SPARSH AGARWAL

CS 576

HW 6

AI) (a) $P(A, B, C)$

A	B	C	
T	T	T	0.024
T	T	F	0.096
T	F	T	0.196
T	F	F	0.084
F	T	T	0.084
F	T	F	0.336
F	F	T	0.126
F	F	F	0.054

(b) $P(A=T \mid C=F) = \dfrac{P(A=T, C=F)}{P(C=F)} = \dfrac{0.084+0.096}{0.84+0.96+0.336+0.054}$

$= \dfrac{0.18}{0.57}$

$= 0.315789474$

(c) $P(A=T \mid B=F) = \dfrac{P(A=T, B=F)}{P(B=F)} = \dfrac{0.28}{0.46}$

$= 0.60869565 2$

(d) $P(A=T \mid B=T, C=F) = \dfrac{P(A=T, B=T, C=F)}{P(B=T, C=F)}$

$= \dfrac{0.096}{0.096 + 0.336}$

$= 0.2222\bar{2}$

(e) $P(C=T \mid A=T) = \dfrac{P(C=T, A=T)}{P(A=T)} = \dfrac{0.024 + 0.196}{0.024 + 0.196 + 0.096 + 0.084}$

~~$P(C=T \mid A=F)$~~

$= \dfrac{0.22}{0.4} = 0.55$

$P(C=T \mid A=F) = \dfrac{P(C=T, A=F)}{P(A=F)} = \dfrac{0.084 + 0.126}{0.6} = 0.35$

C is not independent of A because ~~probability P(C)~~ probability of C = True depends on probability of A being True or false

Or we can check $P(C=T \mid A=T)$ & $P(C=T)$ they will not be equal ∴ C is dependent (of C on A)

SPARSH AGARWAL

~~P(C=T | A=T, B)~~

(f) $P(C, A \mid B) = P(A \mid B) * P(C \mid B)$

if this is true then A & C are independent given B.

$P(C=T, A=T \mid B=T)$

$= \dfrac{0.024}{0.54}$

$= \dfrac{0.024}{0.54}$

$P(A=T \mid B=T) * P(C=T \mid B=T)$

$= \dfrac{0.12}{0.54} * \dfrac{0.108}{0.54}$

$= \dfrac{0.01296}{0.54}$

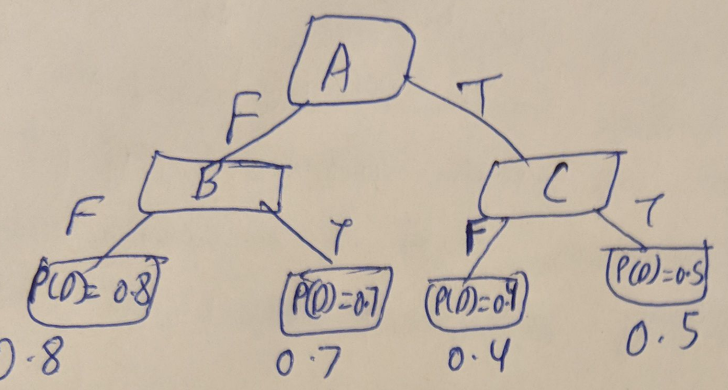
$\therefore P(C, A \mid B) \neq P(A \mid B) * P(C \mid B)$

$\therefore$ C is dependent on A given B.

A2) (a)

A B C D	Y = T	Y = F
T T T T	0.2	0.8
T T T F	0.2	0.8
T T F T	0.9	0.1
T T F F	0.15	0.85
T F T T	0.25	0.75
T F T F	0.25	0.75
T F F T	0.25	0.75
T F F F	0.25	0.75
F T T T	0.2	0.8
F T T F	0.2	0.8
F T F T	0.9	0.1
F T F F	0.15	0.85
F F T T	0.75	0.25
F F T F	0.75	0.25
F F F T	0.75	0.25
F F F F	0.75	0.25

