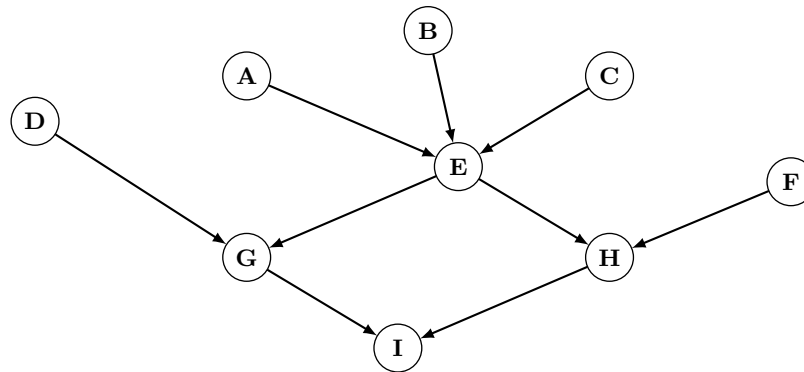


01.112 Machine Learning, Fall 2017
Homework 5

Due Monday 4 Dec 2017, 11.59pm

This homework will be graded by Gary Phua Hwee Leng
Please submit your solutions to eDimension before the deadline.

In this homework, we would like to look at the Bayesian Networks. You are given a Bayesian network as below. All nodes can take 2 different values: $\{1, 2\}$.



- (10 pts) Without knowing the actual value of any node, are node **A** and **F** independent of each other? What if we know the value of node **C** and **I**? **Yes. No.**
- (10 pts) What is the *effective* number of parameters needed **to for** this Bayesian network? What would be the *effective* number of parameters for the same network if node **D** and **F** can take 5 different values: $\{1, 2, 3, 4, 5\}$, and all other nodes can only take 3 different values: $\{1, 2, 3\}$? **2.**
- (10 pts) If we have the following probability tables for the nodes. Compute the following probability. Clearly write down all the necessary steps.

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

A		B		C		D		E	
1	2	1	2	1	2	1	2	1	2
0.2	0.8	0.7	0.3	0.2	0.8	0.5	0.5	0.1	0.9

F		G		H		I	
1	2	1	2	1	2	1	2
0.3	0.7	0.1	0.9	0.1	0.9	0.1	0.9

D		E		F		G		H		I	
1	2	1	2	1	2	1	2	1	2	1	2
0.1	0.9	0.1	0.9	0.1	0.9	0.1	0.9	0.1	0.9	0.1	0.9

4. (10 pts) Now, assume we do not have any knowledge about the probability table for the nodes in the network, but we have the following 12 observations. Find a way to estimate the probability table associated with the nodes **A** and **H**.

A	B	C	D	E	F	G	H	I
1	1	2	2	2	1	1	1	1
1	2	1	1	2	1	1	1	2
2	2	2	1	2	2	1	2	1
1	1	2	1	2	1	1	2	2
1	2	1	1	1	1	2	1	1
2	2	1	2	1	2	2	1	2
2	1	2	2	1	2	2	2	1
2	2	2	1	2	1	2	2	2
1	1	1	1	2	2	1	1	1
1	1	1	1	2	1	1	1	2
1	2	1	2	2	1	2	1	2
2	2	1	2	1	2	2	1	1

5. (20 pts) Based on the above observations, you would like to find a good Bayesian network structure to model the data. You started with the initial structure shown on the previous page, and decided to delete the edge between **H** and **I**. Is the resulting new structure (after deleting the single edge between **H** and **I** from the original graph) better than the original structure in terms of BIC score? Clearly explain the reason. (Hint: Try to find a short answer.)