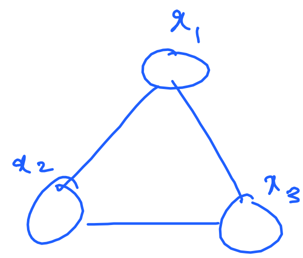
Chapter 13

Factor Graphs and Sum/Max-product Algorithms **

Factor graphs

Sunday, August 27, 2017 4:31 PM

This MRF implies the factorization



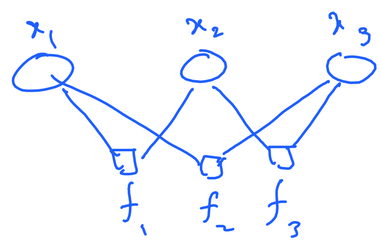
$$p(x_1, x_2, x_3) \propto \psi(x_1, x_2, x_3)$$

But suppose we actually want to represent

$$p(x_1, x_2, x_3) \propto f(x_1, x_2)\, f(x_2, x_3)\, f(x_3, x_1)$$

Is there a way to do this with a graph?

Factor graph:



Two types of nodes:

V : variables : x_1, x_2, x_3

F : factors : f_1, f_2, f_3

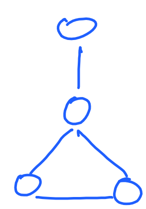
$$p(x_1^m) = \prod_{f_i \in F} f_i(x_{f_i})$$

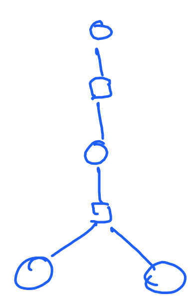
variable nodes adjacent to f_i

MRF $\longrightarrow$ FG

nodes $\longrightarrow$ variable nodes

maximal cliques $\longrightarrow$ factor nodes

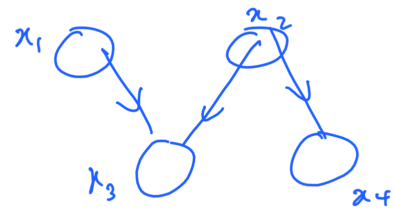


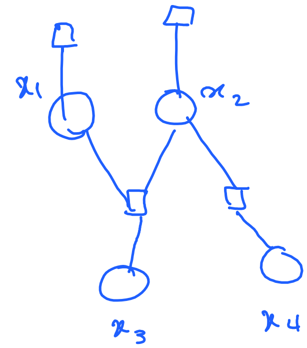


BN $\longrightarrow$ FG

nodes ⟶ variable nodes

(CPDs at) nodes ⟶ factor nodes





Note: even if the original MRF is not a tree, or the original BN has an MRF which is not a tree, the factor graph may still be a tree: that is discarding the types of the nodes in FG leads to a tree.

