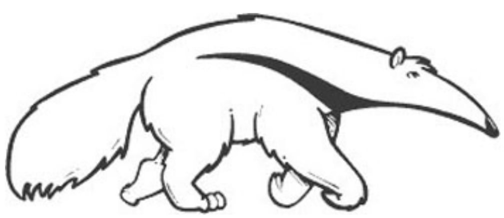
Undirected Graphical Models

Introduction to Graphical Models

Prof. Alexander Ihler



BREN:ICS
INFORMATION AND COMPUTER SCIENCES

UNIVERSITY of CALIFORNIA IRVINE

Gibbs distributions

"α" are sets of variable indices;
X_α are the associated random variables
x_α are their values in a configuration x

- Define $p(X) = \frac{1}{Z} \prod_\alpha f_\alpha(X_\alpha)$
  - where $f_\alpha(x_\alpha) \geq 0 \quad \forall x_\alpha$

- We call $f(X) = \prod_\alpha f_\alpha(X_\alpha)$ "unnormalized measure"

  $Z = \sum_x \prod_\alpha f_\alpha(x_\alpha)$ "partition function" (normalizes p(x) to sum to 1)

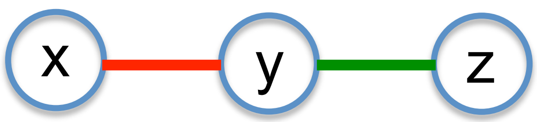
- Chain rule expansion (Bayes net) is a special case with Z=1
- Called "Gibbs" or "Boltzmann" distributions from physics:

$$p(X) = \frac{1}{Z(\tau)} \exp\left[-\sum_\alpha E_\alpha(X_\alpha)/k\tau \right] \qquad -\frac{1}{k\tau} E_\alpha(x_\alpha) = \log f_\alpha(x_\alpha)$$

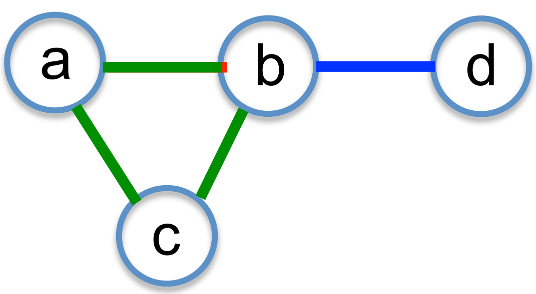
Markov graphs

- Undirected graphical model
- Variables represented by nodes
- Connect a variables if they appear in the same scope

- Ex: $p(x, y, z) = p(x)\, p(y \,|\, x)\, p(z \,|\, y)$



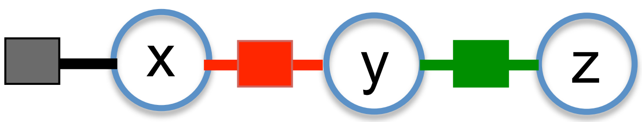
- Ex: $p(a, b, c, d) = p(a)\, p(b \,|\, a)\, p(c \,|\, a, b)\, p(d \,|\, b)$



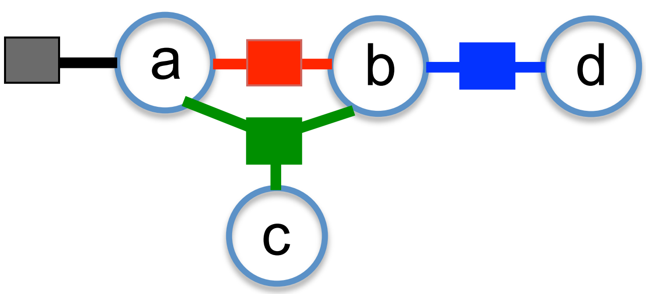
Factor graphs

- Undirected graphical model
- Variables, factors are each represented by nodes
- Connect a factor to the variables in its scope

