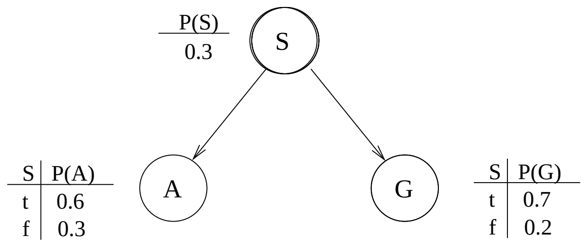
# CMSC 421 Homework 4

Due Date: Thursday, December 4 before 5PM

1. **Bayesian Networks [15 points]**

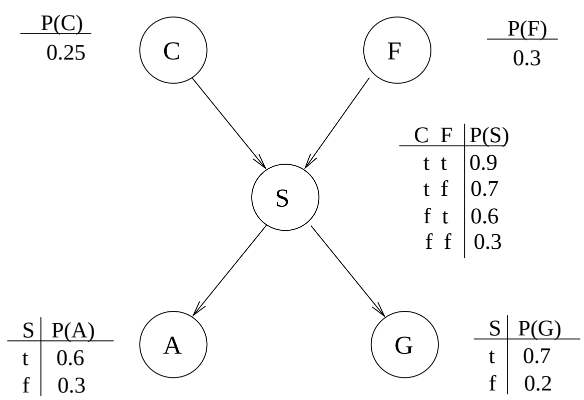
   A cmsc421 student notices that people that drive SUVs (S) consume large amounts of gas (G) and are involved in more accidents than the national average (A). They have constructed the following Bayesian network:




   (a) Compute $P(A)$ in two ways:
      i. By generating the *entire joint distribution* over these variables and explicitly summing the appropriate entries.
      ii. Using the variable elimination algorithm

   (b) Using conditional independence, compute $P(\bar{g}, a \mid s)$ and $P(\bar{g}, a \mid \bar{s})$. Then use Bayes rule to compute $P(s|\bar{g}, a)$.

   (c) The enterprising cmsc421 student notices that there are two types of people that drive SUVs, people from California (C) and people with large families (F). After collecting some statistics, the student arrives at the following form for the Bayesian network:




   Using the chain rule from probability compute the probability $P(\bar{g}, a, s, c, \bar{f})$.

   (d) Without explicitly generating the entire probability distribution, compute $P(s)$ and $P(s \mid c)$ and $P(s \mid \bar{c})$. (Hint: consider different values of F.)

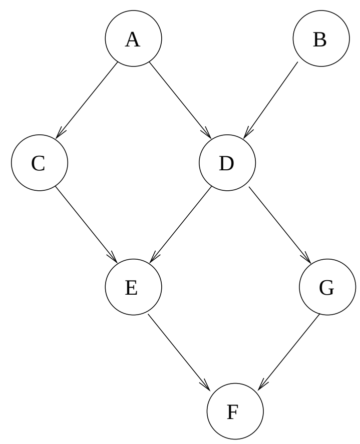
   (e) Show how you can compute the value of $P(G)$ in the complete distribution without doing any additional work (i.e., based solely on the work you have already done). Explain.

   (f) Using the rules for determining when two variables are independent of each other given evidence, answer the following (T/F) for the BN above:

- I(C,G)
- I(F,A | S)
- I(C,F)
- I(A,G)
- I(C,F | S)
- I(C,F | A)

2. **[10 points] Bayesian Networks**

   Given the following Bayesian network:




   (a) Suppose we have the query P(G) and the following elimination ordering: first A, then B, C, D, E, F. Show what the variable elimination algorithm will do on this graph by listing for each variable eliminated, the new factor generated and the factors used to generate the new factor. The factors should be written in the form $f_i(X_1, \ldots, X_k)$ for some set of variables $X_1, \ldots, X_k$.

   (b) Elimination ordering matters to the complexity of inference. Assume all variables have the same domain size $m > 2$. Find an elimination ordering for the same query whose time complexity as a function of m is worse than the ordering given above. As in the above problem, show what the factors generated at each step.

3. **[15 points] Utilities**

   Russell and Norvig, 16.2.

4. **[15 points] Decision Making Under Uncertainty**

   Russell and Norvig, 16.11.

