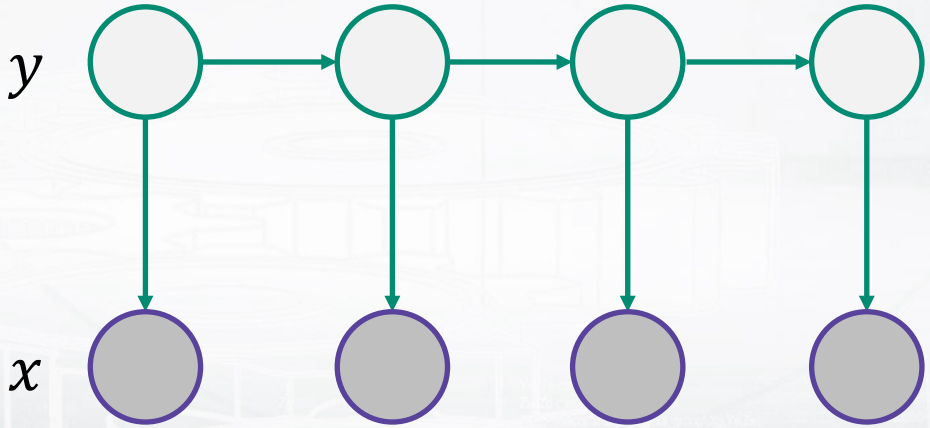
MEMMs, CRFs and other sequential models for Named Entity Recognition

Hidden Markov Model

$$p(\mathbf{y}, \mathbf{x}) = \prod_{t=1}^{T} p(y_t|y_{t-1})p(x_t|y_t)$$

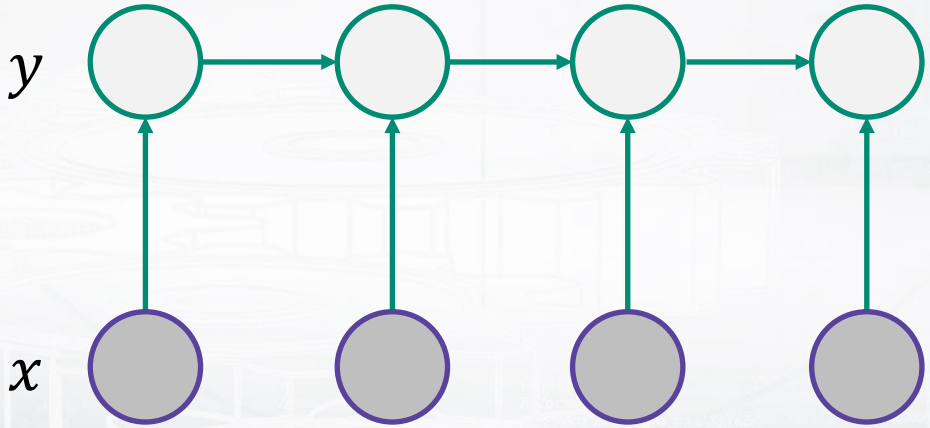
Generative model



Maximum Entropy Markov Model

$$p(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^{T} p(y_t|y_{t-1}, x_t)$$

Discriminative model



Maximum Entropy Markov Model

$$p(y_t|y_{t-1}, x_t) = \frac{1}{Z_t(y_{t-1}, x_t)} \exp\left(\sum_{k=1}^{K} \theta_k f_k(y_t, y_{t-1}, x_t)\right)$$

Normalization constant

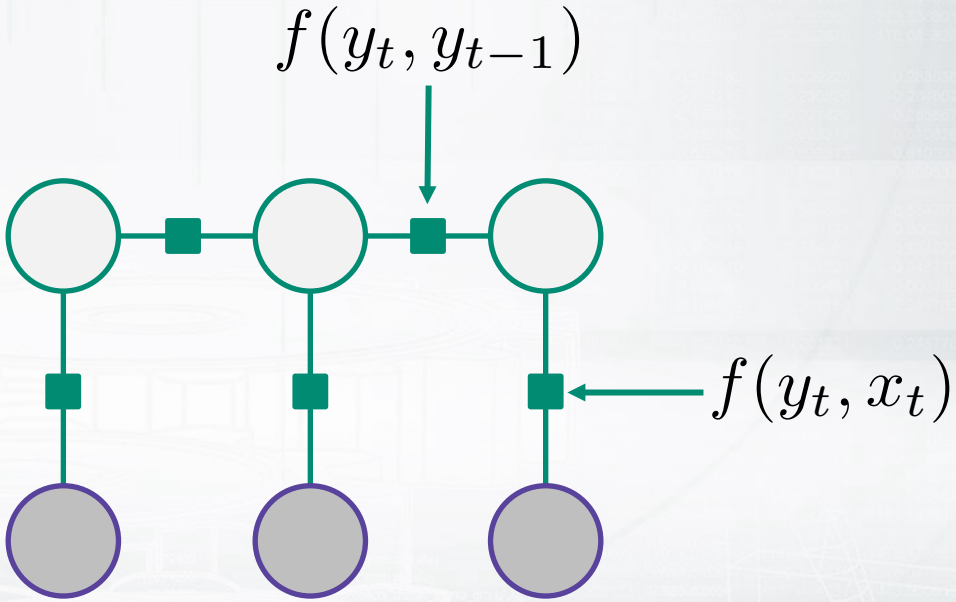
weight

feature

Conditional Random Field (linear chain)