This is good because ↴

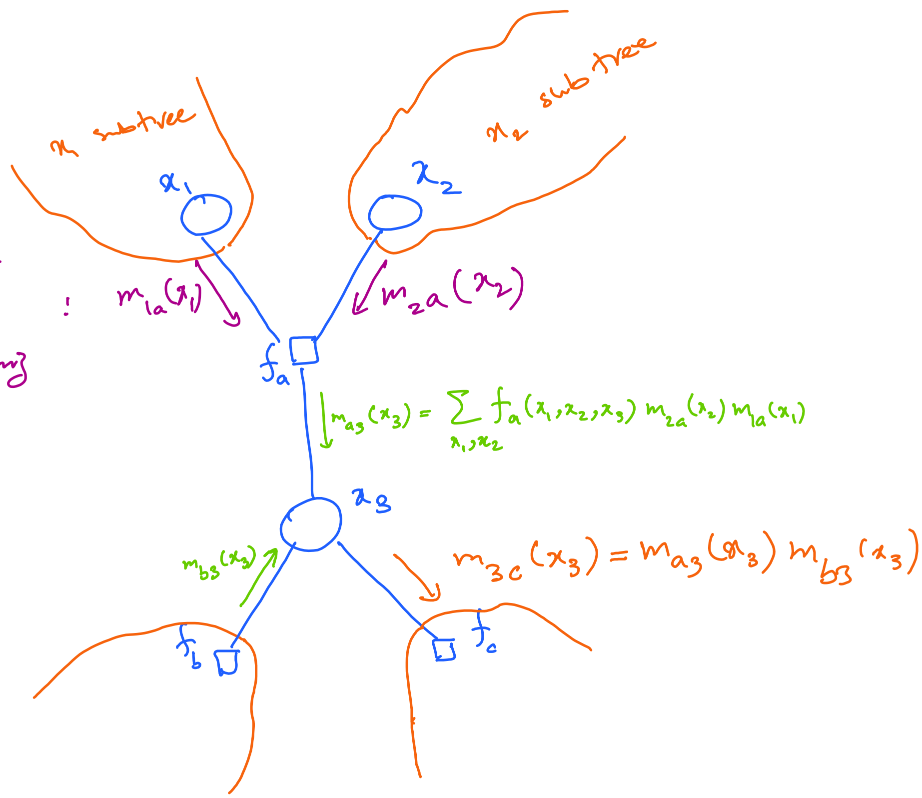
Sum-product for factor tree graphs:

sum of every factor on x_1 subtree assuming $x_1 = x_1$

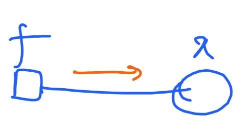
1

2

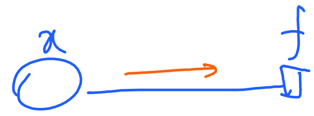
3



Starting steps at leaves:



$m_{fx}(x) = f(x)$

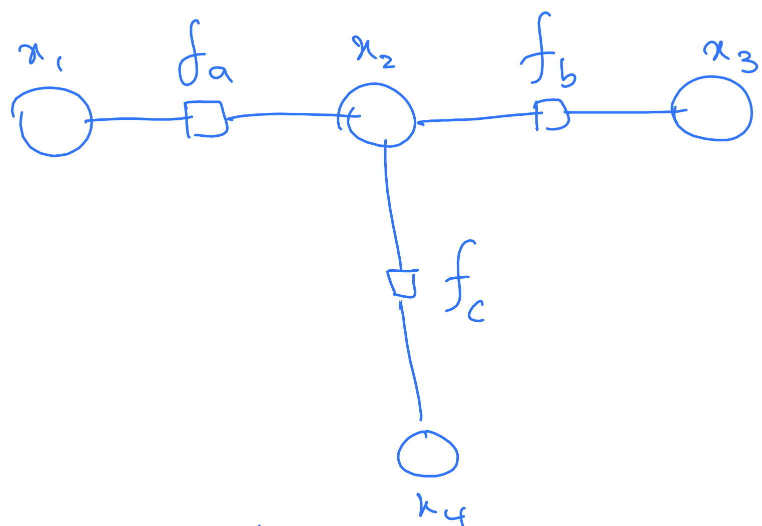


$m_{xf} = 1$

Marginal at each node: product of msgs it receives

marginals at all nodes takes twice as much of that of a single node: two msgs per link as opposed to one.

Example:



round 1

$\mu_{1a}(x_1) = 1$

$\mu_{4c}(x_4) = 1$

$\mu_{3b}(x_3) = 1$

round 2

$$\mu_{a2}(x_2) = \sum_{x_1} f_a(x_1, x_2)\, \mu_{1a}(x_1)$$

$$\mu_{b2}(x_2) = \sum_{x_3} f_b(x_3, x_2)\, \mu_{3b}(x_3)$$

$$\mu_{c2}(x_2) = \sum_{x_4} f_c(x_2, x_4)\, \mu_{4c}(x_4)$$

Max-product

Sunday, August 27, 2017 6:53 PM

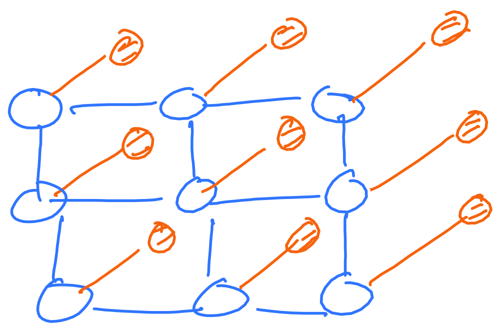
Problem: Identify the most likely configuration:

Find a set of values $x_1^*, \dots, x_m^*$ s.t.

