Bayesian network. Graphical models

Virgil Pavlu September 30, 2008

1 Introduction to probabilities, statistics

- random variables X,Y and density functions
- conditional probability $\mathbf{P}[X|Y]$
- joint probability $\mathbf{P}[X,Y] = \mathbf{P}[X|Y] \cdot \mathbf{P}[Y]$
- Bayes rule $\mathbf{P}[X|Y] \cdot \mathbf{P}[Y] = \mathbf{P}[Y|X] \cdot \mathbf{P}[X]$
- independence $\mathbf{P}[X,Y] = \mathbf{P}[X] \cdot \mathbf{P}[Y]$
- marginalization $\mathbf{P}[X] = \sum_{Y=y} \mathbf{P}[X|Y=y] \cdot \mathbf{P}[Y=y]$

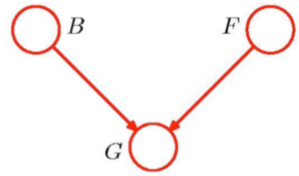
	red	blue	green	
square	0.25	0.10	0.21	0.56
round	0.17	0.04	0.23	0.44
	0.42	0.14	0.44	

Figure 1: probabilities

2 Bayesian networks. Inference example

Also called *Belief networks* or *Probabilistic networks*.

$$
\begin{aligned}
p(G=1|B=1,F=1) &= 0.8\\
p(G=1|B=1,F=0) &= 0.2\\
p(G=1|B=0,F=1) &= 0.2\\
p(G=1|B=0,F=0) &= 0.1
\end{aligned}
$$



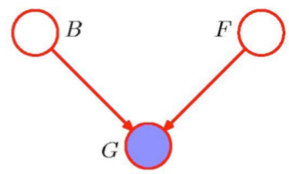
$$
\begin{aligned}
p(B=1) &= 0.9\\
p(F=1) &= 0.9
\end{aligned}
$$

and hence

$$p(F=0) = 0.1$$

B = Battery (0=flat, 1=fully charged)
F = Fuel Tank (0=empty, 1=full)
G = Fuel Gauge Reading (0=empty, 1=full)

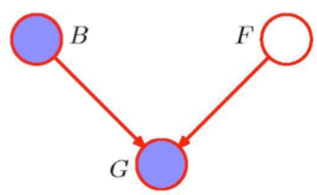
Figure 2: bayesian computation



$$
\begin{aligned}
p(F=0|G=0) &= \frac{p(G=0|F=0)p(F=0)}{p(G=0)}\\
&\simeq 0.257
\end{aligned}
$$

Probability of an empty tank increased by observing $G=0$.

Figure 3: bayesian computation



$$
\begin{aligned}
p(F=0|G=0,B=0) &= \frac{p(G=0|B=0,F=0)p(F=0)}{\sum_{F\in\{0,1\}} p(G=0|B=0,F)p(F)}\\
&\simeq 0.111
\end{aligned}
$$

Probability of an empty tank reduced by observing $B=0$.
This referred to as "explaining away".

