50.007 Machine Learning, Fall 2015
Homework 5

Due Monday 7 Dec 2015, 5pm

This homework will be graded by Dinh Quang Thinh.
Please submit the hard copy of your solutions to 2.716-S2.

Note that no late submissions will be allowed for this homework
as the solutions will be released shortly after the deadline for final exam preparations.

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: $\{1, 2\}$.
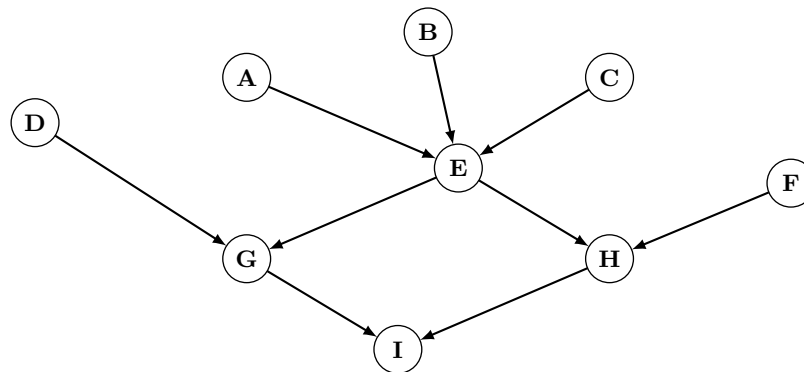
1. (10 pts) How many effective parameters are needed to for this Bayesian network? What would be the number of effective parameters for the same network if node **D** and **F** can take 4 different values: $\{1, 2, 3, 4\}$, and all other nodes can only take 2 different values: $\{1, 2\}$?

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**?

3. (10 pts) If we have the following probability tables for the nodes. Compute the following probability. Clearly write down all the necessary steps.

$$P(\mathbf{E} = 1 | \mathbf{C} = 2)$$

			E	
A	B	C	1	2
1	1	1	0.1	0.9
1	1	2	0.3	0.7
1	2	1	0.5	0.5
1	2	2	0.0	1.0
2	1	1	0.9	0.1
2	1	2	0.6	0.4
2	2	1	0.4	0.6
2	2	2	0.5	0.5

A	
1	2
0.2	0.8

B	
1	2
0.5	0.5

C	
1	2
0.2	0.8

D	
1	2
0.1	0.9

F	
1	2
0.3	0.7

		G	
D	E	1	2
1	1	0.1	0.9
1	2	0.4	0.6
2	1	0.5	0.5
2	2	0.5	0.5

		H	
E	F	1	2
1	1	0.1	0.9
1	2	0.4	0.6
2	1	0.5	0.5
2	2	0.5	0.5

		I	
G	H	1	2
1	1	0.1	0.9
1	2	0.9	0.1
2	1	0.1	0.9
2	2	0.9	0.1

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	1	2	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.)