CS4786/5786: Machine Learning for Data Science, Spring 2016
4/27/2016: Assignment 3: Graphical models

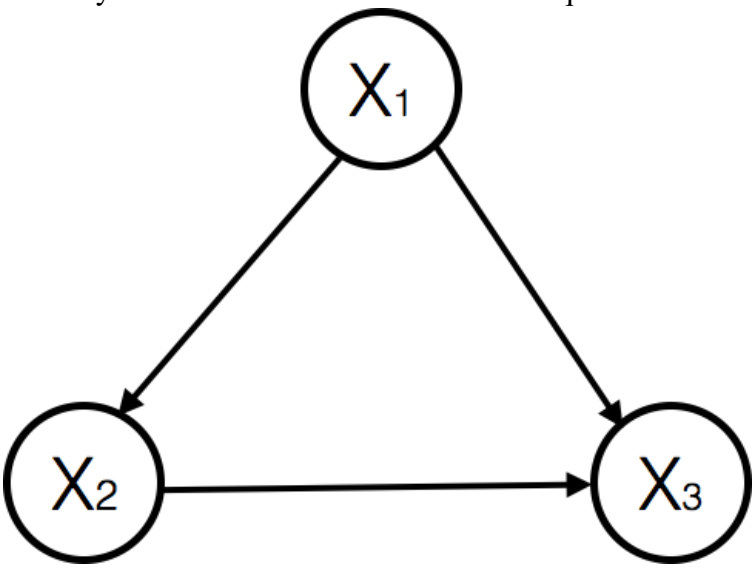
Due Tuesday May 9th at 11:59pm on CMS. Submit what you have at least once by an hour before that deadline, even if you havent quite added all the finishing touches CMS allows resubmissions up to, but not after, the deadline. If there is an emergency such that you need an extension, contact the professors.

You may work in groups of one up to four, for the same reasons as have been explained in previous assignments. Please ensure that each member of the group can individually defend or explain your groups submission equally well.

Students may discuss and exchange ideas with students not in their group, but only at the conceptual level. As discussed in previous assignments, we distinguish between "merely" violating the rules and violating academic integrity. The way to avoid violating academic integrity is to always document any portions of work you submit that are due to or influenced by other sources, even if those sources werent permitted by the rules.

1 Questions

The Bayesian network we consider for this question is the following:



along with the following probability tables:

$X_1 = 1$	$X_1 = 0$
0.1	0.9

X_1	$X_2 = 1$	$X_2 = 0$
1	0.3	0.7
0	0.4	0.6

X_1	X_2	$X_3 = 1$	$X_3 = 0$
1	1	0.2	0.8
1	0	0.5	0.5
0	1	0.5	0.5
0	0	0.8	0.2

*Q1 (***Variable Elimination).**

1. **Write down the steps for variable elimination generally in terms of** $P(\text{Variable}|\text{Parents})$ **(ie. don't plug in the values of probabilities yet)**

2. **Plug in values from the table and compute marginal probabilities** $P(X_1|X_2 = 1)$ **and** $P(X_3|X_2 = 1)$ **for observation** $X_2 = 1$

*Q2 (***Message Passing).** Given observation, $X_2 = 1$:

1. **Write down the messages passed between these three nodes for 4 rounds (not including round 0 which is trivial). Write them down in terms of** $P(\text{Variable}|\text{Parents})$

2. **Next plug in the values from tables and write down the values of all the 6 messages passed on round 4**

*Q3 (***Inference through sampling).** Again the observation is $X_2 = 1$.

1. **Use the Likelihood weighted (importance weighted) sampling procedure for this Bayesian Network. Submit** 100 **samples in file "sample.csv" where each line is a vector of length** 3 **separated by commas and you will provide** 100 **lines, one for each sample.**

2. **Based on the sample, compute empirical marginal distributions for** $P(X_1|X_2 = 1)$**,** $P(X_3|X_2 = 1)$ **and report them.**