CSL302– Artificial Intelligence
Lab 5

Due on 21/4/2016 11.55pm

Instructions: Upload to your moodle account one zip file containing the following. Please do not submit hardcopy of your solutions. In case moodle is not accessible email the zip file to the instructor at ckn@iitrpr.ac.in. Late submission is not allowed without prior approval of the instructor. You are expected to follow the honor code of the course while doing this homework.

1. **You are allowed to work in teams of size at most 2 for this programming lab**.

2. A neatly formatted PDF document with your answers for each of the questions in the homework. You can use latex, MS word or any other software to create the PDF.

3. Include a separate folder named as 'code' containing the scripts for the homework along with the necessary data files.

4. Include a README file explaining how to execute the scripts.

5. Name the ZIP file using the following convention rollnumber1rollnumber2hwnumber.zip

---

1. Probabilistic Inference using Bayesian Networks

In this lab you will implement two techniques for drawing inference from a Bayes Net.

- Exact inference using variable elimination
- Approximate inference using rejection sampling

For implementing exact inference, you will have to implement the following functions

- reduce – retains only those entries in the factor that support the evidences.
- join – joins two factors
- sum– sums out a variable from the factor
- normalize – normalizes the factor

For implementing rejection sampling, you will have to write a function that selects the value for a variable from a given probability distribution aka sampling!

Given a text file containing the description of the Bayes network, your code must first internalize the network and answer queries provided in a separate file. For simplicity we will assume the Bayes net to represent only Boolean variables.

Input

The first input to your code will be the name of the text file containing the description of the Bayes network. This is as follows

N
X1 Parents of X1 separated by space
Conditional probability table
X2 Parents of X2 separated by space
Conditional probability table
…

The first line indicates the number of random variables in the network. The variables and their parents, are presented in the second line, followed by the conditional probability table. We will use integers to represent the random variables. For example

3 – there are three random variables in the network
1 2 – Random variable 1 has a single parent 2
0.8 0.2 – P(1=truel2=true) = 0.8 and P(1=falsel2=true) = 0.2
0.4 0.6 – P(1=truel2=false) = 0.4 and P(1=falsel2=false) = 0.6
3 – Random variable 3 has no parents
0.2 0.8 – P(3=true) = 0.2 and P(3=false) = 0.8
2 – Random variable 2 has no parents
0.6 0.4 – P(2=true) = 0.6 and P(2=false) = 0.4

The second input to your code will be another text file containing inference queries and the choice of the inference technique. Each line in the text file will contain an inference query. This line adheres to the following format

technique q query variables separated by space e evidence variables separated by space

For example, if we want to use variable elimination to estimate P(1=true,2=truel3=false) for the previous network the query line will look like

ve q 1 2 e ~3

If we have to perform the same inference using rejection sampling, then the query would be

rs q 1 2 e ~3

We will use ~ symbol to indicate that the variable is false. To make it simple, we will also assume that we are only interested in obtaining probability values instead of a distribution.

The output of your code will be probability value, one per line for every query.

Perform a study to investigate the convergence of the probabilities estimated from rejection sampling as a function of number of samples generated. The true probability can be obtained via variable elimination. Include as a pdf, your analysis of this experiment.

Included as part of the assignment are sample Bayesian networks and the corresponding queries

Please use Python or C++ as the language of implementation.