10-601 HOMEWORK-6

SAM DAVID CHRIST DOSS PUSHPAM

ANDREW ID: schristd.

1. Fundamentals.

a) 1. False, Since S to C to R is not blocked.
2. False, Since S to C to PU is not blocked.
3. True
4. False Since S→C ~~to~~ PU is not blocked.
5. False, S to T to C is not blocked.
6. False. S to T to PU is not blocked.
7. True
8. True
9. False S to ~~P~~ to PU is not blocked.
10. True

b) 1. No variables are d-separated from R. C is parent and PU is children. Since, C is unknown its not d-separated from S. And not d-separated from T because S is unknown.

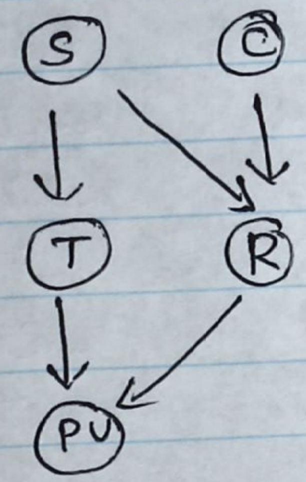
2. T is d-separated from R given S

3.

3. T and R are d-separated given S. knowing C, doesnot d-separate T and R. Similarly, knowing PU doesnot d-separate T and R, because aagain S is unknown.

c) $P(S, T, C, R, PU) = P(S)\,P(T|S)\,P(C|S)\,P(R|C) \cdot P(PU|T, C, R)$

d)



e) 1. $P(C = yes \mid summer) = 0.7$

2. $P(PU = yes \mid C = yes)$

$= \sum_{T,C,R} \cancel{P(PU = yes \mid T, R, C = yes)\,P(T|S)\,P(S)\,P(R|C = yes)}$

$= \sum_{C,T} P(PU = yes \mid R = yes, C, T)\,P(C \mid summer)\,P(T \mid summer)$

$= 0.8 \times 0.7 \times 0.2 + 0.3 \times 0.05 \times 0.8 + 0.9 \times 0.7 \times 0.2 + 0.01 \times 0.2 \times 0.3$

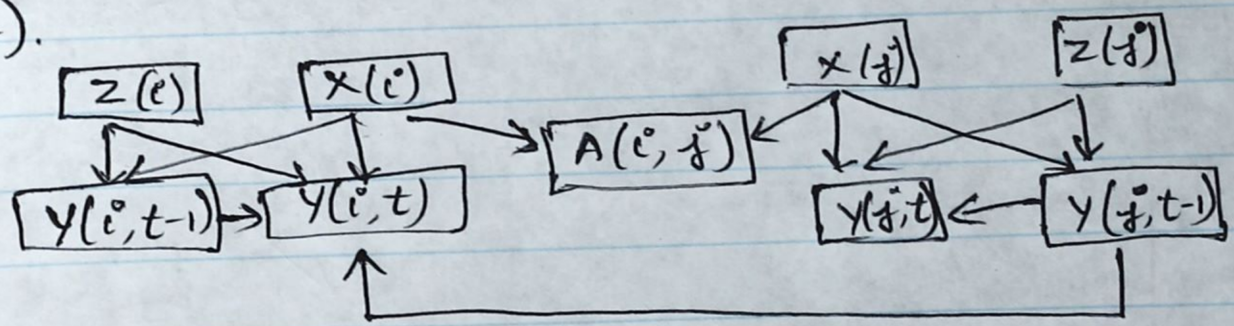
$= 0.1946.$

3. $P(PU = yes \mid C = yes) = \sum_{T,S,R} P(PU = yes \mid T, R, C = yes) \cdot P(T \mid S)\, P(S)\, P(R \mid C = yes)$

$= 0.719225$

2. Homophily or Contagion.

a).

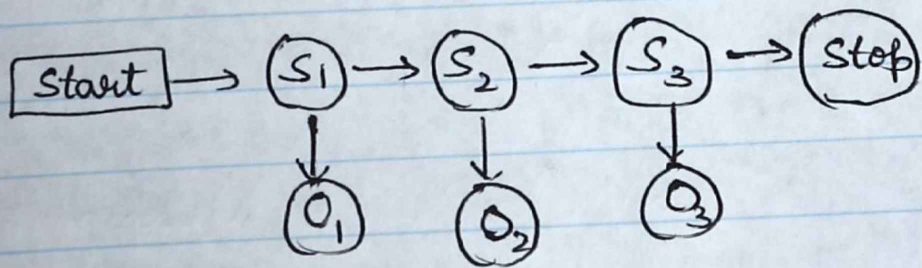


b) The rephrased question is, given $Y(j, t-1)$, is $Y(i, t)$ d-separated from $Y(i, t-1)$

This is equivalent because, if $Y(i, t-1) \perp Y(i, t) \mid Y(j, t-1)$ any change in $Y(i, t)$ given $Y(j, t-1)$ is solely because of $Y(j, t-1)$.
Other boxes doesnot change.

c) no. Since, the change of $Y(i, t)$ also depends on $Y(i, t-1)$, we can't guarantee that this is due to social link.

