

Neural networks

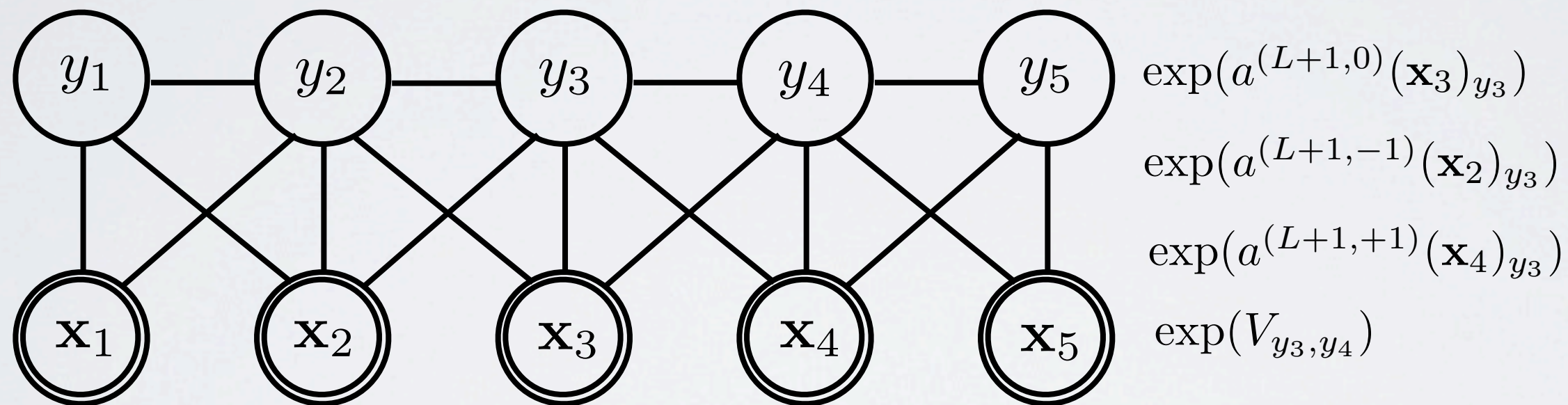
Training CRFs - maximum-entropy Markov model

LINEAR CHAIN CRF

Topics: Markov network

- Illustration for $K=5$

 = observed



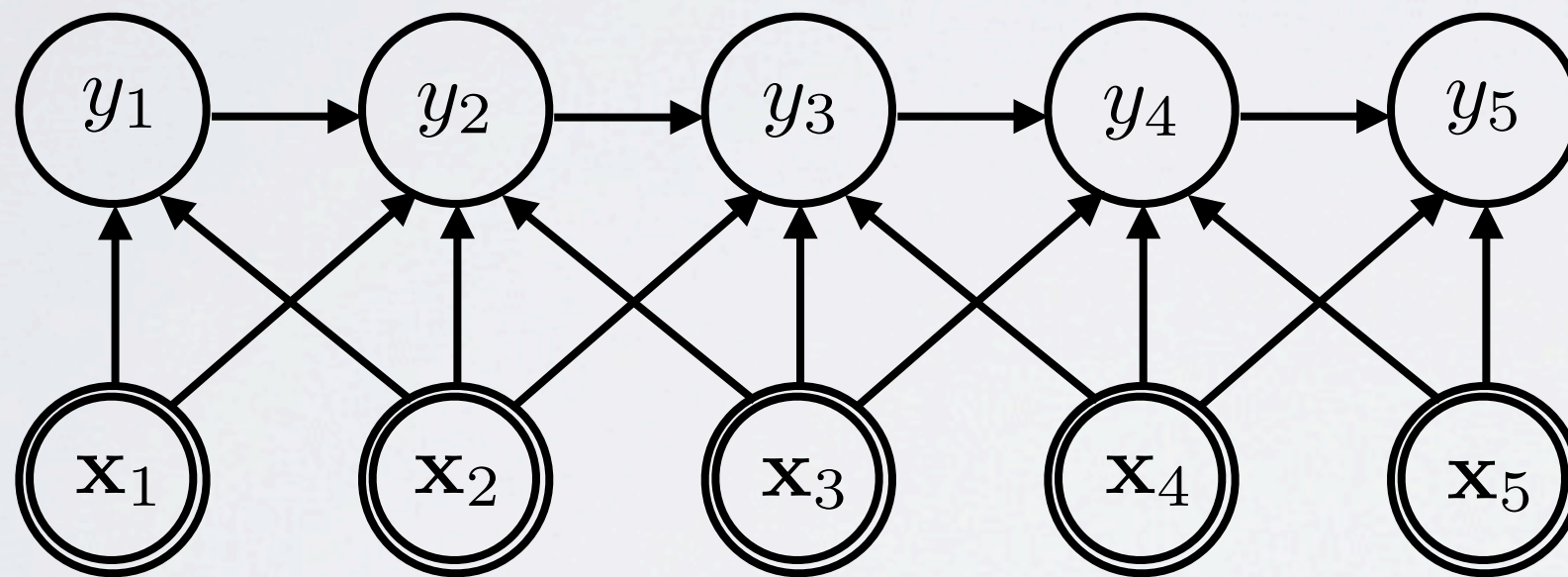
- Conditional random field are discriminatively trained, which should work better with more data
- Other alternative discriminatively trained sequence model?

