CS221 Final Exam Review

Week 10

Stanford University

Questions encountered

- What is the difference between a Markov net, Bayesian net, HMM and Markov model?
- Why Gibbs sampling? How to compute $P(x_i \mid X_{-i})$?
- 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

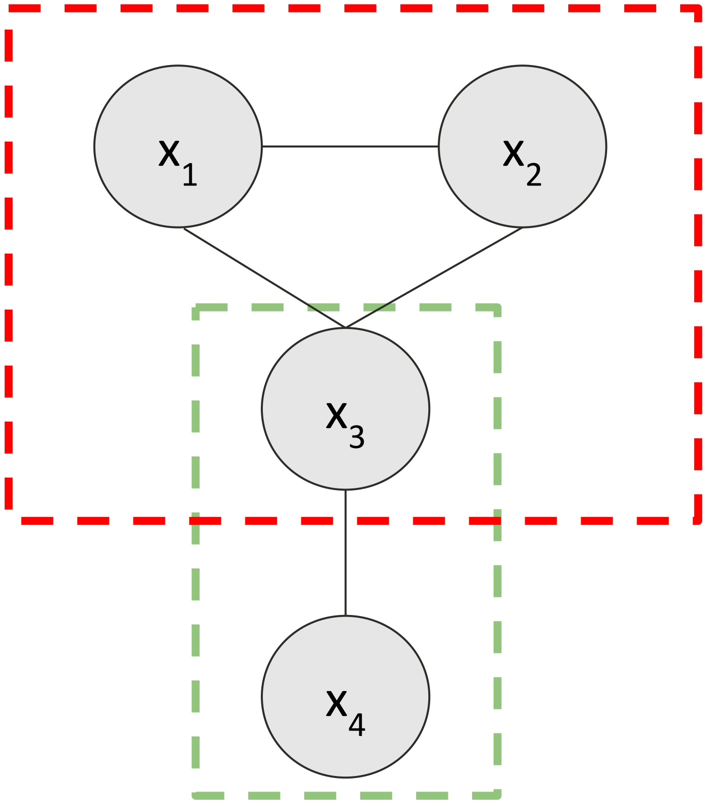
- **Markov networks vs Bayesian networks vs Markov models vs HMMs**
- Markov networks
  - Gibbs sampling
- Bayesian networks
  - Forward backward algorithm

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 >= 0

Bayesian networks: g = P, and f_i's are conditional Ps, hence directed and Z = 1



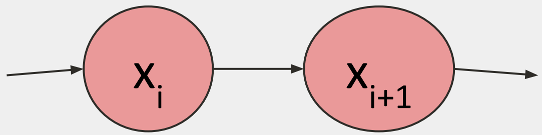
All the different nets

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

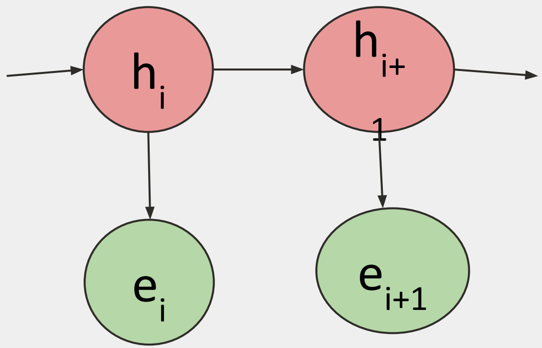
Markov networks: g = P when normalized, f_i's >= 0

Bayesian networks: g = P, and f_i's are conditional Ps, hence directed and Z = 1

Markov models:



HMMs:

