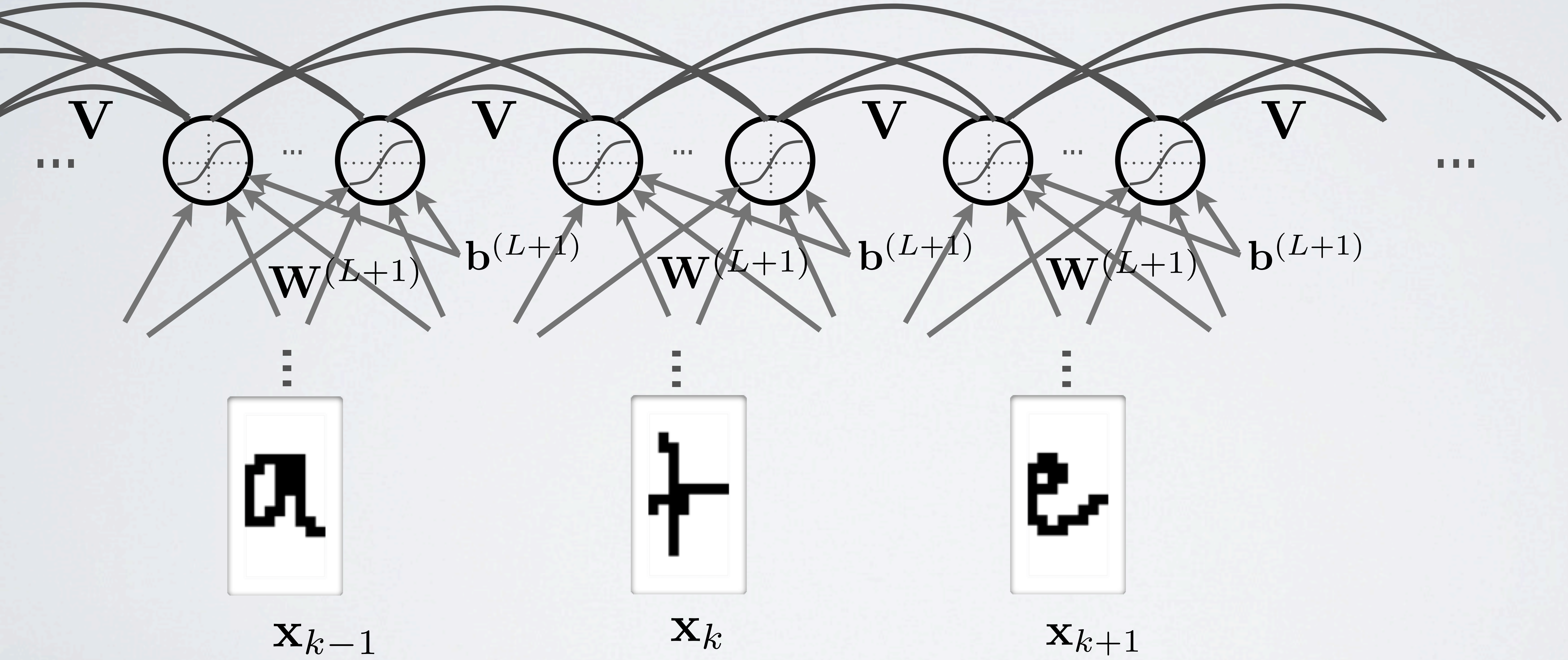
Neural networks

Conditional random fields - context window

LINEAR CHAIN CRF

Topics: lateral weights

- Sequence classification with linear chain:



LINEAR CHAIN CRF

Topics: context window

- Could incorporate a context window to the prediction at each position
  - e.g. context window of radius 1

$$p(\mathbf{y}|\mathbf{X}) \quad = \quad \exp\left(\sum_{k=1}^{K} a^{(L+1,0)}(\mathbf{x}_k)_{y_k} + \sum_{k=1}^{K-1} V_{y_k,y_{k+1}} + \right.$$

