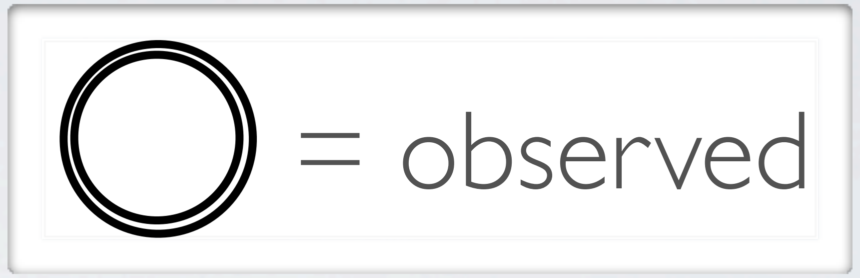
Neural networks

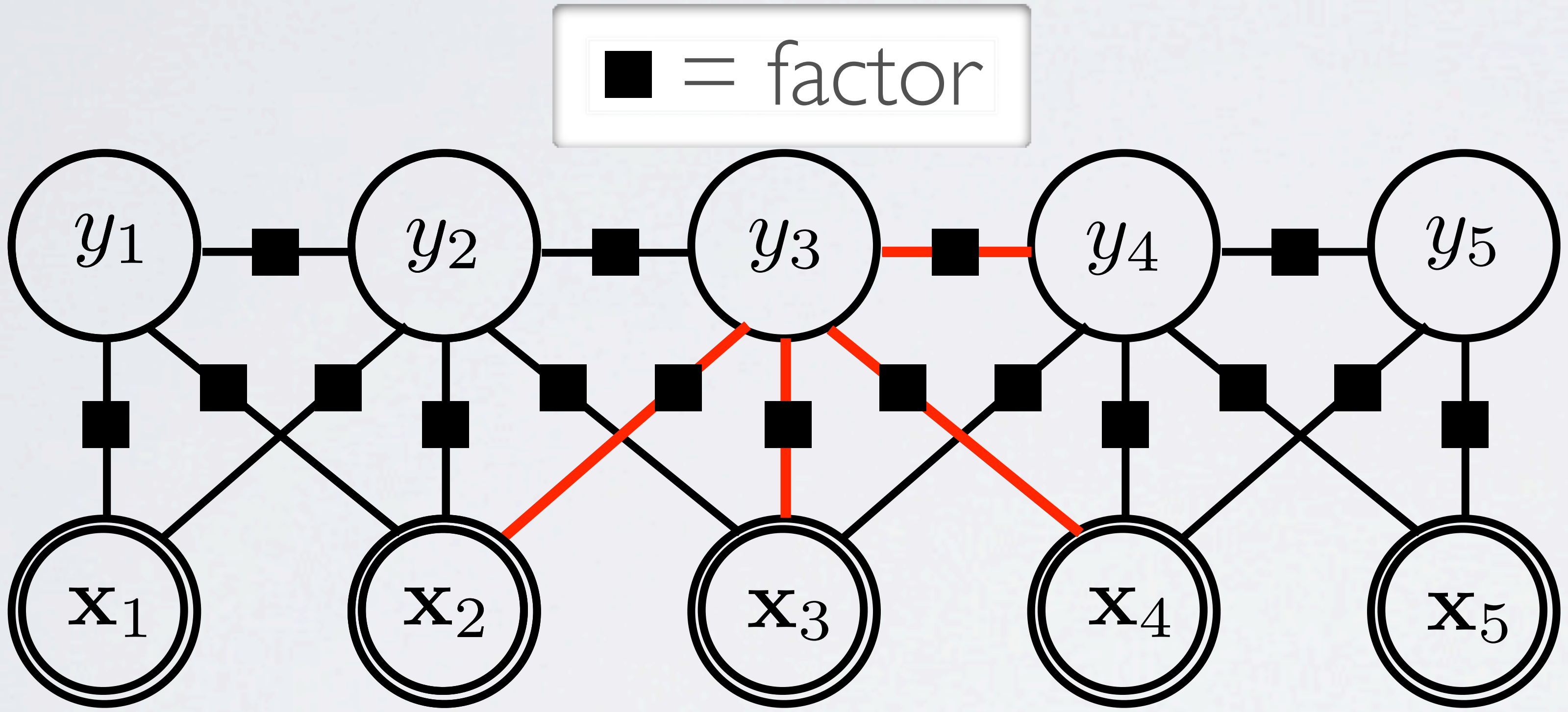
Conditional random fields - belief propagation

FACTOR GRAPH VISUALIZATION

Topics: factor graph



- Factor graphs better represent factors



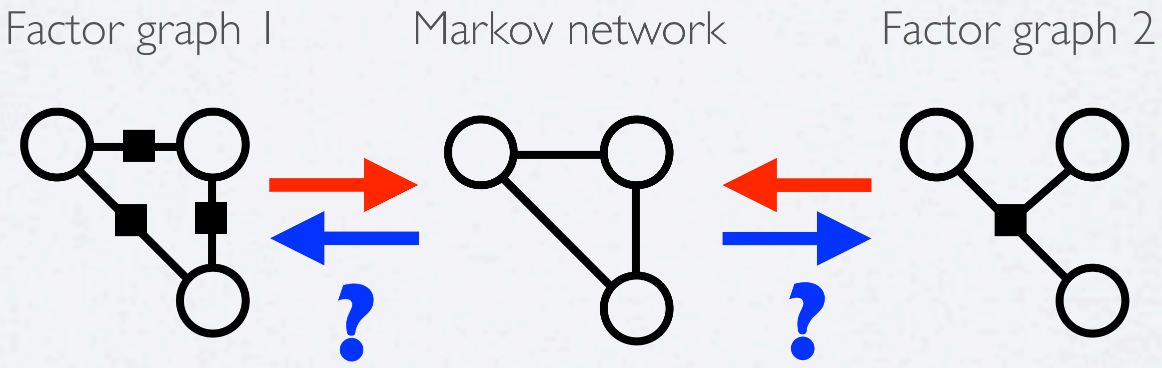
$\exp(a^{(L+1,0)}(\mathbf{x}_3)_{y_3})$

$\exp(a^{(L+1,-1)}(\mathbf{x}_2)_{y_3})$

$\exp(a^{(L+1,+1)}(\mathbf{x}_4)_{y_3})$

$\exp(V_{y_3,y_4})$

- This is less ambiguous:

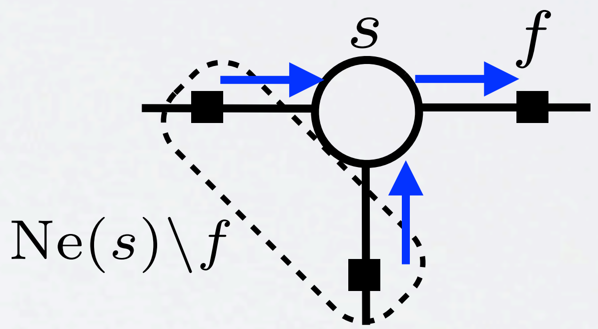


BELIEF PROPAGATION

Topics: belief propagation, message passing

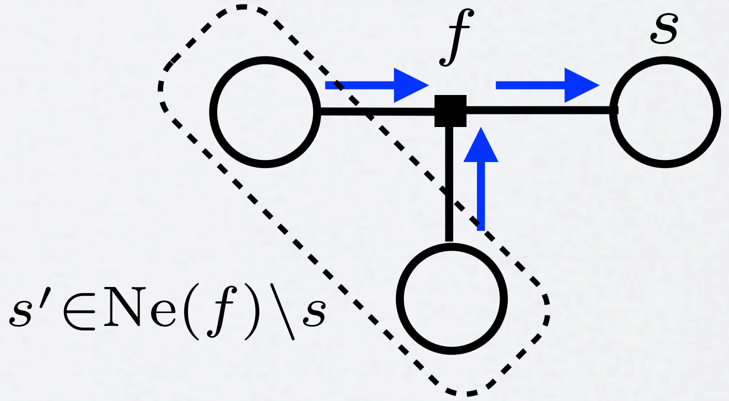
- Factor graphs better represent the computations needed to do inference
  - we can write the forward-backward algorithm seen before into a general message passing form
  - there are two types of message:
    - from a variable node (○) to its neighbor factor nodes (■):

$$\mu_{s \to f}(i) = \prod_{f' \in \mathrm{Ne}(s) \setminus f} \mu_{f' \to s}(i)$$