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(x)} \prod_{t=1}^{T} \exp\left(\sum_{k=1}^{K} \theta_k f_k(y_t, y_{t-1}, x_t) \right)$$

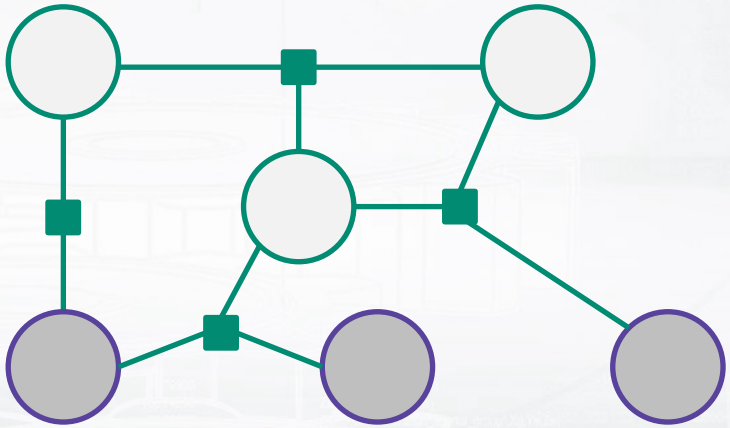
Undirected graph:



Conditional Random Field (general form)

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(x)} \prod_{a=1}^{A} \Psi_a(y_a, x_a)$$

Arbitrary factors



Black-box implementations

CRF++	https://sourceforge.net/projects/crfpp/
MALLET	http://mallet.cs.umass.edu/
GRMM	http://mallet.cs.umass.edu/grmm/
CRFSuite	http://www.chokkan.org/software/crfsuite/
FACTORIE	http://www.factorie.cc

http://homepages.inf.ed.ac.uk/csutton/publications/crftut-fnt.pdf

Features engineering

Label-observation features:

- $f(y_t, y_{t-1}, x_t) = [y_t = y]\, g_m(x_t)$
- $f(y_t, y_{t-1}, x_t) = [y_t = y][y_{t-1} = y']$
- $f(y_t, y_{t-1}, x_t) = [y_t = y][y_{t-1} = y']\, g_m(x_t)$

Observation function examples

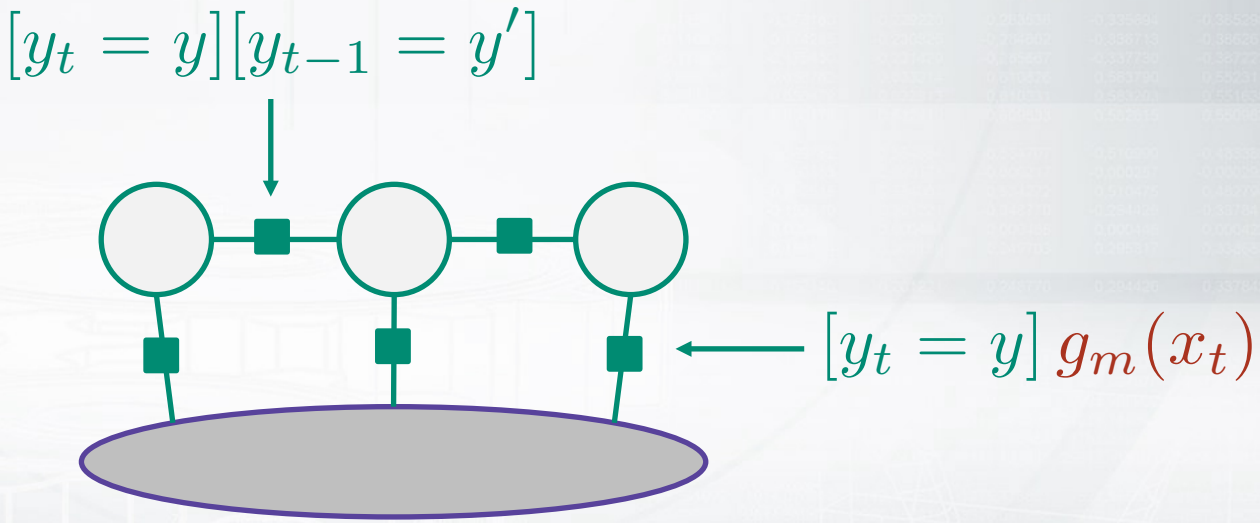
	$w_t = v$	$\forall v \in$
	part-of-speech tag for w_t is j	$\forall$ tags j
	w_t is in a phrase of syntactic type j	$\forall$ tags j
Capitalized	w_t matches [A-Z][a-z]+	
AllCaps	w_t matches [A-Z]+	
EndsInDot	w_t matches [^\.]+.*\.	
	w_t matches a dash	
	w_t appears in a list of stop words	
	w_t appears in list of capitals	

http://homepages.inf.ed.ac.uk/csutton/publications/crftut-fnt.pdf

Dependencies on input

Trick: Pretend the current input x_t contains not only the current word w_t, but also w_{t-1} and w_{t+1} and build observation functions for them as well.

The model is discriminative, so we can use the whole input:



Resume for the lesson

Probabilistic graphical models:

- Hidden Markov Models (generative, directed)
- Maximum Entropy Markov Models (discriminative, directed)
- Conditional Random Field (discriminative, undirected)

Tasks:

- Training – fit parameters (Baum-Welch for HMM)
- Decoding – find the most probable tags (Viterbi for HMM)

Practice:

- Features engineering + black-box implementation