CS221 Final Exam Review

Week 10
Questions encountered

- What is the difference between a Markov net, Bayesian net, HMM and Markov model?
- Why Gibbs sampling? How to compute $P(x_i | X_{-i})$?
- What is the FB algorithm? What does F and B mean? Why is it prob? Why do we only we use it for HMMs?
Outline

- Markov networks vs Bayesian networks vs Markov models vs HMMs
- Markov networks
  - Gibbs sampling
- Bayesian networks
  - Forward backward algorithm

What are we leaving out?

- More about Bayesian networks - PS week 8
- Logic - PS week 9
Outline

- Markov networks vs Bayesian networks vs Markov models vs HMMs
- Markov networks
  - Gibbs sampling
- Bayesian networks
  - Forward backward algorithm
All the different nets

\[ g(x_1, x_2, x_3, x_4) = f_1(x_1, x_2, x_3) \times f_2(x_3, x_4) \]

Markov networks: \( g = P \) when normalized, \( f_i \)'s \( \geq 0 \)

Bayesian networks: \( g = P \), and \( f_i \)'s are conditional Ps, hence directed and \( Z = 1 \)
All the different nets

\[ g(x_1, x_2, x_3, x_4) = f_1(x_1, x_2, x_3) \times f_2(x_3, x_4) \]

Markov networks: \( g = P \) when normalized, \( f_i \)'s \( \geq 0 \)

Bayesian networks: \( g = P \), and \( f_i \)'s are conditional Ps, hence directed and \( Z = 1 \)
All the different nets

\[ g(x_1, x_2, x_3, x_4) = f_1(x_1, x_2, x_3) \times f_2(x_3, x_4) \]

Why factorize?

- Simplifies \( g \)
- Reduced \# \ text{ params}
- Table size: \( O(|\text{domain}|^4) \)
  reduced to \( O(|\text{domain}|^3) \)
Outline

- Markov networks vs Bayesian networks vs Markov models vs HMMs
- **Markov networks**
  - Gibbs sampling
- **Bayesian networks**
  - Forward backward algorithm
Markov nets

\[\begin{array}{ccc}
\kappa_1 & \kappa_2 & \text{weight}(\kappa_1, \kappa_2) \\
0 & 0 & 2 \\
0 & 1 & 0 \\
1 & 0 & 1 \\
1 & 1 & 2 \\
\end{array} \]

\[\begin{array}{ccc}
\kappa_3 & \kappa_4 & \text{weight}(\kappa_3, \kappa_4) \\
0 & 0 & 1 \\
0 & 1 & 2 \\
1 & 0 & 2 \\
1 & 1 & 1 \\
\end{array} \]
Markov nets

\[
\begin{array}{ccc}
\kappa_1 & \kappa_3 & \text{weight}(\kappa_1, \kappa_3) \\
0 & 0 & 2 \\
0 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 2 \\
\end{array}
\quad \begin{array}{ccc}
\kappa_1 & \kappa_2 & \text{weight}(\kappa_1, \kappa_2) \\
0 & 0 & 1 \\
0 & 1 & 2 \\
1 & 0 & 2 \\
1 & 1 & 1 \\
\end{array}
\quad \begin{array}{ccc}
\kappa_3 & \kappa_n & \text{weight}(\kappa_3, \kappa_n) \\
0 & 0 & 1 \\
0 & 1 & 2 \\
1 & 0 & 2 \\
1 & 1 & 1 \\
\end{array}
\]

What is \( P(\kappa_1, \kappa_2, \kappa_3, \kappa_n) \)?

\[
\frac{1}{Z} \times \frac{\text{weight}(\kappa_1, \kappa_2) \times \text{weight}(\kappa_1, \kappa_3) \times \text{weight}(\kappa_3, \kappa_n)}{2}
\]
Markov nets

\[
\begin{array}{ccc}
\pi_1 & \pi_3 & \text{weight}(\pi_1, \pi_3) \\
0 & 0 & 2 \\
0 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 2 \\
\end{array}
\]

\[
\begin{array}{ccc}
\pi_1 & \pi_2 & \text{weight}(\pi_1, \pi_2) \\
0 & 0 & 1 \\
0 & 1 & 2 \\
1 & 0 & 2 \\
1 & 1 & 1 \\
\end{array}
\]

\[
\begin{array}{ccc}
\pi_3 & \pi_n & \text{weight}(\pi_3, \pi_n) \\
0 & 0 & 1 \\
0 & 1 & 2 \\
1 & 0 & 2 \\
1 & 1 & 1 \\
\end{array}
\]

what is \( P(\pi_1) \)?