    - from a factor node (■) to its neighbor variable nodes (○):

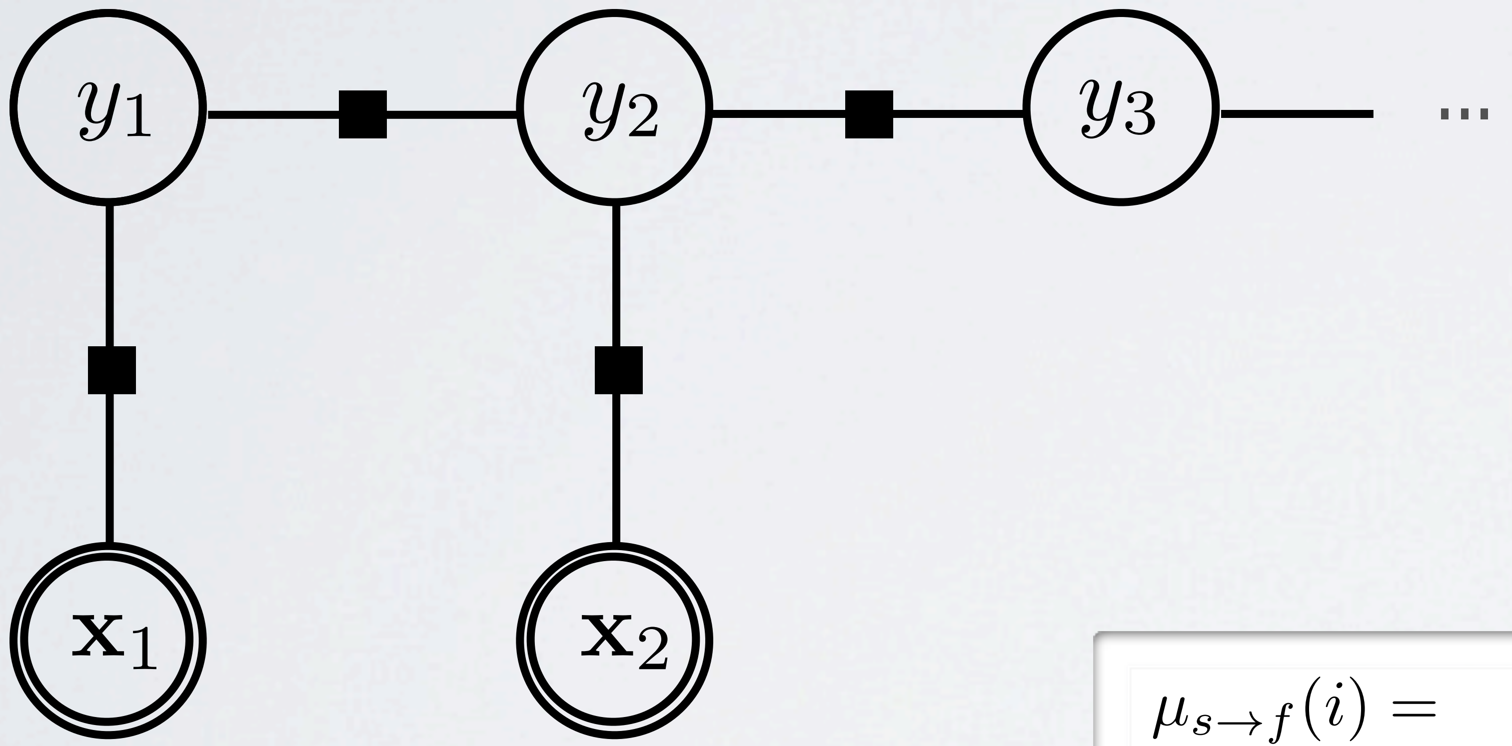
$$\mu_{f \to s}(i) = \sum_{\substack{\mathbf{z} \text{ is unobs.} \\ \text{and } z_s = i}} \phi_f(\mathbf{z}) \prod_{s' \in \mathrm{Ne}(f) \setminus s} \mu_{s' \to f}(z_{s'})$$



BELIEF PROPAGATION

Topics: belief propagation, message passing

- Example: our (simplified) linear chain CRF



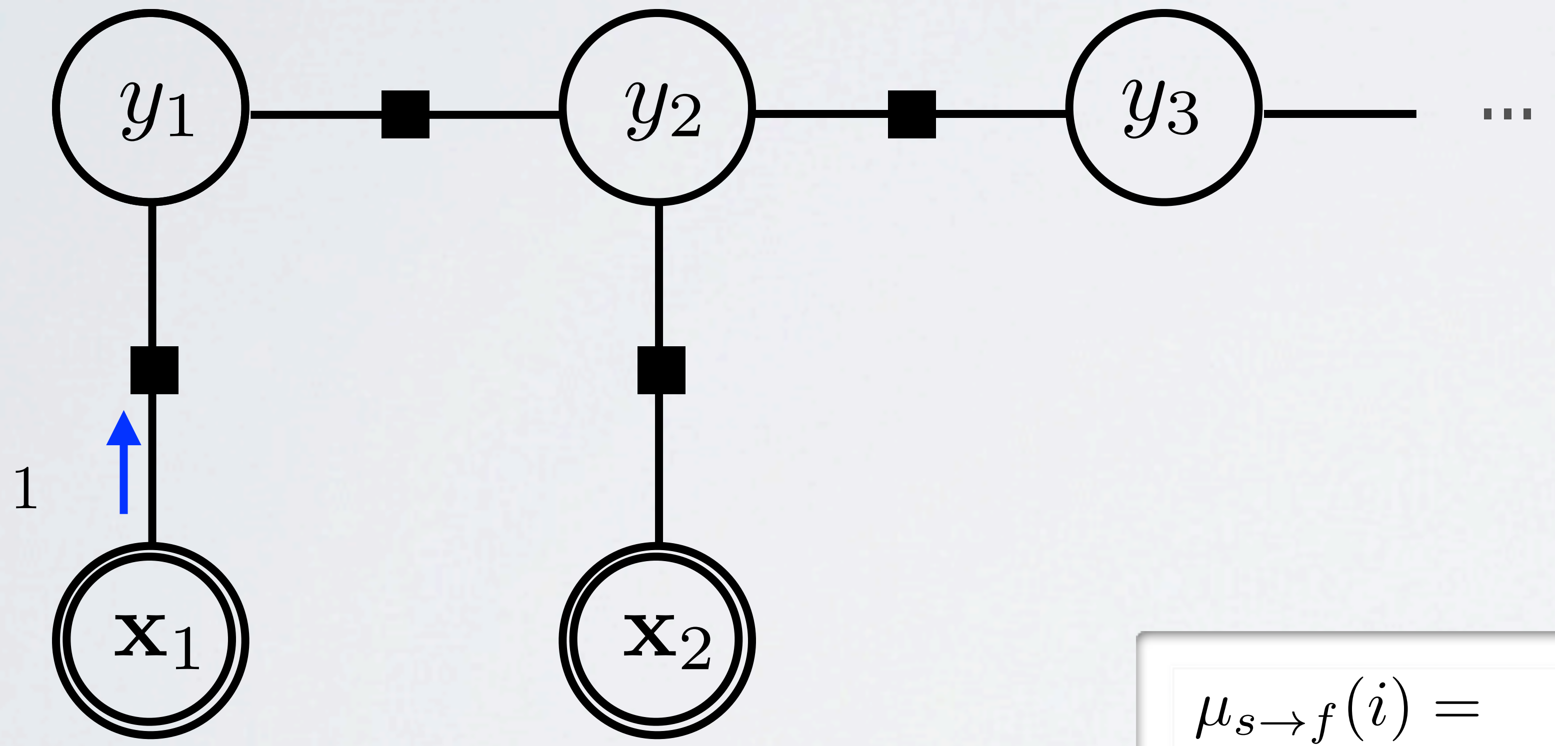
$$\mu_{s \to f}(i) = \prod_{f' \in \mathrm{Ne}(s) \setminus f} \mu_{f' \to s}(i)$$

