Machine Learning 2

Graphical Models
- Exact Inference
- Mode Finding
-Ideas and Problems

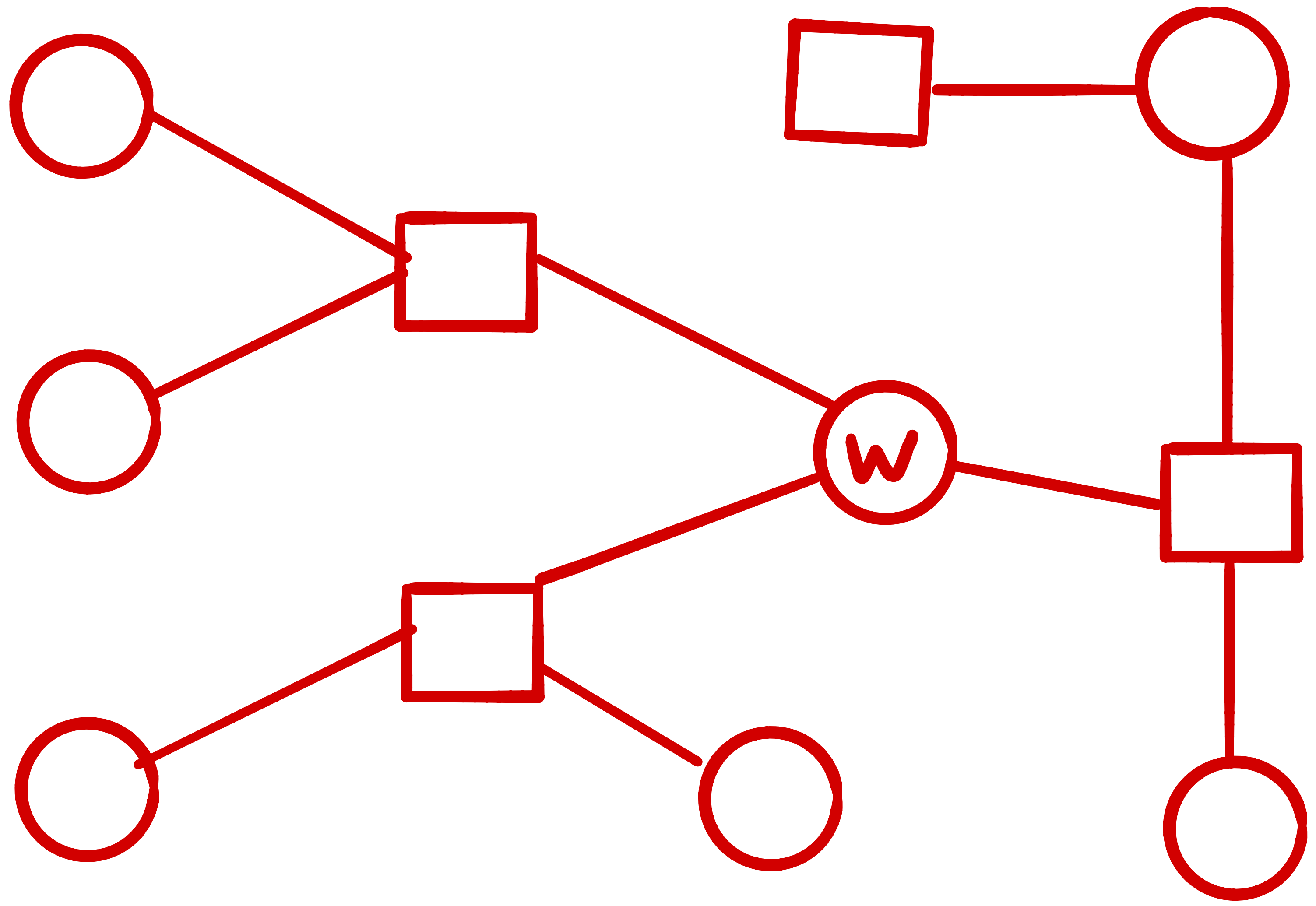
Patrick Forré

The objective - find a mode

- $G = (V, F, E)$ Factor Tree with $p(x_V) = \frac{1}{Z}\prod_{\alpha \in F} \psi_\alpha(x_\alpha)$.
- Goal: Find mode: $x^* \in \arg\max_x p(x) = \arg\max_x \log p(x)$

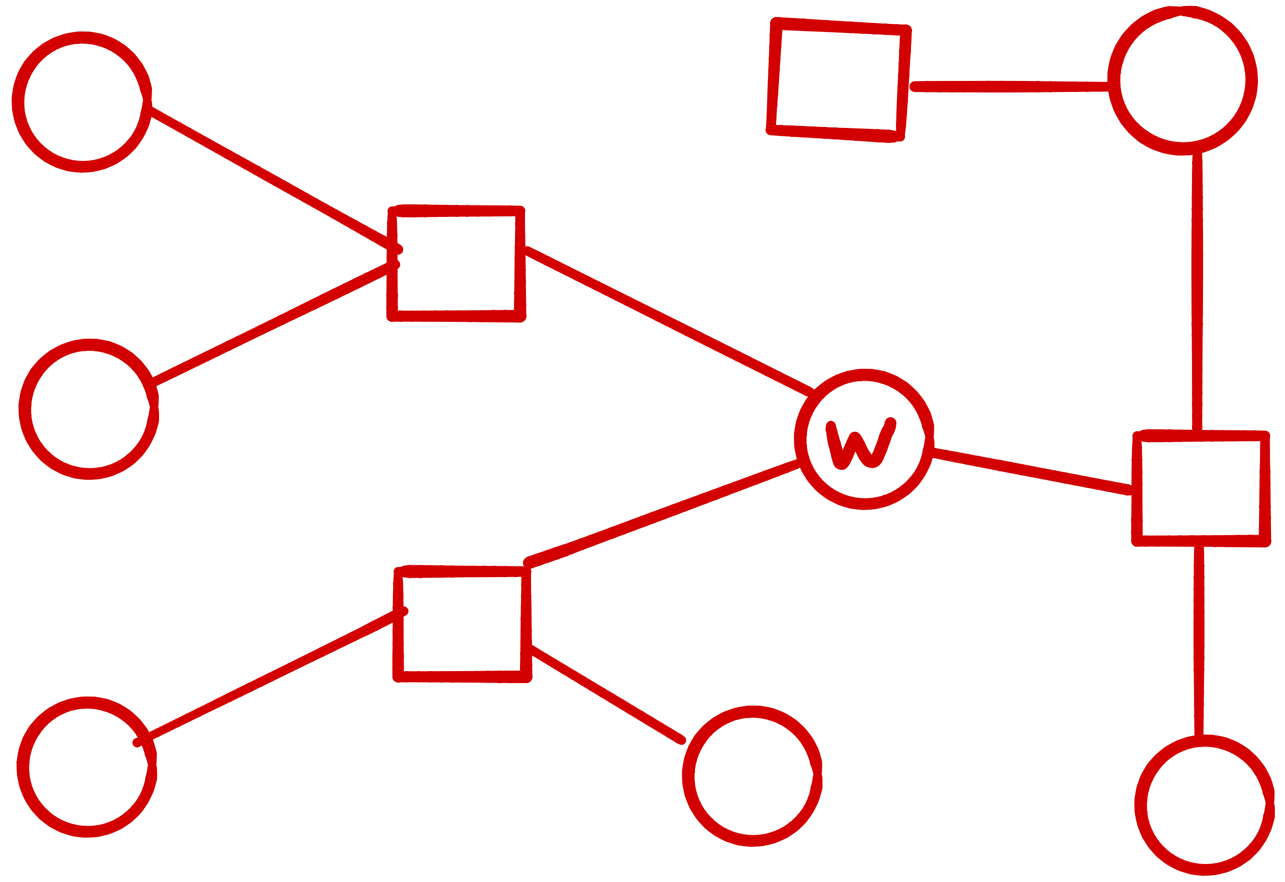
$$= \arg\max_x \frac{1}{Z} \prod_{\alpha \in F} \psi_\alpha(x_\alpha)$$

$$= \arg\max_x \sum_{\alpha \in F} \log \psi_\alpha(x_\alpha) \; (-\log Z)$$



- How to find a global mode component by component?