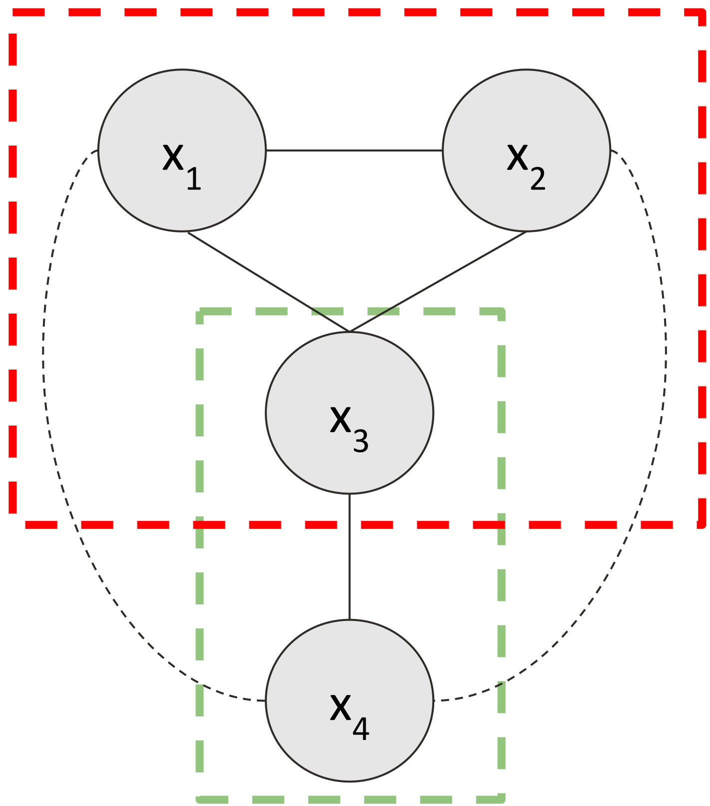


All the different nets

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

Why factorize?

- Simplifies g
- Reduced # params
- Table size: $O(|domain|^4)$ reduced to $O(|domain|^3)$



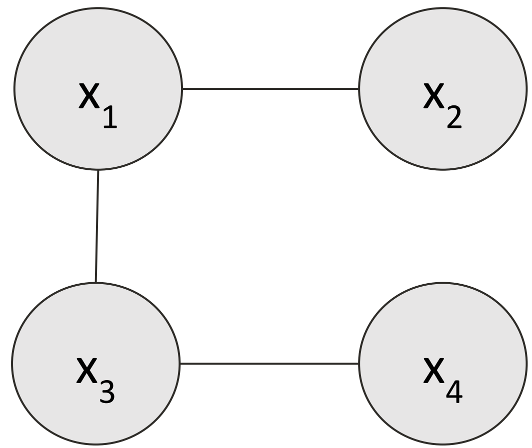
Outline

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

Markov nets

x_1	x_2	weight(x_1, x_2)
0	0	1
0	1	2
1	0	2
1	1	1

x_1	x_3	weight(x_1, x_3)
0	0	2
0	1	1
1	0	1
1	1	2



x_3	x_4	weight(x_3, x_4)
0	0	1
0	1	2
1	0	2
1	1	1

Markov nets

x_1	x_3	weight(x_1, x_3)
0	0	2
0	1	1
1	0	1
1	1	2

x_1	x_2	weight(x_1, x_2)
0	0	1
0	1	2
1	0	2
1	1	1

x_3	x_4	weight(x_3, x_4)
0	0	1
0	1	2
1	0	2
1	1	1

What is $P(x_1, x_2, x_3, x_4)$?

$$\frac{1}{Z}\, \text{weight}(x_1, x_2)\, \text{weight}(x_1, x_3)\, \text{weight}(x_3, x_4)$$

Markov nets

x_1	x_3	weight(x_1, x_3)
0	0	2
0	1	1
1	0	1
1	1	2

x_1	x_2	weight(x_1, x_2)
0	0	1
0	1	2
1	0	2
1	1	1

x_3	x_4	weight(x_3, x_4)
0	0	1
0	1	2
1	0	2
1	1	1

what is $P(x_1)$?

$$\sum_{x_2, x_3, x_4} P(x_1, x_2, x_3, x_4)$$

Expensive !!

Markov nets

x_1	x_3	weight(x_1, x_3)
0	0	2
0	1	1
1	0	1
1	1	2

x_1	x_2	weight(x_1, x_2)
0	0	1
0	1	2
1	0	2
1	1	1

x_3	x_4	weight(x_3, x_4)
0	0	1
0	1	2
1	0	2
1	1	1

What is $P(x_1 \mid x_2 = 0, x_3 = 1, x_4 = 0)$?

