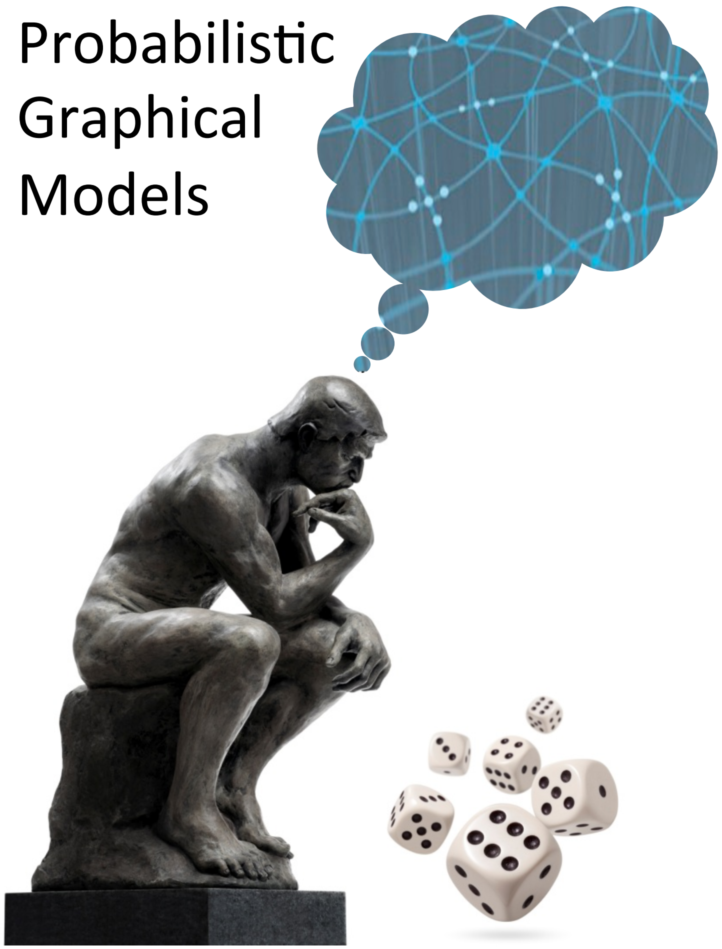
Probabilistic
Graphical
Models



Representation

Markov Networks

Conditional Random Fields

Motivation

- Observed variables **X**
- Target variables **Y**

$X \longrightarrow Y$

$P(X,Y)$ joint

$P(Y|X)$ conditional

	X	Y
image segmentation	pixel values	pixel labels
text processing	words in a sentence	parts of speech

CRF Representation

$$\phi_1(\boldsymbol{D}_1), \ldots, \phi_k(\boldsymbol{D}_k)$$

$\tilde{P}(\bar{X}, \bar{Y}) = \prod_{i=1}^{k} \phi_i(\bar{D}_i)$ unnormalized measure

$Z(\bar{X}) = \sum_{\bar{Y}} \tilde{P}(\bar{X}, \bar{Y})$ different $Z(\bar{X})$ for every assignment $\bar{x}$ to the obs. variables $\bar{X}$

$P(\bar{Y} \mid \bar{X}) = \frac{1}{Z(\bar{X})} \tilde{P}(\bar{X}, \bar{Y})$ $\left(\sum_{\bar{Y}} P(\bar{Y} \mid \bar{X}) = 1 \text{ for all } \bar{X} \right)$

