Lecture 1 - scribbles

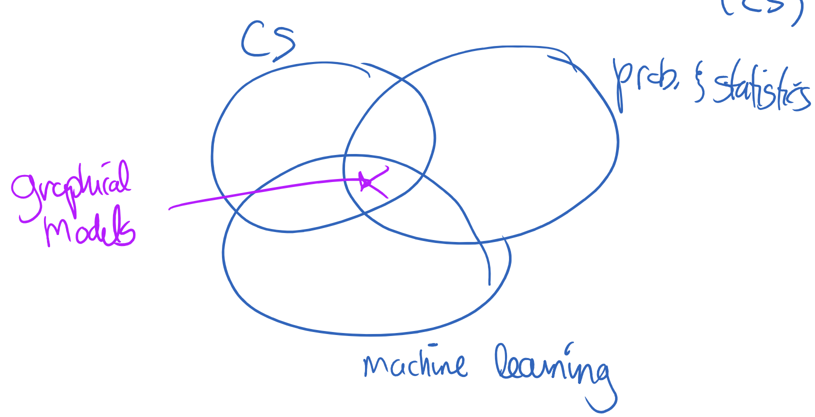
Tuesday, September 4, 2018 14:48

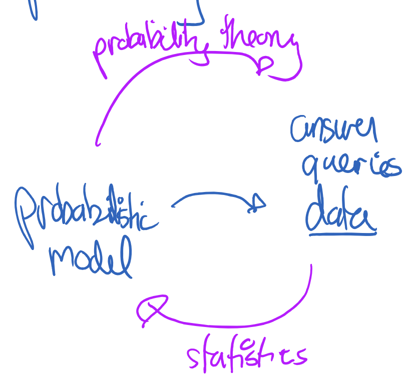
email sign-up by tomorrow : bit.ly/IFT6269-F18

probabilistic graphical model

model multivariate data

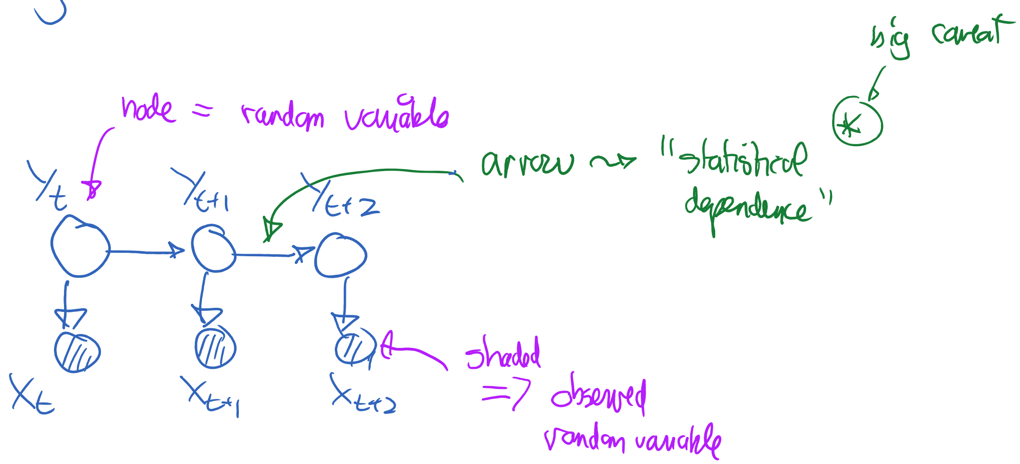
graphical model → mix of graph theory (CS) + prob. theory





Applications:

hidden Markov model



a) speech recognition: $Y_t \to$ phoneme

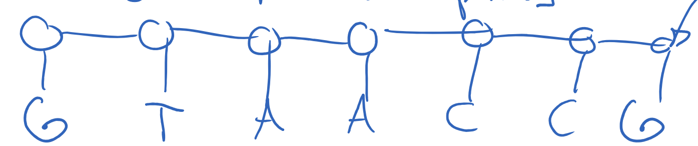
$X_t \to$ sound wave

b) part-of-speech tagging: Y_t's: DT — V — DT — ADJ — N $\quad Y_t$: part-of-speech

This is a red box $\quad X_t$: words

(*) [normally → multiple nucleotides per X_t]

c) Gene finding $\quad Y_t$ $\quad X_t$:



coding vs non-coding (i.e. $Y_t \in \{0,1\}$)

d) <u>control</u>

$$y_{t+1} = \underbrace{A}_{\text{matrices}} \underbrace{y_t}_{\text{latent state}} + \underbrace{B}_{\text{matrices}} \underbrace{V_t}_{\text{control term (given to system)}} + \underbrace{\varepsilon_t}_{\text{noise}}$$

$$x_t = C y_t + \varepsilon_t'$$

here x_t & y_t are cts. vectors

if noise is Gaussian, HMM ⇝ Kalman filter

Why graphical models?

POS tagging $\quad$ observation $(x_1, \ldots, x_T) \triangleq x_{1:T}$

$x_t \in \{1, \ldots, K\}$

size of vocabulary

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

issue: exponential size state space

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

trick: make a factorization assumption about p

$$p(x_1, \ldots, x_T) = f_1(x_1) \underbrace{f_2(x_2|x_1)}_{\text{factor}} f_3(x_3|x_2) \cdots f_T(x_T|x_{T-1})$$

factor $\rightarrow$ 2 variables

$\approx K^2$ parameters

clique in graphical model

T factors $\Rightarrow$ $\boxed{T \cdot K^2 \text{ parameters}}$ $\ll K^T$

theme: representation

computation:

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

distributivity:

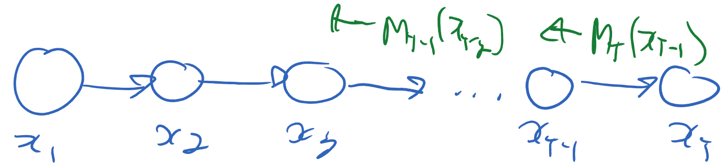
$a \cdot (b+c) = a \cdot b + a \cdot c$

$$p(x_1) = \sum_{x_2, \ldots, x_T} p(x_{1:T})$$ ← exponential sum??

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

$x_2,\dots,x_T$

$$= f_1(x_1)\left(\sum_{x_2} f_2(x_2|x_1)\left[\sum_{x_3} f_3(x_3|x_2)\cdots\left(\underbrace{\sum_{x_T} f(x_T|x_{T-1})}_{m_T(x_{T-1})}\right)\cdots\right]\right)$$



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

Key themes:

representation: how to represent structured prob distributions? → parameterization
semantic?
full table paramet.
or
"exponential family"

estimation: given data, how do learn/estimate the parameters of dist. → learning e.g. maximum likelihood

inference: answer questions about data → computation
e.g. message passing

e.g. computing $p(y|x)$, or $p(z)$
conditional marginal

also approximate inference

conditional   marginal   also approximate inference
e.g. sampling
variational methods