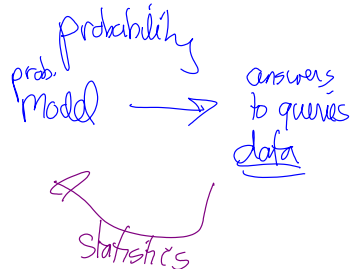
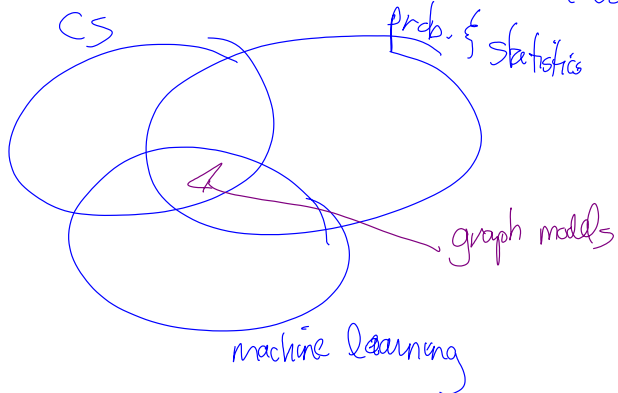


email list sign-up: [bit.ly/IFT6269-A17](http://bit.ly/IFT6269-A17)  
fill this by tomorrow? ↑  
hyper

### Probabilistic graphical model

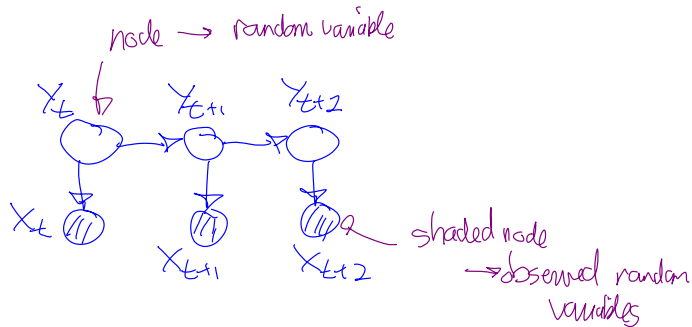
model multivariate data

graphical model → mix of graph theory + prob. theory  
(CS)



### Applications

hidden Markov model  
(HMM)

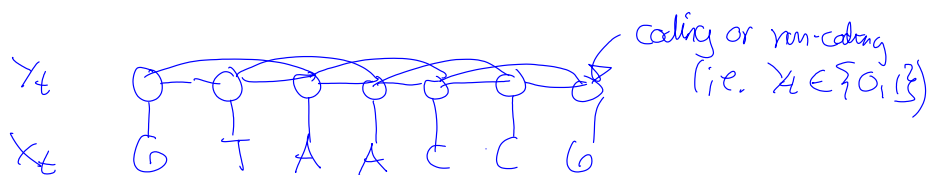


a) speech recognition:  $Y_t \rightarrow$  phoneme  
 $X_t \rightarrow$  sound wave

b) part-of-speech tagging  $\{ \text{DT} \rightarrow \text{V} \rightarrow \text{DT} \rightarrow \text{ADJ} \rightarrow \text{N} \}$   $Y_t$ : part of speech

$x_t$ : 'word'  $x_t$ : This is a red box

c) gene finding



d) control

$$y_{t+1} = A y_t + B v_t + \epsilon_t$$

latent state matrix
control term (given to system)
noise

$x_t = C y_t + \epsilon'_t$  here  $y_t$  &  $x_t$  are continuous vectors  
 if noise is Gaussian, HMM  $\rightarrow$  is Kalman filter

Why graphical models?

pos tagging observation:  $(x_1, \dots, x_T) \triangleq x_{1:T}$

$x_t \in \{1, \dots, K\}$   
 size of vocabulary

want to model  $p(x_{1:T}) \rightarrow$  exponential size state space

$\approx K^T$  parameters  
 to define a distribution on  $x_{1:T}$

trick: make a factorization assumptions about  $p$

$$p(x_1, \dots, x_T) = f_1(x_1) \underbrace{f_2(x_2|x_1)}_{\substack{\text{factor} \\ \rightarrow \text{choice in } M_0}} f_3(x_3|x_2) \dots \underbrace{f_T(x_T|x_{T-1})}_{\substack{\text{needs } \approx K^2 \text{ numbers} \\ \text{to specify}}}$$

T factors

→ clique in my graph model

T factors to specify

Go from  $\downarrow$  parameters to  $\approx T \cdot K^2$

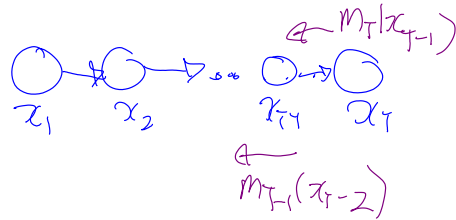
say want to compute

$$p(x_1) = \sum_{x_2, \dots, x_T} p(x_{1:T})$$

$$= \sum_{x_2, \dots, x_T} f_1(x_1) f_2(x_2|x_1) \dots f_T(x_T|x_{T-1})$$

$$= f_1(x_1) \left( \sum_{x_2} f_2(x_2|x_1) \left( \sum_{x_3} f_3(x_3|x_2) \dots \left( \sum_{x_T} f_T(x_T|x_{T-1}) \right) \dots \right) \right)$$

$M_T(x_{T-1})$



"message passing algorithm" to compute efficiently marginal  $p(x_1)$

Key themes:

representation: how to represent structured prob. distributions?  
semantic?

→ parametrization

full table  
parametrization  
exponential family

estimation: given data, how do we learn the parameters of distribution

→ learning

e.g.  
maximum likelihood

inference: answer questions about data

→ computation

eg computing  $p(y|x)$ , or  $p(x|y)$   
conditional marginal

eg message  
passing