

Neural networks

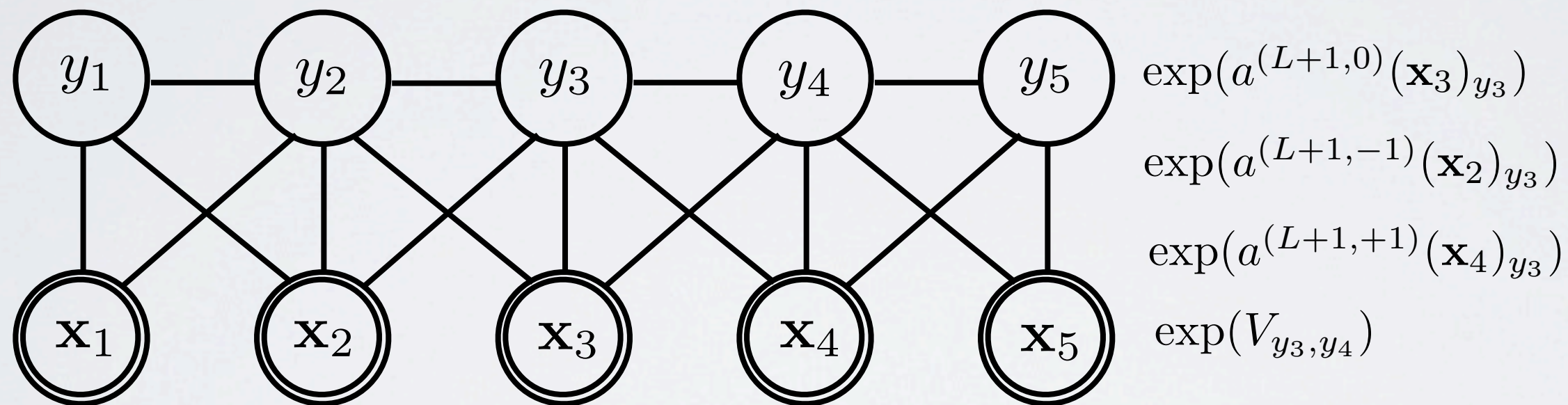
Training CRFs - hidden Markov model

LINEAR CHAIN CRF

Topics: Markov network

- Illustration for $K=5$

 = observed



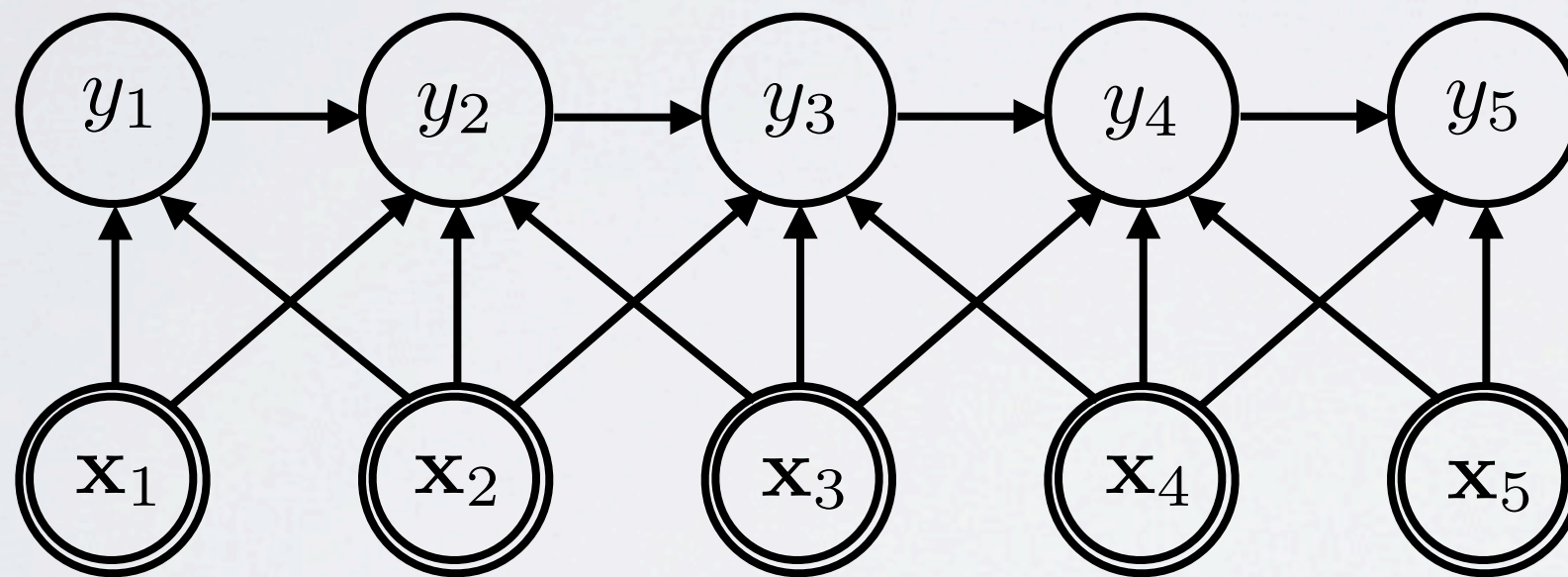
- Conditional random field are discriminatively trained, which should work better with more data
- Other alternative discriminatively trained sequence model?

