Neural networks

Training CRFs - hidden Markov model
**Topics:** Markov network

- Illustration for $K=5$

- Conditional random field are discriminatively trained, which should work better with more data

- Other alternative discriminatively trained sequence model?
**Topics:** MEMM

- MEMM is directed and discriminative:

  - «label bias» problem: observations far away don’t impact early predictions
    - example: \( p(y_3 | X) = p(y_3 | x_1, \ldots, x_4) \)
    - observations after \( x_4 \) do not change our decision about \( y_3 \)!
**Topics:** discriminative HMM

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

- 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