Figure 4: bayesian computation

3 Bayesian networks. Factorization

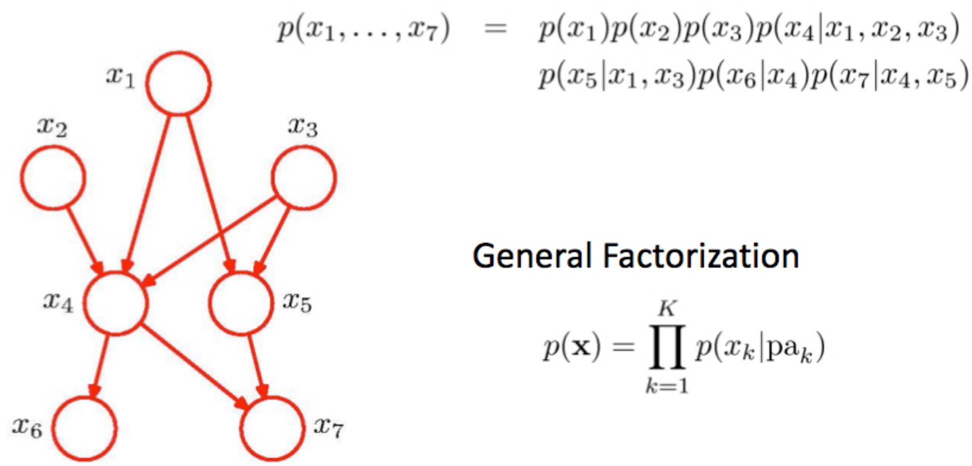


Figure 5: bayesian network factorization

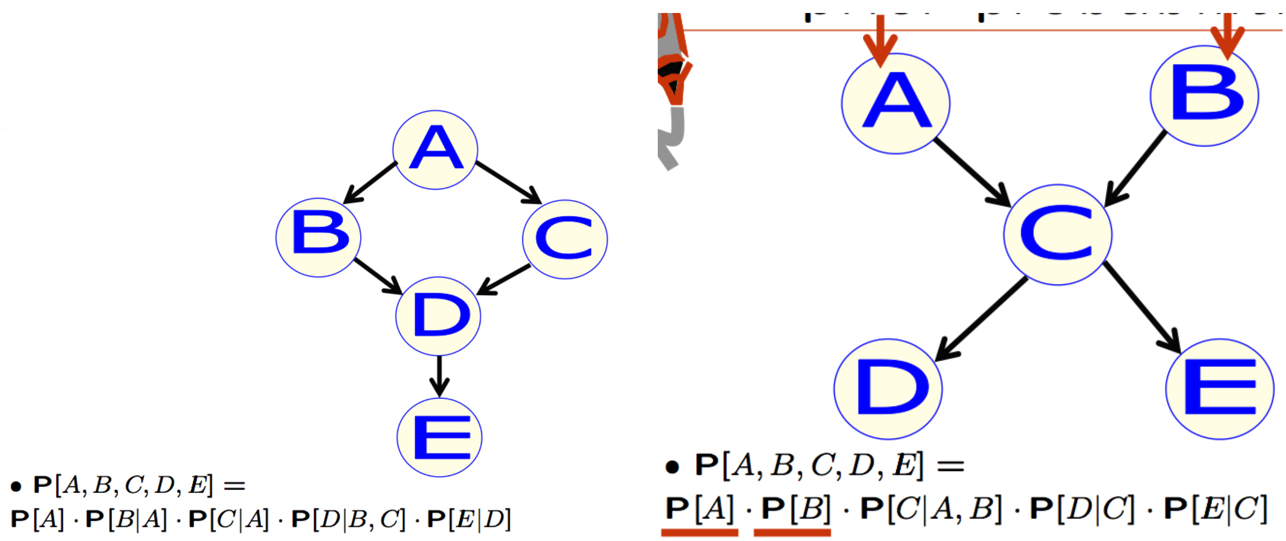


Figure 6: bayesian network factorization

4 Complexity of generative models

- cut links or assume independence of components

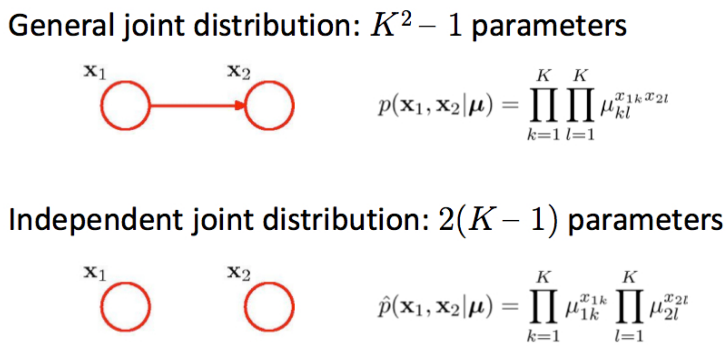


Figure 7: reduce parameters

- share parameters

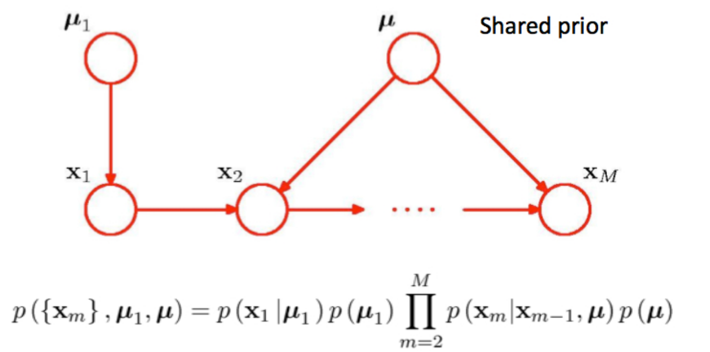


Figure 8: share parameters

- use parametrized models for conditional distribution instead of tables

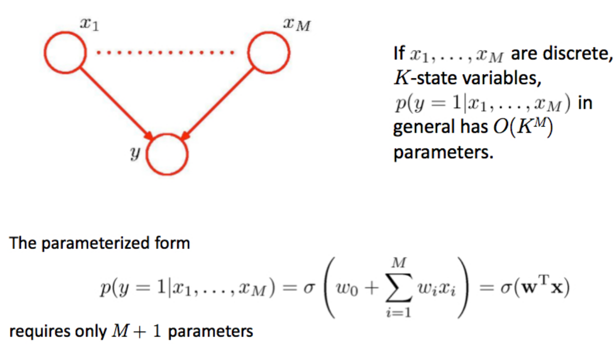


Figure 9: parametrize conditional

5 Conditional independence

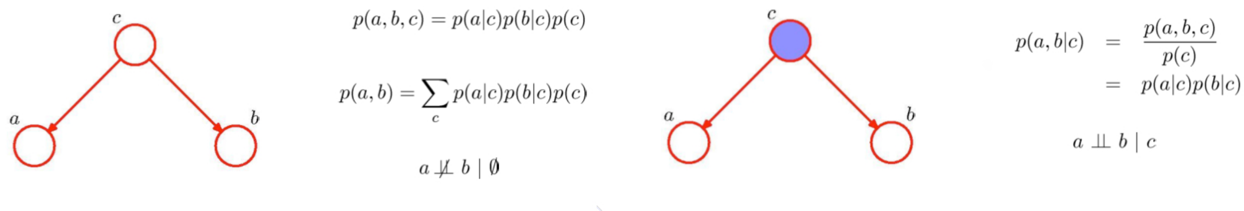


Figure 10: conditional independence