$$\mu_{f \to s}(i) = \sum_{\substack{\mathbf{z} \text{ is unobs.} \\ \text{and } z_s = i}} \phi_f(\mathbf{z}) \prod_{s' \in \mathrm{Ne}(f) \setminus s} \mu_{s' \to f}(z_{s'})$$

BELIEF PROPAGATION

Topics: belief propagation, message passing

- Example: our (simplified) linear chain CRF



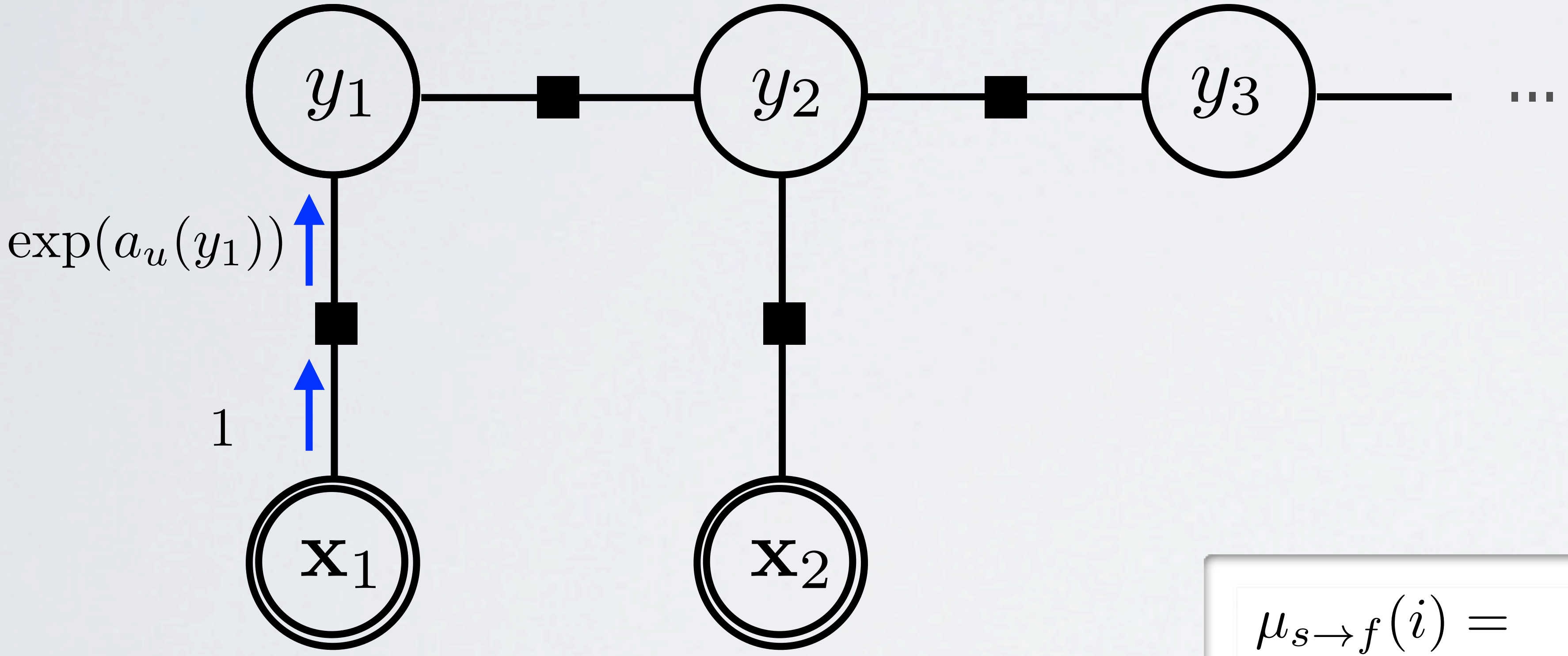
$$\mu_{s \to f}(i) = \prod_{f' \in \mathrm{Ne}(s) \setminus f} \mu_{f' \to s}(i)$$

$$\mu_{f \to s}(i) = \sum_{\substack{\mathbf{z} \text{ is unobs.} \\ \text{and } z_s = i}} \phi_f(\mathbf{z}) \prod_{s' \in \mathrm{Ne}(f) \setminus s} \mu_{s' \to f}(z_{s'})$$