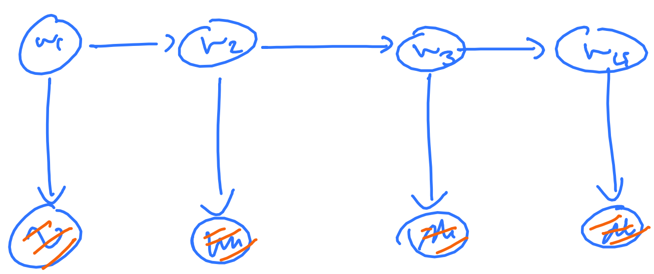
$$P(x_1^*, \dots, x_m^*) \geq P(x_1^m) \quad \text{for all } x_1^m$$

Applications:

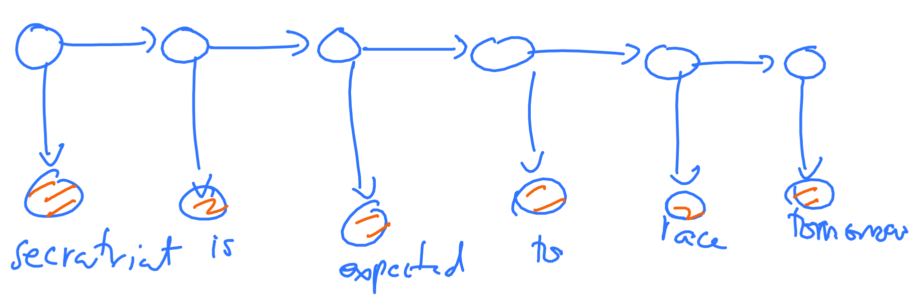
- Image denoising (Lab 1)



- Voice recognition



- Part of speech tagging



Do we already know how to solve this?

Finding most probable state for a node:

run sum-product and find state with max prob.

Finding most probable configuration for graph:

max-product

$$x_{max} = \arg\max \; p(x)$$

May not be the same.

	$x=0$	$x=1$
$y=0$	0.3	0.4
$y=1$	0.3	0

$x=0$ max for x

$y=0$ max for y

$(x,y)=(0,1)$ max for (x,y)

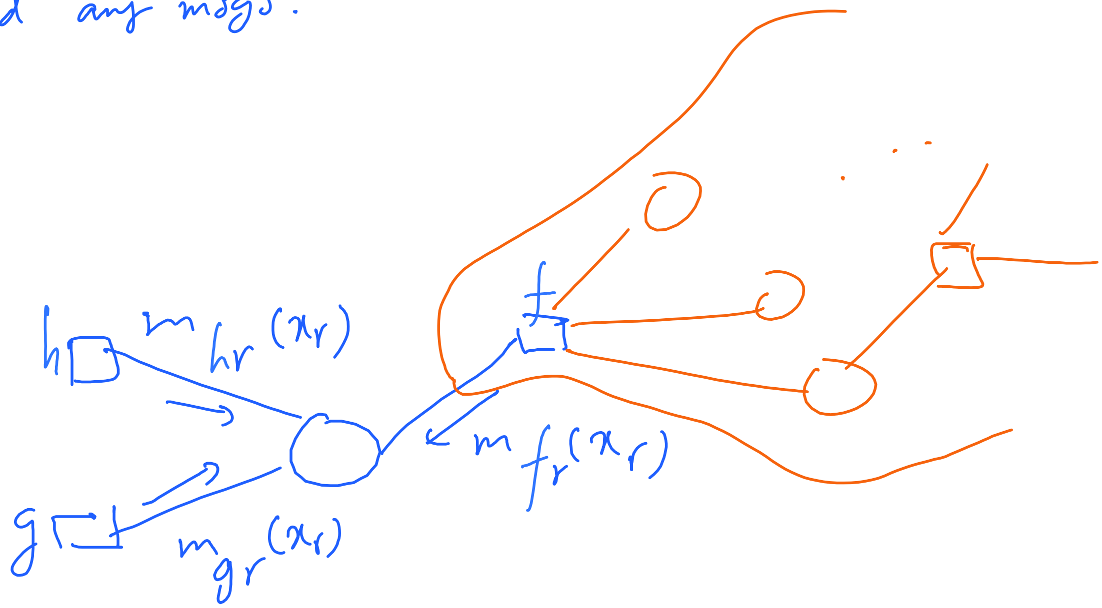
Let's instead try to find $\max_{x_1^m} p(x_1^m)$