CRFs and Logistic Model

$$\phi_i(X_i, Y) = \exp\{w_i \mathbf{1}\{X_i = 1, Y = 1\}\}$$

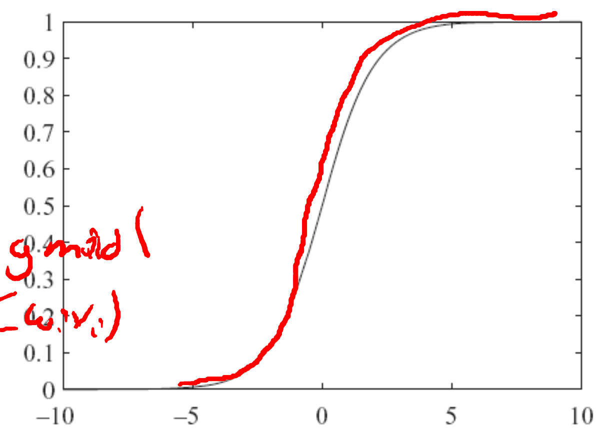
$\phi_i(x_i, Y=1) = \exp\{w_i x_i\}$

$\phi_i(x_i, Y=0) = 1$

$\tilde{P}(Y=1, x_1, \ldots, x_n) = \exp(\sum_i w_i x_i)$

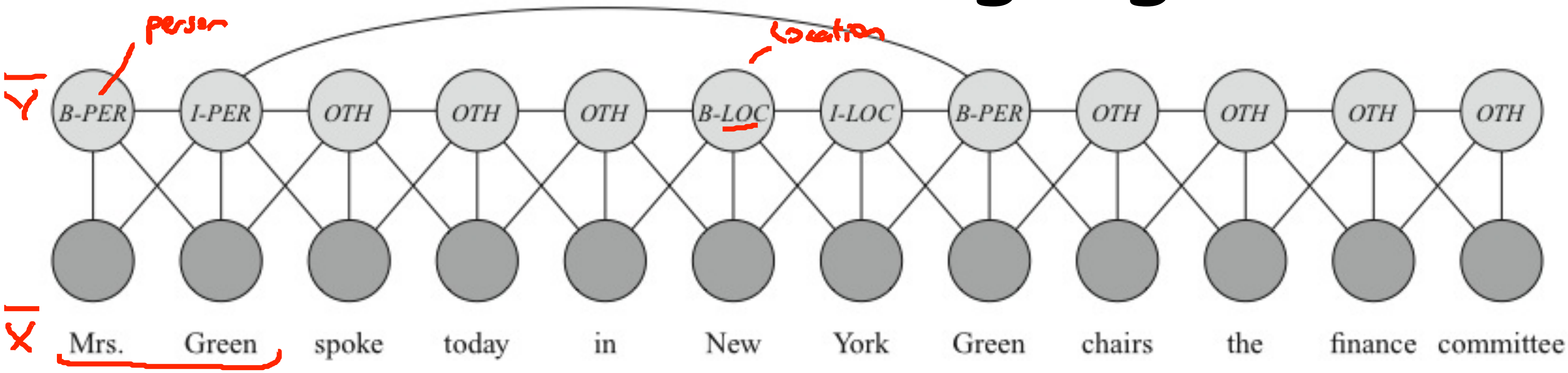
$\tilde{P}(Y=0, x_1, \ldots, x_n) = 1$

$P(Y=1 \mid x_1, \ldots, x_n) = \dfrac{\exp(\sum_i w_i x_i)}{1 + \exp(\sum_i w_i x_i)} = \text{sigmoid}(\sum w_i x_i)$



Daphne Koller

CRFs for Language



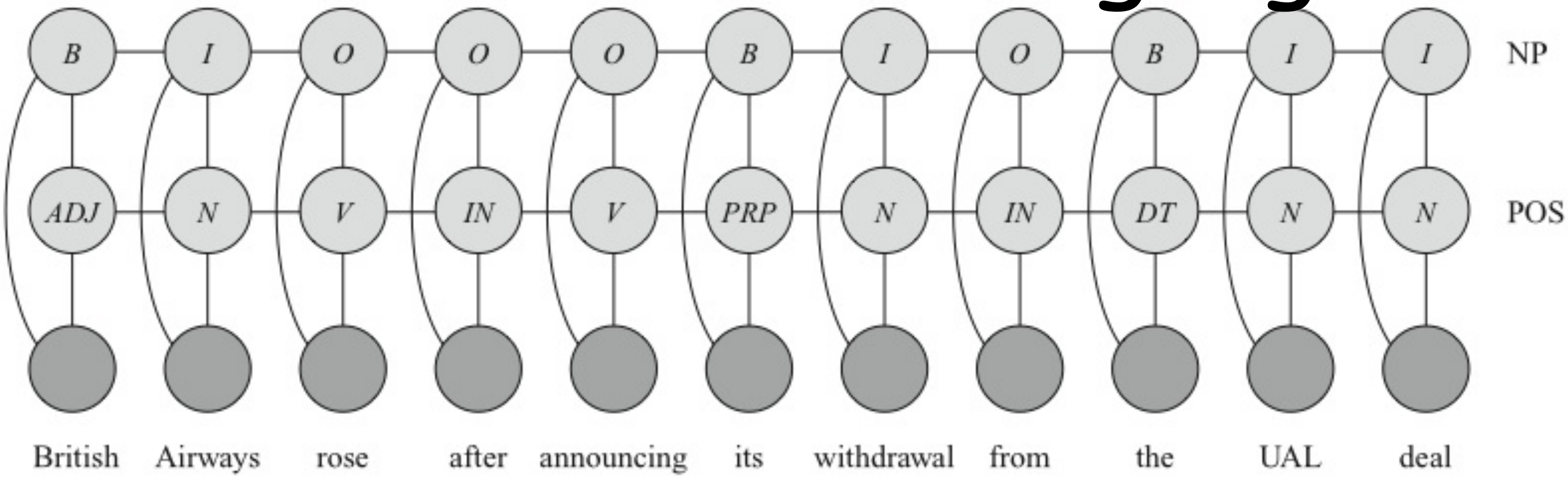
Features: word capitalized, word in atlas or name list, previous word is "Mrs", next word is "Times", …

$\tilde{P}(\bar{X},\bar{Y})$

$\Rightarrow P(\bar{Y}|\bar{X})$

Daphne Koller

More CRFs for Language



KEY

B	Begin noun phrase	*V*	Verb
I	Within noun phrase	*IN*	Preposition
O	Not a noun phrase	*PRP*	Possesive pronoun
N	Noun	*DT*	Determiner (e.g., a, an, the)
ADJ	Adiective		

Summary

- A CRF is parameterized the same as a Gibbs distribution, but normalized differently
- Don't need to model distribution over variables we don't care about
- Allows models with highly expressive features, without worrying about wrong independencies