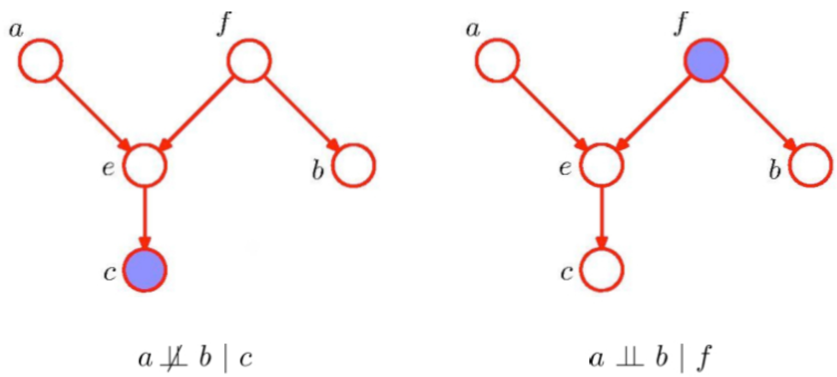
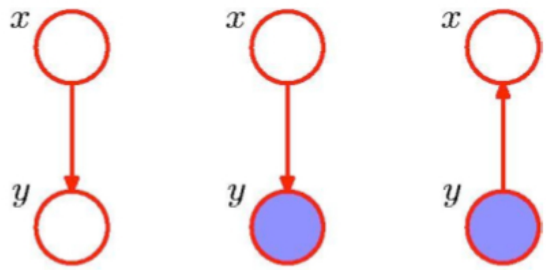


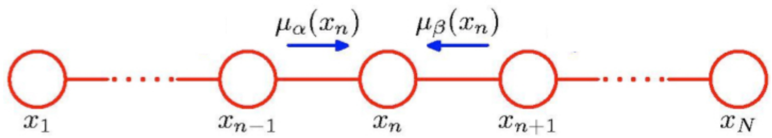
Figure 11: conditional independence

6 Inference in graphical models



$$p(y) = \sum_{x'} p(y|x')p(x') \qquad\qquad p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

Figure 12: graphical inference

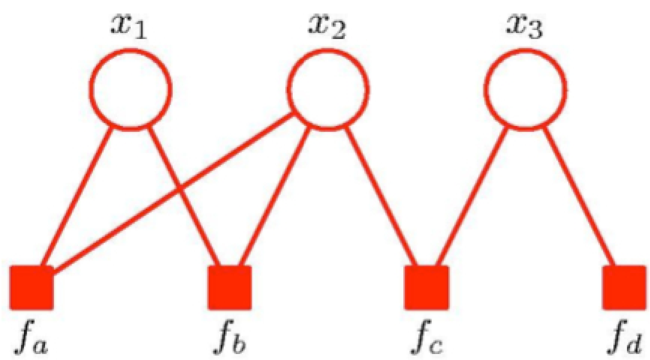


$$p(x_n) = \frac{1}{Z} \underbrace{\left[\sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \cdots \left[\sum_{x_1} \psi_{1,2}(x_1, x_2)\right] \cdots \right]}_{\mu_\alpha(x_n)}$$

$$\underbrace{\left[\sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \cdots \left[\sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N)\right] \cdots \right]}_{\mu_\beta(x_n)}$$