BELIEF PROPAGATION

Topics: belief propagation, message passing

- Example: our (simplified) linear chain CRF



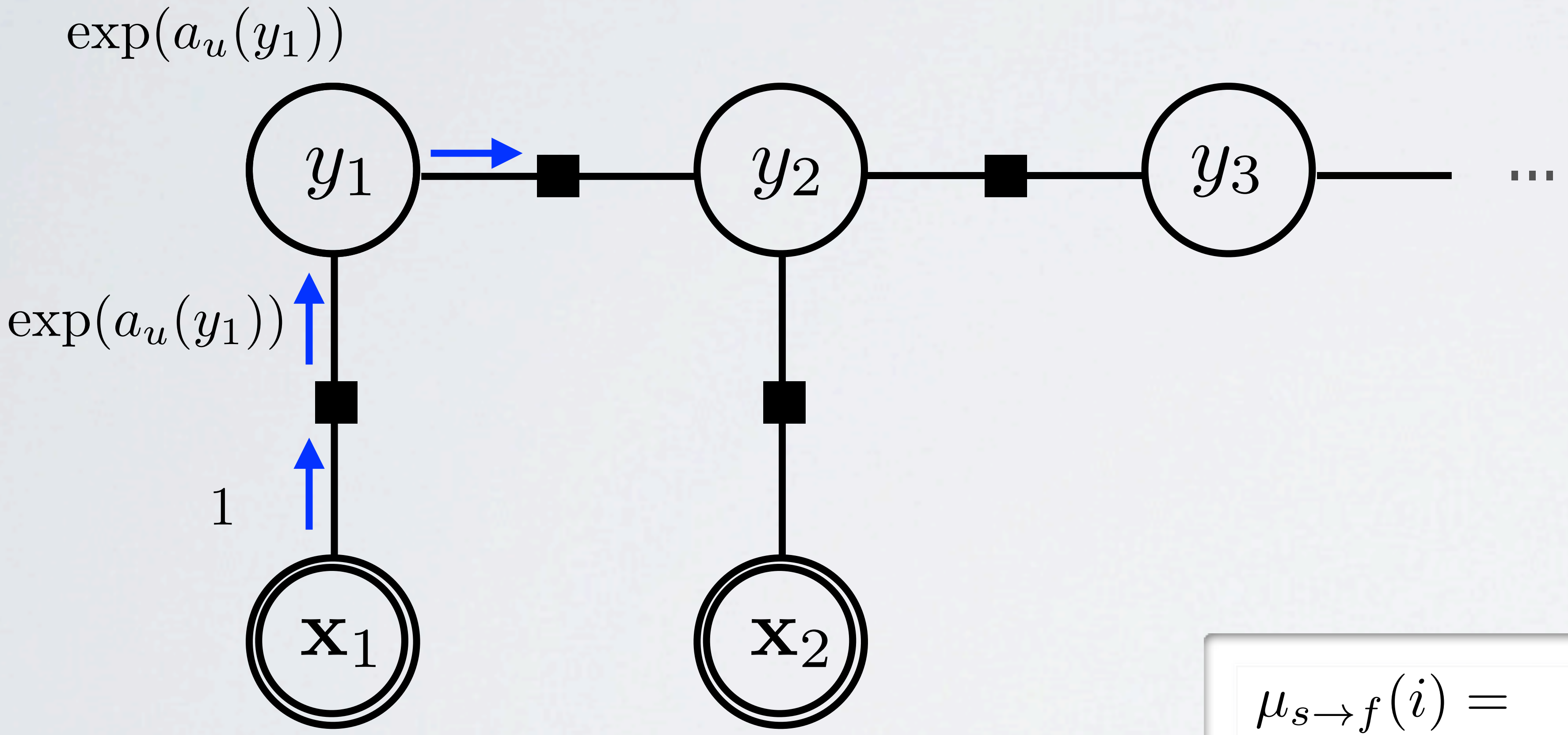
$$\mu_{s \to f}(i) = \prod_{f' \in \mathrm{Ne}(s) \setminus f} \mu_{f' \to s}(i)$$

$$\mu_{f \to s}(i) = \sum_{\substack{\mathbf{z} \text{ is unobs.} \\ \text{and } z_s = i}} \phi_f(\mathbf{z}) \prod_{s' \in \mathrm{Ne}(f) \setminus s} \mu_{s' \to f}(z_{s'})$$

BELIEF PROPAGATION

Topics: belief propagation, message passing

- Example: our (simplified) linear chain CRF



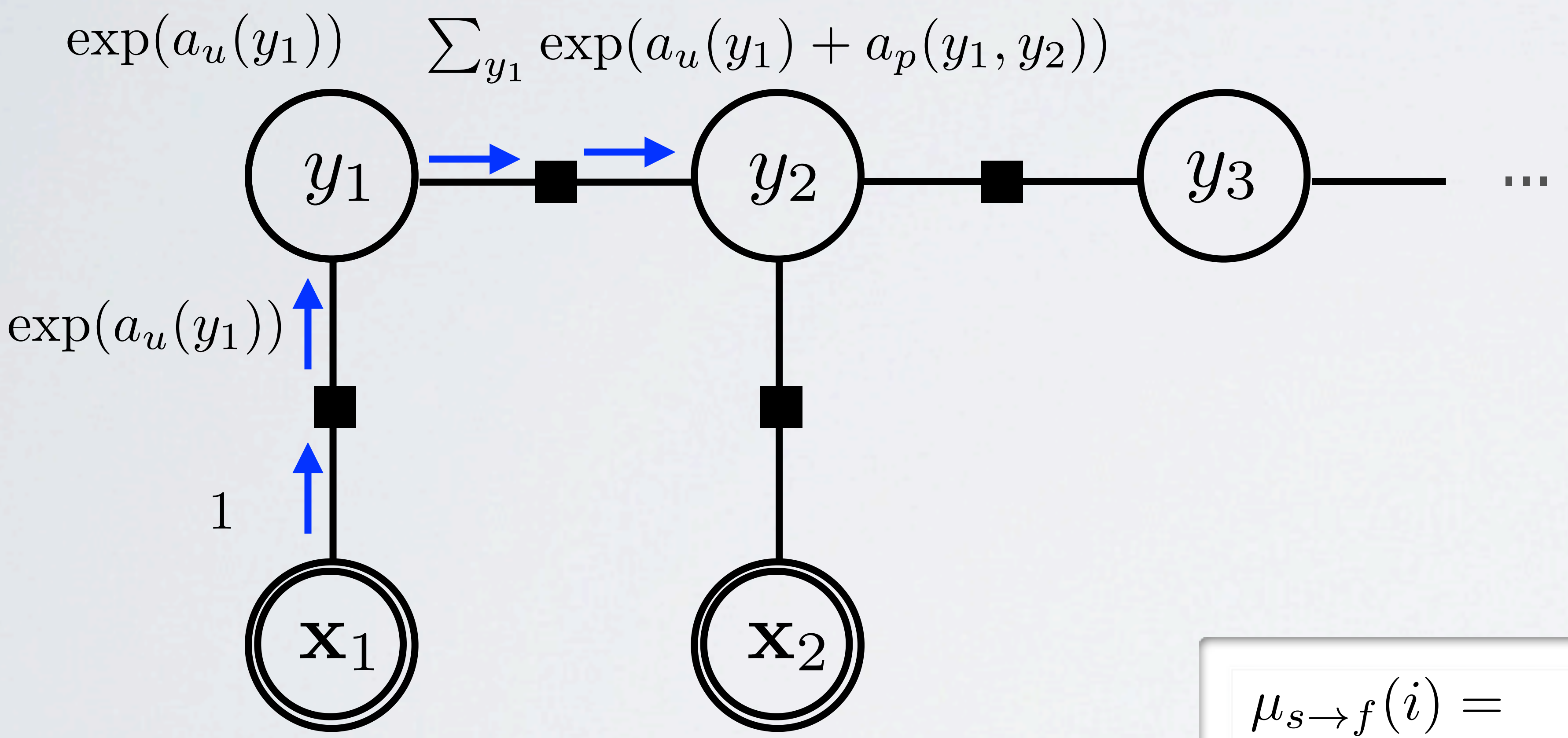
$$\mu_{s \to f}(i) = \prod_{f' \in \mathrm{Ne}(s) \setminus f} \mu_{f' \to s}(i)$$

$$\mu_{f \to s}(i) = \sum_{\substack{\mathbf{z} \text{ is unobs.} \\ \text{and } z_s = i}} \phi_f(\mathbf{z}) \prod_{s' \in \mathrm{Ne}(f) \setminus s} \mu_{s' \to f}(z_{s'})$$