Marginal vs Mode



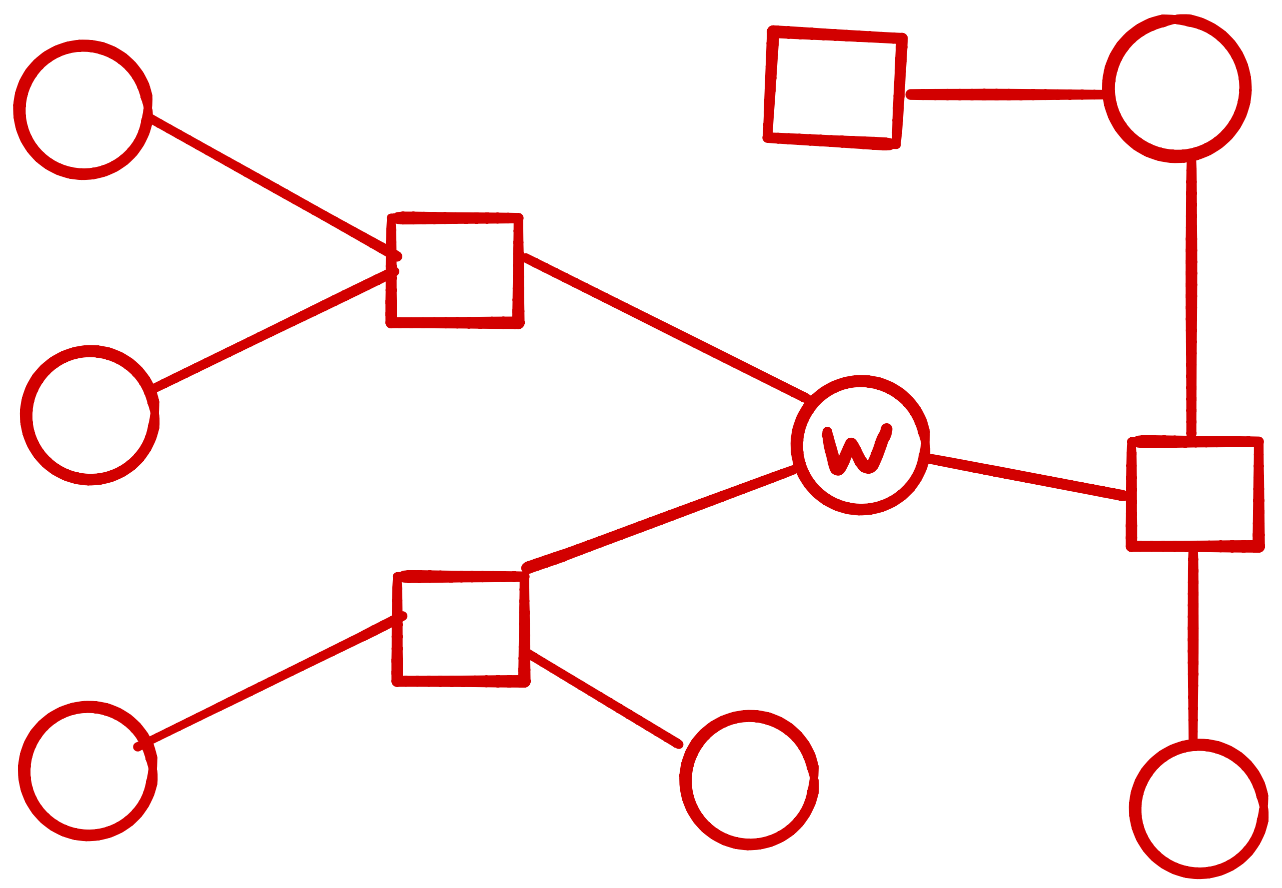
- $G = (V, F, E)$ Factor Tree with $p(x_V) = \frac{1}{Z}\prod_{\alpha\in F}\psi_\alpha(x_\alpha)$.

- Marginal:
- $p(x_v) = \sum_{x_{\neg v}} \frac{1}{Z}\prod_{\alpha\in F}\psi_\alpha(x_\alpha)$
- $\sum$-$\prod$

- Mode:
- $p(x^*) = \max_x \frac{1}{Z}\prod_{\alpha\in F}\psi_\alpha(x_\alpha)$
- $\max$-$\prod$, but better in $\log$-space:
- $\log p(x^*) = \max_x \sum_{\alpha\in F} \log \psi_\alpha(x_\alpha) - \log Z$
- $\max$-$\sum$-algorithm

Idea: Max-Sum Algorithm

- $G = (V, F, E)$ Factor Tree with $p(x_V) = \frac{1}{Z} \prod_{\alpha \in F} \psi_\alpha(x_\alpha)$.
- Goal: Find a Mode x^*: $\log p(x^*) = \max_x \sum_{\alpha \in F} \log \psi_\alpha(x_\alpha) - \log Z$
- Idea: Just run the Sum-Product Algorithm with the following changes:
  - take the $\log$ for all quantities
  - replace all $\sum$'s by $\max$'s.
  - replace all $\prod$'s by $\sum$'s