- Ex: $p(x, y, z) = p(x)\, p(y \mid x)\, p(z \mid y)$

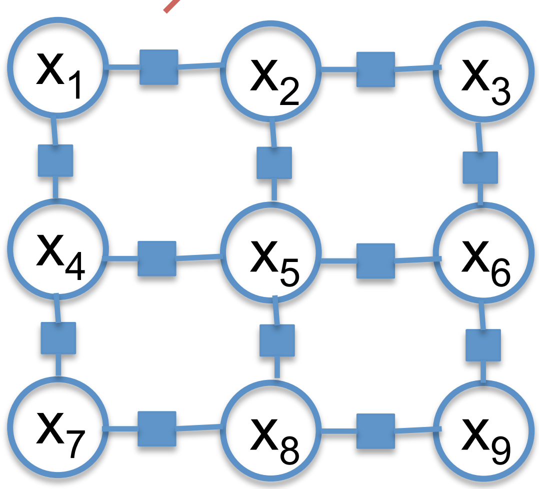


- Ex: $p(a, b, c, d) = p(a)\, p(b \mid a)\, p(c \mid a, b)\, p(d \mid b)$



Factors

- Factors do not need to be conditional probabilities
  - Or even probabilities at all
  - Values correspond to “relative” probability for certain configurations



X_1	X_2	F(.)
0	0	3
0	1	1
1	0	1
1	1	3

“Equal values are more likely than non-equal values”

Same relationship holds for all neighboring pairs of pixels

Factors & Exponential Family Models

- Table-based functions as (overcomplete) exp. family models

X_1	X_2	$f_{12}(.)$
0	0	2.71
0	1	1
1	0	0.37
1	1	7.39

$\overset{\log(\cdot)}{\Longrightarrow}$

X_1	X_2	$\log f_{12}(.)$	
0	0	1.0	$= \theta_{12;00}$
0	1	0.0	$= \theta_{12;01}$
1	0	-1.0	$= \theta_{12;10}$
1	1	2.0	$= \theta_{12;11}$

$$f_{12}(X_1 = a, X_2 = b) = \exp\left(\theta_{12;ab}\, \mathbb{1}[X_1 = a, X_2 = b]\right)$$

$$\Rightarrow \quad f_{12}(X_1, X_2) = \exp\Big(\sum_{a,b} \theta_{12;ab}\, \mathbb{1}[X_1 = a, X_2 = b]\Big)$$

$$= \exp\big(\theta \cdot u(X)\big)$$

$$\theta = [\,\theta_{12;00} \ldots \theta_{12;11}\,]$$

$$u(X) = [\,\mathbb{1}[X_1 = 0, X_2 = 0] \ldots]$$

More generally,

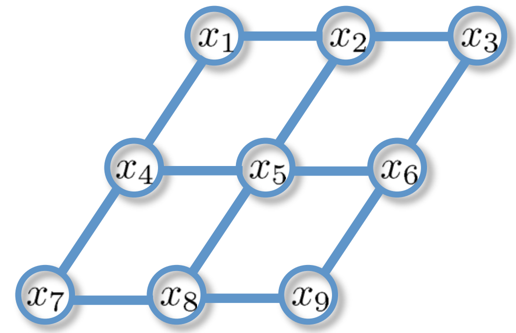
$$f(X) = \prod_{\alpha} f_\alpha(X_\alpha) = \exp\Big(\sum_{\alpha}\sum_{x_\alpha} \theta_{\alpha;x_\alpha}\, \mathbb{1}[X_\alpha = x_\alpha]\Big)$$

Example

- Ising model: $X_i \in \{-1, +1\}$

$$p(X) = \frac{1}{Z} \exp\Big[\sum_i \theta_i X_i + \sum_{ij \in E} \theta_{ij} X_i X_j\Big]$$

$$= \frac{1}{Z} \prod_\alpha f_\alpha(X_\alpha)$$



X_i	$f(X_i)$
-1	$\exp(-\theta_i)$
1	$\exp(\theta_i)$

X_i	X_j	$f(X_i,X_j)$
-1	-1	$\exp(\theta_{ij})$
-1	1	$\exp(-\theta_{ij})$
1	-1	$\exp(-\theta_{ij})$
1	1	$\exp(\theta_{ij})$

