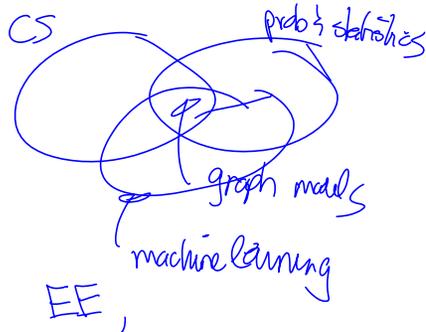


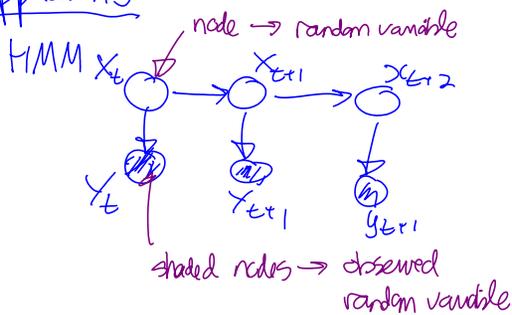
Probabilistic graphical models

model multivariate data

graphical models \rightarrow graph + prob theory



Applications



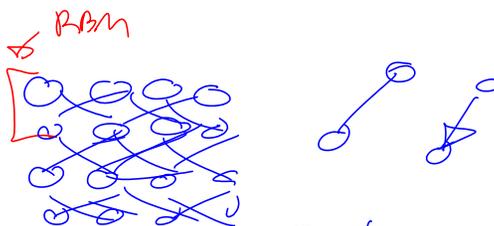
speech recognition $Y_t \rightarrow$ MM sound waves

$X_t \rightarrow$ phonemes

80s \rightarrow NN Hinton, Sejnowski

\approx 2007-2008 Hinton was back?

deep RBMs



\approx 2012 \rightarrow huge gain in performance $\left\{ \begin{array}{l} \text{speech} \\ \text{computer vision} \\ \text{image net} \end{array} \right.$

part-of-speech tagging

part-of-speech tagging

y's N-V-DT-ADJ-N
 x's this is a red box } words

part-of-speech

gene finding

coding vs non-coding regions

G T A A C C G

control

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

latent state, control term, matrix, noise

$$x_t = C y_t + \epsilon'_t \text{ } \} \text{ observed terms}$$

→ HMM here y's & x's continuous
 Kalman filter

Why graphical models?

POS tagging



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

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

→ exponential state space

$\sim K^T$ parameters to define distribution over $x_{1:T}$

→ trick here make factorization assumptions about distributions

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

→ factors → degree in the graph
 size of this is K^2

→ T factors
 #parameters $\sim TK^2$

Key themes:

representation : how to represent structured proba distributions w/ semantic.

estimation : how do we learn parameters from data?

inference : answer questions about data

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

→ ... $M_T(x_{T-1})$

