Announcements

• Next lecture Friday 12/8: in CIT 367

• Plan for the remainder of the term:
  – Today: graphical models
  – Friday 12/8: beyond this course.
  – Monday 12/11: advanced applications.
  – Wednesday 12/13: final review.
Review: Viterbi decoding

- Max-probabilities:

\[
\delta_t(s) \triangleq \max_{s_1, \ldots, s_{t-1}} p(x_1, \ldots, x_t, s_1, \ldots, s_{t-1}, s_t = s)
\]

- Computed through forward recursion:

\[
\delta_1(s) = p(s_1 = s)p(x_1 | s_1 = s), \\
\delta_t(s) = \max_{s'} [\delta_{t-1}(s') p(s' \rightarrow s)] p(x_t | s_t = s)
\]

- Viterbi decoding: backtrack through the most likely sequence of states:

\[
s^*_N = \arg \max_s \delta_N(s), \\
s^*_t = \arg \max_s \delta_t(s)p(s \rightarrow s^*_{t+1})
\]
Input-Output HMM

- Model a state-dependent mapping from input $x_t$ to output $y_t$.

$$p(x_1, \ldots, x_N, y_1, \ldots, y_N, s_1, \ldots, s_N) = p(x_1)p(s_1 | x_1)p(y_1 | x_1, s_1) \times \prod_{t=2}^{N} p(x_t)p(s_t | x_t, s_{t-1})p(y_t | x_t, s_t)$$

- Can be seen as a “recurrent mixture of experts” model.
String-edit HMMs

- Modeling sequences of symbols, corresponding to template and potentially corrupted by noise.

- Three kinds of states:
  
  \( m_t \) match the observed symbol \( x_t \),
  
  \( d_t \) delete the symbol at position \( t \),
  
  \( i_t \) insert a symbol at position \( t \).
String-edit HMM: example
String-edit HMM: example

AGAA-C
AGAA-C
TCAGCATC
String-edit HMM: example
Directed graphical models

- A directed acyclic graph (DAG) on the random variables.
- Lack of directed edge from $X_i$ to $X_j$ means they are independent.
- Joint distribution can be factorized:

$$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p \left( X_i \mid X_{\text{Parents}(X_i)} \right).$$

```
p(A)p(B)
```
Directed graphical models

- A directed acyclic graph (DAG) on the random variables.
- Lack of directed edge from $X_i$ to $X_j$ means they are independent.
- Joint distribution can be factorized:

$$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p\left(X_i \mid X_{\text{Parents}(X_i)}\right).$$

$$p(A)p(B)p(C \mid A, B)$$
Directed graphical models

- A directed acyclic graph (\textit{DAG}) on the random variables.
- Lack of directed edge from $X_i$ to $X_j$ means they are independent.
- Joint distribution can be factorized:

\[
p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i \mid X_{\text{Parents}(X_i)})
\]

\[
p(A)p(B)p(C \mid A, B)p(E \mid B, C)
\]
Directed graphical models

- A directed acyclic graph (DAG) on the random variables.
- Lack of directed edge from $X_i$ to $X_j$ means they are independent.
- Joint distribution can be factorized:

$$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i \mid X_{\text{Parents}(X_i)}) .$$

\[
p(A)p(B)p(C \mid A, B)p(E \mid B, C)p(D \mid E, C)
\]
\[ p(A)p(B)p(C \mid A, B)p(E \mid B, C)p(D \mid E, C) \]

• Suppose each variable has \( K \) values.

• Naive representation (fully connected graph): \( K^5 \)-entry table.

• Factorized representation:
  - A \( 1 \times K \) vector in \( A, B \)
  - A \( K \times K \times K \) table in \( C, E \) and \( D \).

