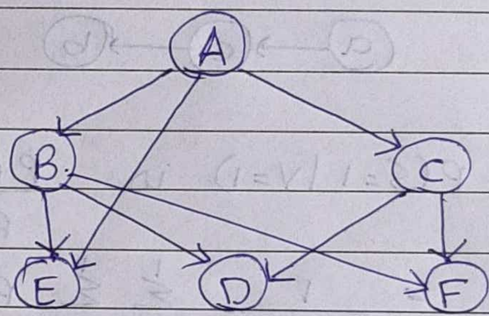
Machine Learning

HOMEWORK-VI

1. A graph with no links is a trivial D-Map.
- True.
A graph is a D-map of distribution if every conditional independence statement satisfied by the distribution is reflected in the graph. Therefore, a graph with no links is a trivial D-Map.

2. a) Yes, A is conditionally independent of D given {B, C} as $A \to B \to D$ and $A \to C \to D$ are Head to tail connections



b) No, E is not marginally independent of F.

c) We can delete the edge $A \to C$ to make A independent of C.

3. From the given table, we get the following tables for $p(a)$, $p(c|a)$, $p(b|c)$ by marginalizing and conditioning the joint distribution from the given table.

a	P(a)
0	0.6
1	0.4

a	c	P(c|a)
0	0	0.4
0	1	0.6
1	0	0.6
1	1	0.4

c	b	P(b|c)
0	0	0.8
0	1	0.2
1	0	0.4
1	1	0.6

Multiplying the three distribution together we recover the joint distribution $p(a,b,c)$ given in the table, thereby allowing us to verify the validity of the decomposition

$$p(a,b,c) = p(a) * p(c|a) * p(b|c)$$

We can express the distribution using the graph

$$a \rightarrow c \rightarrow b$$

4. a) $P(S=1|V=1)$ is $\dfrac{P(S=1, V=1)}{P(V=1)}$

$$= \frac{1}{P(V=1)} \sum_{R=0}^{1} \sum_{G=0}^{1} P(V=1) * P(G) * P(R|V=1, G) * P(S=1|G)$$

$$= \sum_{RG} P(G) * P(R|V=1, G) * P(S=1|G)$$

$$= \sum_{G} P(G) * P(S=1|G) \sum_{R} P(R|V=1, G)$$

$$= \sum P(G) * P(S=1|G) * 1$$

$$= P(G=0) * P(S=1|G=0) + P(G=1) * P(S=1|G=1)$$

$$= \alpha(1-\gamma) + (1-\alpha) * (1-\beta)$$

$$= \alpha - \alpha\gamma + 1 + \alpha\beta - \alpha - \beta$$

$$= 1 - \alpha\gamma + \alpha\beta - \beta$$

b.) $P(S=1|V=1)$ and $P(S=1|V=0)$ are same because they are independent of V, so the expression is same as above.

c.) MLE can be estimated by counting events. Thus, $\alpha = 1/3$, $\beta = 0$ and $\gamma = 1$.

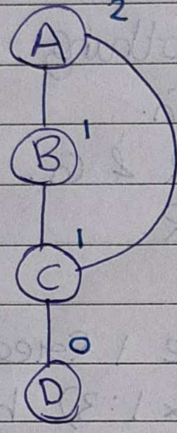
V and S are independent given nothing.
V and S are independent given G.
V and S are independent given R & G.
V and S are independent given R.

5. a.) The CPDs for nodes 1,2,3 have 1 free parameters each (since they are Bernoulli). $P(H|x1:3)$ has 8 parameters one per conditioning case.
$P(x_i|H)$ for $i=4:6$ are $2*2$ tables, but due to the sum of one constraint, only have 2 free parameters.
So total of $(3)(1) + 8 + 3(2) = \underline{17}$.

b.) For the graph on the right, the CPDs for nodes 1,2,3 have 1 free parameters each (since they are Bernoulli).
$P(x_4|x_{1:3})$ has 8 parameters, one per conditioning case.
$P(x_5|x_{1:4})$ has 16 parameters.
$P(x_6|x_{1:5})$ has 32 parameters.
In total, $3+8+16+32 = 59$ parameters.

6. The functions after instantiating evidence variable are $P(B|A)$, $P(C|A,B)$, $P(D|C)$ and function of D. The induced graph along the ordering ABCD is shown below. The number of children for A, B, C & D are 2, 1, 1 & 0 resp., so the width of the tree is 2. The complexity of the variable elimination algorithm is $O(n \exp(w+1))$, where

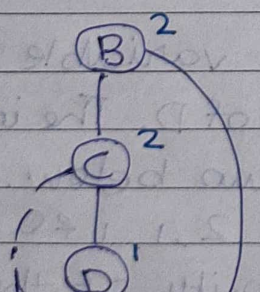
n is number of non-evidence variables and 'w' is the width of the ordering. Therefore the complexity is O(4 exp(3)).



Max width = 2

7. The functions after instantiating evidence variable are P(B|A), P(C|B), P(D|C), P(A) and function of D.
The induced graph along the ordering BCDA is shown below.

The number of children for B, C, D and A are 2, 2, 1 and 0 respectively, so the tree-width is 2. The complexity of the variable elimination algorithm is O(n exp(w+1)), where n is number of non-evidence variables and w is the width of the ordering.
Thus, the complexity is O(4 exp(3)).



Max width = 2