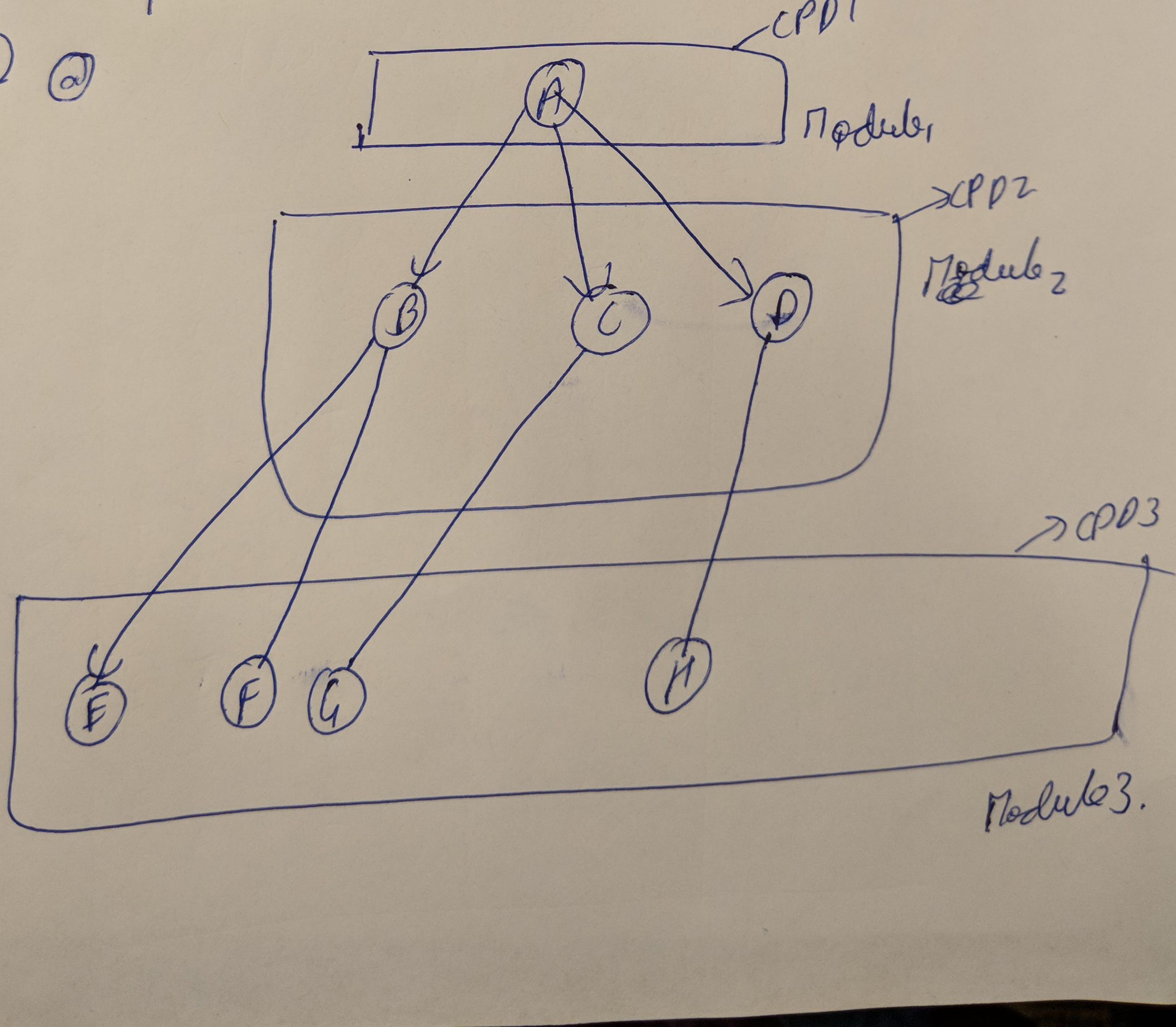
(b)



(c) We can get the independence relations of parameters directly through tree representation.

∴ using tree we can get additional properties of independencies.

A4 (a)



(b) A learned module network is better model than a learned ~~bea~~ bayesian network. ~~~~ Parameter sharing in the same module allows each parameter to be estimated based on much larger sample. Moreover, this allows us to learn dependencies that are considered ~~be~~ too weak based on statistics of single variable

A3) a) $$I(X, Z) = \sum_{(x,z)} P(x,z) \log \frac{P(x,z)}{P(x)\cdot P(z)}$$

$$= P(x=T, z=T) \log \left(\frac{P(x=T, z=T)}{P(x=T)\cdot P(z=T)} \right) +$$

$$P(x=F, z=T) \log \left(\frac{P(x=F, z=T)}{P(x=F) P(z=T)} \right) +$$

$$P(x=T, z=F) \log \frac{P(x=T, z=F)}{P(x=T) P(z=F)} +$$

$$P(x=F, z=F) \log \frac{P(x=F, z=F)}{P(x=F) P(z=F)}$$

$$= 0.348 \log \frac{(0.348)}{0.6 \times 0.608} +$$

$$0.26 \log \frac{0.26}{0.4 \times 0.608} + 0.252 \log \frac{0.252}{0.6 \times 0.392}$$

$$+ 0.14 \log \frac{0.14}{0.4 \times 0.392}$$

$$-0.0164070$$

$$= \cancel{-0.0171478} + 0.0173673 + 0.0173862 + $$

$$- 0.015866$$

$$I(X,Z) = 0.0024805$$

(b) $I(Y,Z) = \sum_{(y,z)} P(y,z) \log \frac{P(y,z)}{P(y)\,P(z)}$

$= \overset{(T,T)}{0.162} \log \left(\frac{0.162}{0.54 \times 0.392} \right) + \overset{(TF)}{0.378} \log \left(\frac{0.378}{0.54 \times 0.608} \right) +$

$\overset{(FT)}{0.23} \log \left(\frac{0.23}{0.46 \times 0.392} \right) + \overset{(FF)}{0.23} \log \left(\frac{0.23}{0.46 \times 0.608} \right)$

$= -0.04331657 + 0.053262261 +$

$0.055969639 - 0.04498036$

$$I(Y,Z) = 0.02093497$$

(c) Based on these values, Y should be selected as the candidate parent for Z because there ~~exg~~ exists more ~~mutal~~ mutual information between Y & Z as compared to X & Z.

(d) $P_{net}(X) = \frac{90 + 210 + 50 + 50}{1000} = 0.4$

$P_{net}(Y|X) = P_{net}(Y,X) \,/\, P_{net}(X)$

$= P(Y=T, X=T) \,/\, P_{net}(X)$

$= 0.3 / 0.4 = 0.75$

$P_{net}(Z|Y) = P_{net}(Z, Y) / P_{net}(Y)$

$= 0.162 / 0.54$

$= 0.3$

(e)

$$D_{KL}(\hat{P}(X,Z) \| P_{net}(X,Z)) = \sum \hat{P}(X,Z) \log \frac{\hat{P}(X,Z)}{P_{net}(X,Z)}$$

$D_{KL} = 0$

(f) since we do not gain any information by considering X as parent of Z ($D_{KL} = 0$)

$\therefore$ we cannot consider X to be parent of Z