$$\left. \right) / Z(\mathbf{X})$$

LINEAR CHAIN CRF

Topics: context window

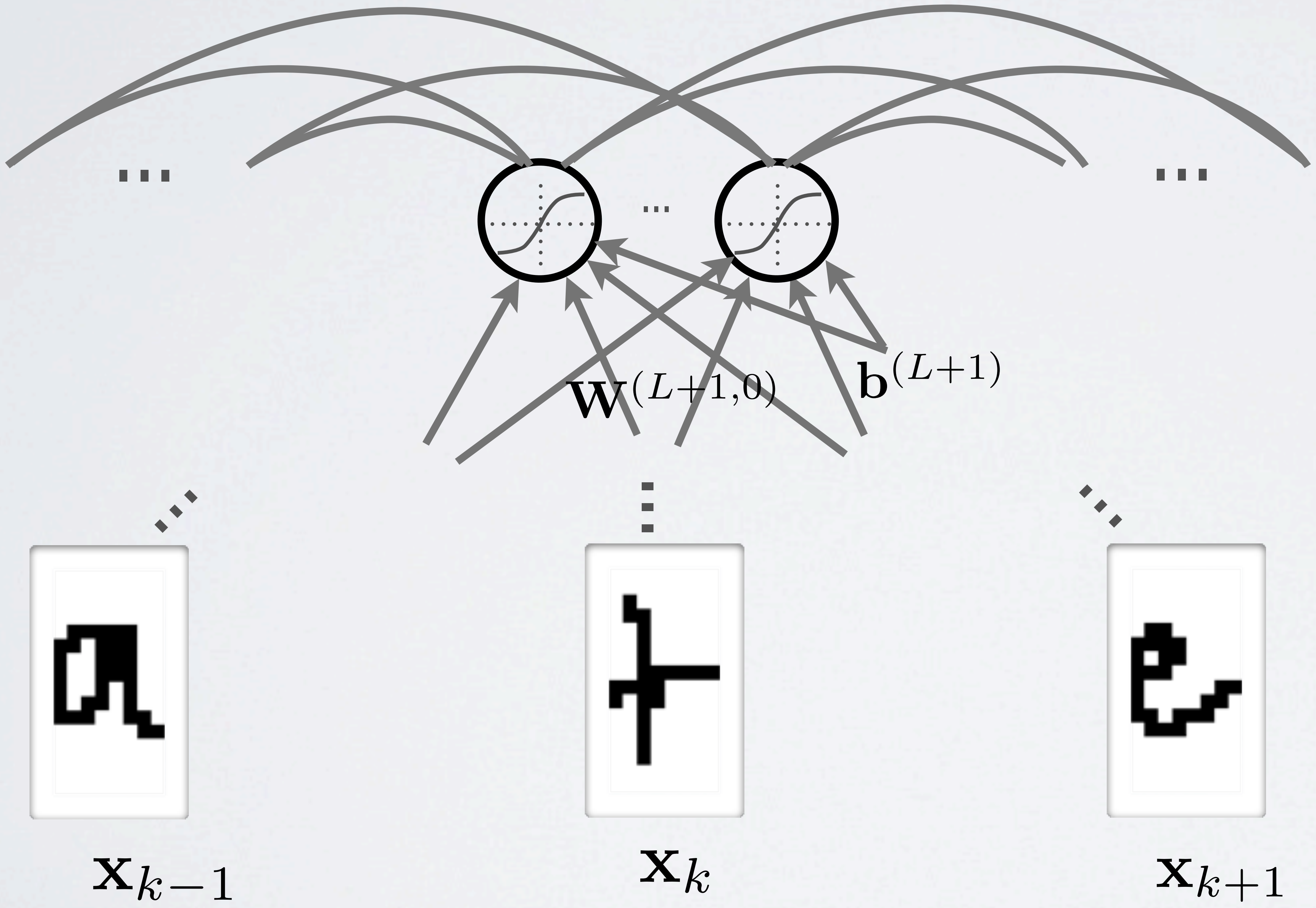
- Could incorporate a context window to the prediction at each position
  - e.g. context window of radius 1

$$p(\mathbf{y}|\mathbf{X}) = \exp\left(\sum_{k=1}^{K} a^{(L+1,0)}(\mathbf{x}_k)_{y_k} + \sum_{k=1}^{K-1} V_{y_k,y_{k+1}} + \underbrace{\sum_{k=2}^{K} a^{(L+1,-1)}(\mathbf{x}_{k-1})_{y_k}}_{\text{is } y_k \text{ likely given input on the left ?}} + \underbrace{\sum_{k=1}^{K-1} a^{(L+1,+1)}(\mathbf{x}_{k+1})_{y_k}}_{\text{is } y_k \text{ likely given input on the right ?}}\right) / Z(\mathbf{X})$$

