MachineLearning Homework05

Yuan Bowei 1001916

1. Effective parameters are those parameters whose probability do not depend on other probabilities. So when each node can only take two possible values, there are 25 effective parameters, they are:

node	effective parameters	number
A	P(A=1)	1
B	P(B=1)	1
C	P(C=1)	1
D	P(D=1)	1
E	P(E=1 given A,B,C)	8
F	P(F=1)	1
G	P(G=1 given D,E)	4
H	P(H=1 given E,F)	4
I	P(I=1 given G,H)	4

when D and F can take 4 possible values, the number of effective parameters is 37, they are

node	effective parameters	number
A	P(A=1)	1
B	P(B=1)	1
C	P(C=1)	1
D	P(D=1), P(D=2), P(D=3)	3
E	P(E=1 given A,B,C)	8
F	P(F=1), P(F=2), P(F=3)	3
G	P(G=1 given D,E)	8
H	P(H=1 given E,F)	8
I	P(I=1 given G,H)	4

2. Without knowing the value of any other nodes, A and F are independent of each other.

   If the value of C and I are given, A and F are dependent of each other. We can imagnine how a Bayes Ball starts from A: Obviously it could reach G as A and G are dependent, as I is given, G and H are also dependent, so this Bayes Ball could reach H, H and F are apparantly dependent, so A and F are dependent of each other.

3. The procedure is:

$P(E=1|C=2)$

$= \frac{P(E=1,C=2)}{P(C=2)}$

$= \frac{\sum_{A\in(1,2)}\sum_{B\in(1,2)} P(E=1,A,B,C=2)}{P(C=2)}$

$= \frac{\sum_{A\in(1,2)}\sum_{B\in(1,2)} P(E=1|A,B,C=2)P(A)P(B)P(C=2)}{P(C=2)}$

$= P(E=1|A=1,B=1,C=2)P(A=1)P(B=1) + P(E=1|A=1,B=2,C=2)P(A=1)P(B=2)$
$+ P(E=1|A=2,B=1,C=2)P(A=2)P(B=1) + P(E=1|A=2,B=2,C=2)P(A=2)P(B=2)$
$= 0.3*0.2*0.5+0+0.6*0.8*0.5+0.5*0.8*0.5 = 0.03+0+0.24+0.2 = 0.47$

4. As node A doesn't depend on any other nodes, so the probability table of A is calculated with the formula:

$$P(A=i) = \frac{count(A=i)}{count(A)}$$

Therefore, $P(A=1) = \frac{7}{12}$, $P(A=2) = \frac{5}{12}$

A	Probability
1	7/12
2	5/12

However, node H is dependent on E and F, thus the probability table of H involves the calculation of conditional probability, that is

$$P(H=i|E=j,F=k) = \frac{count(E=j,F=k,H=i)}{count(E=j,F=k)}$$

.

Therefore, $P(H=1|E=1,F=1) = \frac{1}{1} = 1$, and we can analogously compute all other three probabilities:

condition	*	*
E F	H=1	H=2
1 1	1	0
1 2	3/4	1/4
2 1	3/5	2/5
2 2	1/2	1/2

5. We need to calculate the change of BIC after removing the edge between H and I. And the change of BIC is equal to the change of the value of node I's score function, which is:

before the change:

$Score(I|G, H; G)$

$= 2 * log(I = 1|G = 1, H = 1) + 2 * log(I = 2|G = 1, H = 1)$

$+log(I = 1|G = 1, H = 2) + log(I = 2|G = 1, H = 2) + 2 * log(I = 1|G = 2, H = 1) + 2 * log(I = 2|G = 2, H = 1)$

$+log(I = 1|G = 2, H = 2) + log(I = 2|G = 2, H = 2) - 2 * log(12)$

$= 8 * log(\frac{1}{2}) + 4 * log(\frac{1}{2}) - log(144)$

$= 12 * log(\frac{1}{2}) - 2 * log(12)$

after the change: $Score(I|G; G)$

$= 3 * log(I = 1|G = 1) + 3 * log(I = 2|G = 1) + 3 * log(I = 1|G = 2) + 3 * log(I = 2|G = 2)$

$-log(12)$

$= 12 * log(\frac{1}{2}) - log(12)$

So we can see after the change the BIC is greater than the BIC before the change by log(12). So after removing the edge between H and I the structure is better.