analyze the graph
Reachability (D-Separation)

Question: are X and Y conditionally independent given evidence vars \{Z\}?
- Yes, if X and Y “separated” by Z
- Look for active paths from X to Y
- No active paths = independence!

A path is active if each triple is active:
- Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
- Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
- Common effect $A \rightarrow B \leftarrow C$ where B or one of its descendents is observed

All it takes to block a path is a single inactive segment
Example

\[ R \perp B \quad \text{Yes!} \]
\[ R \perp B \mid T \]
\[ R \perp B \mid T' \]
Example

$L \perp T'|T$

$L \perp B$

$L \perp B|T$

$L \perp B|T'$

$L \perp B|T,R$

Yes!

Yes!

Yes!
Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I am sad

- Questions:
  
  \[
  T \perp D
  \]
  
  \[
  T \perp D \mid R
  \text{ Yes!}
  \]
  
  \[
  T \perp D \mid R, S
  \]