MAXIMUM-ENTROPY MARKOV MODEL

Topics: MEMM

- MEMM is directed and discriminative:

 = observed



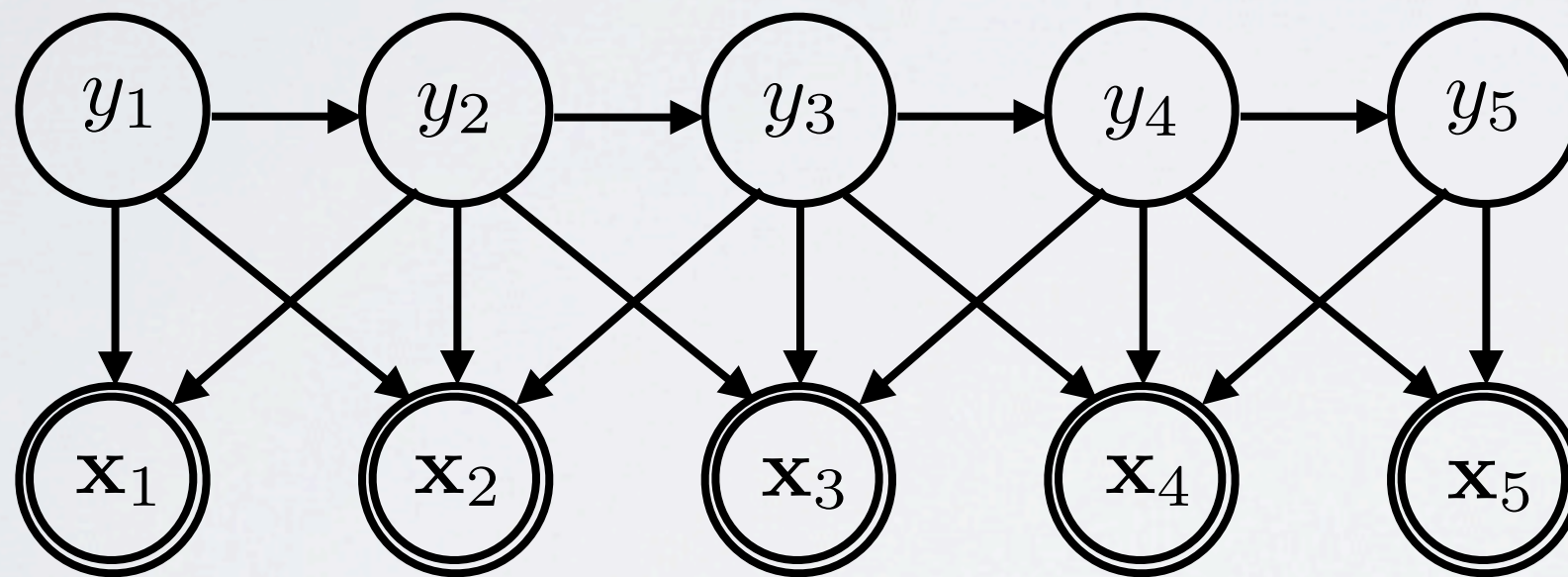
- ▶ «label bias» problem: observations far away don't impact early predictions
 - example: $p(y_3|\mathbf{X}) = p(y_3|\mathbf{x}_1, \dots, \mathbf{x}_4)$
 - observations after \mathbf{x}_4 do not change our decision about y_3 !

HIDDEN MARKOV MODEL

Topics: discriminative HMM

- HMMs can be trained discriminatively (i.e. minimize $-\log p(\mathbf{y}|\mathbf{X})$)

 = observed



- ▶ used a lot in speech recognition (called «maximum mutual information training»)
- ▶ we don't have the same label bias anymore
- ▶ however, optimization might be more complicated, since factors most correspond to normalized probabilities