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

probabilistic graphical model
model multivariate data

graphical model $\rightarrow$ mix of graph theory + prob. theory
(CS)

graphical model

prob & statistics

machine learning

Applications:

hidden Markov model

node = random variable

node $\rightarrow$ observed random variable

big caveat
a) speech recognition: $X_t \rightarrow \text{phoneme}$
   $X_t \rightarrow \text{sound wave}$

b) part-of-speech tagging: $X_t \in \{\text{DT, V, Di-AQI, N}\} \rightarrow \text{part of speech}$

   This is a red box $X_t$: words

   (x) [normally multiple nucleotides per $X_t$]

   coding vs non-coding (i.e. $X_t \in \{\text{G, C}\}$)

c) Gene finding $X_t$

   $X_t$:
   $\text{C T A A A C C C}$

   control region (given by system)

   $X_t^i$ = latent state

   noise

   We $X_t, Y_t$ are also vectors

   if noise is Gaussian, HMM $\rightarrow$ Kalman filter

   Why graphical models?

   pos tagging observation $(x_1, \ldots, x_M) \in X_{1:T}$

   $x_t \in \{\text{C, G, A, T}\}$
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 \( p \)

\[
p(x_1, \ldots, x_T) = \prod_{i=1}^T f_i(x_i) f_2(x_1|x_i) f_3(x_2|x_2) \cdots f_T(x_T|x_{T-1})
\]

factor \( \rightarrow \) 2 variables

\( \approx K^2 \) parameters

clique in

graphical model

\( T \) factors \( \approx (T \cdot K^2 \text{ parameters}) \)

\( \ll K^T \)

home: representation

computation:

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

distributivity:

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

\[
p(x_1) = \sum_{x_1, \ldots, x_T} p(x_{1:T}) \quad \text{exponential sum}\?
\]

\[
= \sum_{x_1, \ldots, x_T} \prod_{i=1}^T f_i(x_i) f_2(x_1|x_i) \cdots f_T(x_T|x_{T-1})
\]
\[
\prod_{i=1}^{T} \frac{f_i(x_i)}{f_2(x_2|x_1)} \prod_{i=2}^{T} \frac{f_i(x_i|x_{i-1})}{f_2(x_2|x_1)} \frac{f_T(x_T|x_{T-1})}{f_T(x_T|x_{T-1})} 
\]

"message passing algorithm" to compute efficiently marginal \( p(x_t) \)

**key themes:**

representation: how to represent structural prob distributions?

estimation: given data, how do learn/estimate the parameters of dist.

inference: answer questions about data

e.g., computing \( p(y|x) \) or \( p(x) \)

paramaterization

full table param,
or
"exponential family"

learning e.g., maximum likelihood

computation e.g., message passing

also approximate inference
conditional marginal

also approximate inference e.g. sampling
variational methods