\[
\sum_{\pi_2, \pi_3, \pi_n} P(\pi_1, \pi_2, \pi_3, \pi_n)
\]

expensive!!
Markov nets

\[
\begin{array}{ccc}
\pi_1 & \pi_3 & \text{weight}(\pi_1, \pi_3) \\
0 & 0 & 2 \\
0 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 2 \\
\end{array}
\quad \begin{array}{ccc}
\pi_1 & \pi_2 & \text{weight}(\pi_1, \pi_2) \\
0 & 0 & 1 \\
0 & 1 & 2 \\
1 & 0 & 1 \\
1 & 1 & 1 \\
\end{array}
\quad \begin{array}{ccc}
\pi_3 & \pi_\eta & \text{weight}(\pi_3, \pi_\eta) \\
0 & 0 & 1 \\
0 & 1 & 2 \\
1 & 0 & 1 \\
1 & 1 & 1 \\
\end{array}
\]

What is \( P(\pi_1 | \pi_2 = 0, \pi_3 = 1, \pi_4 = 0) \)?

\[
P(\pi_1 = 0 | 0, 1, 0) = \frac{1}{s + n}
\]

\[
P(\pi_1 = 1 | 0, 1, 0) = \frac{n}{1 + n}
\]

\( x_1 = 0 \quad 1 \times 1 \)

\( x_1 = 1 \quad 2 \times 2 \)

Cheap!

Sum over \( x_1 \) only!!
Markov nets: Why Gibbs?

- Compute $P(x_i)$ using $P(x_i \mid X_{-i})$ instead of summing over $X_{-i}$

**Algorithm: Gibbs sampling**

- Initialize $x$ to a random complete assignment
- Loop through $i = 1, \ldots, n$ until convergence:
  - Set $x_i = v$ with prob. $P(X_i = v \mid X_{-i} = x_{-i})$
    - ($X_{-i}$ denotes all variables except $X_i$)
  - Increment count$_i(x_i)$
- Estimate $\hat{P}(X_i = x_i) = \frac{\text{count}_i(x_i)}{\sum_v \text{count}_i(v)}$

For more practice on Gibbs sampling: refer PS week 7
Outline

- Markov networks vs Bayesian networks vs Markov models vs HMMs
- Markov networks
  - Gibbs sampling
- Bayesian networks
  - Forward backward algorithm
Forward backward algorithm

- Compute $P(h_i \mid e's)$
- Applicable only to HMMs or similar
- What special about markov models?
  - One parent for each node

Intuition:

$F(h_i) = P(h_i, e_i \mid e_{<i})$
$B(h_i) = P(e_{>i} \mid h_i)$
$S(h_i) \propto P(h_i, e_i \mid e_{<i}) \times P(e_{>i} \mid h_i) \propto P(h_i \mid e's)$

For more details refer: wiki
Problem: P2, Winter 2021 Exam 2

Pick a coin

Toss it

H / T?

\[ P_0(C_X) = \lambda_0 \]

With probability \( \lambda \) coin at “t” is same as “t-1”

\[ P_C(H) = p_C \]
How does the bayesian net look?

\[ p(c_i) = \begin{cases} 0 & \text{if } c_i = x \\ 1-\gamma & \text{else} \end{cases} \]

\[ p(c_{i+1} | c_i) = \begin{cases} \gamma & \text{if } c_{i+1} = c_i \\ 1-\gamma & \text{else} \end{cases} \]

\[ p(o_i | c_i) = \begin{cases} p_{c_i} & \text{if } o_i = H \\ 1-p_{c_i} & \text{else} \end{cases} \]
Inference: FB in practice

- Stop after two steps, observe \{H,H\}
Inference: FB in practice

- Stop after two steps, observe \{H,H\}
- Draw FB lattice representation
Inference: FB in practice

- Stop after two steps, observe \{H,H\}
- What are the weights of edges?
Inference: FB in practice

- Stop after two steps, observe \{H,H\}
- Compute forward passes and backward passes
Inference: FB in practice

- Stop after two steps, observe \{H,H\}
- Compute forward passes and backward passes

\[
\begin{align*}
F_1(X) &= \lambda_0 p_X \\
F_1(Y) &= (1 - \lambda_0) p_Y \\
B_2(X) &= 1 \\
B_2(Y) &= 1
\end{align*}
\]

\[
\begin{align*}
F_2(X) &= w(X, X) \cdot F_1(X) + w(Y, X) \cdot F_1(Y) \\
F_2(Y) &= w(X, Y) \cdot F_1(X) + w(Y, Y) \cdot F_1(Y) \\
B_1(X) &= w(X, X) \cdot B_2(X) + w(X, Y) \cdot B_2(Y) \\
B_1(Y) &= w(Y, X) \cdot B_2(X) + w(Y, Y) \cdot B_2(Y)
\end{align*}
\]

For more about Bayesian networks: **PS Week 8**
Thank you!

Good luck on the exam!