$x_1 = 0 \mid 1 \times 1$

$x_1 = 1 \mid 2 \times 2$

$\therefore P(x_1 = 0 \mid 0, 1, 0) = \frac{1}{1+4}$

$P(x_1 = 1 \mid 0, 1, 0) = \frac{4}{1+4}$

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. $\mathbb{P}(X_i = v \mid X_{-i} = x_{-i})$

    (X_{-i} denotes all variables except X_i)

  Increment $\text{count}_i(x_i)$

Estimate $\hat{\mathbb{P}}(X_i = x_i) = \frac{\text{count}_i(x_i)}{\sum_v \text{count}_i(v)}$

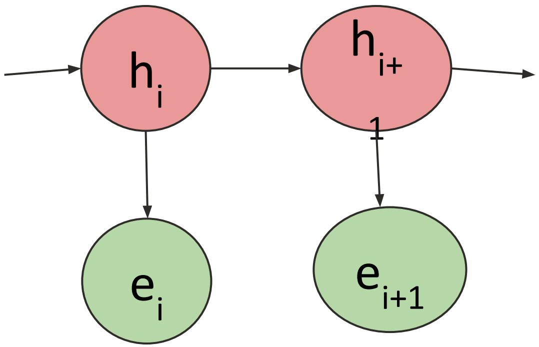
For more practice on Gibbs sampling: refer [PS week 7]

Outline

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

Forward backward algorithm

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



Intuition:

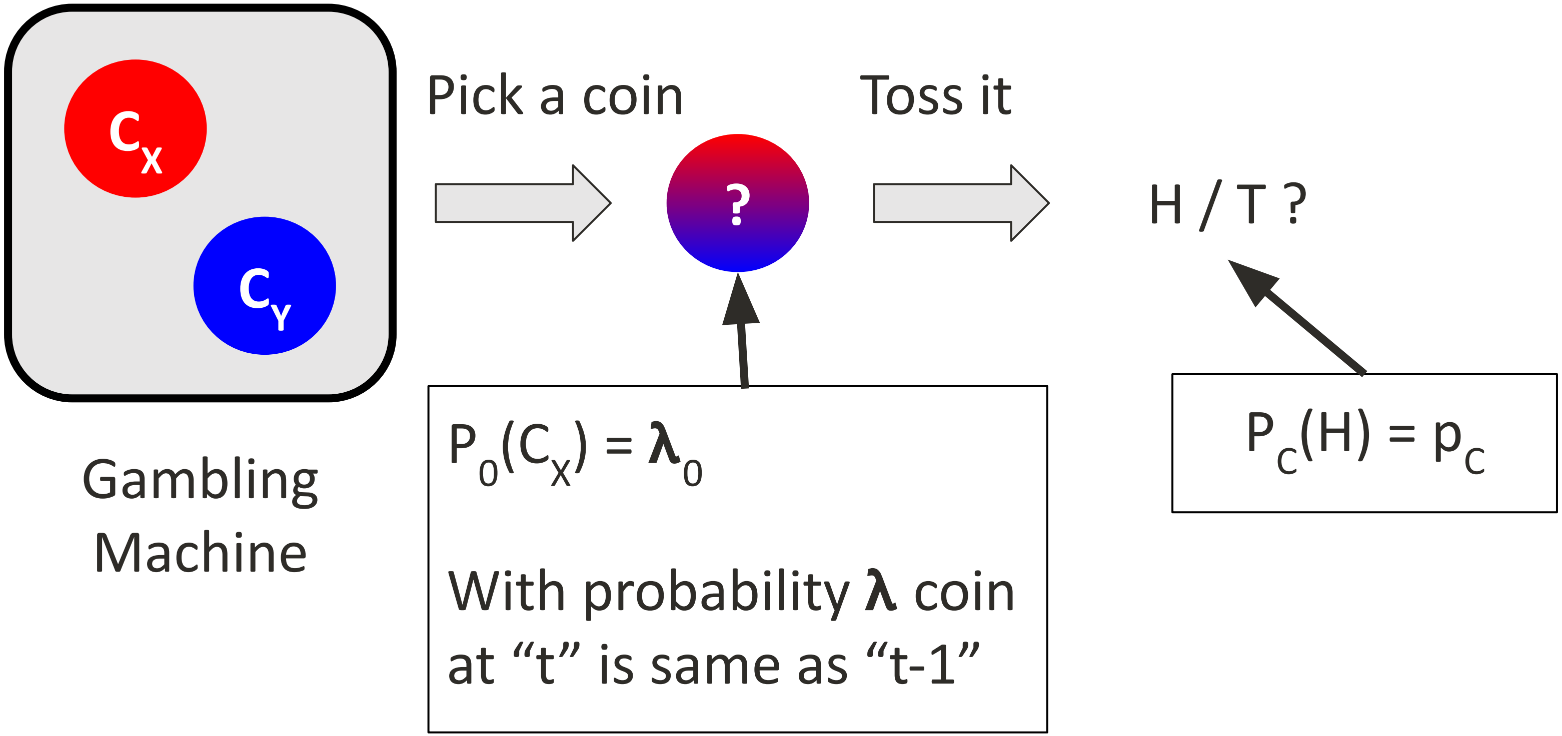
$F(h_i) = P(h_i, e_i | e_{<i})$

$B(h_i) = P(e_{>i} | h_i)$

$S(h_i) \propto P(h_i, e_i | e_{<i}) \times P(e_{>i} | h_i) \propto P(h_i | e\text{'s})$

For more details refer: wiki

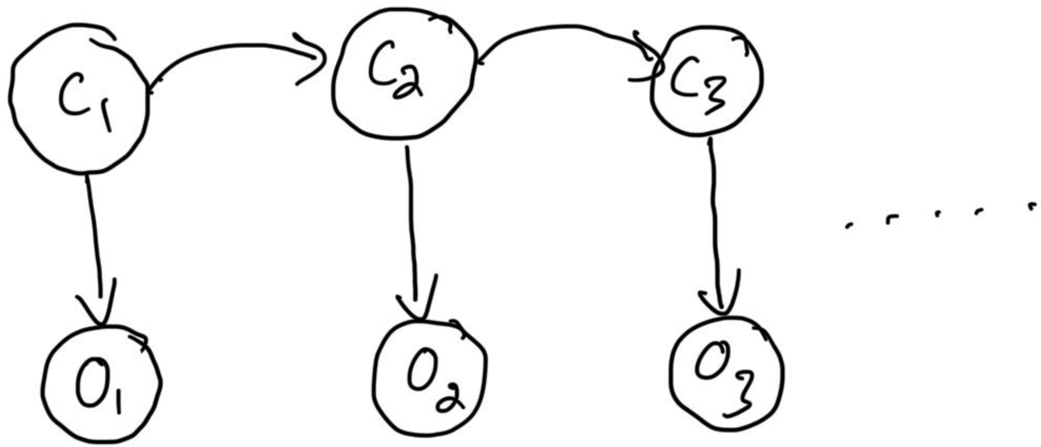
Problem: P2, Winter 2021 Exam 2



How does the bayesian net look?

$$P(C_1) = \begin{cases} \lambda_0 & \text{if } C_1 = X \\ 1-\lambda_0 & \text{else} \end{cases}$$