Figure 13: graphical inference

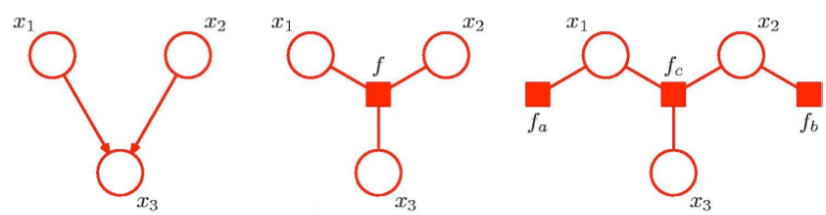
7 Factor graphs



$$p(\mathbf{x}) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3)$$

$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s)$$

Figure 14: Factor graph



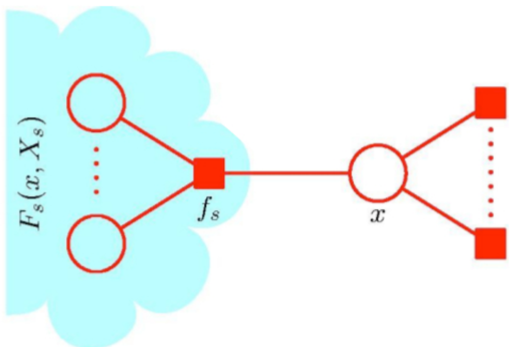
$$p(\mathbf{x}) = p(x_1)p(x_2)\, p(x_3|x_1, x_2)$$

$$f(x_1, x_2, x_3) = p(x_1)p(x_2)p(_3|x_1, x_2)$$

$$\begin{aligned} f_a(x_1) &= p(x_1) \\ f_b(x_2) &= p(x_2) \\ f_c(x_1, x_2, x_3) &= p(x_3|x_1, x_2) \end{aligned}$$

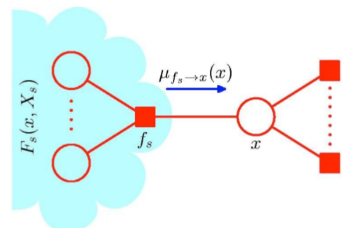
Figure 15: Factor graph

8 Belief propagation: sum-product algorithm



$$p(x) = \sum_{\mathbf{x}\setminus x} p(\mathbf{x})$$

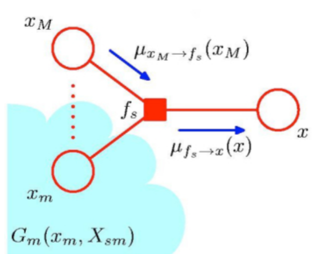
$$p(\mathbf{x}) = \prod_{s\in\mathrm{ne}(x)} F_s(x, X_s)$$



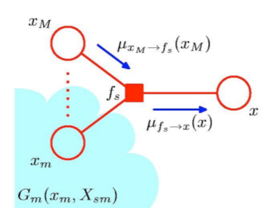
$$\begin{aligned} p(x) &= \prod_{s\in\mathrm{ne}(x)} \left[\sum_{X_s} F_s(x, X_s)\right] \\ &= \prod_{s\in\mathrm{ne}(x)} \mu_{f_s\to x}(x). \end{aligned}$$

$$\mu_{f_s\to x}(x) \equiv \sum_{X_s} F_s(x, X_s)$$

Figure 16: Sum-product algorithm

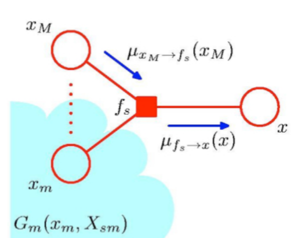


$$F_s(x, X_s) = f_s(x, x_1, \dots, x_M) G_1(x_1, X_{s1}) \dots G_M(x_M, X_{sM})$$



$$\begin{aligned} \mu_{f_s\to x}(x) &= \sum_{x_1}\dots\sum_{x_M} f_s(x, x_1, \dots, x_M) \prod_{m\in\mathrm{ne}(f_s)\setminus x} \left[\sum_{X_{sm}} G_m(x_m, X_{sm})\right] \\ &= \sum_{x_1}\dots\sum_{x_M} f_s(x, x_1, \dots, x_M) \prod_{m\in\mathrm{ne}(f_s)\setminus x} \mu_{x_m\to f_s}(x_m) \end{aligned}$$

