Neural networks

Conditional random fields - computing the partition function

LINEAR CHAIN CRF

Topics: unary and pairwise log-factors

- For brevity, let's assume this notation:
  - unary log-factors

$$a_u(y_k) = a^{(L+1,0)}(\mathbf{x}_k)_{y_k} + 1_{k>1}\; a^{(L+1,-1)}(\mathbf{x}_{k-1})_{y_k} + 1_{k<K}\; a^{(L+1,+1)}(\mathbf{x}_{k+1})_{y_k}$$

or

$$a_u(y_k) = a^{(L+1)}(\mathbf{x}_{k-1}, \mathbf{x}_k, \mathbf{x}_{k+1})_{y_k}$$

  - pairwise log-factors

$$a_p(y_k, y_{k+1}) = 1_{1\leq k<K}\; V_{y_k, y_{k+1}}$$

- Then we have:

$$p(\mathbf{y}|\mathbf{X}) = \exp\left(\sum_{k=1}^{K} a_u(y_k) + \sum_{k=1}^{K-1} a_p(y_k, y_{k+1})\right)/Z(\mathbf{X})$$

INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

- Then we have:

$$p(\mathbf{y}|\mathbf{X}) = \exp\left(\sum_{k=1}^{K} a_u(y_k) + \sum_{k=1}^{K-1} a_p(y_k, y_{k+1})\right) / Z(\mathbf{X})$$

where

$$Z(\mathbf{X}) = \sum_{y'_1} \sum_{y'_2} \cdots \sum_{y'_K} \exp\left(\sum_{k=1}^{K} a_u(y'_k) + \sum_{k=1}^{K-1} a_p(y'_k, y'_{k+1})\right)$$

hard to compute

INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

$$
\begin{aligned}
Z(\mathbf{X}) \quad = \quad & \sum_{y'_K} \exp(a_u(y'_K)) \\
& \left(\sum_{y'_{K-1}} \exp(a_u(y'_{K-1}) + a_p(y'_{K-1}, y'_K)) \right. \\
& \quad \ldots \\
& \left(\sum_{y'_2} \exp(a_u(y'_2) + a_p(y'_2, y'_3)) \right. \\
& \left. \left. \left(\sum_{y'_1} \exp(a_u(y'_1) + a_p(y'_1, y'_2)) \right) \right) \ldots \right)
\end{aligned}
$$

INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

$$Z(\mathbf{X}) = \sum_{y'_K} \exp(a_u(y'_K))$$

$$\left(\sum_{y'_{K-1}} \exp(a_u(y'_{K-1}) + a_p(y'_{K-1}, y'_K))\right.$$

$$\ldots$$

$$\left(\sum_{y'_2} \exp(a_u(y'_2) + a_p(y'_2, y'_3))\right.$$

$$\alpha_1(y'_2) \quad \left(\sum_{y'_1} \exp(a_u(y'_1) + a_p(y'_1, y'_2))\right)\Bigg) \ldots \Bigg)$$

INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

$$Z(\mathbf{X}) \quad = \quad \sum_{y'_K} \exp(a_u(y'_K))$$

$$\left(\sum_{y'_{K-1}} \exp(a_u(y'_{K-1}) + a_p(y'_{K-1}, y'_K)) \right.$$

$$\ldots$$

$\alpha_2(y'_3)$

$$\left(\sum_{y'_2} \exp(a_u(y'_2) + a_p(y'_2, y'_3)) \right.$$

$\alpha_1(y'_2)$

$$\left(\sum_{y'_1} \exp(a_u(y'_1) + a_p(y'_1, y'_2)) \right) \Bigg) \ldots \Bigg)$$

INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

$$Z(\mathbf{X}) = \sum_{y'_K} \exp(a_u(y'_K))$$

$$\alpha_{K-1}(y'_K) \quad \left(\sum_{y'_{K-1}} \exp(a_u(y'_{K-1}) + a_p(y'_{K-1}, y'_K)) \right.$$

$$\dots$$

$$\alpha_2(y'_3) \quad \left(\sum_{y'_2} \exp(a_u(y'_2) + a_p(y'_2, y'_3)) \right.$$

$$\alpha_1(y'_2) \quad \left(\sum_{y'_1} \exp(a_u(y'_1) + a_p(y'_1, y'_2)) \right) \Bigg) \dots \Bigg)$$

INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

$$Z(\mathbf{X}) \quad = \quad \sum_{y'_K} \exp(a_u(y'_K))$$

$$\alpha_{K-1}(y'_K) \quad \left(\sum_{y'_{K-1}} \exp(a_u(y'_{K-1}) + a_p(y'_{K-1}, y'_K)) \right.$$

$$\dots$$

$$\alpha_2(y'_3) \quad \left(\sum_{y'_2} \exp(a_u(y'_2) + a_p(y'_2, y'_3)) \right.$$

$$\alpha_1(y'_2) \quad \left(\sum_{y'_1} \exp(a_u(y'_1) + a_p(y'_1, y'_2)) \right) \Bigg) \dots \Bigg)$$

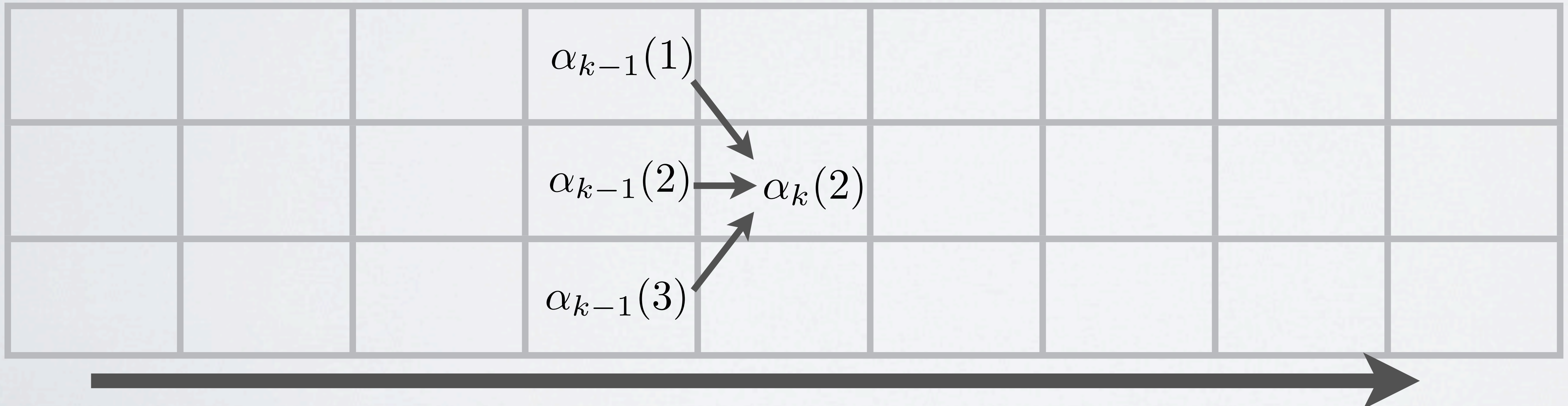
INFERENCE

nb. of classes

Topics: computing $p(\mathbf{y}|\mathbf{X})$

$\boxed{\text{Complexity in } \mathrm{O}(KC^2)}$

- Algorithm goes as follows:
  - initialize, for all values of y'_2 : $\alpha_1(y'_2) \Longleftarrow \sum_{y'_1} \exp(a_u(y'_1) + a_p(y'_1, y'_2))$
  - for $k = 2$ to K-1, for all values of y'_{k+1} :
    - $\alpha_k(y'_{k+1}) \Longleftarrow \sum_{y'_k} \exp(a_u(y'_k) + a_p(y'_k, y'_{k+1}))\ \alpha_{k-1}(y'_k)$
  - $Z(\mathbf{X}) \Longleftarrow \sum_{y'_K} \exp(a_u(y'_K))\ \alpha_{K-1}(y'_K)$



INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

$$
\begin{aligned}
Z(\mathbf{X}) \quad = \quad & \sum_{y'_1} \exp(a_u(y'_1)) \\
& \left(\sum_{y'_2} \exp(a_u(y'_2) + a_p(y'_1, y'_2)) \right. \\
& \qquad \dots \\
& \quad \left(\sum_{y'_{K-1}} \exp(a_u(y'_{K-1}) + a_p(y'_{K-2}, y'_{K-1})) \right. \\
& \qquad \left. \left. \left(\sum_{y'_K} \exp(a_u(y'_K) + a_p(y'_{K-1}, y'_K)) \right) \right) \dots \right)
\end{aligned}
$$

INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

$$
\begin{aligned}
Z(\mathbf{X}) \quad = \quad & \sum_{y'_1} \exp(a_u(y'_1)) \\
& \quad \left(\sum_{y'_2} \exp(a_u(y'_2) + a_p(y'_1, y'_2)) \right. \\
& \qquad \ldots \\
& \qquad \left(\sum_{y'_{K-1}} \exp(a_u(y'_{K-1}) + a_p(y'_{K-2}, y'_{K-1})) \right. \\
& \overbrace{\left(\sum_{y'_K} \exp(a_u(y'_K) + a_p(y'_{K-1}, y'_K)) \right)}^{\beta_K(y'_{K-1})} \left. \right) \ldots \left. \right)
\end{aligned}
$$

INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

$$Z(\mathbf{X}) = \sum_{y'_1} \exp(a_u(y'_1))$$

$$\left(\sum_{y'_2} \exp(a_u(y'_2) + a_p(y'_1, y'_2)) \right.$$

$$\dots$$

$$\underbrace{\left(\sum_{y'_{K-1}} \exp(a_u(y'_{K-1}) + a_p(y'_{K-2}, y'_{K-1})) \underbrace{\left(\sum_{y'_K} \exp(a_u(y'_K) + a_p(y'_{K-1}, y'_K)) \right)}_{\beta_K(y'_{K-1})} \right)}_{\beta_{K-1}(y'_{K-2})} \dots \Bigg)$$

INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

$$Z(\mathbf{X}) = \sum_{y'_1} \exp(a_u(y'_1))$$

$\beta_2(y'_1)$

$$\left(\sum_{y'_2} \exp(a_u(y'_2) + a_p(y'_1, y'_2)) \right.$$

$$\ldots$$

$\beta_{K-1}(y'_{K-2})$

$$\left(\sum_{y'_{K-1}} \exp(a_u(y'_{K-1}) + a_p(y'_{K-2}, y'_{K-1})) \right.$$

$\beta_K(y'_{K-1})$

$$\left. \left(\sum_{y'_K} \exp(a_u(y'_K) + a_p(y'_{K-1}, y'_K)) \right) \right) \ldots \right)$$

INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

$$Z(\mathbf{X}) \;=\; \sum_{y'_1} \exp(a_u(y'_1))$$

$\beta_2(y'_1)$

$$\left(\sum_{y'_2} \exp(a_u(y'_2) + a_p(y'_1, y'_2)) \right.$$

$$\dots$$

$\beta_{K-1}(y'_{K-2})$

$$\left(\sum_{y'_{K-1}} \exp(a_u(y'_{K-1}) + a_p(y'_{K-2}, y'_{K-1})) \right.$$

$\beta_K(y'_{K-1})$

$$\left.\left.\left(\sum_{y'_K} \exp(a_u(y'_K) + a_p(y'_{K-1}, y'_K)) \right) \right) \dots \right)$$