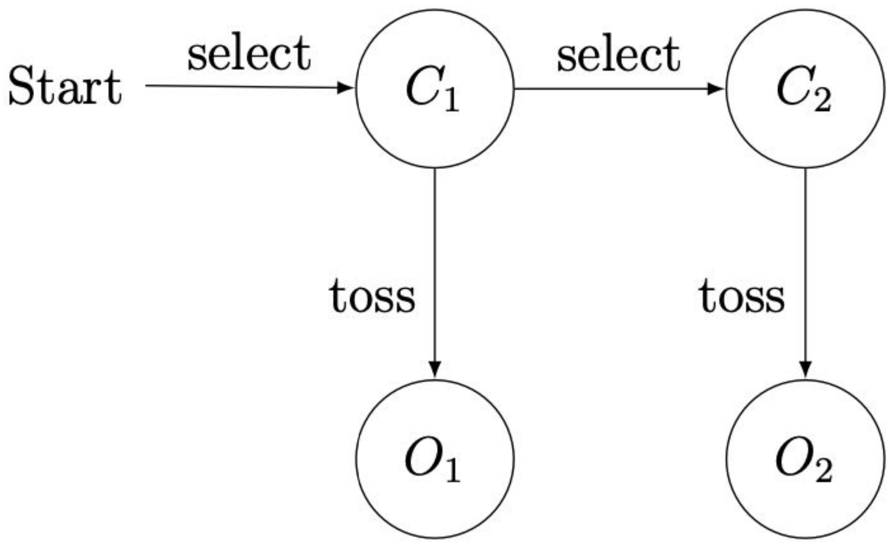
$$P(C_{i+1} \mid C_i) = \begin{cases} \lambda & \text{if } C_{i+1} = C_i \\ 1-\lambda & \text{else} \end{cases}$$



$$P(O_i \mid C_i) = \begin{cases} p_{C_i} & \text{if } O_i = H \\ 1-p_{C_i} & \text{else} \end{cases}$$