- If θ_{ij} positive: encourages X_i, X_j to have same sign

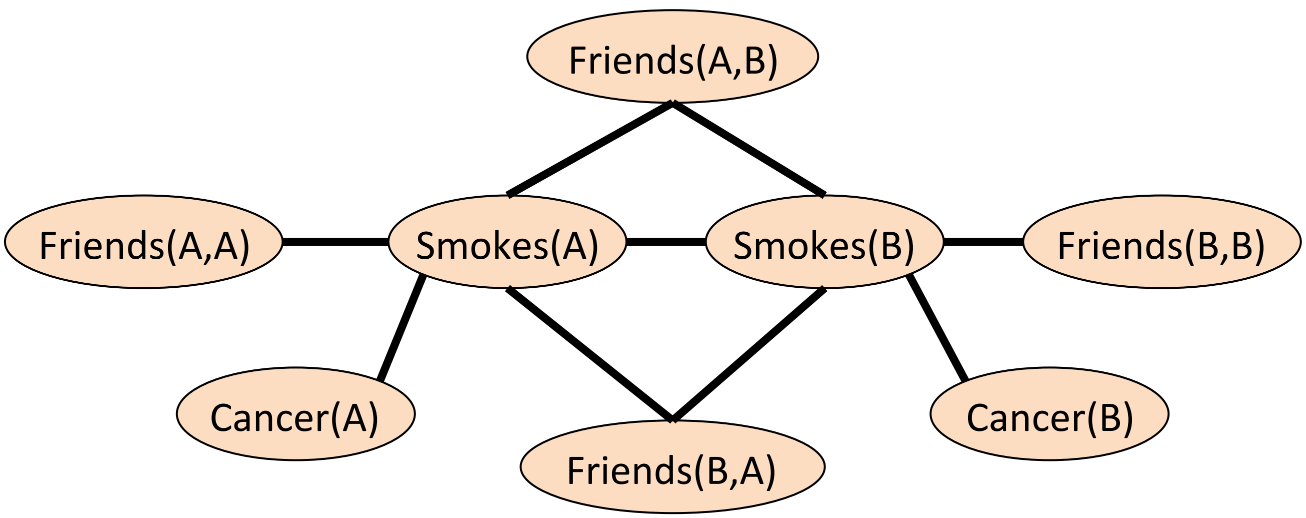
Example: Markov logic

[Richardson & Domingos 2005]

1.5	$\forall x\ Smokes(x) \Rightarrow Cancer(x)$
1.1	$\forall x, y\ Friends(x,y) \Rightarrow \left(Smokes(x) \Leftrightarrow Smokes(y)\right)$

Two constants: **Anna** (A) and **Bob** (B)

S_A	C_A	$f(S_A, C_A)$
0	0	exp(1.5)
0	1	exp(1.5)
1	0	1.0
1	1	exp(1.5)



F_{AB}	S_A	S_B	f(.)
0	0	0	exp(1.1)
0	0	1	exp(1.1)
0	1	0	exp(1.1)
0	1	1	exp(1.1)
1	0	0	exp(1.1)
1	0	1	1.0
1	1	0	1.0
1	1	1	exp(1.1)