• Total of \( 2K + 3K^3 \) instead of \( K^5 \).
  - Less storage;
  - Fewer parameters to estimate!
Bayes net: example

- Binary variables:
  - $E$: Earthquake
  - $B$: Burglary
  - $A$: Alarm went off
  - $J$: Neighbor Jim called
  - $M$: Neighbor Mary called

- Joint: $p(B, E, A, J, M) = p(E)p(B)p(A \mid E, B)p(J \mid A)p(M \mid A)$

- $E$ and $B$ are independent;

- $J$ and $B$ are dependent.
Suppose we hear the alarm, $A = T$.

$$p(A) = \sum_{B,E} p(A, B, E)$$
• Suppose we hear the alarm, $A = T$.

$$p(A) = \sum_{B,E} p(A, B, E)$$

$$= \sum_{B,E} p(A | B, E) p(B, E)$$

$$= \sum_{B,E} p(A | B, E) p(B) p(E)$$
Bayes net: inference

- Suppose we hear the alarm, $A = T$.

\[
p(A) = \sum_{B,E} p(A, B, E)
= \sum_{B,E} p(A | B, E) p(B, E)
= \sum_{B,E} p(A | B, E) p(B)p(E) = 0.0025
\]
Bayes net: inference

- Suppose we hear the alarm, \( A = T \).

\[
p(A) = \sum_{B,E} p(A, B, E) \\
= \sum_{B,E} p(A | B, E) p(B, E) \\
= \sum_{B,E} p(A | B, E) p(B)p(E) = 0.0025
\]

\[
p(E | A) = \frac{p(A | E) p(E)}{p(A)}
\]
Bayes net: inference

- Suppose we hear the alarm, $A = T$.

\[
p(A) = \sum_{B,E} p(A, B, E) \\
= \sum_{B,E} p(A \mid B, E) p(B, E) \\
= \sum_{B,E} p(A \mid B, E) p(B) p(E) = 0.0025
\]

\[
p(E \mid A) = \frac{p(A \mid E) p(E)}{p(A)}
\]

\[
p(A \mid E) = \sum_{B} p(A, B \mid E)
\]
Bayes net: inference

- Suppose we hear the alarm, $A = T$.

\[
p(A) = \sum_{B,E} p(A, B, E)
\]
\[
= \sum_{B,E} p(A \mid B, E) p(B, E)
\]
\[
= \sum_{B,E} p(A \mid B, E) p(B) p(E) = 0.0025
\]
\[
p(E \mid A) = \frac{p(A \mid E) p(E)}{p(A)}
\]
\[
p(A \mid E) = \sum_{B} p(A, B \mid E) = \sum_{B} p(A \mid B, E) p(B \mid E)
\]
Bayes net: inference

• Suppose we hear the alarm, \( A = T \).

\[
p(A) = \sum_{B,E} p(A, B, E)
\]
\[
= \sum_{B,E} p(A | B, E) p(B, E)
\]
\[
= \sum_{B,E} p(A | B, E) p(B) p(E) = 0.0025
\]

\[
p(E | A) = \frac{p(A | E) p(E)}{p(A)}
\]

\[
p(A | E) = \sum_B p(A, B | E) = \sum_B p(A | B, E) p(B | E) = \sum_B p(A | B, E) p(B)
\]
Bayes net: inference

\[
p(A \mid E) = \sum_B p(A \mid B, E) p(B) = 0.2913
\]

\[
p(E \mid A) = \frac{p(A \mid E) p(E)}{p(A)} = \frac{0.2913 \times 0.001}{0.0025} \approx 0.1165
\]

- Note: we did not have to consider \( J, M \) since given \( A, E \) is independent of them!
Bayes net: explaining away

\[ p(E | A) \approx 0.1165 \]

- Now suppose \( A = T \) and we know there was a burglary, \( B = T \).

\[
p(E | A, B) = \frac{p(A, B, E)}{p(A, B)} = \frac{p(A | B, E) p(B) p(E)}{p(A | B) p(B)}
\]
Bayes net: explaining away

\[ p(E \mid A) \approx 0.1165 \]

- Now suppose \( A = T \) and we know there was a burglary, \( B = T \).

\[
p(E \mid A, B) = \frac{p(A, B, E)}{p(A, B)} = \frac{p(A \mid B, E)p(B)p(E)}{p(A \mid B)p(B)} \approx 0.001 + \epsilon
\]
Bayes net: explaining away

\[ p(E \mid A) \approx 0.1165 \]

- Now suppose \( A = T \) and we know there was a burglary, \( B = T \).

\[
p(E \mid A, B) = \frac{p(A, B, E)}{p(A, B)} = \frac{p(A \mid B, E) p(B)p(E)}{p(A \mid B)p(B)} \approx 0.001 + \epsilon
\]

- Given \( A \), the variables \( B \) and \( E \) are no longer independent!
• $J$ and $M$ are not independent:
  – If we know (only) $J$, it changes our belief about $M$.
  – If we are told $A$ it changes our belief about both $J$ and $M$.

