PROJECT #4: UNCERTAIN REASONING

Please take care to complete all parts of this project independently. Do not use any solutions found online or written by other students. If you are struggling please go to office hours, post to Piazza, and/or send us email. Please submit a single PDF file `project4.pdf` to Canvas for all handwritten problems (1-6). The PDF file can contain a scan of handwritten or typewritten/typeset solutions, and any supporting graphics. Please also submit a single code file `project4.tar.gz` to Canvas containing your Python or C++ code solution to Problem #7.

1. **R&N Problem 13.10:** (10 points)
   Deciding to put probability theory to good use, we encounter a slot machine with three independent wheels, each producing one of the four symbols $BAR, BELL, LEMON,$ or $CHERRY$ with equal probability. The slot machine has the following payout scheme for a bet of 1 coin (where "?" denotes that we don't care what comes up for that wheel):

   > $BAR/BAR/BAR$ pays 20 coins
   > $BELL/BELL/BELL$ pays 15 coins
   > $LEMON/LEMON/LEMON$ pays 5 coins
   > $CHERRY/CHERRY/CHERRY$ pays 3 coins
   > $CHERRY/CHERRY/?$ pays 2 coins
   > $CHERRY/?/?$ pays 1 coin

   a. Compute the expected "payback" percentage of the machine. In other words, for each coin played, what is the expected coin return?
   b. Compute the probability that playing the slot machine once will result in a win.
   c. Estimate the mean and median number of plays you can expect to make until you go broke, if you start with 10 coins. You can run a simulation to estimate this, rather than trying to compute an exact answer.
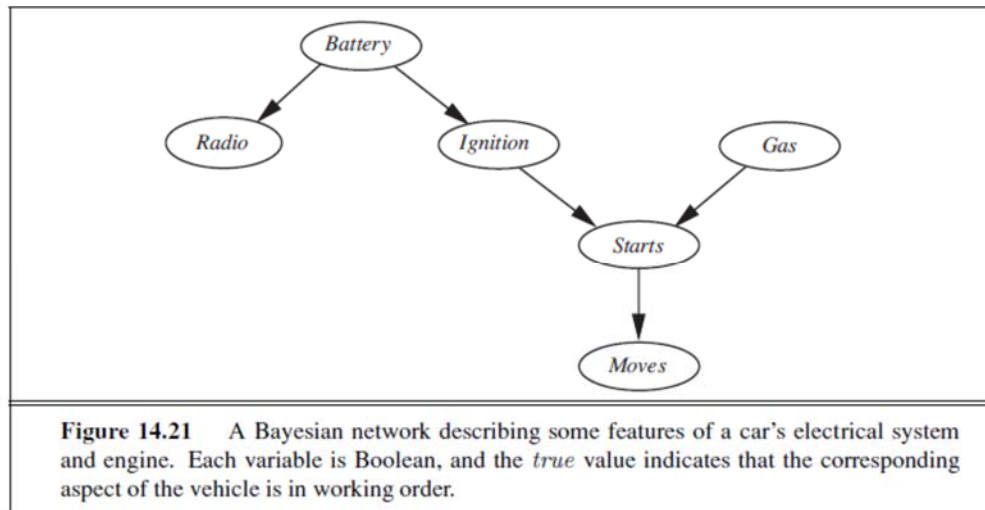
2. **R&N Problem 13.13:** (10 points)
   Consider two medical tests, A and B, for a virus. Test A is 95% effective at recognizing the virus when it is present, but has a 10% false positive rate (indicating that the virus is present, when it is not). Test B is 90% effective at recognizing the virus, but has a 5% false positive rate. The two tests use independent methods of identifying the virus. The virus is carried by 1% of all people. Say that a person is tested for the virus using only one of the tests, and that test comes back positive for carrying the virus. Which test returning positive is more indicative of someone really carrying the virus? Justify your answer mathematically.

3. **Car Diagnosis Bayes Net (Modified from R&N Problem 14.8):** (10 points)
   Consider the network for car diagnosis shown in Figure 14.21 (reproduced below).
   a. Extend the network with Boolean variables *IcyWeather* and *StarterMotor*. Sketch the updated Bayesian network with eight nodes.
   b. How many independent values are contained in the full joint probability distribution for eight Boolean nodes, assuming that no conditional independence relations are known to hold among them?
   c. How many independent probability values do your Bayesian network tables contain?
   d. What nodes (if any) are conditionally independent of *Starts* given *Moves* and *Battery*?