Conditioning on observations

- Observing a variable's value
  - Reduces the scope of the factor

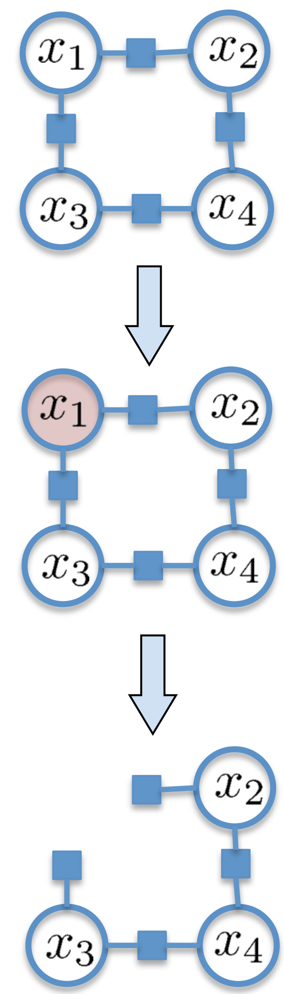
$$p(X) = \frac{1}{Z}\left[f_{12}(X_1, X_2) f_{13}(X_1, X_3) f_{24}(X_2, X_4) f_{34}(X_3, X_4)\right]$$

$$\Downarrow$$

$$p(x_1, X_{2:4}) = \frac{1}{Z}\left[f_{12}(x_1, X_2) f_{13}(x_1, X_3) f_{24}(X_2, X_4) f_{34}(X_3, X_4)\right]$$

$$\Downarrow$$

$$p(X_{2:4} | x_1) = \frac{1}{Z'}\left[\ g_2(X_2) \ \cdot \ g_3(X_3) \ \cdot \ f_{24}(X_2, X_4) f_{34}(X_3, X_4)\right]$$



Conditioning a factor

A	B	f(A,B)
b	b	4
b	g	5
b	r	1
g	b	2
g	g	6
g	r	3
r	b	1
r	g	1
r	r	6

Assign **A=b** $\rightarrow$

B	g(B)
b	4
g	5
r	1

Assign **B=r** $\rightarrow$

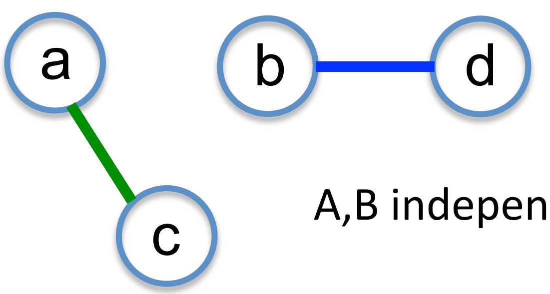
$h_{\varnothing}$
1

Conditional independence

- Undirected graphs have very simple conditional independence
  - X conditionally independent of Y given Z?
  - Check all paths from X to Y
  - A path is “inactive” (blocked) if it passes through a variable node in Z
  - If no path from X to Y, conditionally independent

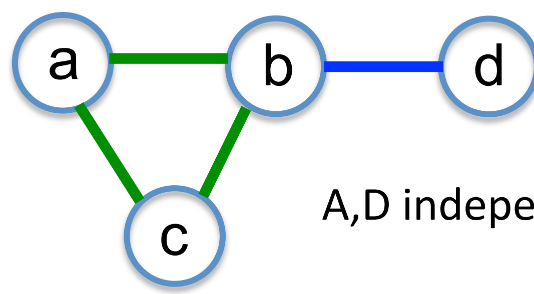
- Examples:

$p(a)\ p(b)\ p(c \mid a)\ p(d \mid b)$



A,B independent

$p(a)\,p(b \mid a)\ p(c \mid a, b)\ p(d \mid b)$



A,D independent given B

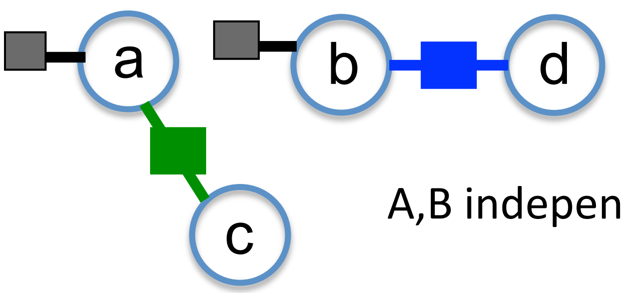
Markov blanket of X: set of variables directly connected to X

Conditional independence

- Undirected graphs have very simple conditional independence
  - X conditionally independent of Y given Z?
  - Check all paths from X to Y
  - A path is "inactive" (blocked) if it passes through a variable node in Z
  - If no path from X to Y, conditionally independent