$$\max_{x_1^m} p(x_1^m) = \max_{x_1} \max_{x_2} \cdots \max_{x_m} p(x_1^m)$$

* We can use a similar approach to elimination except that Σ is replaced with max.

* Similarly sum-product can be turned to max-product to find $\max p(x)$

* Here, we pick an arbitrary root, which does not send any msgs.



The message from f $\mu_{fr}(x_r)$

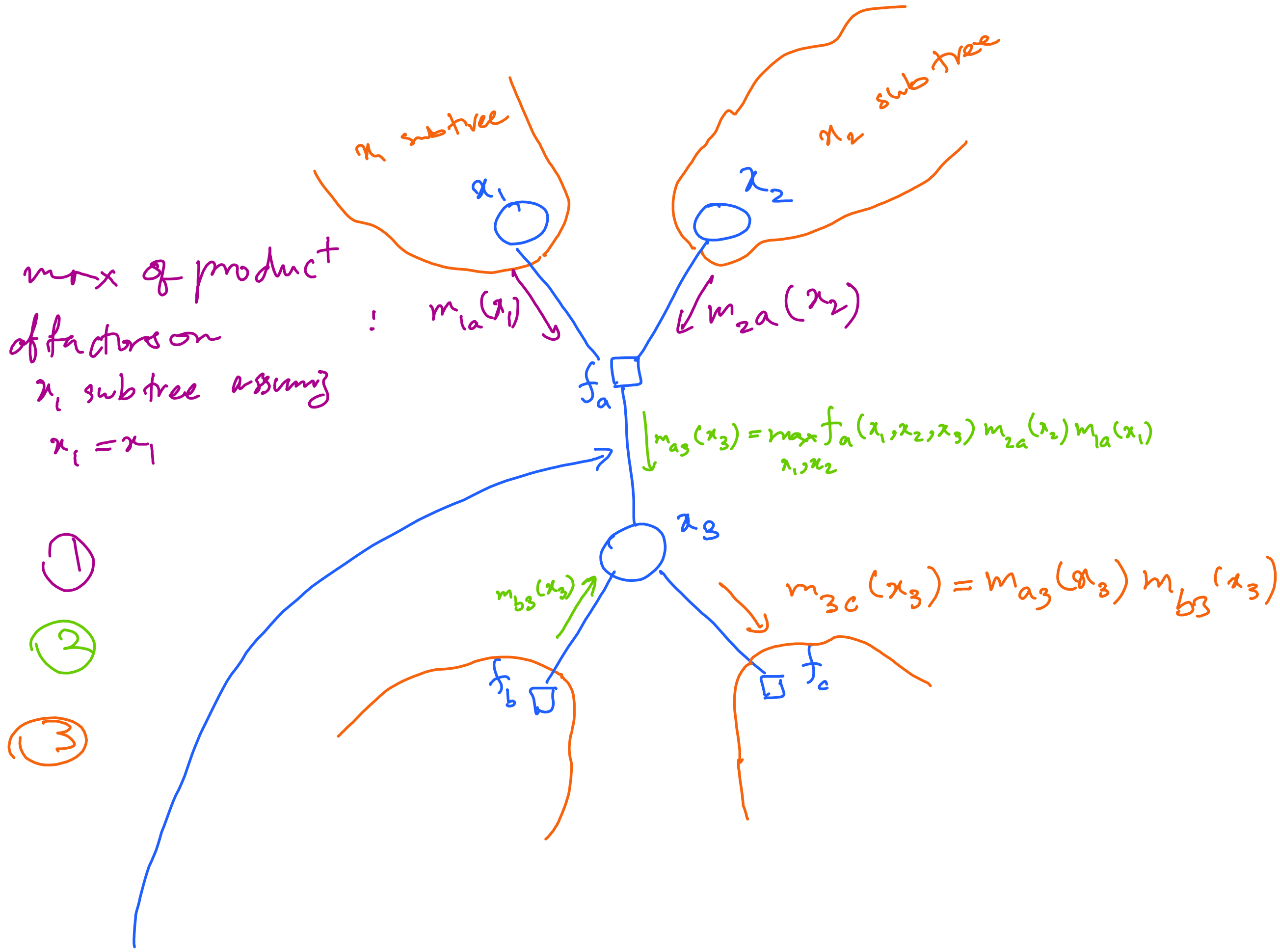
Given the particular value for x_r

* What is the most likely configuration for the subtree of f
* What is the "probability" of this configuration

$$\max p(x_1^m) = \max_{x_r} m_{fr}(x_r)\, m_{gr}(x_r)\, m_{hr}(x_r)$$

from the value of x_r that maximizes this sum and messages we find the most likely configuration

* Finding the maximizing configuration:



For each value of x_3, we also record which values of x_1, x_2 achieved the max.

At the root, we find x_r^* that achieves the max.
Then we back-track and find maximizing values for all nodes.

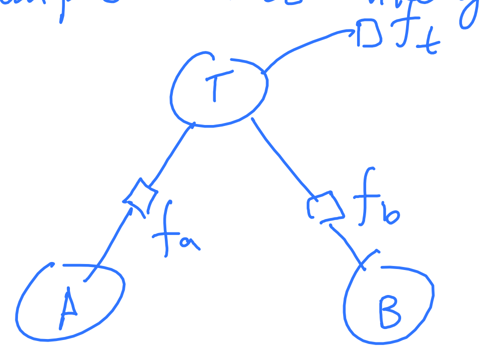