Figure 17: Sum-product algorithm



$$\begin{aligned} \mu_{x_m\to f_s}(x_m) \equiv \sum_{X_{sm}} G_m(x_m, X_{sm}) &= \sum_{X_{sm}} \prod_{l\in\mathrm{ne}(x_m)\setminus f_s} F_l(x_m, X_{ml}) \\ &= \prod_{l\in\mathrm{ne}(x_m)\setminus f_s} \mu_{f_l\to x_m}(x_m) \end{aligned}$$

Figure 18: Sum-product algorithm

Lets look at an example

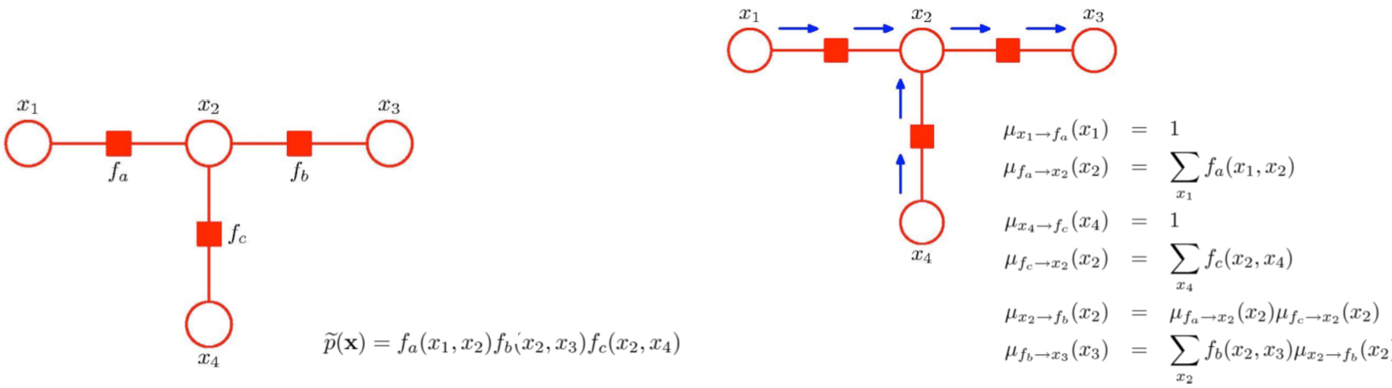


Figure 19: Sum-product example

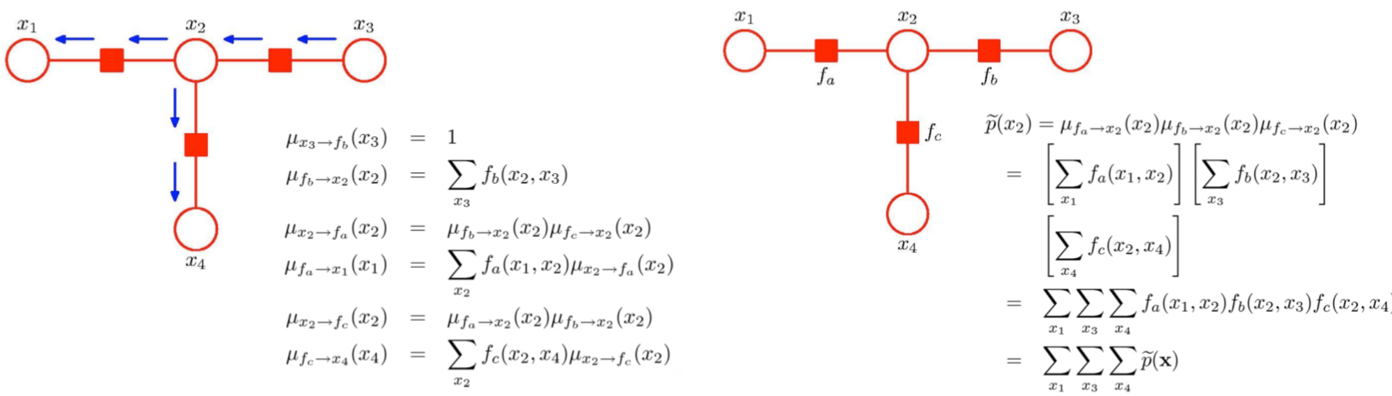


Figure 20: Sum-product example

9 Max-sum algorithm

10 Junction trees. Loopy belief propogation