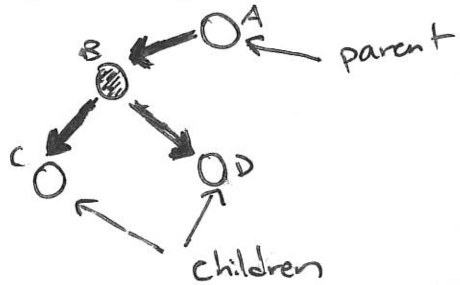
Inference: Bayesian Networks

- definition
- independence
- examples
- Viterbi

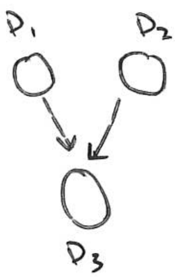
Definition

- directed acyclic graph (DAG)
- probability distribution



$p(A)\,p(B|A)\,p(C|B)\,p(D|B)$

Useful because:
- good model
- massive reduction in parameters
- flexible & adaptable



D_1, D_2 independent dice rolls

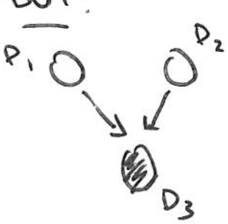
$D_3 = D_1 + D_2$

independent: $p(A,B) = p(A)p(B)$

↳ $p(D_1, D_2) = p(D_1)\,p(D_2)$

so D_1, D_2 independent

BUT!



$p(D_1 D_2 | D_3) \neq p(D_1|D_3)\,p(D_2|D_3)$

i.e. if $D_3 = 6$, then if $D_1 = 2$ then $D_2 = 4$.

thus D_1 and D_2 not independent conditioned on D_3!

"explaining away"

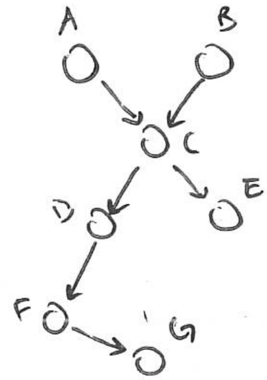
Independence

- marginal independence
- conditional independence
- induced dependence

~same thing: all independence!

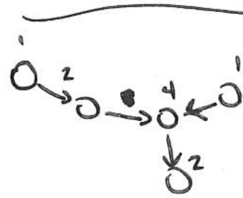
"marginal": sum across all ancestors

"d-separation"

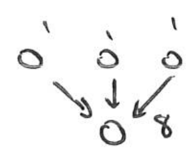


6.034 notes!

1) A, B marginally independent? YES
2) A, B conditionally independent, given C? NO
3) D, E marginally independent? NO
4) A, B conditionally independent, given G? NO
5) D, E conditionally independent, given C? YES
6) D, E conditionally independent, given A, B? NO



= 10 parameters

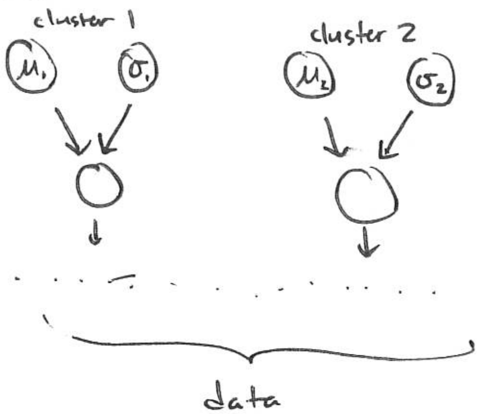


= 11 parameters

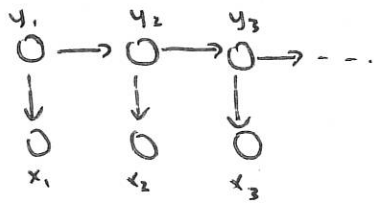
Examples

Mixture Model

$K=2$



Hidden Markov Model



$$\Rightarrow p(Y,X) = \underbrace{p(y_1)p(x_1|y_1)\prod_{i=2}^{n} p(y_i|y_{i-1})p(x_i|y_i)}_{\text{joint distribution}}$$

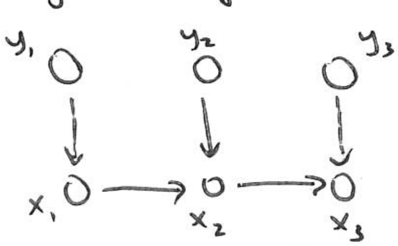
Viterbi Algorithm

POS tagging of sentences!

words: $y_1\ y_2\ y_3 \cdots$

tags: $x_1\ x_2\ x_3 \cdots$

Tag for the current word depends on both the word and on the previous tag! Making Markov assumption.



$$\Rightarrow p(y_1)p(x_1|y_1)\prod_{i=2}^{3} p(y_i)p(x_i|x_{i-1}, y_i)$$

Sample sentence:

"Read your notes."

We are given the sentence, so

$p(y_1) = p(y_2) = p(y_3) = 1$

$$\Rightarrow p(x_1|\text{read})\,p(x_2|x_1,\text{your})\,p(x_3|x_2,\text{notes})$$

Prior information:

$$\begin{array}{l} \\ \text{adj} \\ \text{noun} \\ \text{verb} \end{array}\begin{array}{c} \begin{array}{ccc} \text{read} & \text{your} & \text{notes} \end{array} \\ \begin{bmatrix} 0.4 & 1 & 0 \\ 0 & 0 & 0.5 \\ 0.6 & 0 & 0.5 \end{bmatrix} \end{array}$$

Transition:

$$\begin{array}{l} \\ a \\ n \\ v \end{array}\begin{array}{c} \begin{array}{ccc} a & n & v \end{array} \\ \begin{bmatrix} 0.1 & 0.6 & 0.3 \\ 0.4 & 0.1 & 0.5 \\ 0.4 & 0.6 & 0 \end{bmatrix} \end{array}$$

$\Downarrow$

$$\begin{array}{l} \\ a \\ n \\ v \end{array}\begin{array}{c} \begin{array}{ccc} p(x_1|\text{read}) & p(x_2|x_1,\text{your}) & p(x_3|x_2,\text{notes}) \end{array} \\ \begin{bmatrix} 0.4 & & \\ 0 & & \\ 0.6 & & \end{bmatrix} \end{array}$$

$\max(0.4 \times 1 \times 0.1,\ 0 \times 1 \times 0.4,\ 0.6 \times 1 \times 0.4) = \max(0.04, 0, 0.24)$

$= 0.24$

$\Downarrow$

$$\begin{bmatrix} 0.4 & 0.24 & 0 \\ 0 & 0 & 0.072 \\ 0.6 & 0 & 0.036 \end{bmatrix} \to \text{likelihood!}$$

max likelihoods: $\underbrace{0.072 - 0.24 - 0.6}_{n \quad a \quad v}$

verb $\to$ adj $\to$ noun

"read" your" "notes"! it works :)