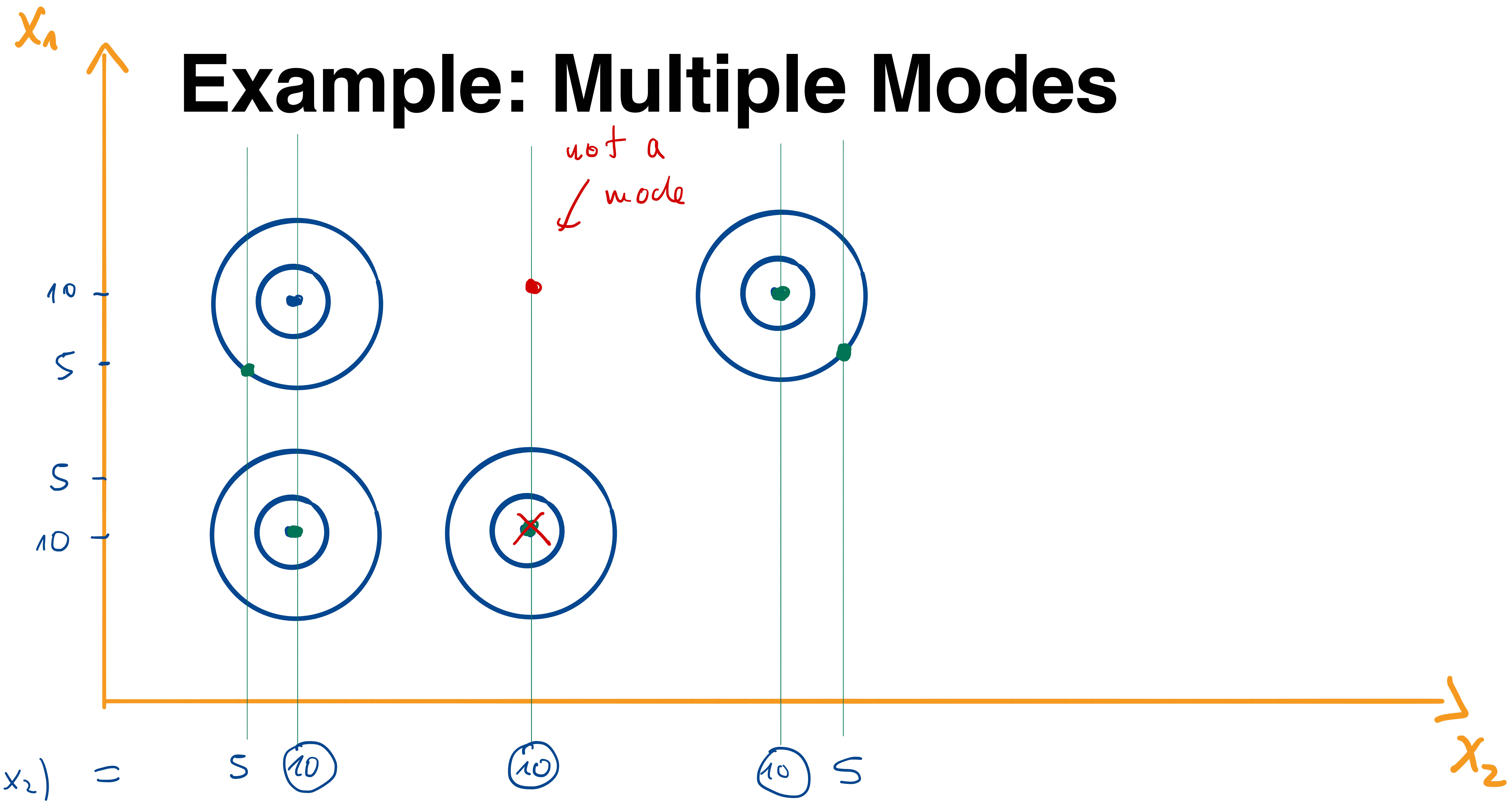


The problem with mode picking

- Let $p(x_1, x_2)$ be probability distribution in two dimensions. Goal: Find $x^* = (x_1^*, x_2^*) \in \arg\max_{x_1, x_2} p(x_1, x_2)$
- It holds: x^* is a mode of $p(x)$ iff $p(x^*) = \max_{x_1} \max_{x_2} p(x_1, x_2)$ iff $p(x^*) = \max_{x_2} \max_{x_1} p(x_1, x_2)$
- BUT, picking $\tilde{x}_2 \in \arg\max_{x_2} \left(\max_{x_1} p(x_1, x_2) \right)$ and $\tilde{x}_1 \in \arg\max_{x_1} \left(\max_{x_2} p(x_1, x_2) \right)$ might lead to

  $(\tilde{x}_1, \tilde{x}_2)$ NOT in $\arg\max_{(x_1, x_2)} p(x_1, x_2)$
- Instead we need book keeping / coordination between coordinates / backtracking:
  - Record: $M_1(x_2) \in \arg\max_{x_1} p(x_1, x_2)$ for every value x_2.
  - Pick: $x_2^* \in \arg\max_{x_2} \left(\max_{x_1} p(x_1, x_2) \right)$
  - Pick: $x_1^* = M_1(x_2^*) \in \arg\max_{x_1} p(x_1, x_2^*)$
  - Then: $x^* := (x_1^*, x_2^*) \in \arg\max_{x_1, x_2} p(x_1, x_2)$

Example: Multiple Modes



$\max_{x_1} p(x_1, x_2) =$ 5 10 10 10 5

x_2^a x_2^b x_2^c x_2^d x_2^e

Machine Learning 2

Graphical Models
- Exact Inference
- Max-Sum Algorithm
(in Factor Trees)

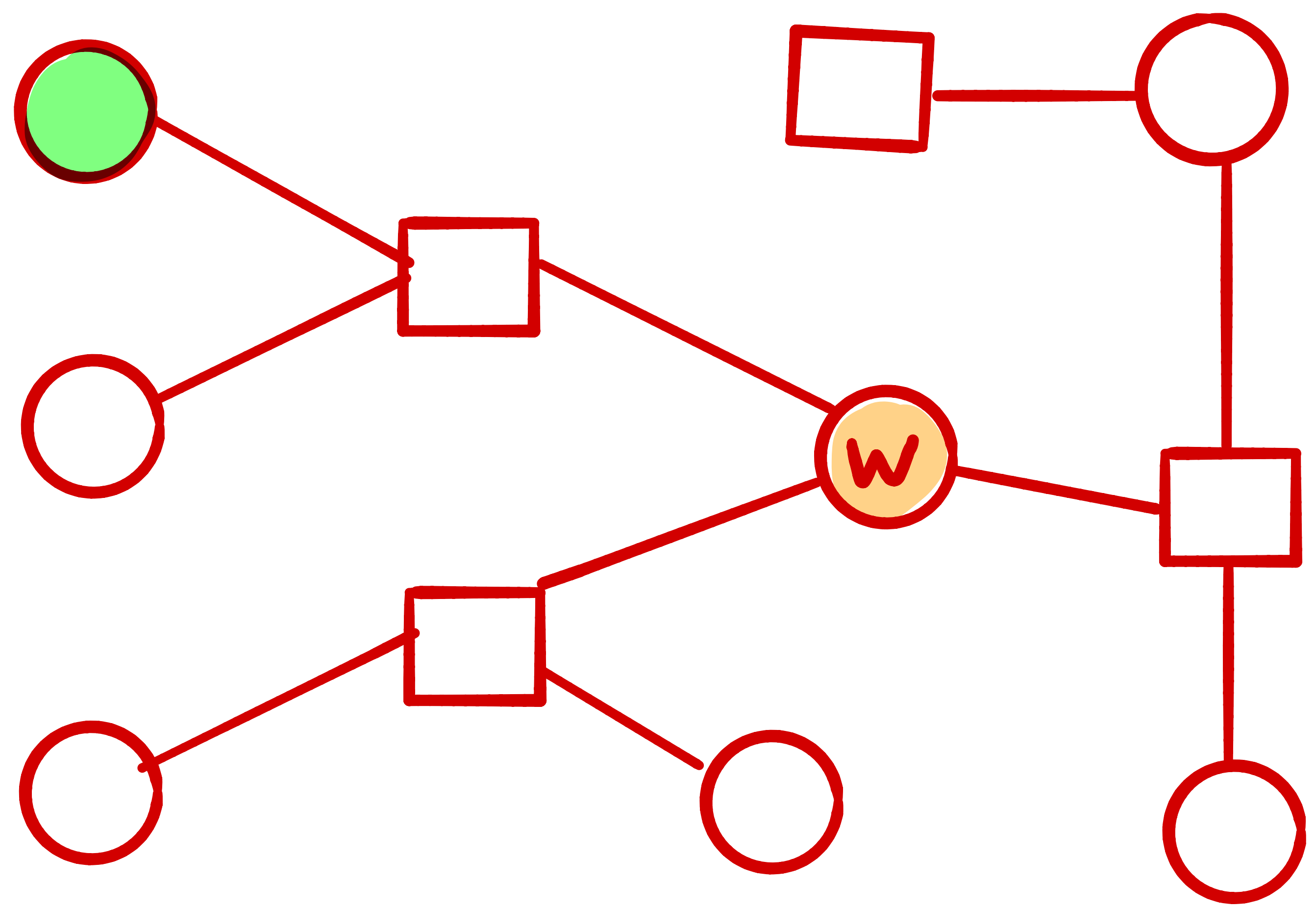
Patrick Forré

Mode Finding in Factor Trees

- $G = (V, F, E)$ Factor Tree with $p(x_V) = \dfrac{1}{Z}\prod_{\alpha \in F} \psi_\alpha(x_\alpha)$.