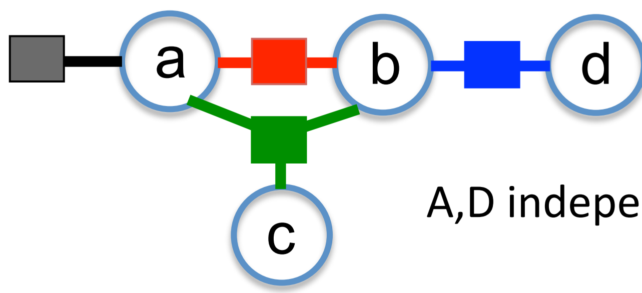
- Examples:

$$p(a)\ p(b)\ p(c \mid a)\ p(d \mid b)$$



A,B independent

$$p(a)\,p(b \mid a)\ p(c \mid a, b)\ p(d \mid b)$$

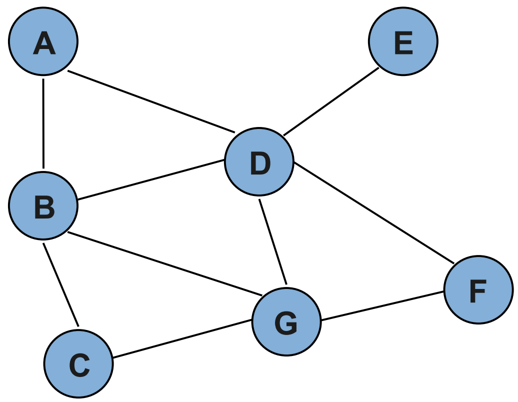


A,D independent given B

Markov blanket of X: set of variables directly connected to X

More examples

- Ex:

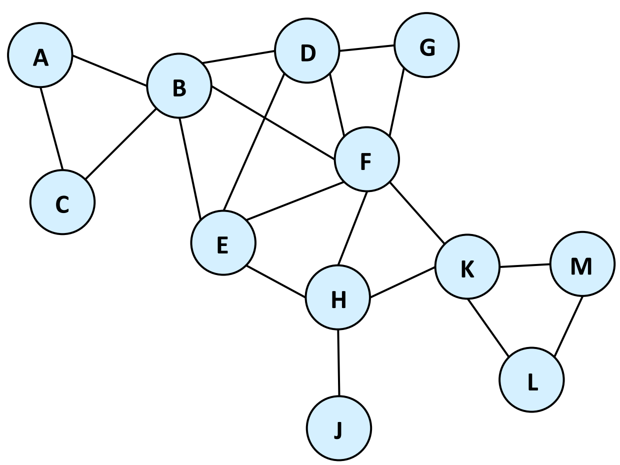


A, E independent given D

A,F independent given D, G

A, G independent given B, D

...



(A,C) , (KL) independent given E,F

(all), G independent given D, F

...

Factorization and Independence

- p(X) is Markov with respect to graph G:
  - p(X) obeys the independence relations in the connectivity of G
- Given a factorization of p(X), can draw Markov graph G
  - p(X) is then Markov with respect to G
- Converse?
- Theorem [Hammersly & Clifford, 1971]
  - If p(X) is Markov with respect to G, and p(x) > 0 for all x,