LINEAR CHAIN CRF

Topics: context window

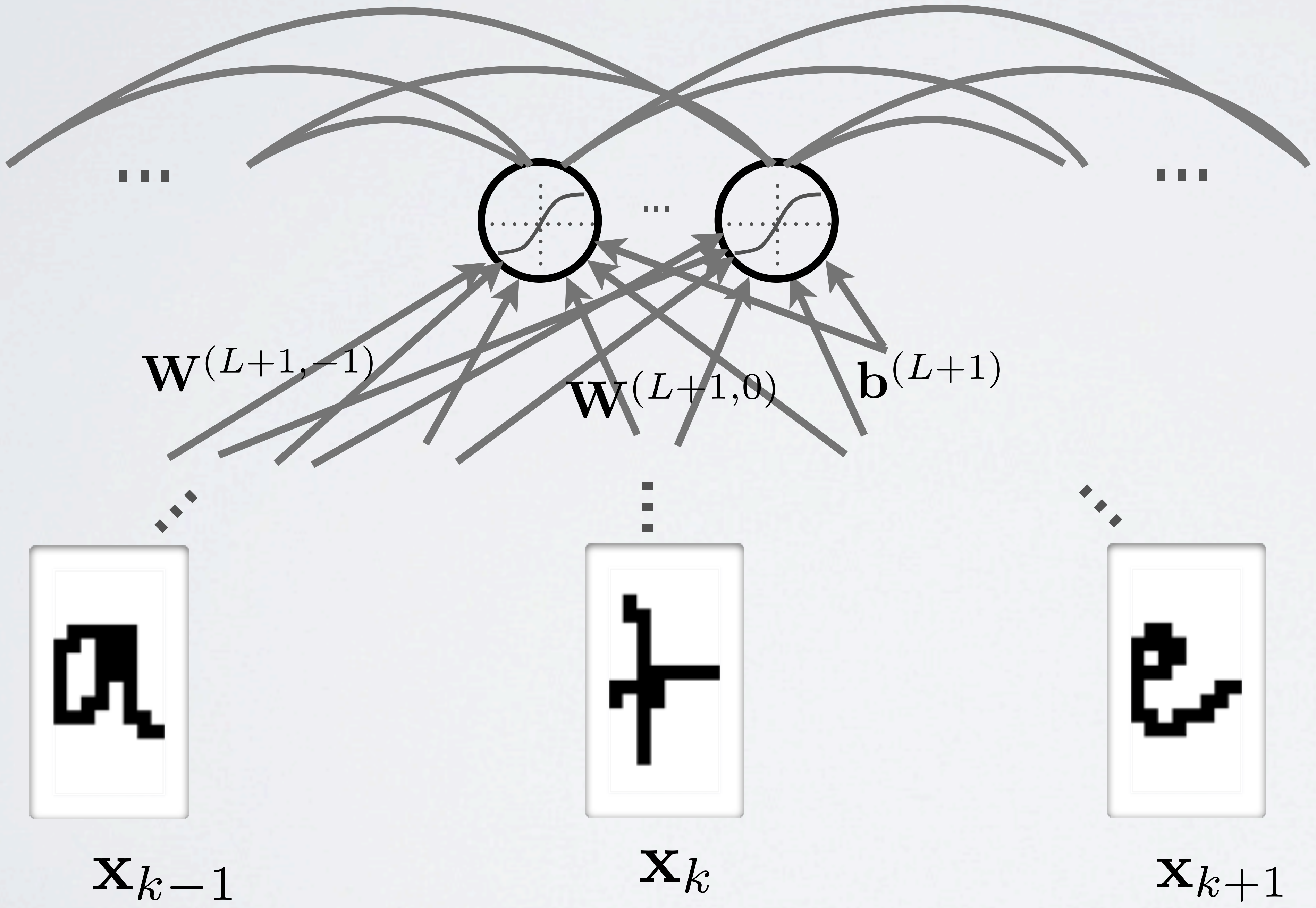
- Sequence classification with linear chain:



LINEAR CHAIN CRF

Topics: context window

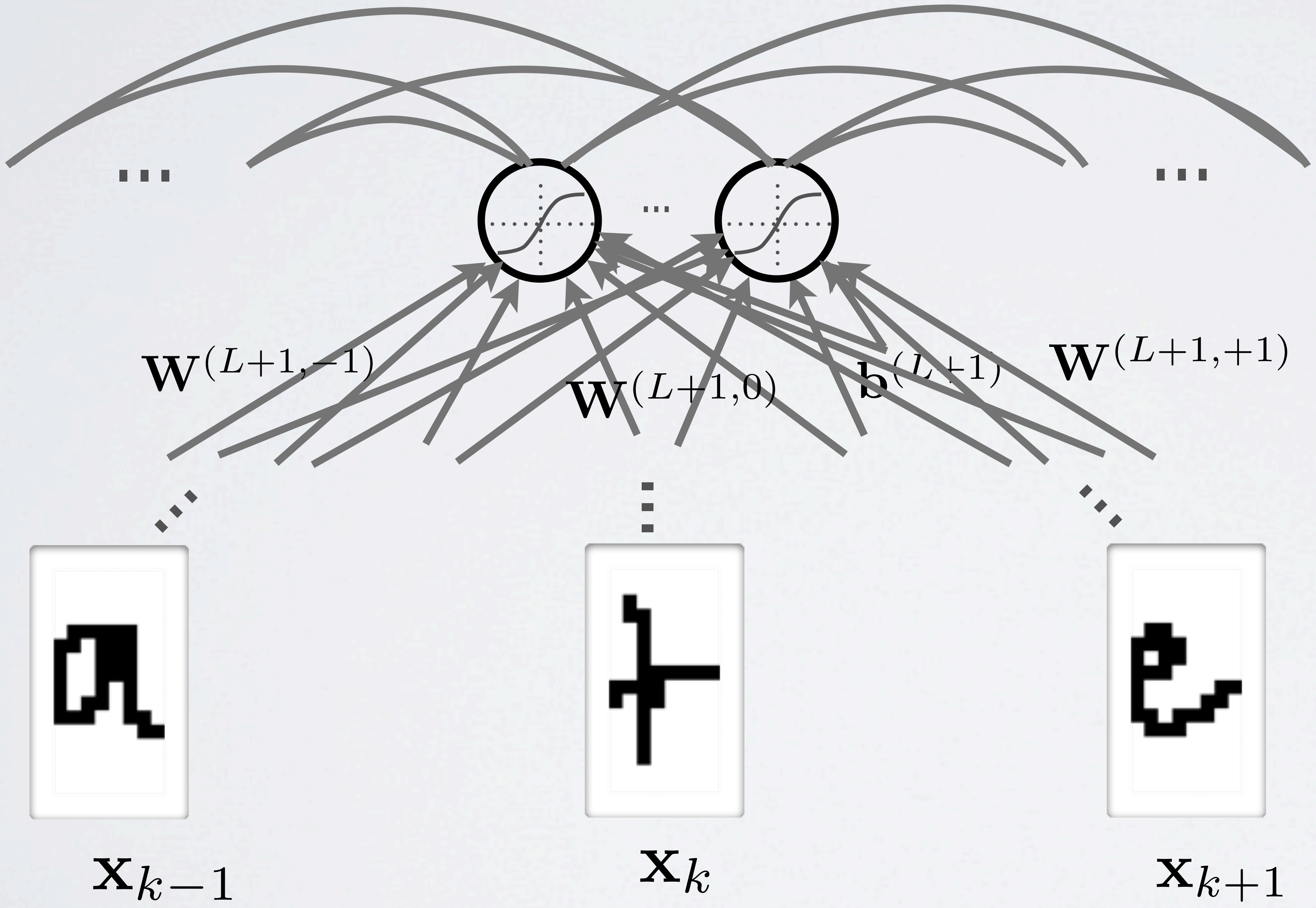
- Sequence classification with linear chain:



LINEAR CHAIN CRF

Topics: context window

- Sequence classification with linear chain:



LINEAR CHAIN CRF

Topics: context window

- Could instead feed the window to a single neural network
  - neural network can learn about the whole context jointly

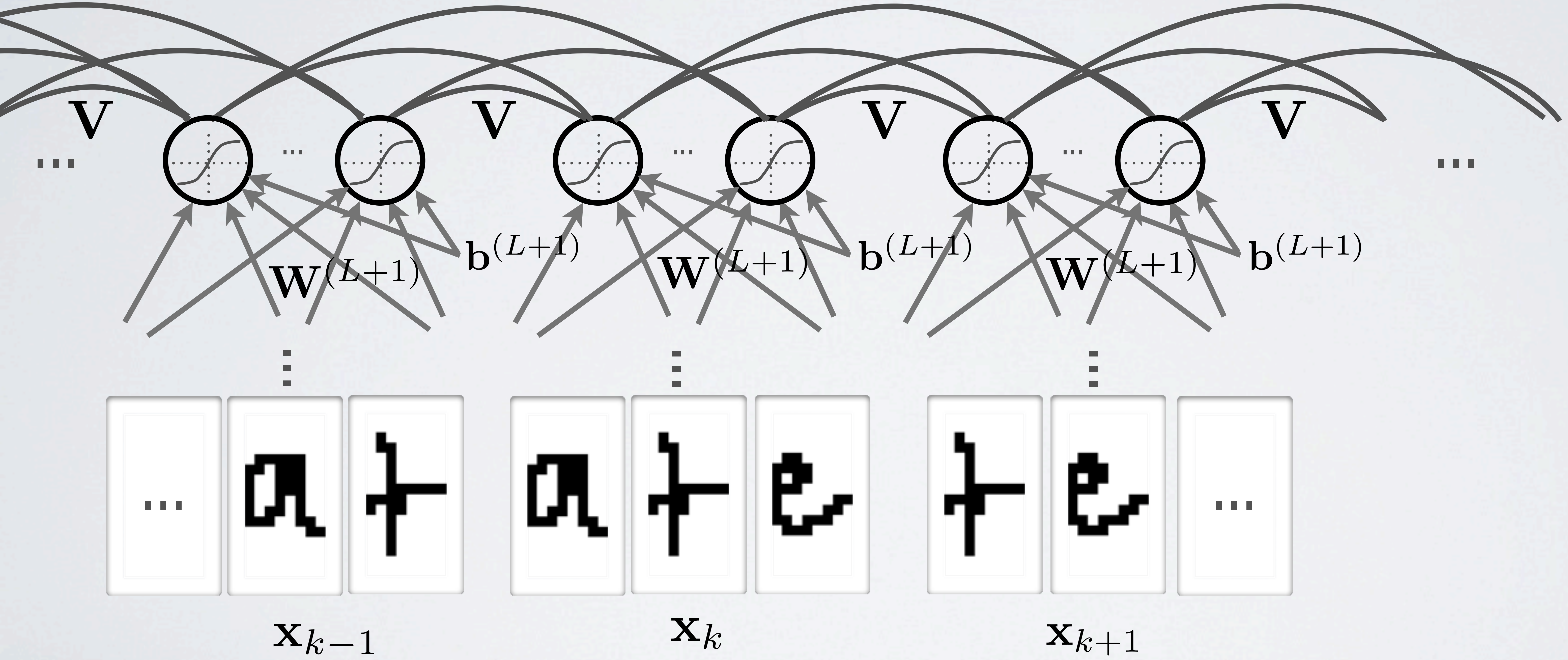
$$p(\mathbf{y}|\mathbf{X}) = \quad \exp\left(\sum_{k=1}^{K} a^{(L+1)}(\mathbf{x}_{k-1}, \mathbf{x}_k, \mathbf{x}_{k+1})_{y_k} + \sum_{k=1}^{K-1} V_{y_k, y_{k+1}}\right) / Z(\mathbf{X})$$

where $\mathbf{x}_0 = 0$ and $\mathbf{x}_{K+1} = 0$ (or some chosen special vectors that indicate beginning/end of sequences)

LINEAR CHAIN CRF

Topics: context window

- Sequence classification with linear chain:



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:

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