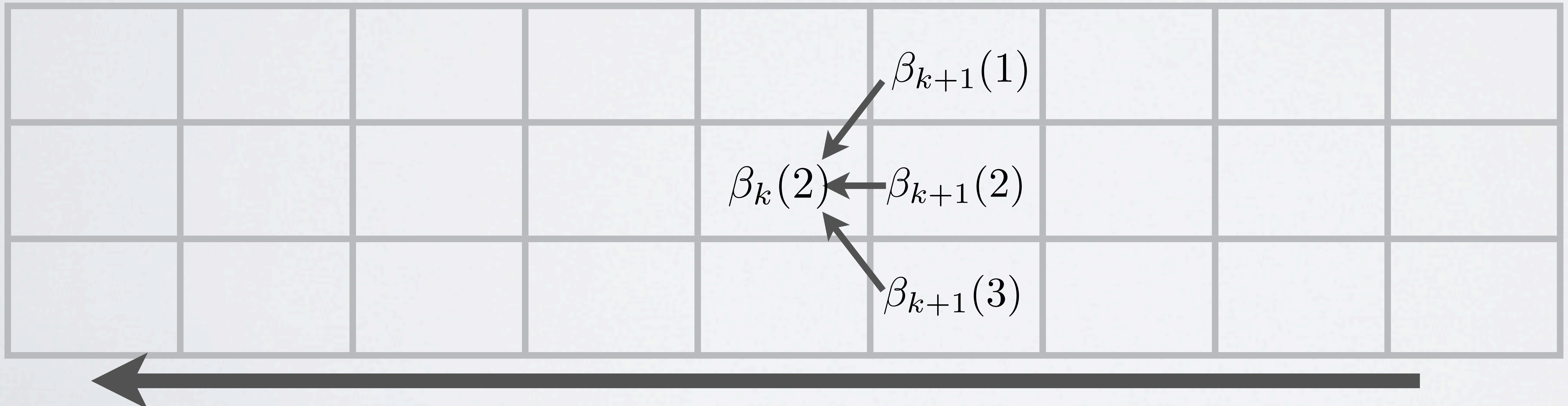
INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

nb. of classes

Complexity in $\mathrm{O}(KC^2)$

- Algorithm goes as follows:
  - initialize, for all values of y'_{K-1} : $\beta_K(y'_{K-1}) \Longleftarrow \sum_{y'_K} \exp(a_u(y'_K) + a_p(y'_{K-1}, y'_K))$
  - for $k = K$-1 to 2, for all values of y'_{k-1} :
    - $\beta_k(y'_{k-1}) \Longleftarrow \sum_{y'_k} \exp(a_u(y'_k) + a_p(y'_{k-1}, y'_k))\ \beta_{k+1}(y'_k)$
  - $Z(\mathbf{X}) \Longleftarrow \sum_{y'_1} \exp(a_u(y'_1))\beta_2(y'_1)$



INFERENCE

Topics: forward/backward or belief propagation

- Computing both tables is often referred to as the forward/ backward algorithm for CRFs
  - α is computed with a forward pass
  - β is computed with a backward pass
- It has other names
  - belief propagation
  - sum-product
- α gives the summation from the left
- β gives the summation from the right

INFERENCE

Topics: stable implementation of belief propagation

- For a stable implementation, should work in log space

$$\log \alpha_k(y'_{k+1}) \Longleftarrow \log \textstyle\sum_{y'_k} \exp\left(a_u(y'_k) + a_p(y'_k, y'_{k+1}) + \log \alpha_{k-1}(y'_k)\right)$$

$$\log \beta_k(y'_{k-1}) \Longleftarrow \log \textstyle\sum_{y'_k} \exp\left(a_u(y'_k) + a_p(y'_{k-1}, y'_k) + \log \beta_{k+1}(y'_k)\right)$$

- Log-sum-exp operations are more stable if computed like this:

$$\log \textstyle\sum_i \exp(z_i) = \underbrace{\max_i(z_i) + \log \textstyle\sum_i \exp(z_i - \max_i(z_i))}_{\text{numerically stable}}$$