Assignment 2: Bayesian Networks

CS486/686 – Winter 2026

Out: January 26, 2026
Due: February 6 (11:59 pm), 2026.

Submit an electronic copy of your assignment via LEARN. Late submissions incur a 2% penalty for every rounded up hour past the deadline. For example, an assignment submitted 5 hours and 15 min late will receive a penalty of ceiling(5.25) * 2% = 12%.

In this assignment, you will implement a simple medical inference tool in 3 steps. First, you will implement the variable elimination algorithm. Second, you will define your medical inference tool as a Bayesian network and answer inference queries regarding the likelihood of some symptoms and a disease. Computations in your Bayesian network should be done with your implementation of the variable elimination algorithm. However, if you are not able to complete your implementation of variable elimination, you can also do the computations by hand since they are relatively simple.

1. **[0 pts]** Implement the variable elimination algorithm in Python by filling in the functions in the skeleton code variable_elimination.py (provided on the course website). Those functions perform operations on factors, which are multidimensional arrays. Hence use numpy multidimensional arrays as your main data structure. If you are not familiar with numpy, go through the following tutorial: `https://numpy.org/doc/stable/user/quickstart.html`.

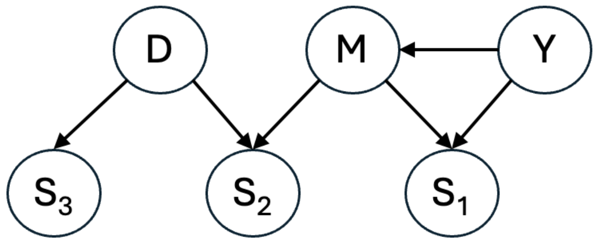
   Tip: test each function individually using simple examples from the lecture slides. If you wait till the end to test your entire program it will be much harder to debug.

   What to hand in: hand in your code. Note that there are no marks given for Question 1. However, in Questions 2, part of the marks will be given for using your variable elimination algorithm to do the computations.

2. **[72 pts]** Suppose you are building a simple medical inference tool. You plan to use the following information:

   - 20% of the population is old. About 10% of older people take medication M, where as only 4% of young people take medication M. medication M can produce symptoms 1 and 2 as side effects. When taking medication M, 20% of old people exhibit symptom 1, whereas only 5% of young people exhibit symptom 1. However, when someone does not take medication M, then the likelihood of exhibiting symptom 1 is only 2%, regardless of age.
   - Disease D often causes symptoms 2 and 3, but medication M can be taken to reduce the occurrence of symptom 2. More precisely, 50% of people suffering from disease D who are not treated with medication M exhibit symptom 2. This decreases to 10% when some suffering from disease D takes medication M. In contrast, only 1% of people who do not suffer from disease D exhibit symptom 2, regardless of whether they take medication M. About 20% of people suffering from disease D exhibit symptom 3 while this drops to 5% among individuals who do not have disease D.

   (a) **[12 pts]** Construct a Causal Bayes Network to answer inference queries. More precisely, define the conditional probability tables for the following Causal Bayesian Network.



Tip: The probabilities given in the text above can be inserted directly into corresponding conditional probability tables (i.e., no computation needed to obtain the conditional probability tables).

What to hand in: Show the conditional probability tables associated with each node in the graph. This network should encode the information stated above. Your network should contain exactly six nodes, corresponding to the following binary random variables:

- D – person suffers from disease D.
- M – person takes medication M.
- Y – person is young.
- S_1 – person exhibits symptom 1.
- S_2 – person exhibits symptom 2.
- S_3 – person exhibits symptom 3.

The arcs defining your Bayes Network should accurately capture the probabilistic dependencies between these variables.

(b) **[30 pts]** What is the prior probability (i.e., before observing anything) that someone has symptom 1? What is the posterior probability of symptom 1 after observing that someone has symptom 3, but not symptom 2?

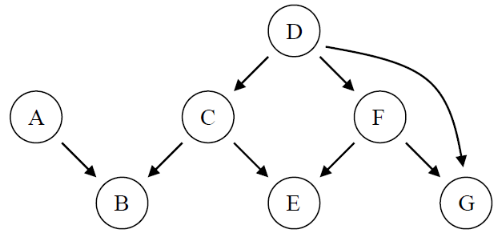
What to hand in: Indicate what queries (i.e., $Pr(variables|evidence)$) you used to compute those probabilities. Whether you answer the queries by hand or using the code you wrote for Question 1, provide a printout of the factors computed at each step of variable elimination (as done in the lecture slides). Use the following variable ordering when summing out variables in variable elimination: Y, S_1, M, S_2, D, S_3. Note that a maximum of two thirds of the marks are earned if you answer correctly the question by doing the computations by hand instead of using your program.

(c) **[30 pts]** Similar to Question 2b, suppose that someone has symptom 3, but not symptom 2, and we are interested in inferring the probability of symptom 1. What is your revised belief about symptom 1 when

- This person tells you that she has been taking medication M.
- Intervening by giving this person medication M.

What to hand in: Same as for Question 2b.

3. **[28 pts]** You have just been hired as a consultant for a car manufacturer. The company would like to improve their maintenance service at its dealerships by assisting the mechanics with an automated fault diagnosis tool. After talking with several experts, you've built the following Bayesian network. Each node is a Boolean variable that corresponds to the status (e.g., working or not) of a car component and each edge indicates a probabilistic dependency for failure.



(a) **[14 pts]** Suppose a mechanics would like to know what components may influence the functioning of other components. Answer the following questions and give a brief justification based on the D-separation rules.

i. Are D and G independent?
ii. Are D and G independent given F?
iii. Are A and G independent?
iv. Are A and G independent given B?
v. Are A and G independent given B and C?
vi. Are A and G independent given B and D?
vii. Are A and G independent given B, D and E?

(b) **[14 pts]** Suppose a mechanics would like to know $\Pr(C|A = true, E = false)$. What is the subset of relevant variables that is sufficient to answer this query? Give a brief justification based on the rules to identify relevant variables.