3. Hidden Markov models.

a) 

b) The new state S_i is generated from the $\theta_{start,i}$. And the corresponding observation is got from $\gamma_{i,o}$ from S_i.
The second state is obtained from $\theta_{i,j}$. The corresponding observation is obtained by $\gamma_{i,o}$.
The end state is obtained by $\theta_{i,end}$. And the corresponding observation using $\gamma_{i,o}$.

c) $O_i \perp S_j | S_i$, O_i is the observation corresponding to S_i, and S_j is some other state.
$S_i \perp S_j | S_{i-1}$, S_j is any state which is the ancestor of S_{i-1}.

d) Markov blanket for a node "p" is the smallest set of nodes "o" which follows the property that, for any node p_1 outside o, $p \perp p_1 | o$.
The nodes given by the markove blanket has the property that, the p is d-separated from all other nodes given the set of nodes o.

ESTIMATION :-

e) $p(D | \theta, \theta_{start}, \theta_{stop}, \gamma)$

$$= \prod_{n=1}^{N} \Big(P(s_1^n | start, \theta_{start}) \cdot P(o_1^n | s_1^n, \gamma) \, P(stop | s_{T_n}^n, \theta_{stop}) \prod_{i=2}^{T_n} P(s_i^n | s_{i-1}^n, \theta) \, P(o_i^n | s_i^n, \gamma) \Big)$$

Taking log gives,

$$\Rightarrow \sum_{n=1}^{N} \log(P(s_1^n | start, \theta_{start})) + \log(P(o_1 | s_1^n, \gamma)) + \log(P(stop | s_{T_n}^n, \theta_{stop}) + \sum_{i=1}^{T_n} \log(P(s_i^n | s_{i-1}^n, \theta)) + \log(P(o_i^n | s_i^n, \gamma))$$

f) From the previous part, let

$$\ell(\theta_{start}) = \sum_{k=1}^{N} \log(P(S_1^n \mid start, \theta_{start}))$$

$$\ell(\theta_{i,:}, \theta_{i,stop}) = \sum_{k=1}^{N} \sum_{t=1}^{T_k} \log(P(S_t^n \mid S_{t-1}^n = i, \theta_{i,:})^{1[S_{t-1}^n = i]} + \log(P(stop \mid S_{T_k}^n = i, \theta_{stop})^{1[S_{T_k}^n = i]}$$

$$\ell(\gamma_i) = \sum_{k=1}^{N} \sum_{t=1}^{T_k} \log(P(O_t^n \mid \gamma_{i,:}, S_t^{(n)} = i)$$

The log likelihood now becomes,

$$\ell(\theta_{start}) + \sum_{i=1}^{n} \ell(\theta_{i,:}, \theta_{i,stop}) + \ell(\gamma_{i,:})$$

Inference :-

g) $$P(Start, S_{1:T}, stop, O_{1:T}) = \frac{P(Start, S_{1:T}, stop, O_{1:T})}{P(O_{1:T})}$$

Since, the $O_{1:T}$ is fixed, max the left side will be

$$\underset{S_{1:t}}{\text{argmax}}\, P(Start, S_{1:T}, Stop, O_{1:T}) = \underset{S_{1:t}}{\text{argmax}}\, P(Start, S_{1:T}, Stop, O_{1:T})$$

h) Total of T random variables $O(T)$, and n choice for each random variable, so the overall complexity $O(Tn^T)$

(i)

Scores $(i,1) = P(s_i|start) * P(O_i|s_i)$

Scores are stored using,

$$scores(i, j+1) = p(O_j|s_i) * \max_k p(s_i|s_k) * scores(k,j)$$

$$backpointers(i, j+1) = \operatorname{argmax}_k p(O_j|s_i) * P(s_i|s_k) * score(k,j)$$

The scores and backpointers, we use backtracking to find the states based on observation.

(j) Computational complexity :-

$O(n^2 T)$.

Every step take $O(n^2)$ and there $O(T)$ steps, so, overall $O(Tn^2)$.

Analysis :-

(o) baseline train :-

train accuracy = 0.8522

Test:-

accuracy : 0.8106

HMM train:-

train accuracy :- 0.94

test accuracy :- 0.92.

The baseline prediction is based only on the prior probability. For HMM conditional probability is also used.

(p) The test accuracy drops down.
This is because prior decreases the influence of the outliers.