BELIEF PROPAGATION

Topics: belief propagation, message passing

- Example: our (simplified) linear chain CRF



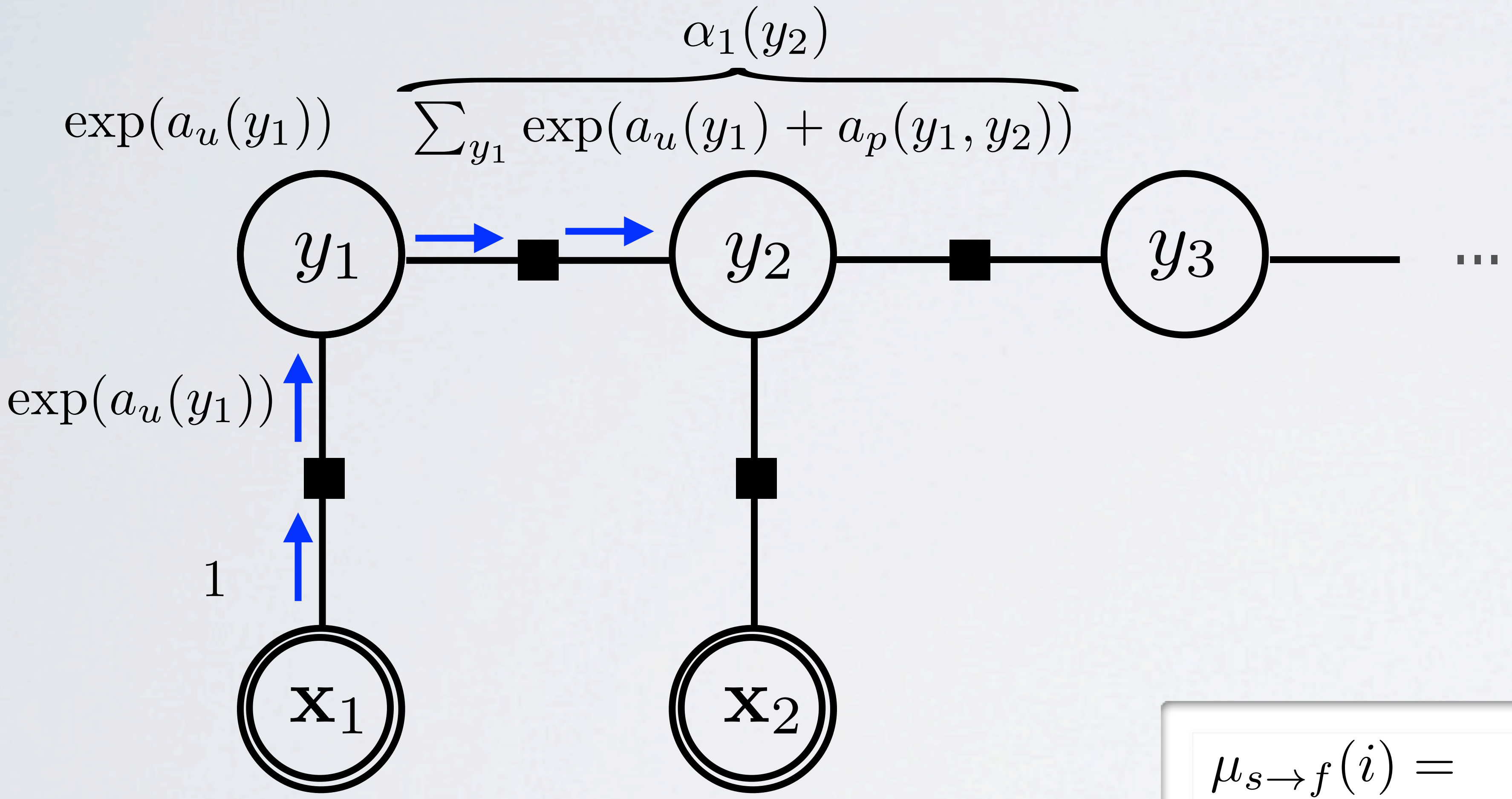
$$\mu_{s \to f}(i) = \prod_{f' \in \mathrm{Ne}(s) \setminus f} \mu_{f' \to s}(i)$$

$$\mu_{f \to s}(i) = \sum_{\substack{\mathbf{z} \text{ is unobs.} \\ \text{and } z_s = i}} \phi_f(\mathbf{z}) \prod_{s' \in \mathrm{Ne}(f) \setminus s} \mu_{s' \to f}(z_{s'})$$

BELIEF PROPAGATION

Topics: belief propagation, message passing

- Example: our (simplified) linear chain CRF



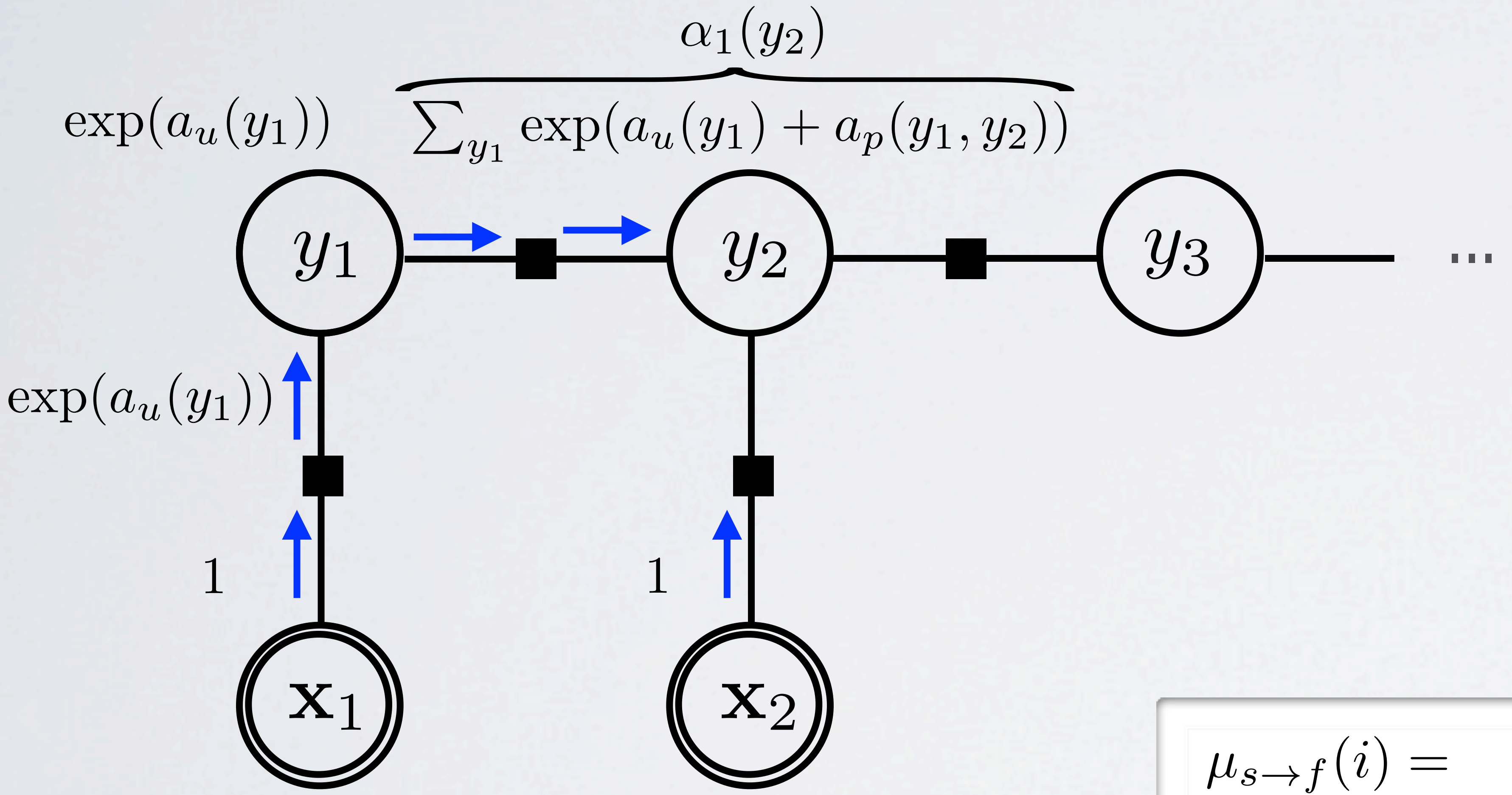
$$\mu_{s \to f}(i) = \prod_{f' \in \mathrm{Ne}(s) \setminus f} \mu_{f' \to s}(i)$$

$$\mu_{f \to s}(i) = \sum_{\substack{\mathbf{z} \text{ is unobs.} \\ \text{and } z_s = i}} \phi_f(\mathbf{z}) \prod_{s' \in \mathrm{Ne}(f) \setminus s} \mu_{s' \to f}(z_{s'})$$