* It may be more convenient to maximize $\ln p(x) \Rightarrow$ max-sum algorithm

$\sum f_i(x_{f_i})$

$\max_{x_r} p(x) = \max_{x_1} \quad x_n \quad \vdots \quad .$

Note: why not continue the alg so that we can find x_i^* for all nodes just like how we found it for x_r? We will find maximizing values, but they may belong to different maximizing configurations.

Example: most-likely configuration



$f_t(T=0) = 0.65$

$f_t(T=1) = 0.35$

$f_a(A=0, T=0) = 0.9$

$f_a(A=0, T=1) = 0.5$

$f_a(A=1, T=0) = 0.1$

$f_a(A=1, T=1) = 0.5$

$f_b(B=0, T=0) = 0.82$

$f_b(B=0, T=1) = 0.15$

$f_b(B=1, T=0) = 0.18$

$f_a(B=1, T=0) = 0.12$

μ_{Aa}: $\max \quad 1$

$A=0 \longrightarrow 1$

$A=1 \longrightarrow 1$

$$\mu_{aT} : \max_{A} f_a(A,T)$$

$T=0 \longrightarrow 0.9$ for $A=0$

$T=1 \longrightarrow 0.5$ for $A=0$ & $A=1$

$$\mu_{tT} : \max f_t(T)$$

$T=0 \longrightarrow 0.65$

$T=1 \longrightarrow 0.35$

$$\mu_{Tb} : \max_{T} \mu_{tT}(T)\, \mu_{aT}(T)\, f_b(B,T)$$

$\ldots$