10-701/15-781, Machine Learning: Homework 3

Eric Xing, Tom Mitchell, Aarti Singh
Carnegie Mellon University

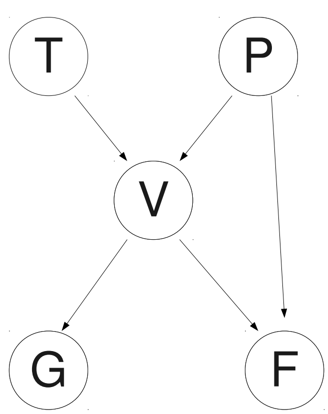
- The assignment is due at 10:30am (beginning of class) on **Wed, Jan 20, 2010**.
- Separate you answers into three parts, one for each TA, and put them into 3 piles at the table in front of the class. Don't forget to put both your name and a TA's name on each part.
- If you have question about any part, please direct your question to the respective TA who design the part.

1 Bayesian Networks [25 pts, Field Cady]

Bayesian networks are a convenient way to represent disitributions where the random variables are only partly dependent on each other.

1.1 Part a : Creating a Network [8 pts]

One useful aspect of Bayesian networks is that they can make use of expert knowledge about a domain. Let's take a (highly) simplified view of land-based ecosystems, and try to characterize all ecosystems in the world. Two of the most important aspects of an ecosystem are temperature(T) and precipition(P); an area can be either hot or cold, and either wet, moist or dry. The foundation of an ecosystem is its vegetation(V), and it is either sparse or dense, and either trees or shrubs/grasses (so 4 possibilities). More water and higher temperatures both make trees more likely and vegetation more dense. The presence grazing animals (G, boolean for whether or not there are grazing animals), like bison, depends on there being a lot of grass, but they don't really need rain or moderate temperatures. Frogs (F, boolean for whether there are frogs), on the other hand, need lots of water and prefer to live near trees. Construct a Bayesian network to describe the relationship between these random variables.



Answer :

1.2 Part b : Memory Efficiency [5 pts]

Perhaps the most important thing about Bayesian networks is that they require vastly less memory to store; in general it takes exponential space to store the dependencies among random variables, but a Bayes net can often reduce the complexity to linear. Calculate

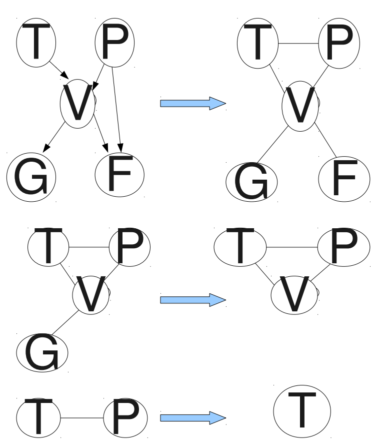
- How many real numbers are required to store the complete probability distribution for the random variables in part a, not making any conditional independence assumptions.
- The number of floats required to store the distribution using our conditional independence assumptions.

Answer :

- T, P, V, F and G have 2, 3, 4, 2, and 2 possibile values respectively, and we need to know the probability of all but one combination of them (the probability of the other combination is known since they all sum to 1). This is $2 \cdot 3 \cdot 4 \cdot 2 \cdot 2 - 1 = 96 - 1 = 95$ real numbers.
- Since T has 2 possibilities, we need 1 float for $PrT = hot$. For precipitation, which has 3 possibilities, we need 2 priors. There are 6 possibile combinations of T and P, and for each one V will need to store 3 floats since it has 4 possible states, totaling 18 floats for the conditional probabilities of V. P and V can have 12 combinations, and for each of them F will need a single float, for 12 total. Finally, G will need a float for each value of V, so 4. All tolled, this is 1+2+18+12+4=37 real numbers.

1.3 Part c : Inference [8 pts]

Bayes nets are fabulous for storing large distributions, but actually using them can be tricky. To get a feel for this, show graphically an elimination sequence for calculating $P(T|F)$.



Answer :

1.4 Part d : Generalizing Complexity [4 pts]

Imagine a Bayes net with binary random variables X_1, X_2,... , X_n. Let X_i be conditionally dependent on all X_j with $j < i$. Find a simple formula, as a function of n, for the number of real numbers needed to store this distribution.

This distribution is problematic because X_n and its neighbors have so many conditional dependencies. What if each X_i was conditionally dependent only on the one before it? What would the formula be then?

Answer : For the first distribution, we have the recurrence relation $N_{n+1} = N_n + 2^n = N_{n-1} + 2^{n-1} + 2^n = 2^0 + 2^1 + ... + 2^{n-1} = 2^n - 1$ so $N_n = 2^n - 1$.

For the second distribution, we need a single number for the first random variable, and 2 for each subsequent one, for a total of $2n - 1$.