Lecture 2 - scribbles

Friday, September 4, 2020 13:27

Why graphical models?

POS tagging observation: $(x_1, x_2, \dots, x_T) \triangleq x_{1:T}$ $\quad x_{\{1,3,5\}} \quad (x_1, x_3, x_5)$

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

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

issue: exponential size of state space
↳ in length in input

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

trick: make a factorization assumption about p

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

factor → 2 variables
$\Rightarrow \approx k^2$ parameters to specify

clique in graph model
theme: representation

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

computation?

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

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

exponential sum!?

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

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

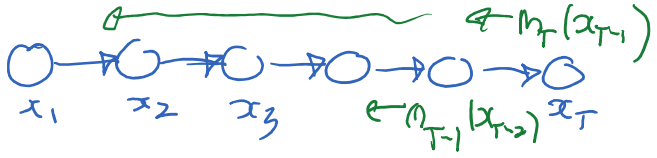
$$= f_1(x_1) \left(\sum_{x_2} f_2(x_2|x_1) \left(\sum_{x_3} f_3(x_3|x_2) \left(\cdots \left(\sum_{x_T} f_T(x_T|x_{T-1}) \right) \cdots \right) \right) \right)$$

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

$$\sum_{x_{T-1}} f_{T-1}(x_{T-1}|x_{T-2}) \cdot M_T(x_{T-1})$$

$\sum_{x_{T-1}} \psi_{T-1}(x_{T-1}, x_{T-2}) \cdot m_T(x_{T-1})$

$m_{T-1}(x_{T-2})$



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

→ $T \cdot k^2$ time

Key themes:

I) representation : how to represent structured prob. dist.?

- graph → factorization
- parameterization — full table param. vs. "exponential family"

II) estimation : given data, how to learn / estimate the parameters of dist?

↝ learning e.g. - maximum likelihood
- max. entropy
- moment matching

III) "probabilistic" inference : answer questions about data

e.g. compute $p(y|x)$ (conditional: query y, observation x) or $p(x)$ (marginal)

↝ computation e.g. • message passing
• approximate inference
– sampling
– variational methods

next: probability review

why? → principled framework to model uncertainty

Sources of uncertainty

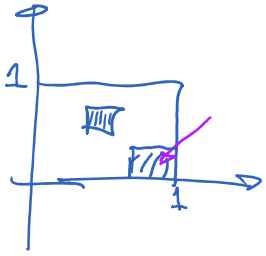
1) intrinsic uncertainty → quantum mechanics

2) partial information / observation :
- card games
- rolling a die → don't know the initial conditions

3) incomplete modeling of a complex phenomenon
(computational issues are also important)

example : "most birds can fly"
→ simple rule can be advantageous but then yields uncertainty
- "AI"

probabilities → like areas



Notation : X_1 X_2 X_3 X Y Z

↑ random variables (usually real-valued)

x_1, x_2, x_3 x y z

↑ their "realizations"

X a random variable → represents an uncertain quantity e.g. X "result of a dice roll"

$X = x$ represents the "event" that X takes the value x

Ω → the sample space of "elementary events" possible values of my R.V. $\Omega = \{1,2,3,4,5,6\}$

two types of R.V. : discrete where Ω is countable
continuous " Ω is uncountable

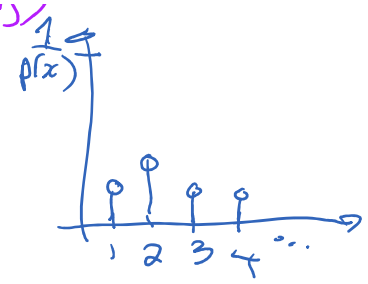
I) [assume Ω is countable → discrete case]

(discrete) R.V. X is characterized by a ~~probability mass fct.~~ (pmf) $\frac{1}{p(x)}$

"lower case" → $p(x)$ for $x \in \Omega$

"lower case" → $p(x)$ for $x \in \Omega$

pmf p is s.t. $\begin{cases} p(x) \geq 0 \quad \forall x \in \Omega \\ \sum_{x \in \Omega} p(x) = 1 \end{cases}$



probability distribution "big" P

is a mapping $P: \mathcal{E} \to [0,1]$

that satisfy the Kolmogorov axioms

$P(E) \geq 0 \quad \forall E \in \mathcal{E}$

$P(\Omega) = 1$

$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ when E_i's are disjoint

$\mathcal{E} = 2^{\Omega} \triangleq$ set of all subsets of Ω

↑ set of "events"

("σ-field" in measure theory needed when Ω is uncountable)

notation: $P\{X = x\} = p(x)$

pmf

$P\{X = x \text{ or } X = x'\} = P\{X = x\} + P\{X = x'\} = p(x) + p(x')$ if $x \neq x'$

$E = \{x, x'\}$

for discrete R.V. $P(E) = \sum_{x \in E} p(x)$

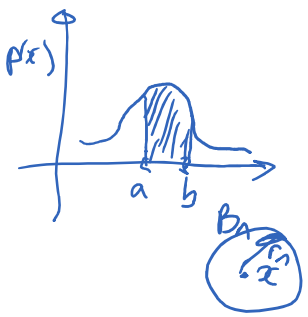
Continuous R.V.:

continuous R.V. is characterized by a probability density function p (pdf)

$p: \Omega \to \mathbb{R}$

$p(x) \geq 0 \quad \forall x \in \Omega$

p is integrable and $\int_{\Omega} p(x)\,dx = 1$



prob. of events:

$\Omega = \mathbb{R}$: $P([a,b]) = \int_a^b p(x)\,dx$

$\mathcal{E} \to$ Borel σ-field

$p(x) = \lim_{n \to \infty} \frac{P(B_n)}{\text{Vol}(B_n)}$

where B_n are balls enclosing x and decreasing radius

and decreasing radius

Recap: discrete R.V. X; pmf $p(x) \leftrightarrow P\{X=x\} = p(x)$

cts R.V. X; pdf $p(x)$ $\quad P\{X=x\} = 0$

"$P\{X \in x \pm \frac{dx}{2}\} \approx p(x)dx$"

$$P\{X \in x \pm \tfrac{a}{2}\} = \int_{x-\frac{a}{2}}^{x+\frac{a}{2}} p(x)dx$$

[sidenote: can change pdf p on countable (has measure zero) # of pts, according Lebesgue measure without changing "anything"]