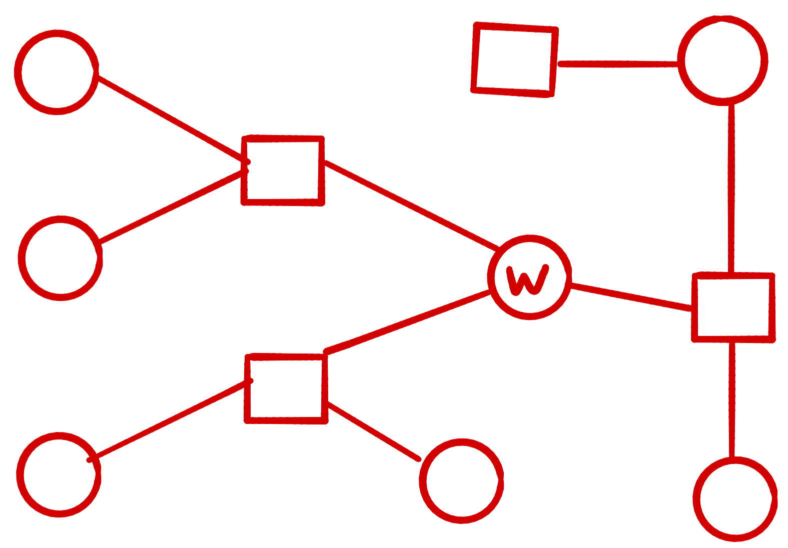
- Goal: Efficiently find mode: $x^* \in \arg\max_x \log p(x)$

  i.e. $\log p(x^*) = \max_x \sum_{\alpha \in F} \log \psi_\alpha(x_\alpha) - \log Z$



- First, pick a Root node $w \in V$.
- Pass Messages from Leaves towards root~~s~~ (while book keeping)
- Back-tracking from Root to Leaves

The Forward Message Passing



- Leaf Variable-to-Factor: $\nu_{u \to \beta}(x_u) = 0$

- Non-leaf Variable-to-Factor: $\nu_{v \to \beta}(x_v) = \sum_{\gamma \in \partial(v) \setminus \{\beta\}} \nu_{\gamma \to v}(x_v)$

- Leaf Factor-to-Variable: $\nu_{\alpha \to v}(x_v) = \log \psi_\alpha(x_v)$

- Non-leaf Factor-to-Variable: $\nu_{\beta \to w}(x_w) = \max_{x_{\beta \setminus w}} \left(\log \psi_\beta(x_\beta) + \sum_{u \in \beta \setminus w} \nu_{u \to \beta}(x_u) \right)$

  - For backtracking keep: $M_{\beta,w}(x_w) \in \arg\max_{x_{\beta \setminus w}} \left(\log \psi_\beta(x_\beta) + \sum_{u \in \beta \setminus w} \nu_{u \to \beta}(x_u) \right)$

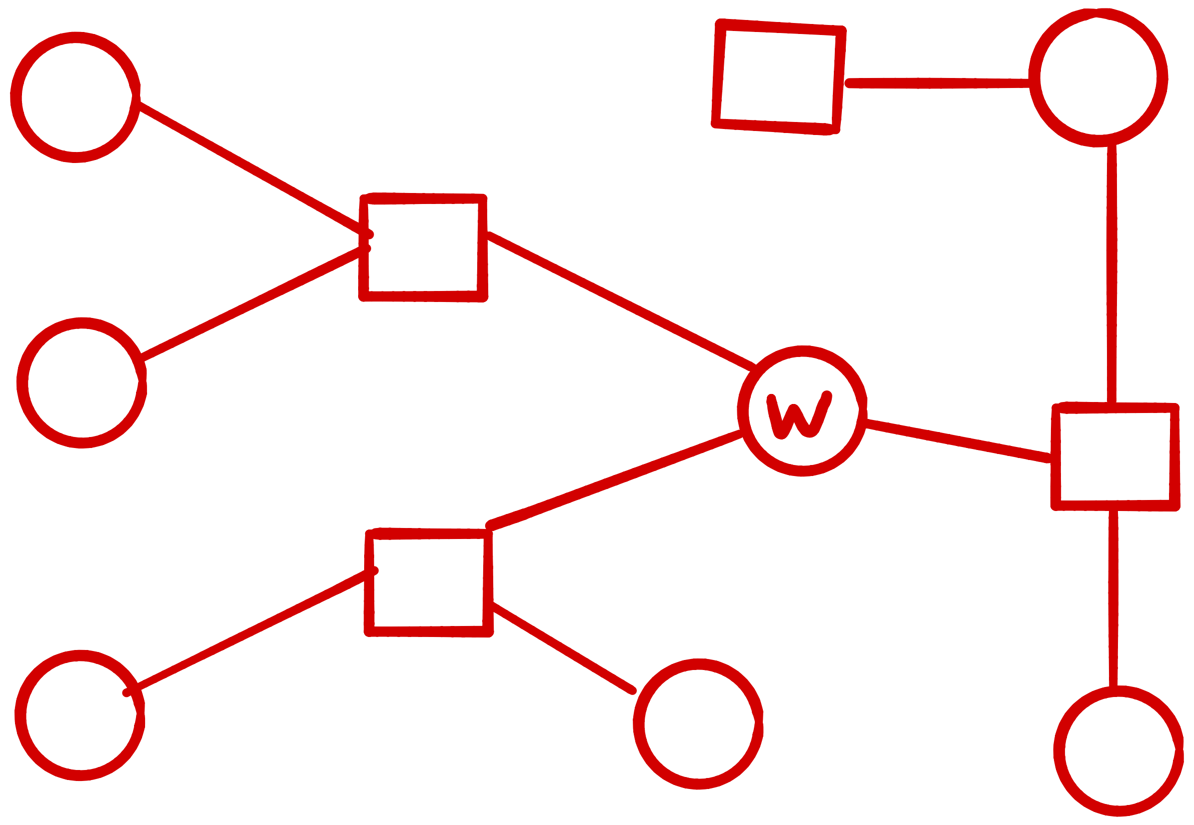
At the Root

- $G = (V, F, E)$ Factor Tree, $p(x_V) = \frac{1}{Z}\prod_{\alpha\in F}\psi_\alpha(x_\alpha)$. Root: variable node $w \in V$