BELIEF PROPAGATION

Topics: belief propagation, message passing

- Example: our (simplified) linear chain CRF



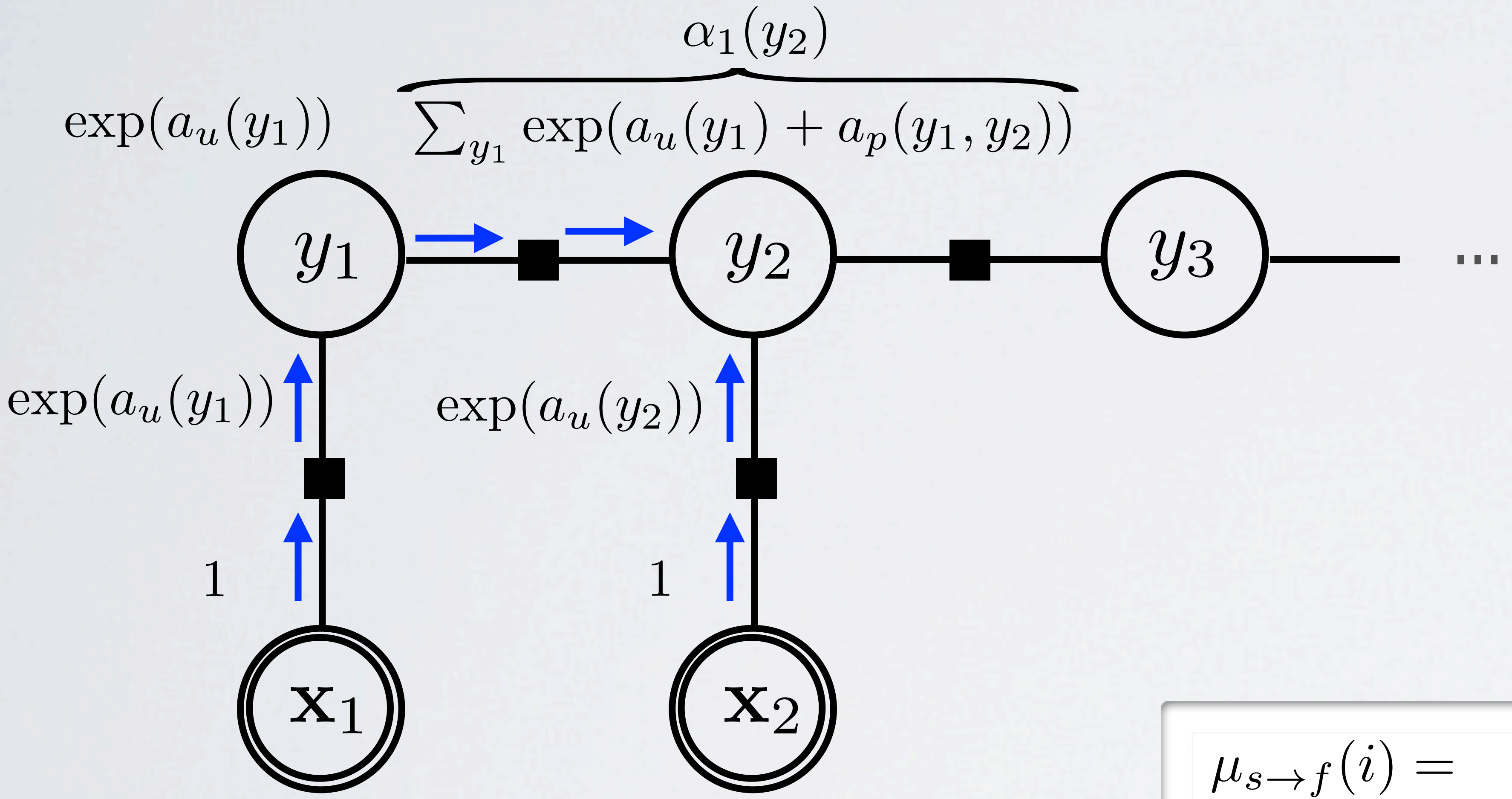
$$\mu_{s \to f}(i) = \prod_{f' \in \mathrm{Ne}(s) \setminus f} \mu_{f' \to s}(i)$$

$$\mu_{f \to s}(i) = \sum_{\substack{\mathbf{z} \text{ is unobs.} \\ \text{and } z_s = i}} \phi_f(\mathbf{z}) \prod_{s' \in \mathrm{Ne}(f) \setminus s} \mu_{s' \to f}(z_{s'})$$

BELIEF PROPAGATION

Topics: belief propagation, message passing

- Example: our (simplified) linear chain CRF



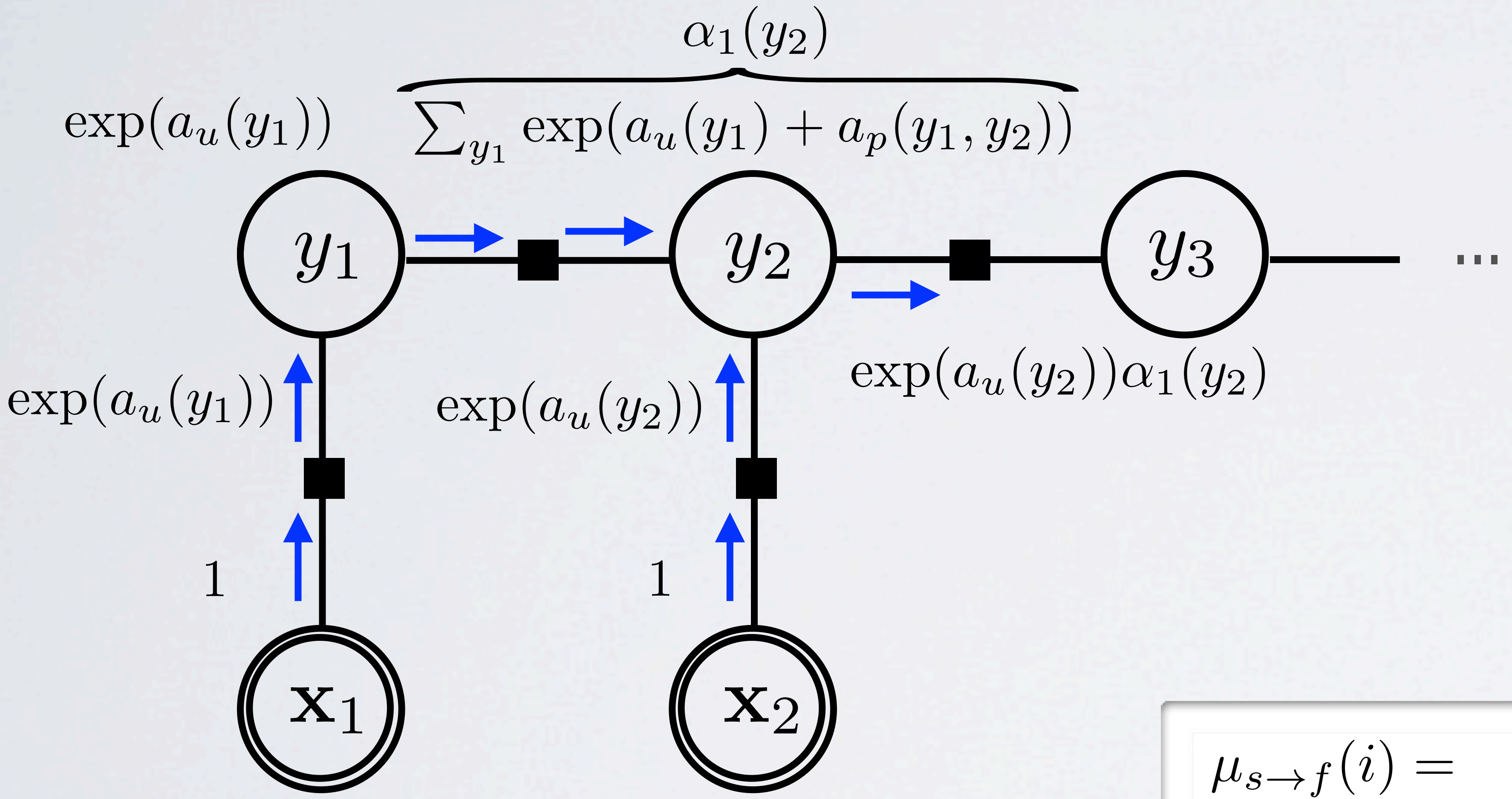
$$\mu_{s \to f}(i) = \prod_{f' \in \mathrm{Ne}(s) \setminus f} \mu_{f' \to s}(i)$$