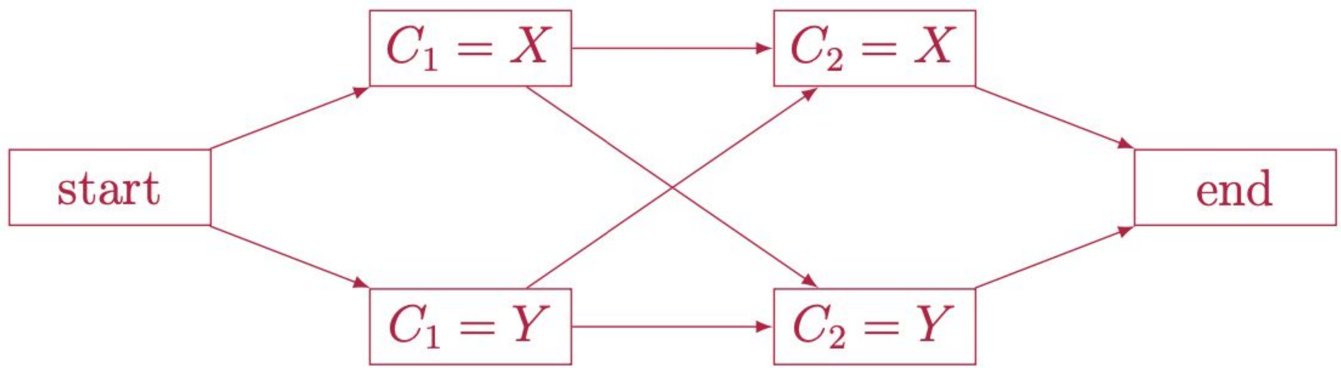
Inference: FB in practice

- Stop after two steps, observe {H,H}



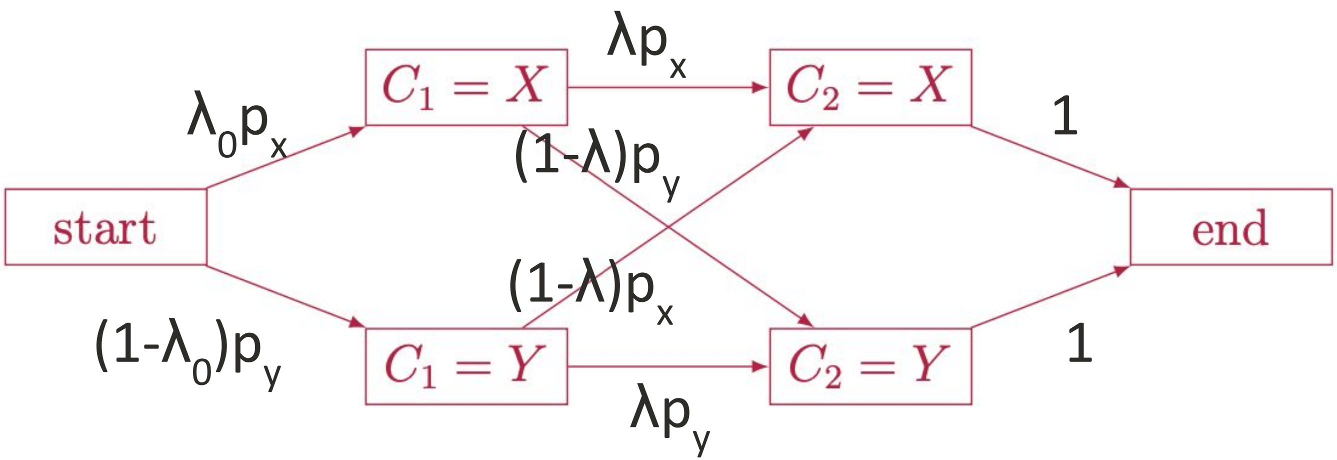
Inference: FB in practice

- Stop after two steps, observe {H,H}
- Draw FB lattice representation



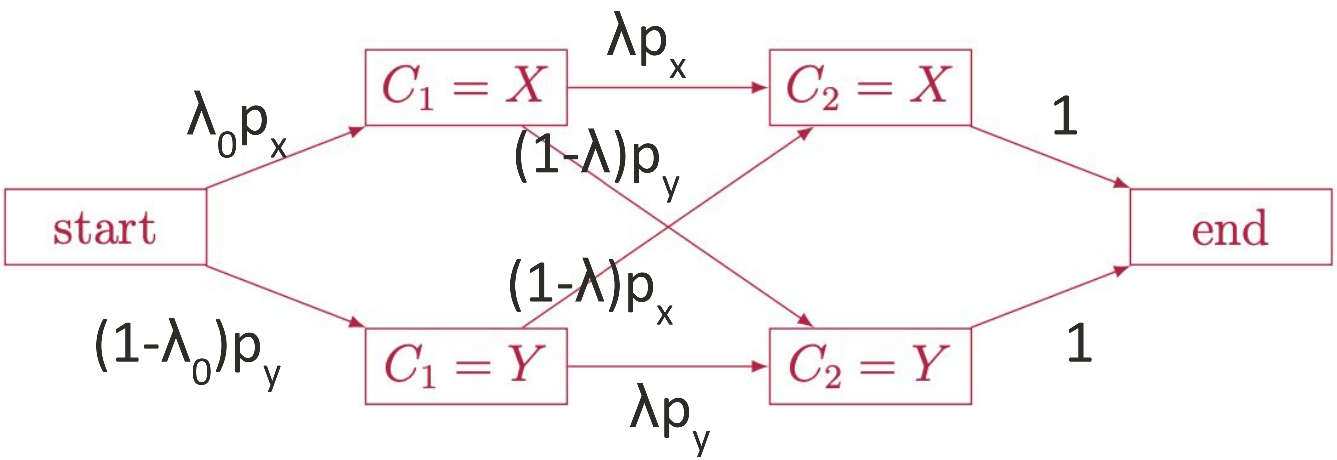
Inference: FB in practice

- 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

$$F_1(X) = \lambda_0 p_X$$
$$F_1(Y) = (1-\lambda_0) p_Y$$
$$B_2(X) = 1$$
$$B_2(Y) = 1$$

$$F_2(X) = w(X,X) \cdot F_1(X) + w(Y,X) \cdot F_1(Y)$$
$$F_2(Y) = w(X,Y) \cdot F_1(X) + w(Y,Y) \cdot F_1(Y)$$
$$B_1(X) = w(X,X) \cdot B_2(X) + w(X,Y) \cdot B_2(Y)$$
$$B_1(Y) = w(Y,X) \cdot B_2(X) + w(Y,Y) \cdot B_2(Y)$$

For more about Bayesian networks: PS Week 8

Thank you!

Good luck on the exam!