- Put: $q(x_w) := \sum_{\gamma \in \partial(w)} \nu_{\gamma\to w}(x_w)$

- One can show: $q(x_w) = \max_{x_{\neg w}} p(x_w, x_{\neg w}) + \log Z$

- Pick maximizing component: $x_w^* \in \operatorname{argmax}_{x_w} q(x_w)$



Backtracking

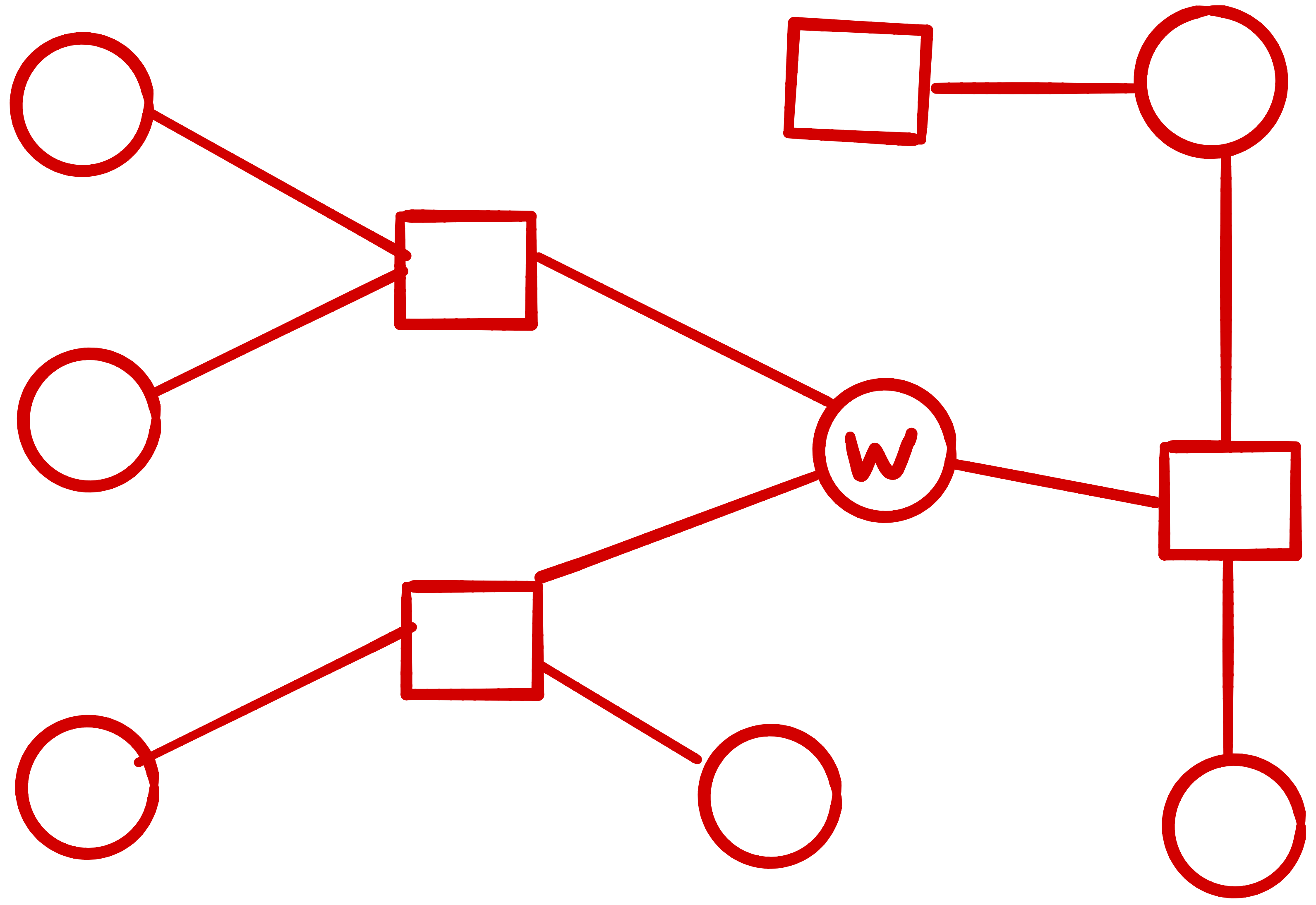
x_w^*

- Starting from Root towards Leaves:

  - pick $x^*_{\alpha\setminus v} = \mu_{\alpha, v}(x_v^*)$

- Then $x^* := (x_v^*)_{v\in V}$ is a mode of p:

$$x^* \in \arg\max_x p(x)$$

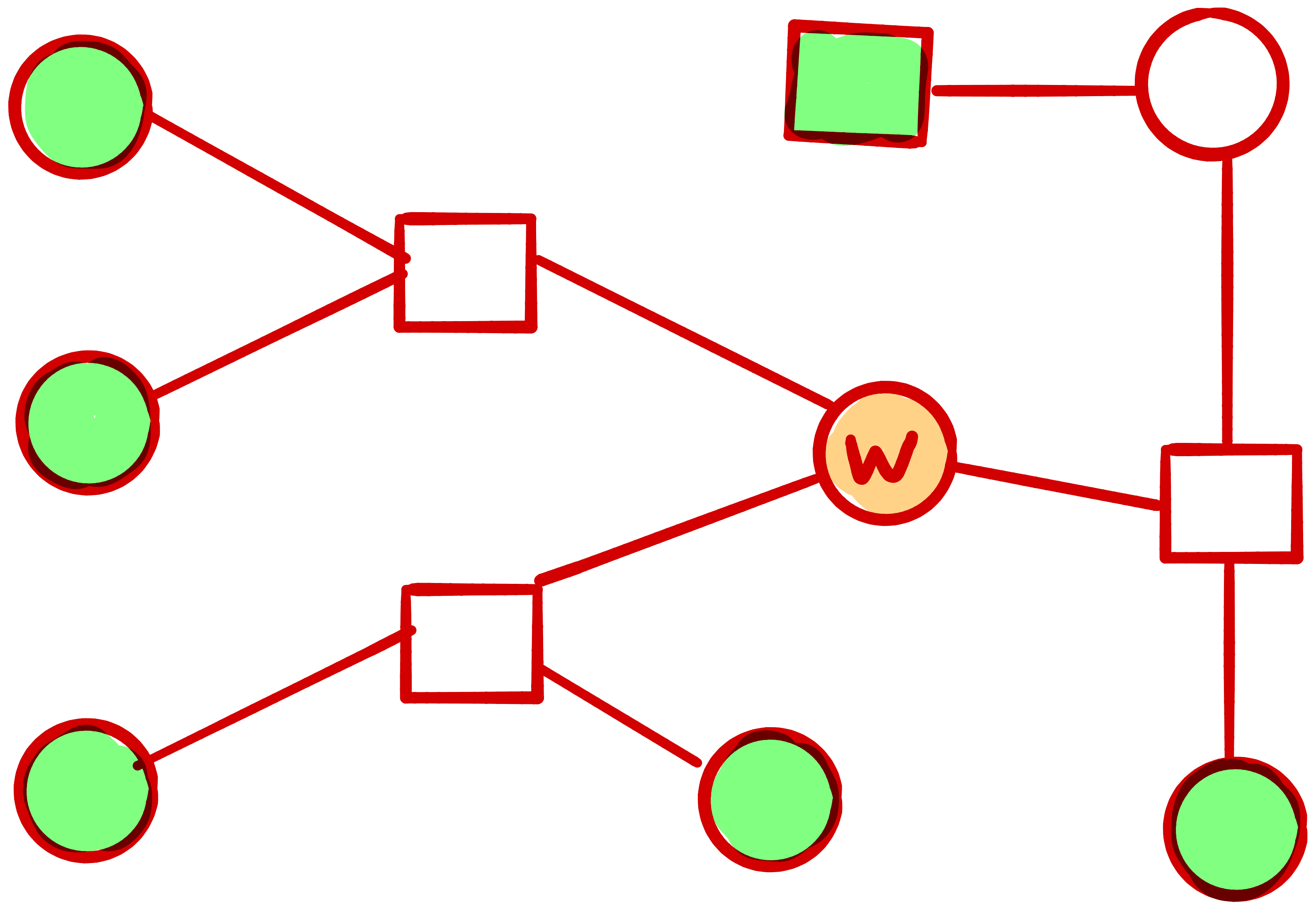


Exercise: Show that the point is a mode

- $G = (V, F, E)$ Factor Tree with $p(x_V) = \frac{1}{Z} \prod_{\alpha \in F} \psi_\alpha(x_\alpha)$.
- Show that the Max-Sum Algorithm outputs a mode of p.

Max-Sum Algorithm

- $G = (V, F, E)$ Factor Tree with $p(x_V) = \frac{1}{Z} \prod_{\alpha \in F} \psi_\alpha(x_\alpha)$. Pick a 'Root' $w \in V$ inside the graph.
- Forward Message Passing towards Root:
  - Variable-to-factor: $\nu_{v \to \beta}(x_v) = \sum_{\gamma \in \partial(v) \setminus \{\beta\}} \nu_{\gamma \to v}(x_v)$, Leaf: $\nu_{u \to \beta}(x_u) = 0$
  - Factor-to-variable: $\nu_{\beta \to w}(x_w) = \max_{x_{\beta \setminus w}} \left(\log \psi_\beta(x_\beta) + \sum_{u \in \beta \setminus w} \nu_{u \to \beta}(x_u) \right)$, Leaf: $\nu_{\alpha \to v}(x_v) = \log \psi_\alpha(x_v)$
  - For backtracking: $M_{\beta, w}(x_w) \in \arg\max_{x_{\beta \setminus w}} \left(\log \psi_\beta(x_\beta) + \sum_{u \in \beta \setminus w} \nu_{u \to \beta}(x_u) \right)$
- At Root: Put $q(x_w) := \sum_{\gamma \in \partial(w)} \nu_{\gamma \to w}(x_w)$ and pick $x_w^* \in \arg\max_{x_w} q(x_w)$.
- Backtracking from Root towards Leaves: Pick $x_{\alpha \setminus v}^* = M_{\alpha, v}(x_v^*)$.
- Then $x^* := (x_v^*)_{v \in V} \in \arg\max_x p(x)$ is a mode.

