Lecture 1 - intro

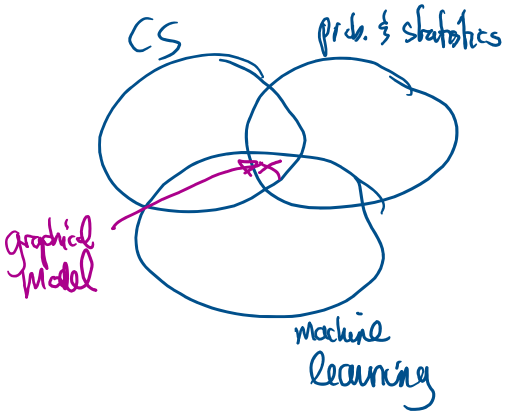
Thursday, September 5, 2024 2:19 PM

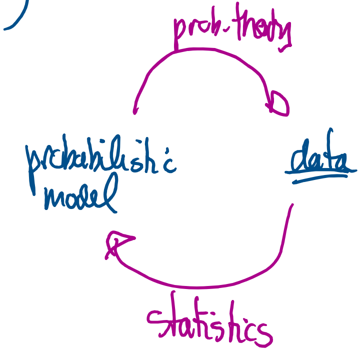
email sign-up by <u>tomorrow 2pm</u> : bit.ly/IFT6269-F24

<u>probabilistic graphical model</u>

model multivariate data

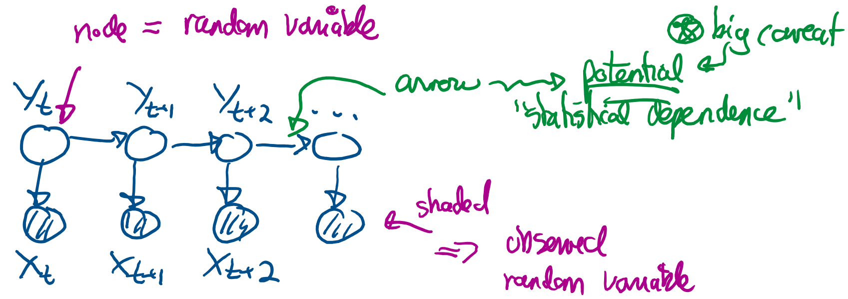
graphical model → mix of graph theory (CS) + probability theory





<u>Applications</u> :

hidden Markov model (HMM)



a) speech recognition : X_t → sound waves (observed)

Y_t → phonemes (latent)

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 X_t → DNA

Y_t → coding vs. non-coding $Y_t \in \{0,1\}$

d) control : $y_{t+1} = A y_t + B v_t + \varepsilon_t$ noise here x_t & y_t are cts. vectors

matrices control term

$x_t = C y_t + \varepsilon_t'$ if ε_t & ε_t' are Gaussian

sensor reading

HMM → "Kalman filter"

e) "GPT" RNN $y_t \to$ "h_t" latent

$x_t \to$ words

15h35

why graphical models?

POS tagging observations $(x_1, x_2, \dots, x_T) \triangleq x_{1:T}$

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

want to model $p(x_{1:T})$ Issue: size of state space is exponential in length of input

$\Rightarrow K^T - 1$ parameters needed to fully describe distribution

trick: make a factorization assumption about p

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

factors → 2 variables $\Rightarrow K^2$ parameters

T factors $\Rightarrow \approx T \cdot K^2$ parameters $<< K^T$

clique in a graph (graphical model) ⤳ theme: representation

computation?: say want to compute $p(x_1)$ "marginal"

$$p(x_1) = \sum_{x_2, x_3, \dots, x_T} p(x_{1:T})$$

$$\hookrightarrow \sum_{x_2 \in \{1,\dots,K\}} \sum_{x_3 \in \{1,\dots,K\}} \cdots$$

exponential distributivity

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

exponential # of terms $\Rightarrow k^{T-1}$

distributivity

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

$$= \sum_{x_{2:T}} f_1(x_1) f_2(x_2|x_1) \cdots f_T(x_T|x_{T-1})$$

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

& $O(k^2)$ to compute.

$$\underbrace{\sum_{x_{T-1}} f_{T-1}(x_{T-1}|x_{T-2}) \cdot m_T(x_{T-1})}_{m_{T-2}(x_{T-2})}$$

$$\overset{m_{T-1}(x_{T-2})}{\longleftarrow} \quad \overset{m_T(x_{T-1})}{\longleftarrow}$$

$$x_1 \to x_2 \to x_3 \;\cdots\; x_T$$

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

$$O(T \cdot k^2) \ll O(k^T)$$