EECS 16ML Fa/Sp 20xx

Machine Learning in Practice
Week 5: Bayesian Networks

Quiz x

1. Why is a Bayesian network a directed graph?

   Answer: The edges in our graph represent the *real-world* dependence of a node on its parent. Within Bayes Nets, parents are not dependent on their children (if one variable somehow determines another in the real world, then the other variable cannot also determine the original), so our edge must have a strict direction from parent to child.

   Key detail: Note the key distinction between probabilistic dependence and real-world dependence. For example, if my sprinkler is on, there's a better-than-not chance that it's *not* raining today, so there is some probabilistic dependence of the rain on my sprinkler. However, my sprinkler has no actual bearing on whether or not it rains, so there is no real-world dependence.

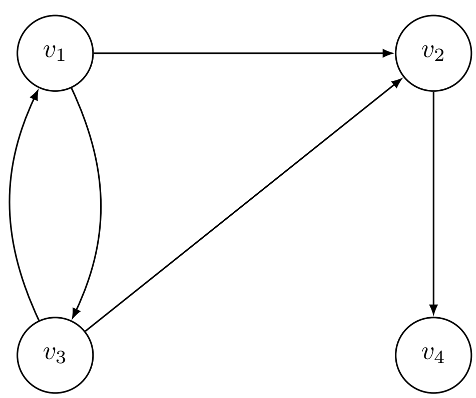
2. What is the reason Bayesian Networks are not allowed to form cycles?

   Answer: Assume there is a cycle between nodes A and B. When performing inference, we take the product of all the conditional probabilities given a data sample. If we were then to allow a cycle in our Bayes Net, when performing inference, we would find that we are continuously finding $P(A|B)$ and $P(B|A)$ which result in circular logic making our calculation impossible.

   More directly, the existence of a path from A to B implies that A has some bearing on B (this is how we interpret Bayesian networks in the real world). If there is a cycle, this means there is also a path from B to A, and we can't have each of A and B affecting the other.

   Key detail: Maybe you've thought of a scenario such as a heating system that only turns on when its cold, so the temperature affects the power state of the heating system, and the heating system affects the temperature, giving a cycle. The reason this doesn't work is because the random variables we would include in our Bayes' net are the power state of the heating system and the temperature *at time* t, $t+1$, *and so on.* The power state of the heating system at time t affects the temperature at time $t+1$ but not at time t, and the temperature at time t affects the power state of the heating system at time $t+1$ but again not at time t.

3. Take a look at the following Bayesian network structure:

   Identify an issue in the structure of this network and describe how the network could be modified to resolve the issue. Then, identify two pairs of independent variables in your modified Bayes' net.

   Answer: There is a cycle consisting of the nodes v_1 and v_3! To fix this, we could remove either of the two edges between these nodes. In either case, the pairs (v_1, v_4) and (v_3, v_4) are independent.

Key detail: One of the most important components of being able to deploy a Bayesian network-based model in industry is having the ability to *interpret* a Bayes' net. Be sure to understand exactly what an arrow in the above graph signifies (recall questions 1 and 2) and how we can tell if two variables are independent. Potential reading: this page.

4. Describe what a conditional probability table (CPT) stores and why this is useful.

   Answer: A conditional probability table stores the likelihood values conditioned for every possible combination of values of the parent or parents of a certain node. This is essential for doing fast inference. In other words, for each possible combination of values of the parents of a certain node, it stores the distribution of that certain node conditional on the parents taking on those particular values.

5. If a node that could take on 5 different values and has 4 parents which can each can take on three possible values, what is the size of the CPT for that node?

   Answer: $5 * 3^4 = 5 * 81 = 405$

6. What is variable elimination and why is it useful?

   Answer: Variable elimination is an algorithm used in computing the marginal distribution of a certain subset of the variables in a Bayesian network. This can be easily adapted to compute arbitrary conditional probabilities within the network. Variable elimination allows us to dramatically increase the speed of our inference calculations by splitting independent variables into different sums and treating them in different steps of the algorithm.

7. Write down the steps of the Maximum Likelihood Parameter Learning Algorithm given complete data.

   Answer: The steps are:

   - Distinguish what your query variables are and what your evidence variables are.
   - Initialize `evidenceSeen` to 0 and `querySeen` to 0.
   - Go through each data point. If it matches the evidence increment `evidenceSeen` by 1. If it also matches the query increment `querySeen` by 1 as well.
   - Return $\frac{\texttt{querySeen}}{\texttt{evidenceSeen}}$.

8. What technique should you use when you observe certain inference values as 0 or/and if you do not have a large data set?

   Answer: Laplace Smoothing.

9. Say your MLE estimate for a certain parameter θ is $\frac{7}{8}$. What does your estimated value change to if you instead used MLE with Laplace Smoothing where $k = 2$. Assume the domain size of the query variable in your parameter is 3.

   Answer: $\frac{7+2}{8+(2*3)} = \frac{9}{14}$

10. Write the equation for determining the branch weight between nodes X and Y in the Chow-Liu algorithm.

    Answer: The equation is:

$$I(u, v) = \Sigma_{u \in U} \Sigma_{v \in V} \mathbb{P}[u, v] \log_2 \frac{\mathbb{P}[u, v]}{\mathbb{P}[u]\mathbb{P}[v]}$$

11. What is the time complexity for the complete search and score structure learning method?

    Answer: Super-exponential