Figure 14.21 A Bayesian network describing some features of a car's electrical system and engine. Each variable is Boolean, and the *true* value indicates that the corresponding aspect of the vehicle is in working order.
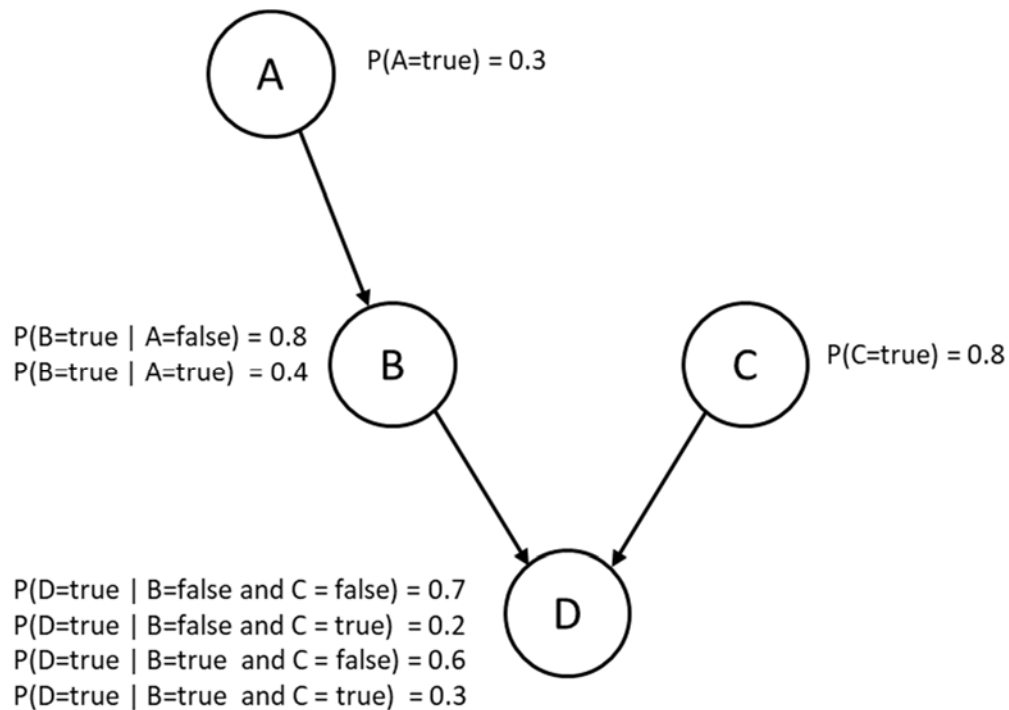
4. **R&N Problem 14.11:** (10 points)
   In your local nuclear power station, there is an alarm that senses when a temperature gauge exceeds a given threshold. The gauge measures the temperature of the core. Consider the Boolean variables A (alarm sounds), F_A (alarm is faulty), and F_G (gauge is faulty) and the multivalued nodes G (gauge reading) and T (actual core temperature).
   a. Draw a Bayesian network for this domain, given that the gauge is more likely to fail when the core temperature gets too high.
   b. Is your network a polytree? Why or why not?
   c. Suppose there are just two possible actual and measured temperatures, normal and high; the probability that the gauge gives the correct temperature is x when it is working, but y when it is faulty. Give the conditional probability table associated with G.
   d. Suppose the alarm works correctly unless it is faulty, in which case it never sounds. Give the conditional probability table associated with A.
   e. Suppose the alarm and gauge are working and the alarm sounds. Calculate an expression for the probability that the temperature of the core is too high, in terms of the carious conditional probabilities in the network.

5. **Simple Bayes Computations:** Consider the Bayesian network drawn below. (10 points)

A   P(A=true) = 0.3

P(B=true | A=false) = 0.8
P(B=true | A=true) = 0.4   B

C   P(C=true) = 0.8

D

P(D=true | B=false and C = false) = 0.7
P(D=true | B=false and C = true) = 0.2
P(D=true | B=true and C = false) = 0.6
P(D=true | B=true and C = true) = 0.3

Show your work for the following calculations.

a. Compute $P(A = true$ and $B = false$ and $C = false$ and $D = true)$.
b. Compute $P(D = true \mid A = false$ and $B = true$ and $C = true)$.
c. Compute $P(A = false \mid B = false$ and $C = true$ and $D = false)$.
d. Compute $P(B = true \mid A = true$ and $C = false)$.
e. Compute $P(B = false)$.