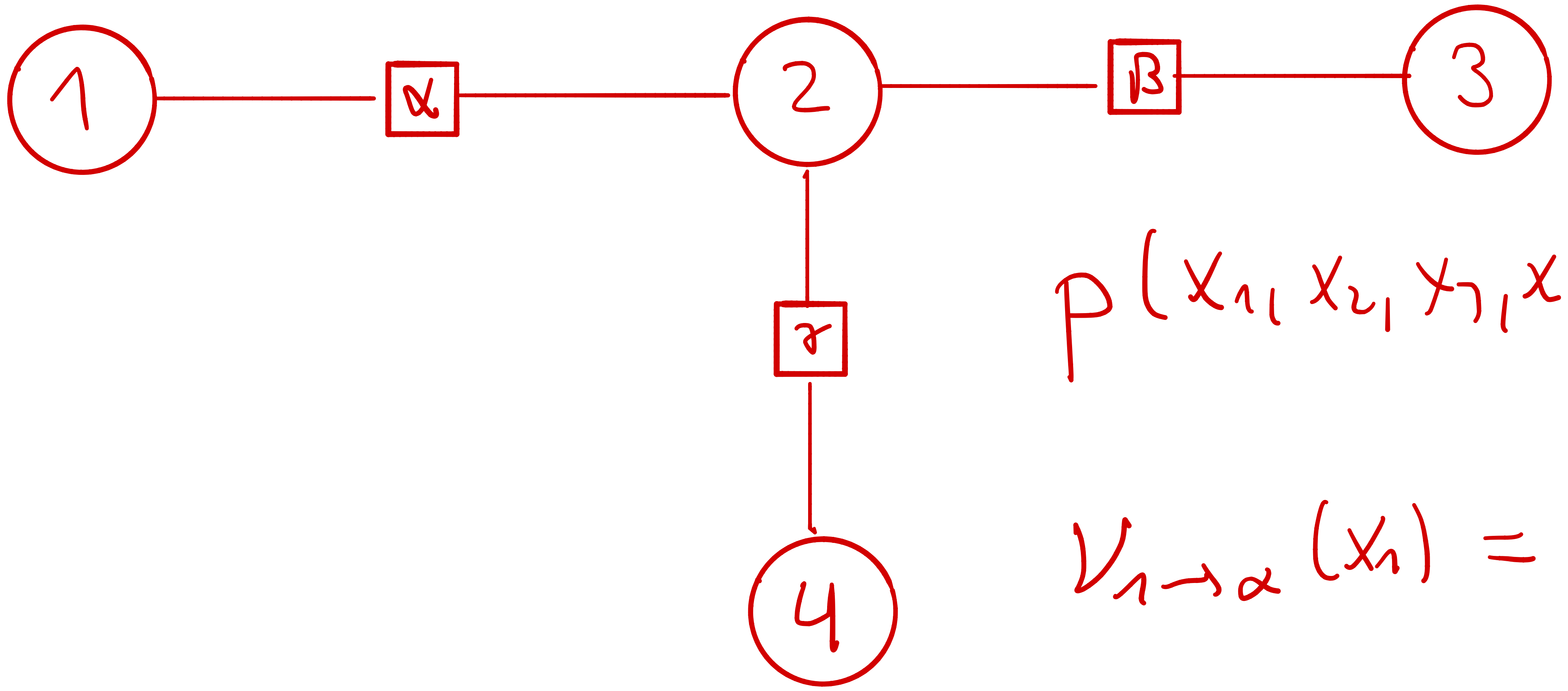


Machine Learning 2

Graphical Models
- Exact Inference
- Max-Sum Algorithm
- Example

Patrick Forré

Example: Max-Sum Algorithm



$$p(x_1, x_2, x_3, x_4) = \frac{1}{Z} \cdot \psi_\alpha(x_1, x_2)\, \psi_\beta(x_2, x_3)\, \psi_\gamma(x_2, x_4)$$

$$\nu_{1\to\alpha}(x_1) = 0$$

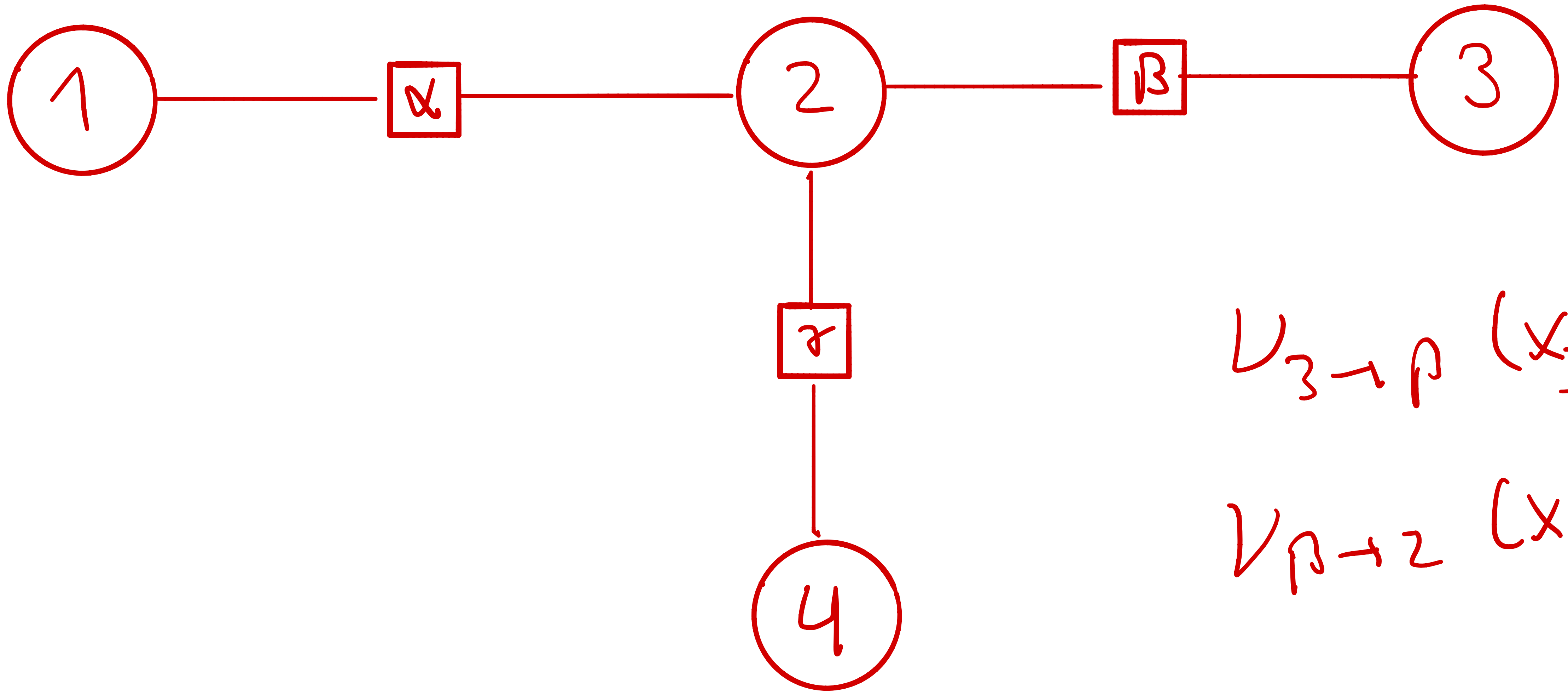
$$\nu_{\alpha\to 2}(x_2) = \max_{x_1} \Big[\log \psi_\alpha(x_1, x_2) + \underbrace{\nu_{1\to\alpha}(x_1)}_{0} \Big]$$

$$= \max_{x_1} \big[\log \psi_\alpha(x_1, x_2) \big]$$

$$x_1^\alpha(x_2) \in \operatorname{argmax}_{x_1} \log \psi_\alpha(x_1, x_2)$$

i.e. $\forall x_2 : \log \psi_\alpha(x_1^\alpha(x_2), x_2) = \max_{x_1} \log \psi_\alpha(x_1, x_2)$

