Lecture 1 - intro

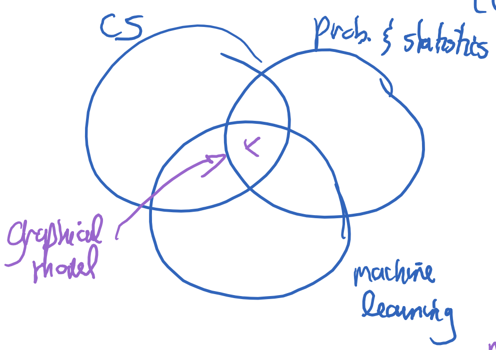
Tuesday, September 5, 2023 12:50 PM

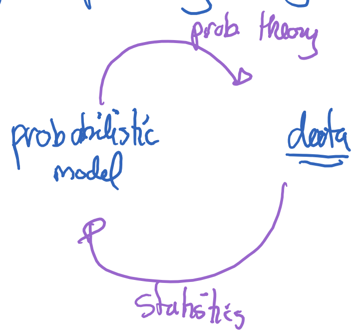
email sign-up by Wednesday evening : bit.ly/IFT6269-F23

probabilistic graphical model :

model multivariate data

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





Applications

hidden Markov model (HMM)

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

Y_t → phonemes (latent)

b) part-of-speech tagging: DT → V → PT → 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 :

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

matrices control term noise

here x_t & y_t are cts. vectors

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

sensor observation

if ε_t & ε_t' are Gaussian

HMM → "Kalman filter"

e) GPT RNN

y_t → "h_t" latent

$x_t \to$ words

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

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

factors $\to$ 2 variables $\Rightarrow \approx k^2$ parameters

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

clique in a graph (graphical model) $\rightsquigarrow$ 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 # of terms (k^{T-1})

distributivity

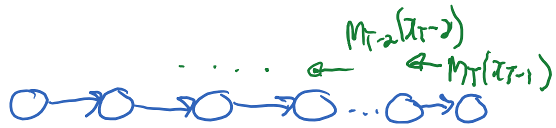
$a \cdot (b+c) = a \cdot b + a \cdot c$

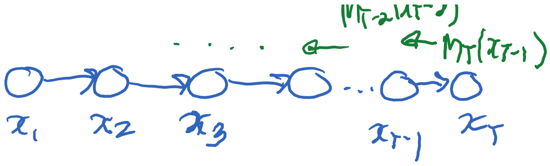
$$= \sum_{x_{2:T}} f_1(x_1) f_2(x_2|x_1) \cdots f_T(x_T|x_{T-1})$$

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

$m_T(x_{T-1}) \leftarrow O(k^2)$

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





"message passing alg." to compute <u>efficiently</u> marginal $p(x_1)$

$$O(T \cdot k^2) \lll O(k^{T-1})$$