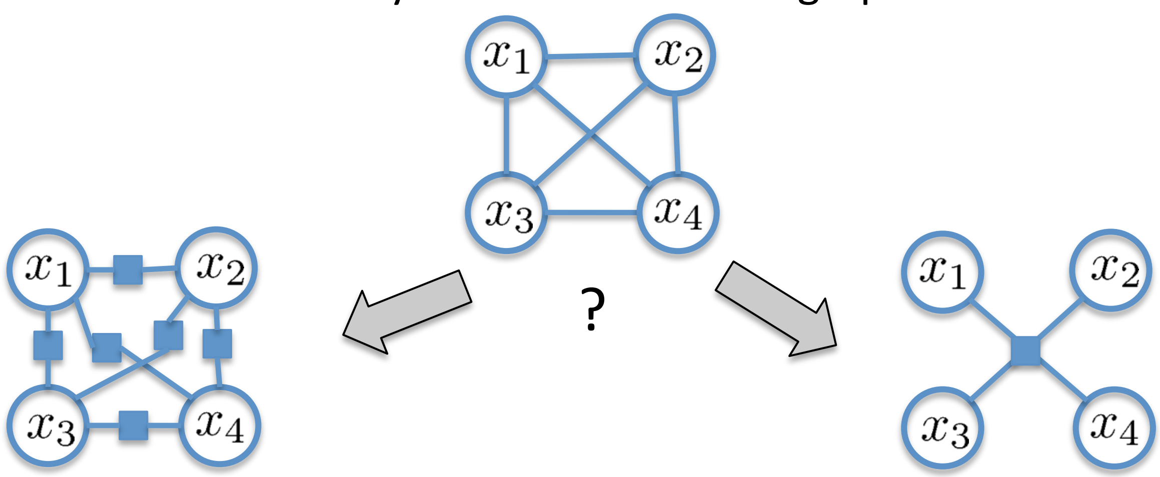
    then p(X) factors as $$p(X) = \frac{1}{Z} \prod_{\alpha} f_\alpha(X_\alpha)$$

    where α are the cliques (fully connected subsets) of G

Pairwise models

- Markov network may mask some structure
- Factor graph shows more detail

Fully connected Markov graph



$$p(X) = \frac{1}{Z} \prod_{i,j} f_{ij}(X_i, X_j)$$

Pairwise model
O(n^2 d^2) parameters

$$p(X) = \frac{1}{Z} f(X_1, X_2, X_3, X_4)$$

Full clique
O(d^n) parameters

Pairwise models: Gaussian

- Exponential family:

$$\mathcal{N}(x\,;\,\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}}\,|\Sigma|^{-\frac{1}{2}}\,\exp\Big[-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T\Big]$$

$$= \frac{1}{Z}\,\exp\Big[-\frac{1}{2}x\Sigma^{-1}x^T + \mu\Sigma^{-1}x^T\Big]$$

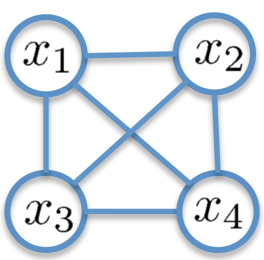
"Information" form

$$= \frac{1}{Z}\,\exp\Big[-\frac{1}{2}x\,J\,x^T + hx^T\Big] \qquad = \mathcal{N}^{-1}(x\,;\,h,J)$$

(Canonical exp. family form)

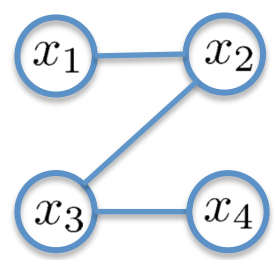
$$= \frac{1}{Z}\prod_i \exp[h_i x_i - .5J_{ii}x_i]\prod_{i<j}\exp[-J_{ij}x_i x_j]$$

Gaussian distribution
= pairwise MRF



$$\Sigma^{-1}_{ij} = 0 \Rightarrow$$

No factor between (i,j)



$$\Sigma^{-1} = ?$$

?	?	0	0
?	?	?	0
0	?	?	?
0	0	?	?

