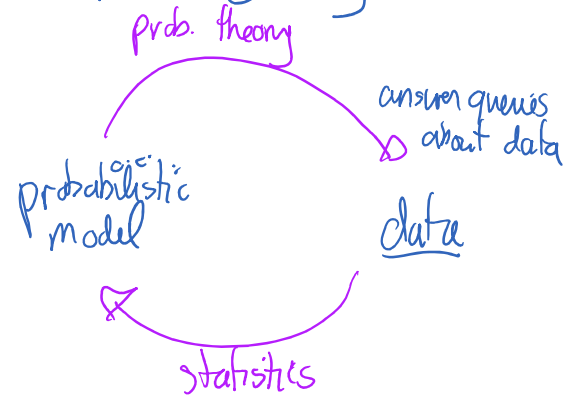
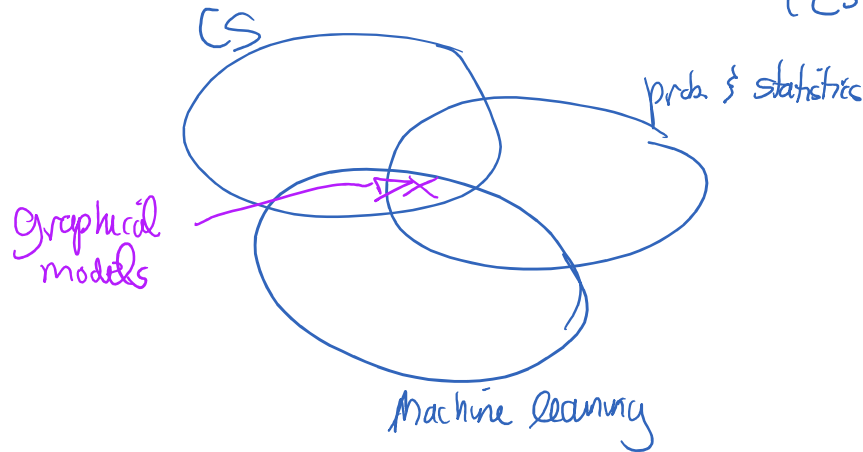


email sign-up by tomorrow: bit.ly/IFT6269-F19

probabilistic graphical model :

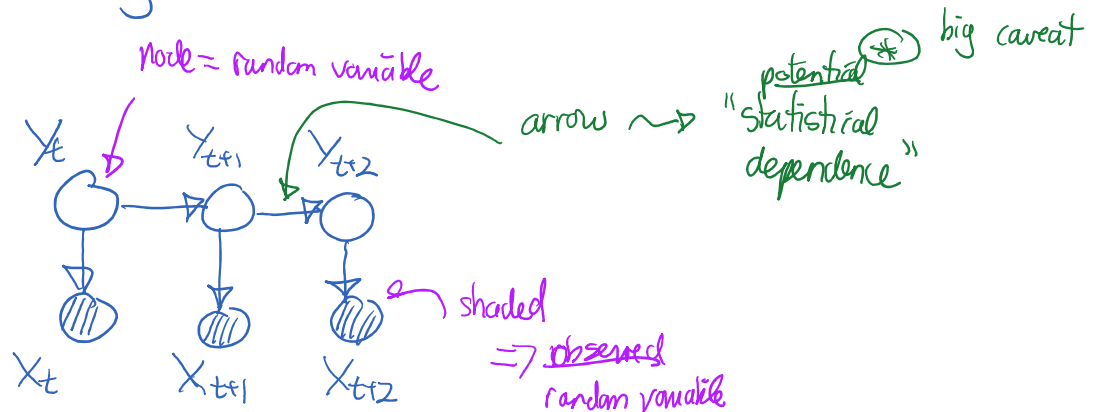
model multivariate data

graphical model \rightarrow mix of graph theory (CS) + probability theory



Applications :

hidden Markov model



a) speech recognition : $Y_t \rightarrow$ "phonemes"

$X_t \rightarrow$ sound waves

b) part-of-speech tagging :

DT — V — DT — ADJ — N
 | | | | |
 This is a red box
 X_1 X_2 X_3 ..

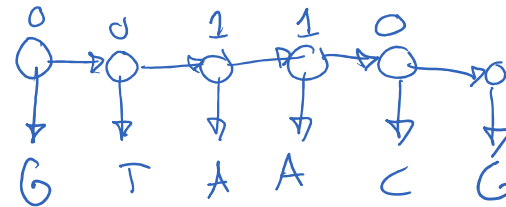
Y_t : part-of-speech

X_t : words

c) gene finding :

Y_t : coding vs non-coding region $Y_t \in \{0,1\}$

$X_t \rightarrow$ DNA



d) control

$$y_{t+1} = Ay_t + Bv_t + \epsilon_t$$

(normally: multiple molecules per X_t)

matrices

control term

noise

here x_t & y_t are cts. vectors

$$x_t = Cy_t + \epsilon_t'$$

if ϵ_t & ϵ_t' is Gaussian

HMM \rightarrow "Kalman filter"

Why graphical models?

Pos tagging

Observation $(x_1, x_2, \dots, x_T) \equiv x_{1:T}$

$x_{\{1,3,5\}}$

$$x_t \in \{1, \dots, k\}$$

↑
size of vocab.

↓
 (x_1, x_3, x_5)

want to model $p(x_{1:T})$

issue: exponential size state space

$\Rightarrow k^T - 1$ parameters needed to specify dist.

trick: make a factorization assumption about p

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

factor \rightarrow 2 variables
 $\Rightarrow \sim k^2$ parameters to specify

clique
in graphical model

T factors \Rightarrow $T \cdot k^2$ parameters
 $\ll k^T$

theme: representation

computation?

say want to compute $p(x_1)$ "marginal"

$$\sum_{x_2 \in \{1, \dots, k\}} \sum_{x_3} \dots \sum_{x_T}$$

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

exponential sum!!

distributivity
 $a \cdot (b+c) \equiv a \cdot b + a \cdot c$

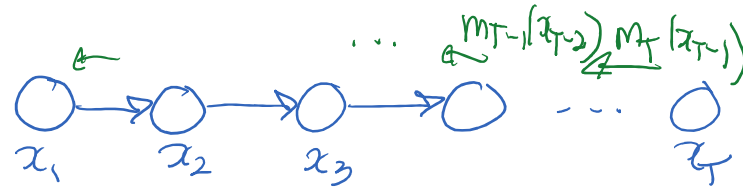
1. $\overline{x_2, \dots, x_T}$ exponential sum?!

distributivity
 $a \cdot (b+c) = a \cdot b + a \cdot c$

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

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

$M_T(x_{T-1})$
 \downarrow
 $O(k^2)$



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