6. **Exam Bayes Net (15 points):** Consider a problem in which a professor wants to determine whether a student has understood material based on an exam score. The below figure illustrates a possible Bayes net that can help answer this question, along with symbol definitions. Associated probabilities are listed below as well.
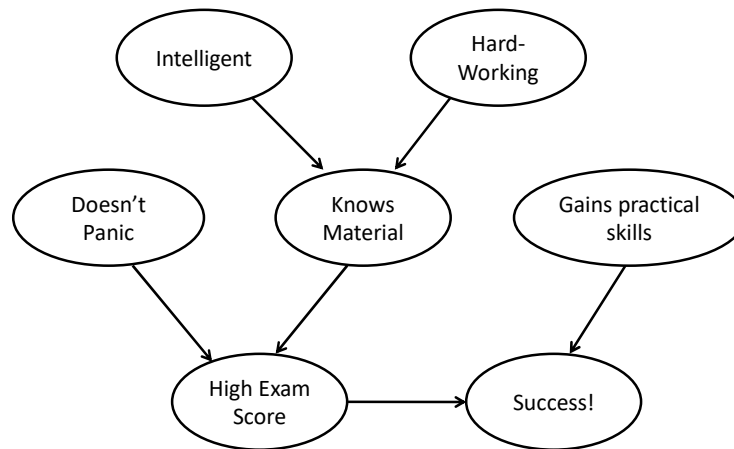
Intelligent = I, Hard working = H, Doesn't panic = DP, High Exam Score = Ex,
Knows material = KM, Gains practical skill = PS, Success! =S

Probability data:
P(I) = 0.75; P(H) = 0.6; P(DP)=0.4; P(PS)=0.8
P(KM|I,H)=1.0; P(KM|I,~H)=0.4; P(KM|~I,H)=0.6; P(KM|~I,~H)=0.05
P(S|PS,Ex)=0.8, P(S|~PS,Ex)=0.7, P(S|PS,~Ex)=0.7, P(S|~PS,~Ex)=0.3
P(Ex|DP,KM)=0.85; P(Ex|~DP,KM)=0.7; P(Ex|DP,~KM)=0.2; P(Ex|~DP,~KM)=0.1

a. Given a high exam score (Ex), which variables are conditionally independent of intelligent (I)?
b. Given knows material (KM) as evidence, which node(s) are conditionally independent of success (S)?
c. Given success (S), which node(s) are conditionally independent of high exam score (Ex)?
d. Given no evidence, compute P(KM).
e. Compute P(S | KM), the probability of Success given Knows Material.
f. Compute P(PS | S), the probability of gaining Practical Skill given exam Success.
g. Compute P(KM | S), the probability of Knows Material given exam Success.

7. **Bayes Net Exact Inference -- Coding** (35 points)

Implement the ENUMERATION-ASK(X, E, BN) function shown in R&N Figure 14.9, reproduced on the next page, within a program or script that manages a single test case and prints the result to the screen. To simplify code for this problem, assume all nodes are labeled by single characters as in Problem 5 above, and assume all variables have Boolean values T (True), F (False). The Bayes Net (BN) will be defined by file bn.txt. The specific test case query will be defined in file input.txt. Output Q(X) for X=T should be printed to the screen. Examples of each formatted input file are provided below for the Problem 5 Bayes Net. The grader will use one easy Bayes Net and one larger Bayes Net. We will only test polytree cases that run successfully given a correct implementation.

```
function ENUMERATION-ASK(X, e, bn) returns a distribution over X
    inputs: X, the query variable
            e, observed values for variables E
            bn, a Bayes net with variables {X} ∪ E ∪ Y   /* Y = hidden variables */

    Q(X) ← a distribution over X, initially empty
    for each value xᵢ of X do
        Q(xᵢ) ← ENUMERATE-ALL(bn.VARS, eₓᵢ)
            where eₓᵢ is e extended with X = xᵢ
    return NORMALIZE(Q(X))
───────────────────────────────────────────────────────────────────
function ENUMERATE-ALL(vars, e) returns a real number
    if EMPTY?(vars) then return 1.0
    Y ← FIRST(vars)
    if Y has value y in e
        then return P(y | parents(Y)) × ENUMERATE-ALL(REST(vars), e)
        else return Σᵧ P(y | parents(Y)) × ENUMERATE-ALL(REST(vars), eᵧ)
            where eᵧ is e extended with Y = y
```

Figure 14.9   The enumeration algorithm for answering queries on Bayesian networks.

Example `bn.txt` file:

```
% Random Variables
A, B, C, D
% Graph Edges (From, To)
A, B
B, D
C, D
% Probability values
P(A=T)=0.4
P(C=T)=0.7
P(B=T|A=F)=0.9
P(B=T|A=T)=0.3
P(D=T|B=F,C=F)=0.8
P(D=T|B=F,C=T)=0.3
P(D=T|B=T,C=F)=0.5
P(D=T|B=T,C=T)=0.1
```

An example `input.txt` file for *P(D/A=T,C=F)*:

```
% Query random variable (print the probability X=T)
D
% Evidence vector (insert blank line when no evidence is given)
A=T, C=F
% End of input
```