Pairwise models: Boltzmann machines

- Boltzmann machines:

$$p(x) = \frac{1}{Z} \exp \Big[\sum_i a_i x_i + \sum_{ij} w_{ij} x_i x_j \Big]$$

Binary pairwise GM:

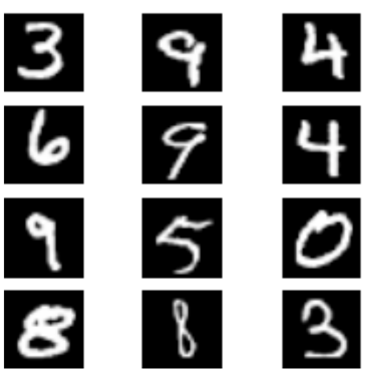
- Restricted Boltzmann machines:

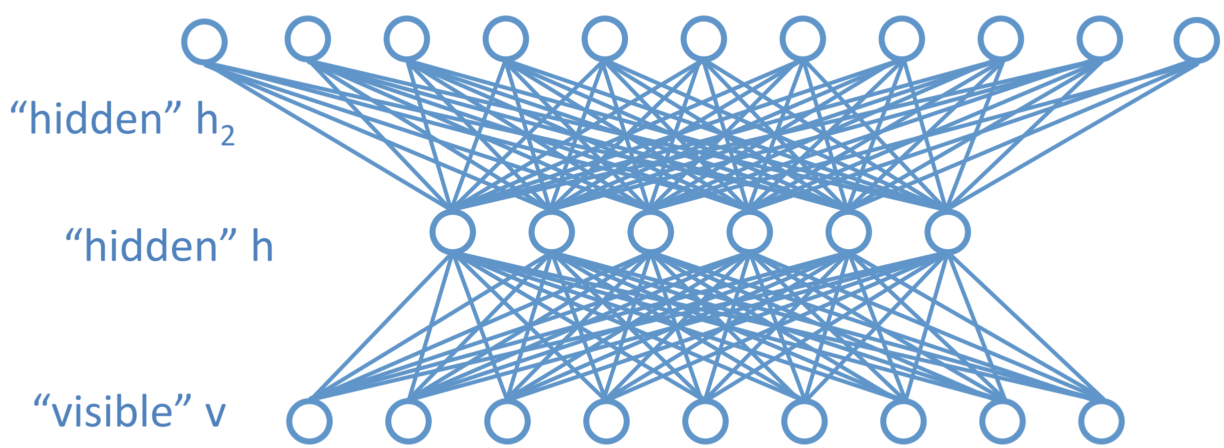
$$p(v,h) = \frac{1}{Z} \exp \Big[\sum_i a_i v_i + \sum_j b_j h_j + \sum_{ij} w_{ij} v_i h_j \Big]$$

- Deep Boltzmann machines:

$$p(v,h_1,h_2) = \frac{1}{Z} \exp \Big[\sum_{ij} w_{1ij} v_i h_{1j} + \sum_{jk} w_{2jk} h_{1j} h_{2k} + \ldots \Big]$$

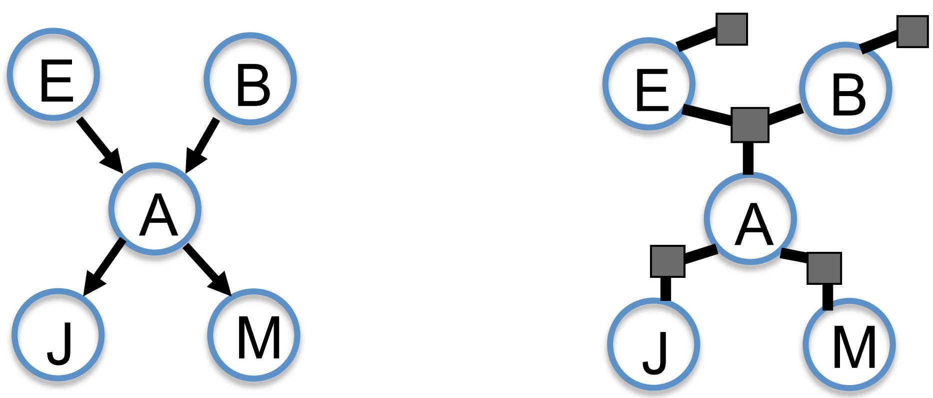
MNIST:





Directed to undirected models

- We can convert directed models to undirected ones:
  - Write factorization associated with Bayes net structure
  - Create factors for each term in the product