• However, $J$ and $M$ are conditionally independent given $A$!

• If we know $A$, and then are told $J$, it doesn’t help us to predict $M$!
Bayes net: larger example

(from Binder et al '97)
Applications of Bayes nets

- Microsoft:
  - The “paper clip”
  - Troubleshooters (printing, etc.)

- Medical diagnosis
  - about 600 diseases, 4,000 symptoms.

- ETS: computer-based adaptive tests.

- NASA: analysis of telemetry and prediction of failures.

- Many, many applications of HMMs and related continuous state-space models.
Bayes Ball: the rational pastime

• Query: is $A$ independent of $B$ given a set $X$

• Shade nodes in $X$

• Try to pass a ball along every undirected path from $A$ to $B$, according to the rules:
Bayes Ball: the rational pastime

• Query: is $A$ independent of $B$ given a set $X$

• Shade nodes in $X$

• Try to pass a ball along every undirected path from $A$ to $B$, according to the rules:
Bayes Ball: the rational pastime

- Query: is $A$ independent of $B$ given a set $X$?

- Shade nodes in $X$.

- Try to pass a ball along every undirected path from $A$ to $B$, according to the rules:

  ![Diagram showing the ball passing along undirected paths]

  - The ball passes along the paths from $A$ to $B$.
  - Nodes in $X$ are shaded.
  - The rules govern how the ball moves along the paths.
Bayes Ball: the rational pastime

- Query: is $A$ independent of $B$ given a set $X$?

- Shade nodes in $X$.

- Try to pass a ball along every undirected path from $A$ to $B$, according to the rules:
Bayes Ball: the rational pastime

- Query: is $A$ independent of $B$ given a set $X$
- Shade nodes in $X$
- Try to pass a ball along every undirected path from $A$ to $B$, according to the rules:
Bayes Ball: the rational pastime

- Query: is $A$ independent of $B$ given a set $X$
- Shade nodes in $X$
- Try to pass a ball along every undirected path from $A$ to $B$, according to the rules:
Bayes Ball: the rational pastime

- Query: is $A$ independent of $B$ given a set $X$

- Shade nodes in $X$

- Try to pass a ball along every undirected path from $A$ to $B$, according to the rules:
Bayes Ball: the rational pastime

• Query: is \( A \) independent of \( B \) given a set \( X \)

• Shade nodes in \( X \)

• Try to pass a ball along every undirected path from \( A \) to \( B \), according to the rules:
Bayes ball: example

- Traffic situation:

  \[ L \] traffic light green
  \[ N \] driver of \( Y \) is nice
  \[ S \] car \( Y \) stops
  \[ T \] car \( X \) turns
  \[ C \] crash!
Bayes ball: example

- Traffic situation:

  \( L \)  traffic light green
  \( N \)  driver of \( Y \) is nice
  \( S \)  car \( Y \) stops
  \( T \)  car \( X \) turns
  \( C \)  crash!

- Suppose we know \( S = 1 \), i.e. car \( Y \) stopped.

- Does knowing \( N \) tell us anything about \( T \)?
Bayes ball: example

• Traffic situation:

- $L$: traffic light green
- $N$: driver of $Y$ is nice
- $S$: car $Y$ stops
- $T$: car $X$ turns
- $C$: crash!

• Suppose we know $S = 1$, i.e. car $Y$ stooped.

• Does knowing $N$ tell us anything about $T$?
Inference in BN

- Simple when the graph has no *undirected* loops
  - Example: forward-backward algorithm in HMM.
  - Generalization of the FB is called *belief propagation*;
  - still can be done in two passes (*message passing* algorithm).

- Loopy graphs: inference is difficult
  - Generally have to resort to approximate inference.