5. **[15 points] Decision tree learning**

   The following table gives a data set for deciding whether to play or cancel a ball game, depending on the weather conditions.

| Outlook | Temp (F) | Humidity (%) | Windy? | Class |
|---|---|---|---|---|
| sunny | 75 | 70 | true | Play |
| sunny | 80 | 90 | true | Don't Play |
| sunny | 85 | 85 | false | Don't Play |
| sunny | 72 | 95 | false | Don't Play |
| sunny | 69 | 70 | false | Play |
| overcast | 72 | 90 | true | Play |
| overcast | 83 | 78 | false | Play |
| overcast | 64 | 65 | true | Play |
| overcast | 81 | 75 | false | Play |
| rain | 71 | 80 | true | Don't Play |
| rain | 65 | 70 | true | Don't Play |
| rain | 75 | 80 | false | Play |
| rain | 68 | 80 | false | Play |
| rain | 70 | 96 | false | Play |

(a) At the root node for a decision tree in this domain, what are the information gains associated with the Outlook and Humidity attributes? (Use a threshold of 75 for humidity (i.e., assume a binary split: humidity $\leq 75$ humidity *ge* 75.)

(b) Again at the root node, what are the gain ratios associated with the Outlook and Humidity attributes (using the same threshold as in (a))?

(c) Suppose you build a decision tree that splits on the Outlook attribute at the root node. How many children nodes are there are at the first level of the decision tree? Which branches require a further split in order to create leaf nodes with instances belonging to a single class? For each of these branches, which attribute can you split on to complete the decision tree building process at the next level (i.e., so that at level 2, there are only leaf nodes)? Draw the resulting decision tree, showing the decisions (class predictions) at the leaves.

6. **[15 points] Version Spaces**

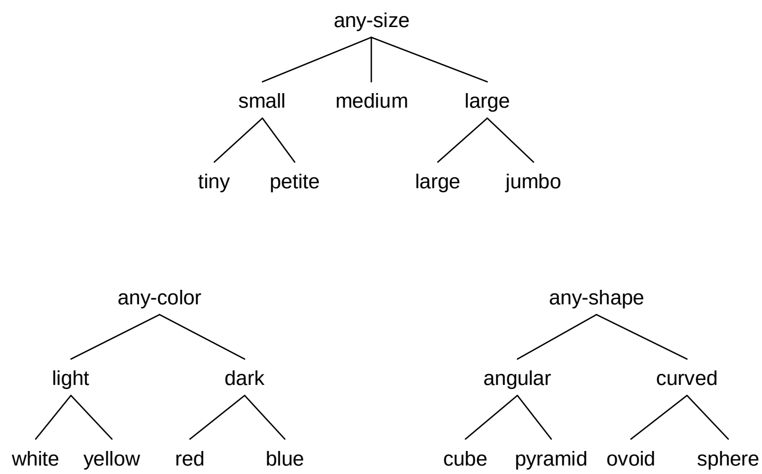
Consider the version space defined by the concepts that are a conjunction of the following three attributes: size, color and shape. Suppose we have the following generalization hierachies for each of the attributes:

Examples of legal concepts in this domain include [medium light angular], [large red sphere], and [any-size any-color any-shape]. No disjunctive concepts are allowed, other than the implicit disjunctions represented by the internal nodes in the attribute hierarchies. (For example, the concept [large [red v yellow] curved] isn't allowed.)

(a) Consider the initial version space for learning in this domain. What is the G set? How many elements are in the S set? Give one representative member of the initial S set.

(b) Suppose the first example I1 is a positive example, with attribute values [jumbo yellow ovoid]. After processing this instance, what are the G and S sets?

(c) Now suppose the learning algorithm receives example I2, a negative example, with attribute values [jumbo red pyramid]. What are the G and S sets after processing this example?




(d) If learning ends at this point, how many possible concepts remain in the version space?

(e) For this question, you should start with the version space that remains after I1 and I2 are processed. For each of the following combinations of instance type and events, give a single instance of the specified type that would cause the indicated event, if such an instance exists. If no such instance exists, explain why.

i. A positive instance that causes the version space to collapse.

ii. A negative instance that causes the version space to collapse.

iii. A positive instance that reduces the size of the version space to a single concept.

iv. A negative instance that reduces the size of the version space to a single concept.

7. **Statistical Learning [15 points]**

Russell and Norvig, 20.4.