Bayesian Networks – Representation

Machine Learning – CSE546
Carlos Guestrin
University of Washington
November 25, 2013

Handwriting recognition

Character recognition, e.g., kernel SVMs
Webpage classification

- Company home page vs Personal home page
- University home page vs...

Handwriting recognition 2

"C" comes after "o" much more than after a "z".

©Carlos Guestrin 2005-2013
Today – Bayesian networks

- One of the most exciting advancements in statistical AI in the last decades
- Generalizes naïve Bayes and logistic regression classifiers
- Compact representation for exponentially-large probability distributions
- Exploit conditional independencies
Causal structure

- Suppose we know the following:
  - The flu causes sinus inflammation
  - Allergies cause sinus inflammation
  - Sinus inflammation causes a runny nose
  - Sinus inflammation causes headaches
- How are these connected?

Possible queries

- Inference
  \[ P(F=\text{t} \mid N=\text{t}) \]
- Most probable explanation
  \[ \max_{f,a,s,h} P(f,a,s,h \mid N=\text{t}) \]
- Active data collection
  given \( N=\text{t} \), which test should I perform
Car starts BN

- 18 binary attributes
  \( 2^{18} \) possibilities
- Inference
  - \( P(\text{BatteryAge}|\text{Starts}=f) \)

\[ P(\text{BA}|S=f) = \sum_{a_i,b_i,c_i,...,\text{BA},S=f} P(a_i,b_i,c_i,...,\text{BA},S=f) \]

- \( 2^{16} \) terms, why so fast?
- Not impressed?
  - HailFinder BN – more than \( 3^{54} = 58149737003040059690390169 \) terms

Factored joint distribution - Preview

\[ P(A, F, S, H, N) = P(F) P(A) P(S|F,A) P(H|S) P(N|H) \]

\[ 2^5 = 32 \text{ terms} \]

\text{many terms}
What about probabilities?
Conditional probability tables (CPTs)

Number of parameters

Flu

Allergy

Sinus

Headache

Nose

$p(f) = 0.05$

$p(f|a) = 0.15$

$p(a) = 0.8$

$p(s|a,f) = 0.08$

$p(s|a,f) = 0.9$

$p(s|a,f) = 0.5$

$p(s|a,f) = 0.1$

$p(s|a,f) = 0.9$

$p(h|s) = 0.1$

$p(h|s) = 0.05$

$p(n|s) = 0.2$

$p(n|s) = 0.8$

$p(n|s) = 0.3$

$p(n|s) = 0.7$

$p(f) \in 1 \text{ param.}$

$p(a) \in 1 \text{ param.}$

$p(s|f,a) \in 9 \text{ params.}$

$p(h|s) \in 2 \text{ params.}$

$p(n|s) \in 7 \text{ params.}$

$2^5 - 1 \text{ params.} = 31 \text{ params.}$

10 params, only numbers, less variance
Key: Independence assumptions

Knowing sinus separates the variables from each other

(Marginal) Independence

- Flu and Allergy are (marginally) independent

\[ F \perp A \]

\[ P(A, F) = P(A) P(F) \]

or

\[ P(A|F) = P(A) \]

<table>
<thead>
<tr>
<th></th>
<th>Flu = t</th>
<th>Flu = f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flu</td>
<td>0.2</td>
<td>0.8</td>
</tr>
<tr>
<td>Allergy</td>
<td>0.4</td>
<td>0.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Flu = t</th>
<th>Flu = f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allergy</td>
<td>0.4 x 2 = 0.8</td>
<td>0.4 x 8</td>
</tr>
<tr>
<td>Allergy</td>
<td>0.6 x 2 = 1.2</td>
<td>0.6 x 6</td>
</tr>
</tbody>
</table>
## Marginally independent random variables

- **Sets** of variables $X, Y$
- $X$ is independent of $Y$ if
  - $P(T(X=x \perp Y=y), \forall x \in \text{Val}(X), y \in \text{Val}(Y))$
  - $P(x=x, y=y) = P(x=x) \cdot P(y=y) \quad \forall x, y$

- Shorthand:
  - **Marginal independence**: $P(T(X \perp Y))$

- **Proposition**: $P$ satisfies $(X \perp Y)$ if and only if
  - $P(X, Y) = P(X) \cdot P(Y)$
  - $\forall y \in \text{Val}(Y), P(X|Y=y) = P(X)$

## Conditional independence

- Flu and Headache are not (marginally) independent
  - $P(H=t | F=t) \neq P(H=t)$
  - $F \perp H$

- Flu and Headache are independent given Sinus infection
  - $F \perp H | S$
  - $P(H=t | S=t) = P(H=t | S=t, F=t)$

- More Generally:
  - $X \perp Y | Z$
  - $P(X|Z) = P(X|Y,Z)$
  - $P(X,Y|Z) = P(X|Z) \cdot P(Y|Z)$
Conditionally independent random variables

- **Sets** of variables $X$, $Y$, $Z$
- $X$ is independent of $Y$ given $Z$ if
  - $P \models (X=x \perp Y=y \mid Z=z)$, $\forall x \in \text{Val}(X)$, $y \in \text{Val}(Y)$, $z \in \text{Val}(Z)$

- **Shorthand:**
  - Conditional independence: $P \models (X \perp Y \mid Z)$
  - For $P \models (X \perp Y \mid \emptyset)$, write $P \models (X \perp Y)$

- **Proposition:** $P$ satisfies $(X \perp Y \mid Z)$ if and only if
  - $P(X,Y|Z) = P(X|Z) P(Y|Z)$

---

The independence assumption

Local Markov Assumption:
A variable $X$ is independent of its non-descendants given its parents (and only its parents)
Explaining away

Local Markov Assumption:
A variable $X$ is independent of its non-descendants given its parents

\[ P(F=t | S=t, A=t) > P(F=t | S=t) \]
if it's not allergy, prob it's flu

Naïve Bayes revisited

Local Markov Assumption:
A variable $X$ is independent of its non-descendants given its parents
Joint distribution

Flu
Allergy
Sinus
Nose
Headache

Why can we decompose? Markov Assumption!

The chain rule of probabilities

\[ P(A,B) = P(A)P(B|A) \]

For any distribution:

\[ P(S,F) = P(S)P(F|S) \]

case:

\[ P(S,F,H) = P(S)P(F|S)P(H|F,S) \]

More generally:

\[ P(X_1, \ldots, X_n) = P(X_1)P(X_2|X_1) \ldots P(X_n|X_1, \ldots, X_{n-1}) \]

Always true
Chain rule & Joint distribution

Local Markov Assumption:
A variable X is independent of its non-descendants given its parents.

The Representation Theorem – Joint Distribution to BN

Encodes independence assumptions

Joint probability distribution:
\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | Pa_{X_i}) \]
Two (trivial) special cases

Edgeless graph

\[ x_1 \quad x_2 \quad x_3 \]

\[ \ldots \quad x_n \]

\( x_i : \{ \text{every body but } \text{else} \} \)

all variables independent.

all the ties in the world

Fully-connected graph

\[ x_1 \rightarrow x_2 \rightarrow x_3 \]

\[ x_4 \]

\( x_i : \{ \text{nobody} \} | \text{parents} \)

no independence assumptions.

no representation anything.

but exponentially many parameters.

lots of variance (no bias).