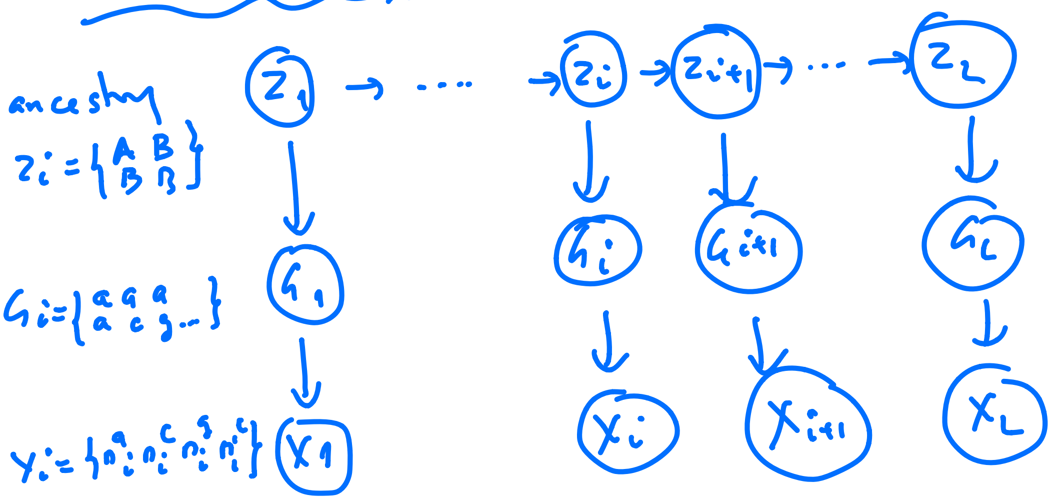
W07 - Probabilistic Models EM

Big Picture Directed Acyclic Graphs (DAG)



$P(Z_{i+1} \mid Z_i)$ → a Markov DAG = HMM

What to do with such model?

1. Sample examples "follow the DAG"
   need to know all probability distributions
   (in our case $p = P(AB \mid AB) = P(BB \mid BB)$)

2. Do Inference (W06)

3. Estimate parameter (W07)

② Do Inference (W06)

we know $D = [X_1 \ldots X_L]$

want to infer the ancestry at each position $[Z_1 \ldots Z_L]$

$P(Z_i = AB \mid X_1 \ldots X_L)$ is the posterior of the 2 possible ancestries at position i given all the data

$P(Z_i = BB \mid X_1 \ldots X_L)$

To calculate these probabilities, it is convenient to calculate by dynamic programming

$\cancel{z_1 \ldots}$ $z_i = 2$

$[x_1 \ldots x_i]$

$$f_2(i) = P(X_1 \ldots X_i, Z_i = 2)$$

forward algorithm

z_i $\cancel{z_{i+1} \ldots z_L}$

x_i $[x_{i+1} \ldots x_L]$

$$b_2(i) = P(X_{i+1} \ldots X_L \mid Z_i = 2)$$

backward algorithm

Using $f_z(i)$ and $b_z(i)$ we can calalate

(2) $$P(z_i = z \mid x_1 \ldots x_L) = \frac{f_z(i)\, b_z(i)}{f_{AB}(i)\, b_{AB}(i) + f_{BB}(i)\, b_{BB}(i)}$$

(3) Expected vales of transitions

$$\begin{array}{cccccc}
 & AB & & AB & & \\
 & \parallel & & \parallel & & \\
\cancel{z_1 \ldots z_{i-1}} & z_i & \rightarrow & z_{i+1} & \rightarrow & \cancel{z_{i+2} \ldots z_L} \\
 & \downarrow & & & & \\
[x_1 \ldots & x_i & & x_{i+1} & & x_L]
\end{array}$$

$$P(x_1 \ldots x_i \; z_i = AB \; x_{i+1} \; z_{i+1} = AB \; x_{i+2} \ldots x_L) =$$

$$= f_{AB}(i)\, P(AB \mid AB)\, P(x_{i+1} \mid AB)\, b_{AB}(i+1)$$

in general

$$E_i(z_i = z, z_{i+1} = z' \mid p)$$

$$= f_z^p(i)\, P(z' \mid z)^p\, P(x_{i+1} \mid z')\, b_{z'}^p(i+1)$$

Then we can do EM

i) start with $p = p^0$ (arbitrary)

ii) calculate Expectations

$$E_i(z_i = z, z_{i+1} = z' \mid p^{(0)})$$

iii) do ML estimation of next $p = p^{(1)}$

$$p^{(1)} = \frac{\sum_z \sum_m \sum_i E_i^m(z_i = z, z_{i+1} = z \mid p^{(0)})}{\sum_z \sum_i \sum_m E_i(z_i = z, z_{i+1} = z \mid p^{(0)}) + \sum_{z \neq z'} \sum_m \sum_i E_i(z_i = z, z_{i+1} = z' \mid p^{(0)})}$$

and iterate

until $p^{(n+1)} \approx p^{(n)}$

Small Details implementing forward/backward

- Allways work in log space
- in log space all quantities are negatives (they are all probabilities)
- $P(X_i \mid Z_i = AB) =$

$$= \sum_{k=1}^{10} P(X_i \mid G_k)\, P(G_k \mid AB)$$

only G_k with $P(G_k \mid AB) \neq 0$ is $G_k = g_i^A g_i^B$, then

$$\underbrace{P(X_i \mid Z_j = AB)}= P(X_i \mid G = g_i^A g_i^B) \cdot \underbrace{P(G = g_i^A g_i^B \mid AB)}_{1}$$

$$= \underbrace{P(X_i \mid G = g_i^A g_i^B)}$$

$$\left\{\begin{array}{l} P(X_i \mid Z = AB) = P(X_i \mid G = g_i^A g_i^B) \\ P(X_i \mid Z = BB) = P(X_i \mid G = g_i^B g_i^B) \end{array}\right\}$$

- If you get errors, $+\infty$, $\log(0)$ ....
  go inside your loop and make print statements

Causal Inference

Judea Pearl (UCLA)

James Robins
Miguel Hernán (Harvard)

$P(A \mid B)$ conditional probability

$P(A \mid do(B))$ the adjustment formula
or causal effect rule

$T = (0,1)$ a treatment

$O = (+,-,0)$ an outcome ($+$ good, $-$ not good)

$$P(O \mid T=1)$$

$$P(O \mid do(T=1))$$

when are they different?