$$\mu_{f \to s}(i) = \sum_{\substack{\mathbf{z} \text{ is unobs.} \\ \text{and } z_s = i}} \phi_f(\mathbf{z}) \prod_{s' \in \mathrm{Ne}(f) \setminus s} \mu_{s' \to f}(z_{s'})$$

BELIEF PROPAGATION

Topics: belief propagation, message passing

- Example: our (simplified) linear chain CRF



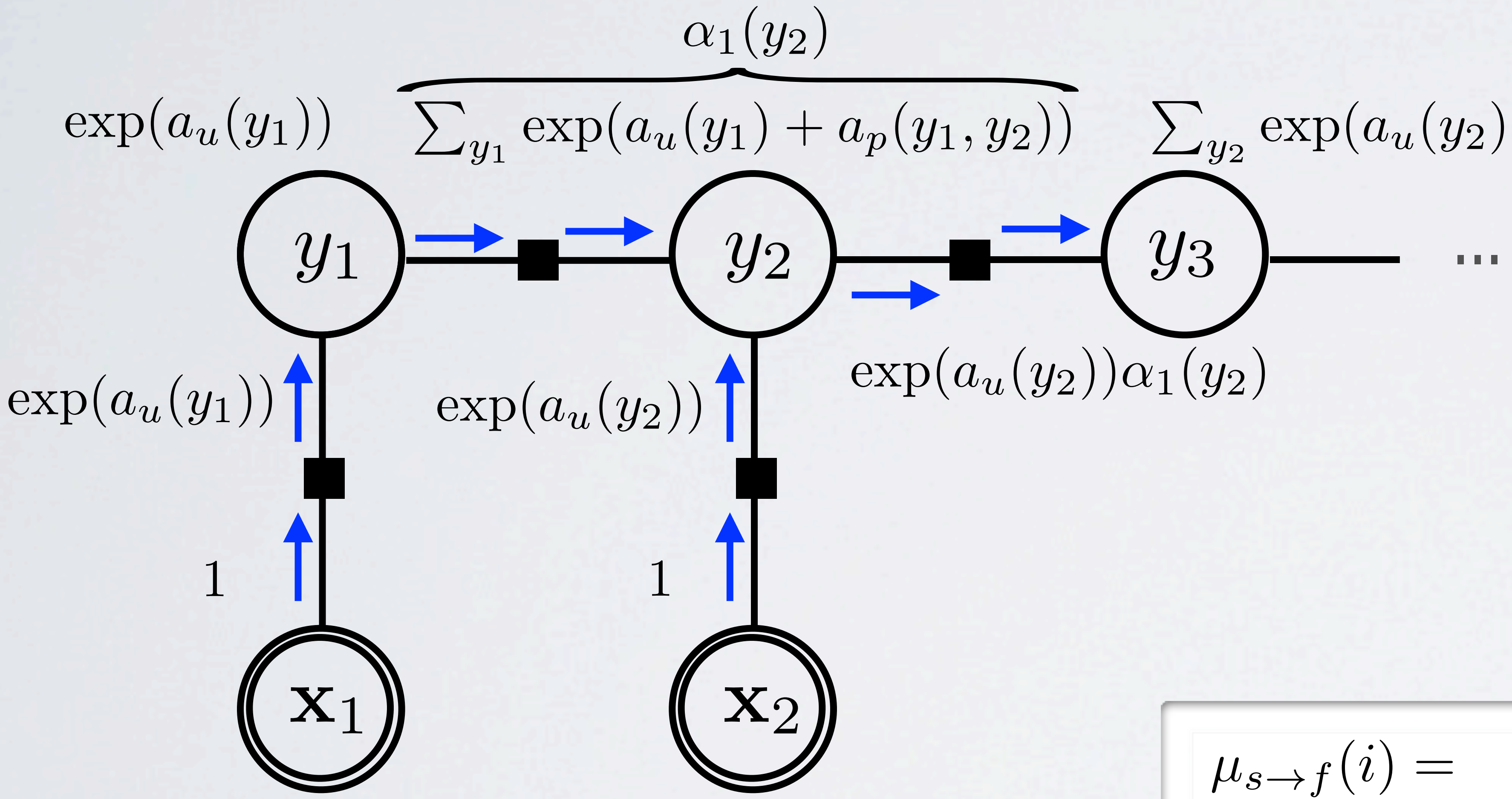
$$\mu_{s \to f}(i) = \prod_{f' \in \mathrm{Ne}(s) \setminus f} \mu_{f' \to s}(i)$$

$$\mu_{f \to s}(i) = \sum_{\substack{\mathbf{z} \text{ is unobs.} \\ \text{and } z_s = i}} \phi_f(\mathbf{z}) \prod_{s' \in \mathrm{Ne}(f) \setminus s} \mu_{s' \to f}(z_{s'})$$