Example: Max-Sum Algorithm



$$\nu_{3\to\beta}(x_3) = 0$$

$$\nu_{\beta\to 2}(x_2) = \max_{x_3} \log \psi_\beta(x_2, x_3)$$

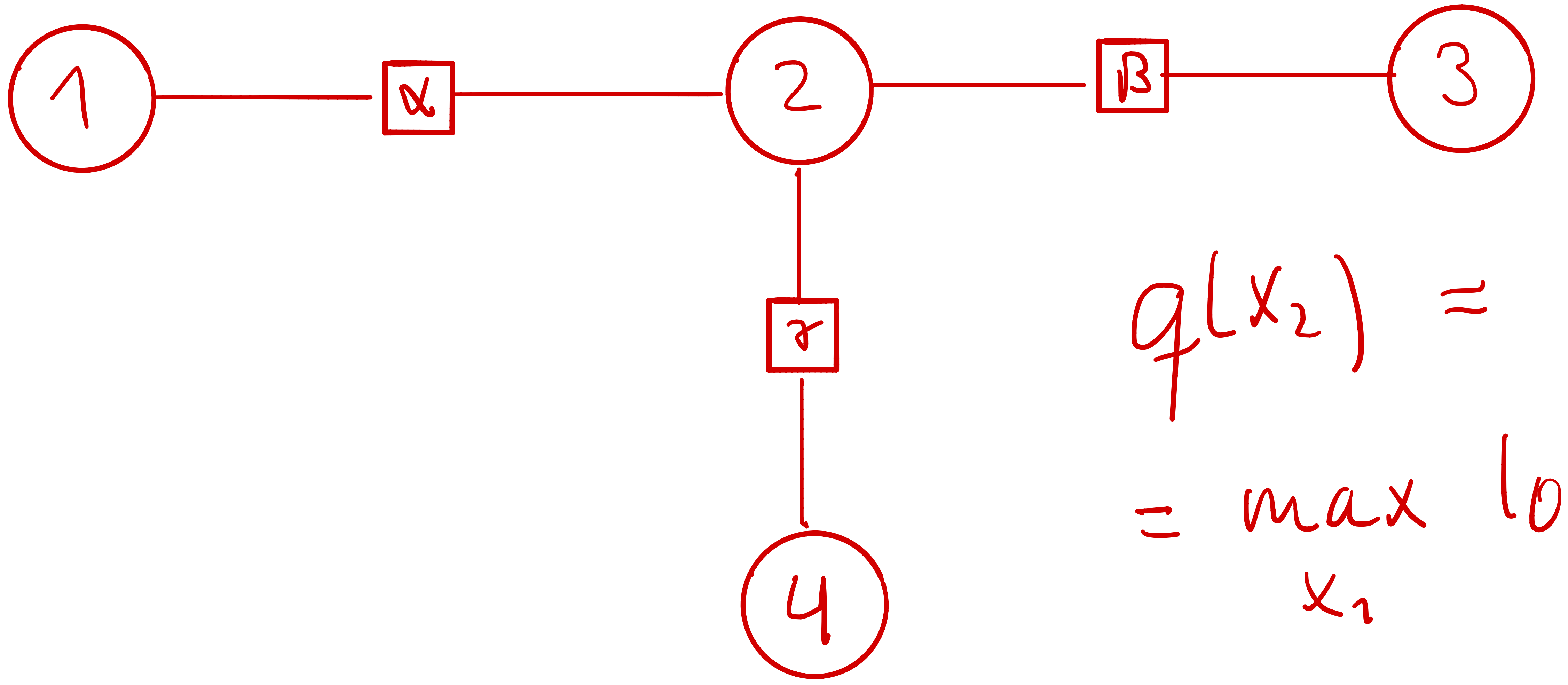
$$x_3^\beta(x_2) \in \operatorname{argmax}_{x_3} \log \psi_\beta(x_2, x_3)$$

$$\nu_{4\to\gamma}(x_4) = 0$$

$$\nu_{\gamma\to 2}(x_2) = \max_{x_4} \log \psi_\gamma(x_2, x_4)$$

$$x_4^\gamma(x_2) \in \operatorname{argmax}_{x_4} \log \psi_\gamma(x_2, x_4)$$

Example: Max-Sum Algorithm



$$q(x_2) = \nu_{\alpha\to 2}(x_2) + \nu_{\beta\to 2}(x_2) + \nu_{\gamma\to 2}(x_2)$$

$$= \max_{x_1} \log \psi_\alpha(x_1,x_2) + \max_{x_3} \psi_\beta(x_2,x_3) + \max_{x_4} \psi_\gamma(x_2,x_4)$$

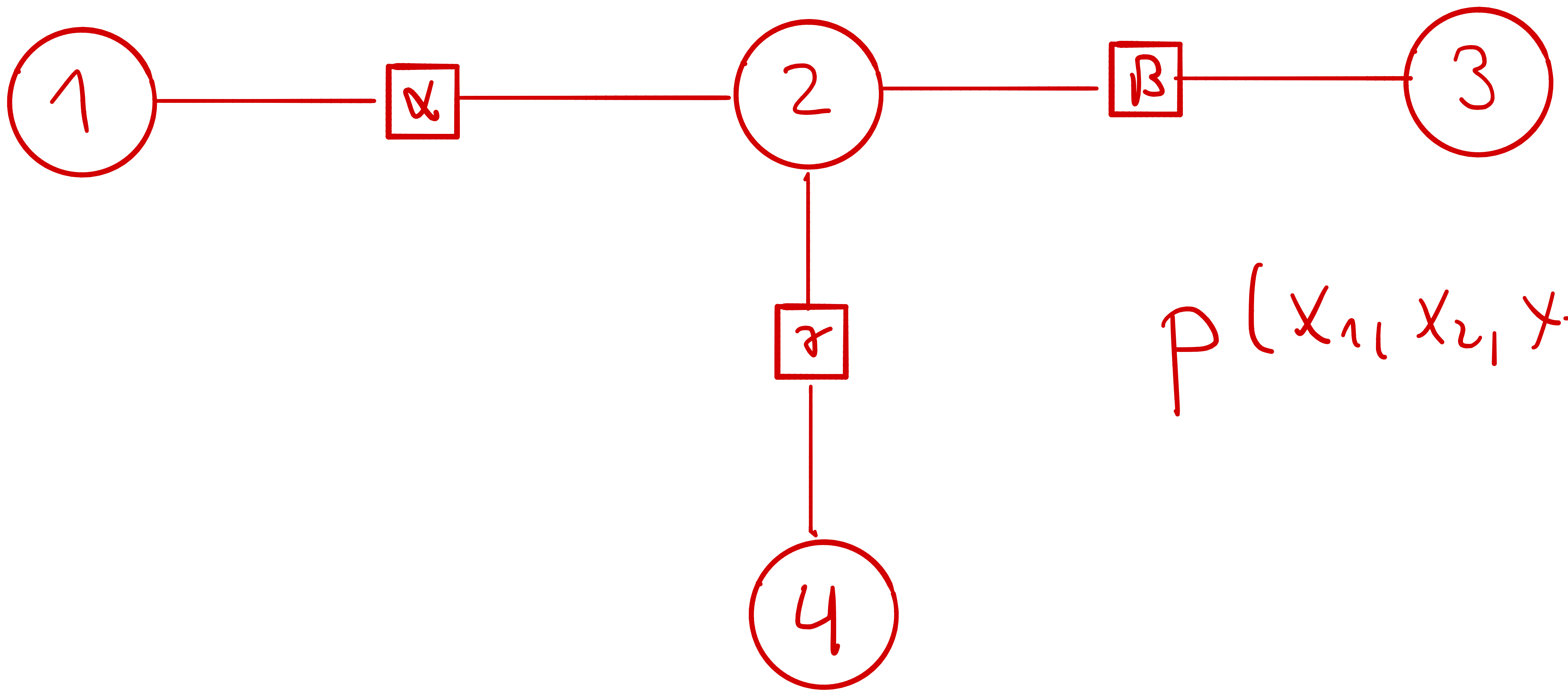
$$x_2^* \in \operatorname{argmax} q(x_2)$$