Use DAGs

$$\begin{tikzcd} & G \arrow[d] \\ T \arrow[r] & O \end{tikzcd}$$

$P(O \mid T) = P(O \mid do(T))$

$$\begin{tikzcd} & G \arrow[dl] \arrow[dr] & \\ T \arrow[rr] & & O \end{tikzcd}$$

$P(O \mid T) \neq P(O \mid do(T))$

DAGs

$$\begin{tikzcd} T_1 & TF_1 \arrow[d] \\ T_2 & TF_2 \arrow[d] \\ & G \end{tikzcd}$$

$T_1 \not\perp G$

$T_1 \perp G \mid T_2$

$$P(G\,T_1 T_2) = P(G \mid T_2)\,P(T_2 \mid T_1)\,P(T_1)$$

chain

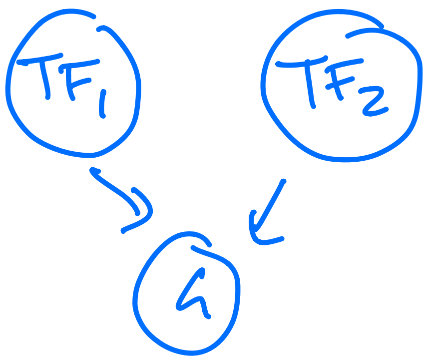
$$\begin{tikzcd} & TF \arrow[dl] \arrow[dr] & \\ G_1 & & G_2 \end{tikzcd}$$

$G_1 \not\perp G_2$

$G_1 \perp G_2 \mid T$

$$P(T\,G_1 G_2) = P(G_1 \mid T)\,P(G_2 \mid T)\,P(T)$$

fork



$T_1 \perp T_2$

$T_1 \not\perp T_2 \mid G$

collider

$$P(G\, T_1\, T_2) = P(G \mid T_1\, T_2)\, P(T_1)\, P(T_2)$$

→ using only the joint probabilities, proof the conditional independence of chains, forks and colliders

Back to the treatment

$$\begin{tikzcd} & G \arrow[dl] \arrow[dr] & \\ T \arrow[rr] & & O \end{tikzcd}$$

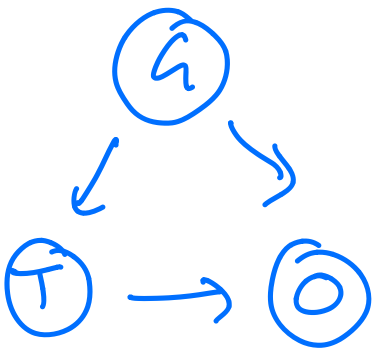
$G = \{M, F\}$ has an influence

$G \not\perp T$ " more F take treatment than M

$G \not\perp O$ " treatment has more effect on M

$$P(T=1 \mid G=M) < P(T=1 \mid G=F)$$

$$P(O=+ \mid G=M) > P(O=- \mid G=F)$$



$$P(O \mid T=1)$$

$$P(O \mid T=1) =$$

$$= \sum_G P(O\, G \mid T=1)$$

$$= \sum_G P(O \mid T=1, G)\, P(G \mid T=1)$$

$$P(O \mid T=1) = P(O \mid T=1\ M)\, P(M \mid T=1) + P(O \mid T=1\ F) \cdot P(F \mid T=1)$$

What is $P(O \mid do(T=1))$?

$$
\begin{tikzcd}
 & G \arrow[dl] \arrow[dr] & & & & G \arrow[d] \\
T \arrow[rr] & & O & \arrow[r, "\text{to}"] & T \arrow[r] & O
\end{tikzcd}
$$

DAG $\qquad\qquad\qquad$ DAG^m

$$P(O \mid do(T=1)) = \underset{DAG^m}{P(O \mid T=1)}$$

$$= \sum_G P^m(O\, G \mid T=1)$$

$$= \sum_G P(O \mid T=1\, G)\, P^m(G \mid T=1)$$

$$= \sum_G P(O \mid T=1\, G)\, P(G)$$

$$= P(O \mid T=1\, M)\, P(M) + P(O \mid T=1\, F)\, P(F)$$

$$
\begin{tikzcd}
 & G \arrow[dl] \arrow[dr] & \\
T \arrow[rr] & & O
\end{tikzcd}
$$

$$
\begin{aligned}
P(O \mid T=1) = {} & P(O \mid T=1\ M)\, P(M \mid T=1) \\
& + P(O \mid T=1\ F)\, P(F \mid T=1)
\end{aligned}
$$

$$
\begin{aligned}
P(O \mid do(T=1)) = {} & P(O \mid T=1\ M)\, P(M) \\
& + P(O \mid T=1\ F)\, P(F)
\end{aligned}
$$

the adjustment formula

In general

$$P(Y \mid X) = \sum_{P_X} P(Y \mid X, P_X)\, P(P_X \mid X)$$

$$P(Y \mid do(X)) = \sum_{P_y} (Y \mid X\, P_x) \cdot P(P_x)$$

were P_X = parents of X in DAG.