BELIEF PROPAGATION

Topics: belief propagation, message passing

- Example: our (simplified) linear chain CRF



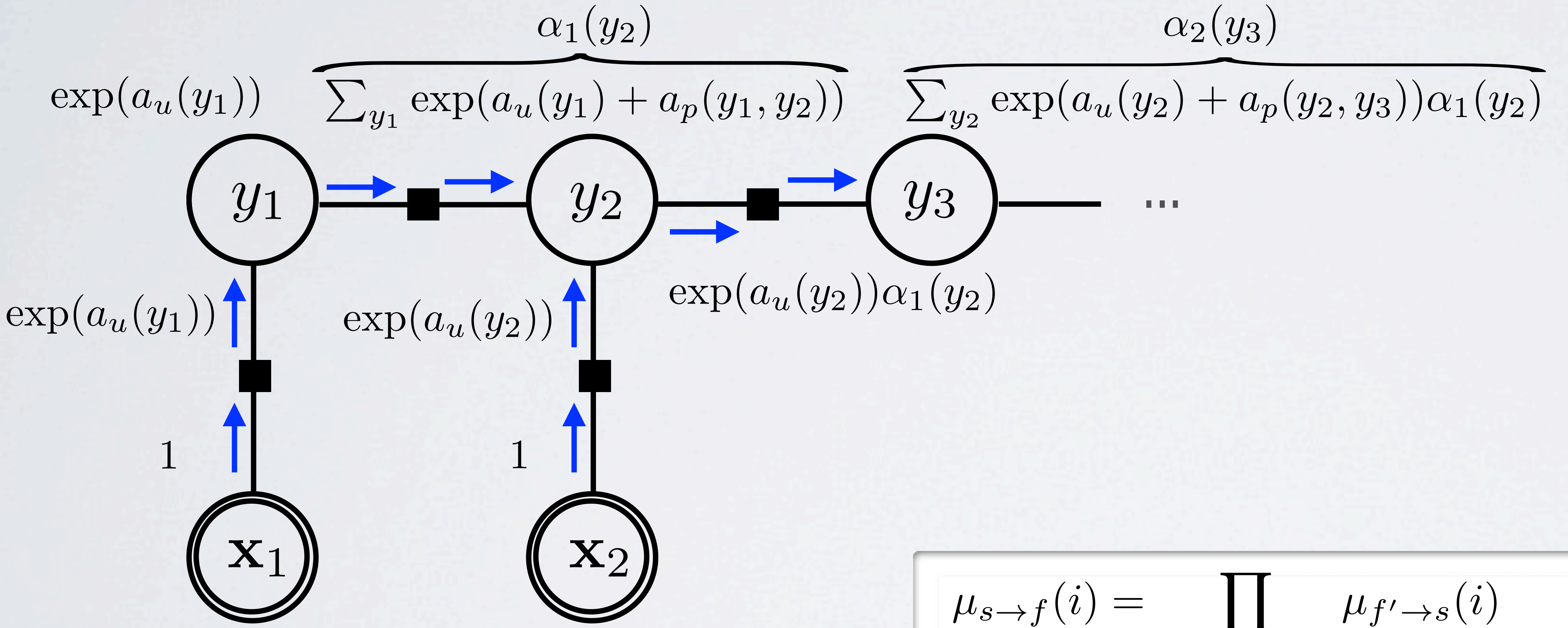
$$\mu_{s \to f}(i) = \prod_{f' \in \mathrm{Ne}(s) \setminus f} \mu_{f' \to s}(i)$$

$$\mu_{f \to s}(i) = \sum_{\substack{\mathbf{z} \text{ is unobs.} \\ \text{and } z_s = i}} \phi_f(\mathbf{z}) \prod_{s' \in \mathrm{Ne}(f) \setminus s} \mu_{s' \to f}(z_{s'})$$

BELIEF PROPAGATION

Topics: belief propagation, message passing

- Example: our (simplified) linear chain CRF



$$\mu_{s \to f}(i) = \prod_{f' \in \mathrm{Ne}(s) \backslash f} \mu_{f' \to s}(i)$$

$$\mu_{f \to s}(i) = \sum_{\substack{\mathbf{z} \text{ is unobs.} \\ \text{and } z_s = i}} \phi_f(\mathbf{z}) \prod_{s' \in \mathrm{Ne}(f) \backslash s} \mu_{s' \to f}(z_{s'})$$

(LOOPY) BELIEF PROPAGATION

Topics: belief propagation, message passing

- On a linear chain graph, belief propagation is the same a forward-backward
  - forward pass of message passing computes the $\alpha_k(y_{k+1})$
  - backward pass of message passing computes the $\beta_k(y_{k-1})$
- For numerical stability, passing log-messages is preferred
- Can do inference on other types of structures
  - belief propagation is also exact on arbitrary trees
  - on a graph with loops, (loopy-)belief propagation can be used to do approximate inference (but can divergence, if not careful)
  - many general purpose libraries are publicly available

LINEAR CHAIN CRF

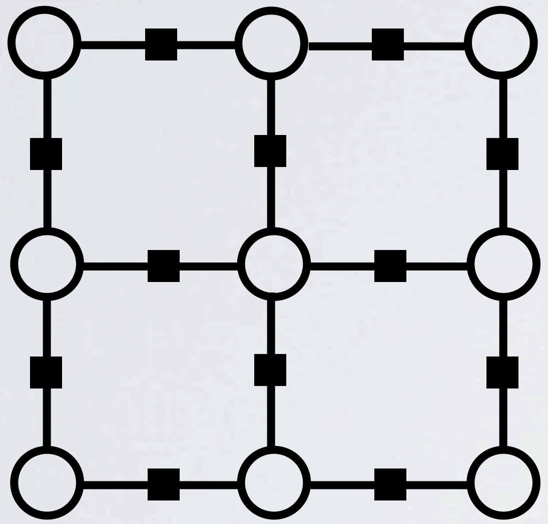
Topics: other variations on linear chain CRF

- We could add lateral connections between labels that are at 2 positions away $\phi_f(y_k, y_{k+2})$
- We could add lateral connections for triplets of labels $\phi_f(y_k, y_{k+1}, y_{k+2})$
- The idea is to add connections between things to model their dependency more directly
- The connectivity can even change between examples

GENERAL CRF

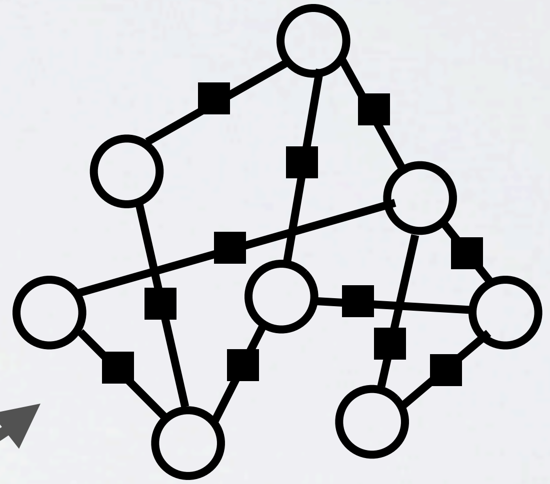
Topics: CRFs in general

- We don't have to restrict the CRF structure to linear chains



Grid structure
(pixels in image)

$$p(\mathbf{y}|\mathbf{X}) = \frac{1}{Z(\mathbf{X})} \prod_f \Psi_f(\mathbf{y}, \mathbf{X})$$



General pair-wise structure
(webpages sharing a link)

- We could also have n-ary factors, with $n>2$

(LOOPY) BELIEF PROPAGATION

Topics: CRFs in general

- Marginals can be approximated with:

$$p(y_k|\mathbf{X}) = \frac{\exp(\log \phi_f(y_k) + \sum_{f' \in \mathrm{Ne}(k) \setminus f} \log \mu_{f' \to k}(y_k))}{\sum_{y'_k} \exp(\log \phi_f(y'_k) + \sum_{f' \in \mathrm{Ne}(k) \setminus f} \log \mu_{f' \to k}(y'_k))}$$

- In general, an approximated marginal is computed by
  1. summing all the log-factors that involve only the y_k variables of interest
  2. summing all the log-messages coming into the y_k variables from other factors
  3. exponentiating
  4. renormalizing