i.e. $q(x_2^*) = \max_{x_2} q(x_2)$

$x_1^* = x_1^\alpha(x_2^*)$, $x_3^* = x_3^\beta(x_2^*)$, $x_4^* = x_4^\gamma(x_2^*)$

$$x^* = (x_1^*, x_2^*, x_3^*, x_4^*)$$

Example: Max-Sum Algorithm



$$p(x_1, x_2, x_3, x_4) = \frac{1}{Z} \cdot \psi_\alpha(x_1, x_2)\, \psi_\beta(x_2, x_3)\, \psi_\gamma(x_2, x_4)$$

$$\log p(x^*) = \log \psi_\alpha(x_1^*(x_2^*), x_2^*) + \log \psi_\beta(x_2^*, x_3^\beta(x_2^*)) + \log \psi_\gamma(x_2^*, x_4^\gamma(x_2^*)) - \log Z$$

$$= \max_{x_1} \log \psi_\alpha(x_1, x_2^*) + \max_{x_3} \log \psi_\beta(x_2^*, x_3) + \max_{x_4} \log \psi_\gamma(x_2^*, x_4) - \log Z$$

$$? \overset{?}{=} \max_{x_2} \left[\max_{x_1} \log \psi_\alpha(x_1, x_2) + \max_{x_3} \log \psi_\beta(x_2, x_3) + \max_{x_4} \log \psi_\gamma(x_2, x_4) \right] - \log Z$$

$$= \max_{x_2} \max_{x_1} \max_{x_3} \max_{4} \Big[\underbrace{\log \psi_\alpha(x_1, x_2) + \log \psi_\beta(x_2, x_3) + \psi_\gamma(x_2, x_4) - \log Z}_{\log p(x)} \Big] = \max_{x} \log p(x) \checkmark$$

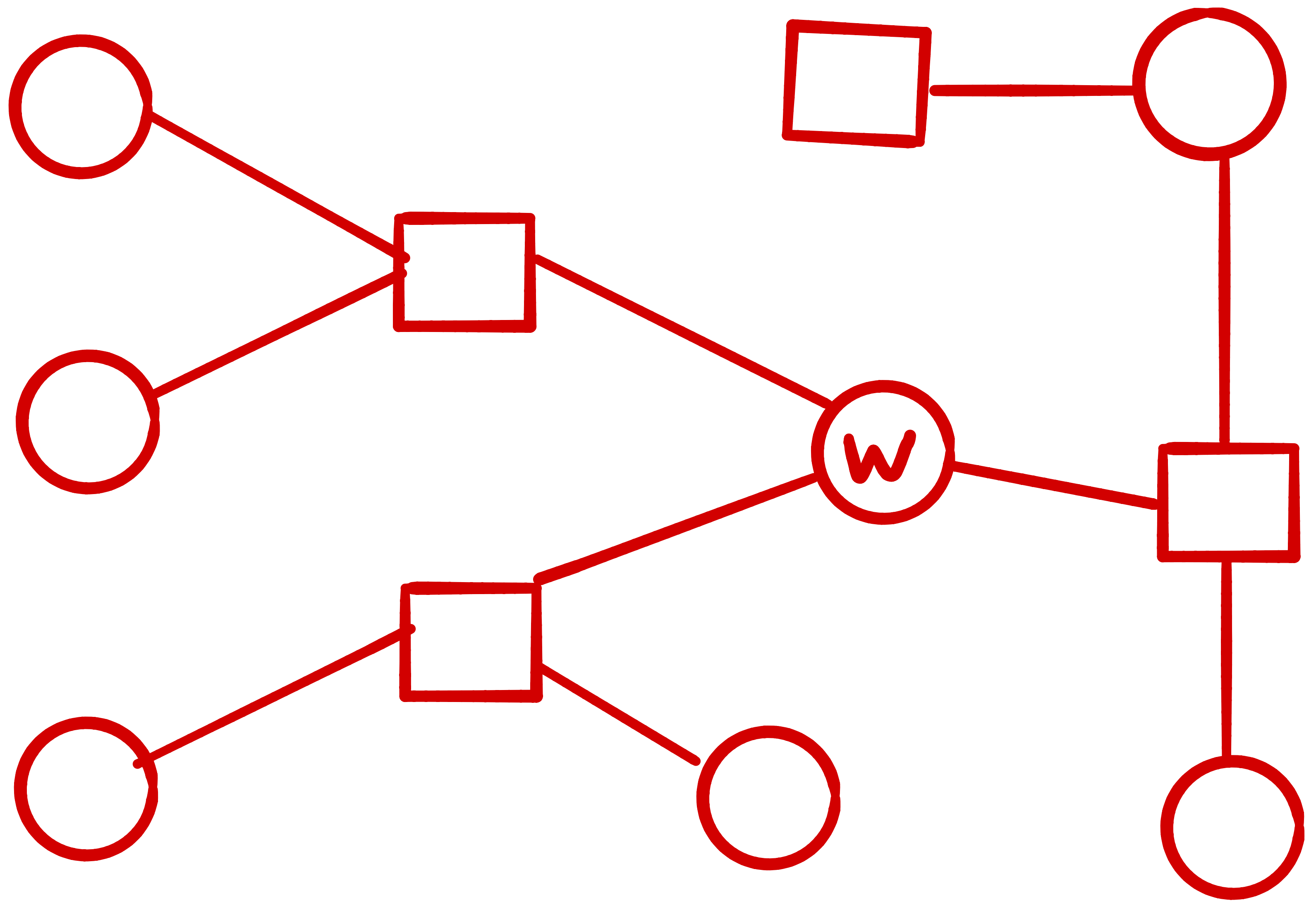
Machine Learning 2

Graphical Models
- Exact Inference
- Max-Sum Algorithm
- Summary

Patrick Forré

Max-Sum Algorithm

- $G = (V, F, E)$ Factor Tree with $p(x_V) = \frac{1}{Z} \prod_{\alpha \in F} \psi_\alpha(x_\alpha)$. Pick a 'Root' $w \in V$ inside the graph.
- Forward Message Passing towards Root:
  - Variable-to-factor: $\nu_{v \to \beta}(x_v) = \sum_{\gamma \in \partial(v) \setminus \{\beta\}} \nu_{\gamma \to v}(x_v)$, Leaf: $\nu_{u \to \beta}(x_u) = 0$
  - Factor-to-variable: $\nu_{\beta \to w}(x_w) = \max_{x_{\beta \setminus w}} \left(\log \psi_\beta(x_\beta) + \sum_{u \in \beta \setminus w} \nu_{u \to \beta}(x_u) \right)$, Leaf: $\nu_{\alpha \to v}(x_v) = \log \psi_\alpha(x_v)$
  - For backtracking: $x_{\beta \setminus w}^{\beta \leftarrow w}(x_w) \in \arg\max_{x_{\beta \setminus w}} \left(\log \psi_\beta(x_\beta) + \sum_{u \in \beta \setminus w} \nu_{u \to \beta}(x_u) \right)$
- At Root: Pick: $x_w^* \in \arg\max_{x_w} \sum_{\gamma \in \partial(w)} \nu_{\gamma \to w}(x_w)$.
- Backtracking from Root towards Leaves: Pick $x_{\alpha \setminus v}^* = x_{\alpha \setminus v}^{\alpha \leftarrow v}(x_v^*)$.
- Then $x^* := (x_v^*)_{v \in V} \in \arg\max_x p(x)$ is a mode.



Remarks

- The Max-Sum Algorithm is similar to the Sum-Product Algorithm but with the following changes:
  - take the $\log$ for all quantities
  - replace all $\sum$'s by $\max$'s.
  - replace all $\prod$'s by $\sum$'s
  - keep track of maximisers.
- If the Factor Graph has cycles use Junction Tree Algorithm to avoid cyclic updates.