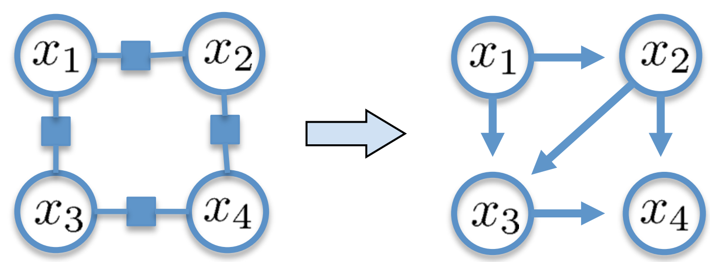
$$p(J, M, A, E, B) = p(E)\ p(B)\ p(A \mid E, B)\ p(J \mid A)\ p(M \mid A)$$

- But, some independence implications are lost
  - Not clear that E, B are independent without observing A, J, or M

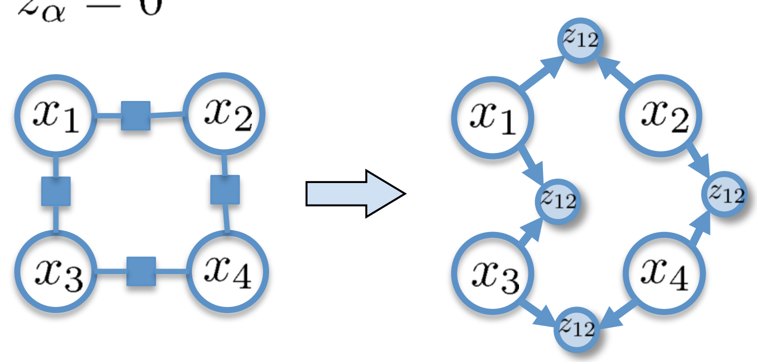
Undirected to directed models

- We can similarly convert undirected to directed
  - Up to a constant, since factors can have values > 1.0

- Method 1: choose an ordering, check for independence
  - Need to choose an ordering
  - Can lose a lot of independence info

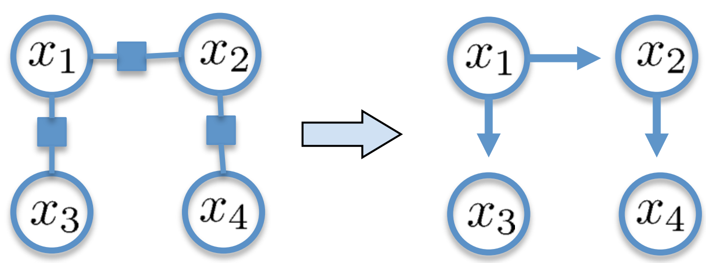


- Method 2: introduce auxiliary variables
  - For each factor, create an observation $z_\alpha = 0$
  - Let $p(Z_\alpha = 0 | X_\alpha) = f_\alpha(X_\alpha) / f_{\max}$
    $p(Z_\alpha = 1 | X_\alpha) = \text{anything}$
  - Very "artificial" construction



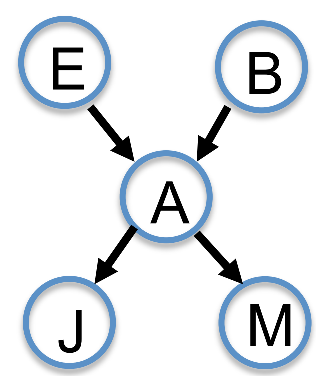
Tree-structured models

- Equivalent
  - Bayesian network is a tree (single root)
  - Markov graph is a tree
  - In each case, factors only involve pairs of variables



  - BN and MG representations interchangeable
  - Unique path between any pair of variables

Note: “Poly-tree”
= more than one root



Summary

- Undirected models
  - Factor graphs
  - Markov graphs
- Conditioning as a graph reduction
- Graph separation $\Rightarrow$ conditional independence
- Directed vs. undirected representations
  - Not equivalent
  - May be more or less appropriate for model assumptions