MAXIMUM-ENTROPY MARKOV MODEL

Topics: MEMM

- MEMM is directed and discriminative:

 = observed



- it's a Markov model where the transition probabilities are given by logistic regressors (or neural networks):

- $p(\mathbf{y}|\mathbf{X}) = \prod_{k=1}^K p(y_k|y_{k-1}, \mathbf{X})$

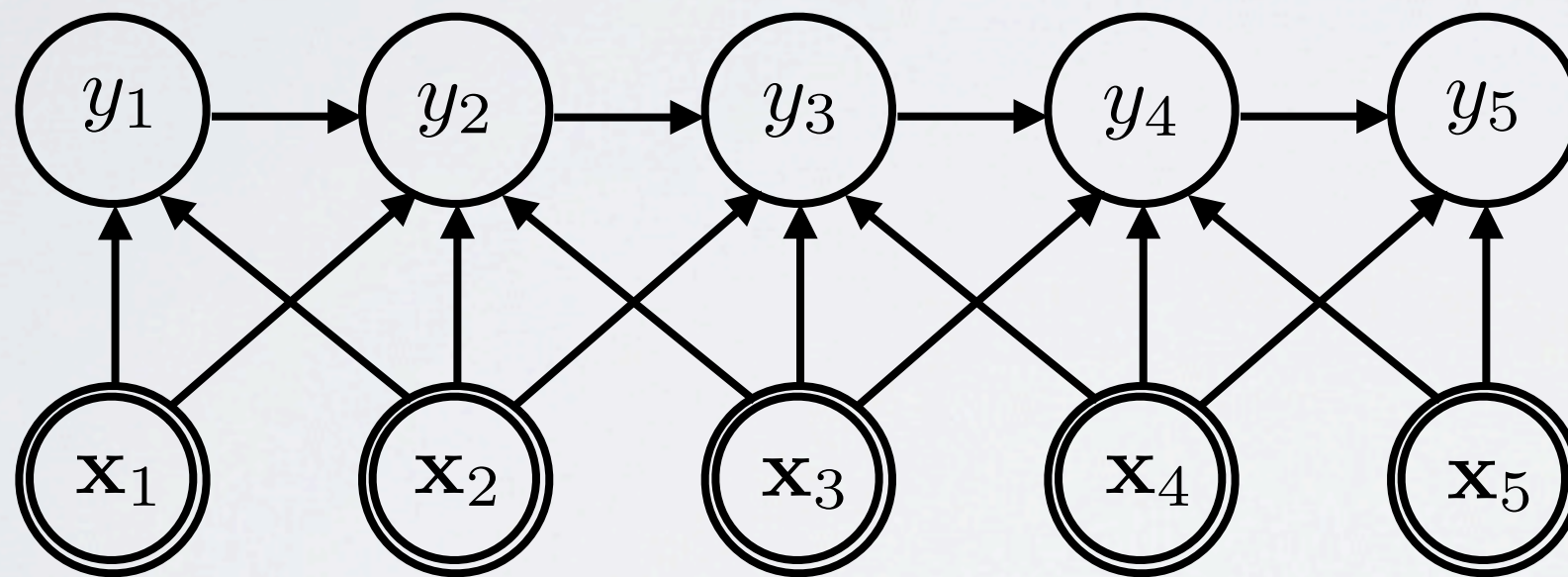
- $p(y_k|y_{k-1}, \mathbf{X}) = \frac{1}{Z(y_{k-1}, \mathbf{X})} \exp(a_u(y_k) + a_p(y_{k-1}, y_k))$

MAXIMUM-ENTROPY MARKOV MODEL

Topics: MEMM

- MEMM is directed and discriminative:

 = observed



- ▶ «label bias» problem: observations far away don't impact early predictions
 - example: $p(y_3|\mathbf{X}) = p(y_3|\mathbf{x}_1, \dots, \mathbf{x}_4)$
 - observations after \mathbf{x}_4 do not change our decision about y_3 !