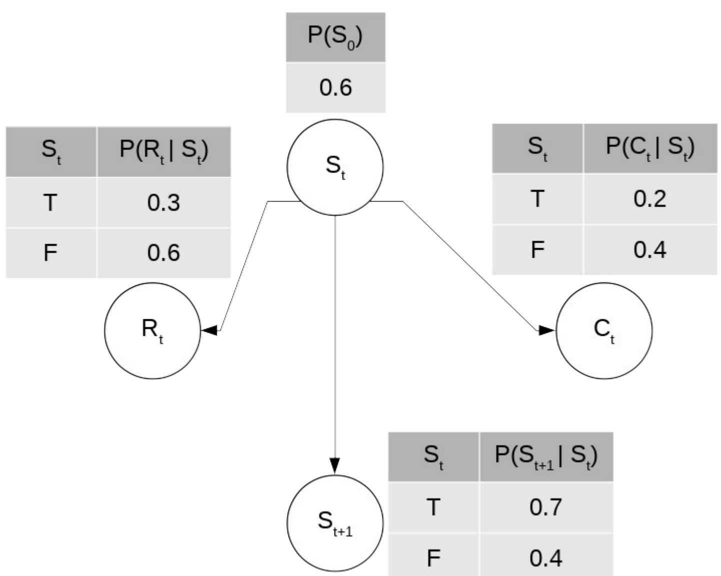
Max O’Cull
0028687211
CS 471
Homework 6
04/22/19

1. A professor wants to know if students are getting enough sleep. Each day, the professor observes whether the students sleep in class, and whether they have red eyes. The professor has the following domain theory:
    - The prior probability of getting enough sleep, with no observations, is 0.6.
    - The probability of getting enough sleep on night t is 0.7 given that the student got enough sleep the previous night, and 0.4 if not.
    - The probability of having red eyes is 0.3 if the student got enough sleep, and 0.6 if not.
    - The probability of sleeping in class is 0.2 if the student got enough sleep, and 0.4 if not.

    Formulate this information as a dynamic Bayesian network that the professor could use to filter or predict from a sequence of observations. Then reformulate it as a hidden Markov model that has only a single observation variable. Give the complete probability tables for the model.

    Dynamic Bayesian Network:

    Let there be 3 variables; S_t for getting enough sleep, E_t for having red eyes, and C_t for sleeping in class.



    To alter this DBN into an HMM, merge red eyes and sleeping into a single variable with four values by multiplying their probabilities together.

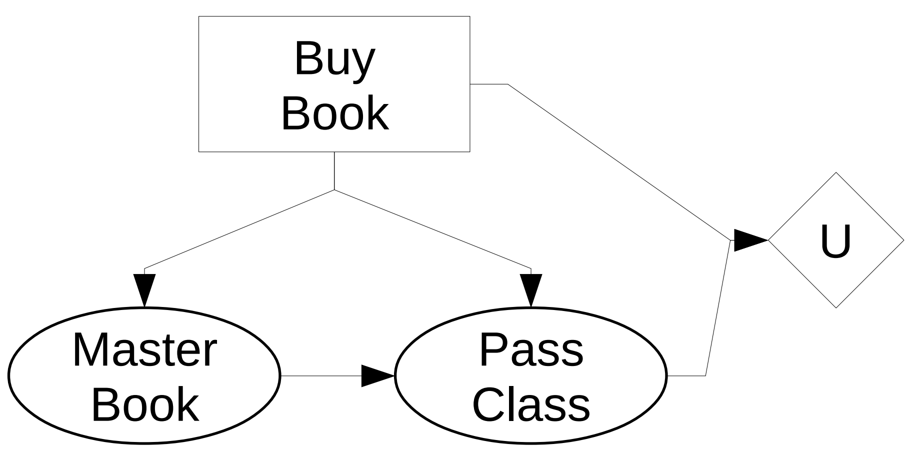
Max O'Cull
0028687211
CS 471
Homework 6
04/22/19

2. Consider a student who has the choice to buy or not buy a textbook for a course. We'll model this as a decision problem with one Boolean decision node, B, indicating whether the agent chooses to buy the book, and two Boolean chance nodes, M, indicating whether the student has mastered the material in the book, and P, indicating whether the student passes the course. Of course, there is also a utility node, U. A certain student, Sam, has an additive utility function: 0 for not buying the book and -$100 for buying it; and $2000 for passing the course and 0 for not passing. Sam's conditional probability estimates are as follows:

   $P(p \mid b, m) = 0.8$
   $P(m \mid b) = 0.8$
   $P(p \mid b, \neg m) = 0.4$
   $P(m \mid \neg b) = 0.6$
   $P(p \mid \neg b, m) = 0.7$
   $P(p \mid \neg b, \neg m) = 0.4$

   You might think that P would be independent of B given M, But this course has an open book final—so having the book helps.

   a. Draw the decision network for this problem.



   b. Compute the expected utility of buying the book and of not buying it.

$$P(p|b) = 0.8*0.8+0.4*0.2 = 0.72$$
$$P(p|\neg b) = 0.7*0.6+0.4*0.4 = 0.58$$

$$EU[b] = 0.72(2000-100)+(1-0.72)(-100) = 1340$$
$$EU[\neg b] = 0.58(2000-100)+(1-0.58)(-100) = 1060$$

Max O'Cull
0028687211
CS 471
Homework 6
04/22/19

Sam's expected utility of buying the book is $1340, while not buying it yields $1060.

c. What should Sam do?

Sam should buy the book because it offers the highest utility.

3. Consider an image classification problem. Suppose an algorithm first splits each image into n = 4 blocks (the blocks are non-overlapping and each block is at the same location and of constant size across all images)and computes some scalar feature value for each of the blocks (e.g., average intensity of the pixels within the block). Suppose that this feature is discrete and can take m = 10 values. The classification function classifies an image as 1 whenever each of the n feature values lies within some interval that is specific to this feature (i.e., the value of the first feature is between a_1 and b_1, the value of the second feature is between a_2 and b_2, and so on), and 0 otherwise. We would like to learn these intervals (a and b values for each interval) automatically based on a training set of images. All the other parameters such as locations and sizes of the blocks are not being learned. The following questions are helpful in understanding the requirements on the size of the training set.

a. What is the size of the hypothesis space H? Assume that only intervals with $a_i <= b_i$ are considered for learning.

When given a value a_i, the values of b_i may be anywhere from a_i to m, meaning there are $m-a_i+1$ values. The quantity of possible intervals is then $m+(m-1)+(m-2)+\cdots+1=\frac{m(m+1)}{2}$.

The problem states there are $n=4$ features, so then the size of the hypothesis space is $\left(\frac{m(m+1)}{2}\right)^n=55^4=9150625$.

b. Assuming noiseless data and that the function we are trying to learn is capable of perfect classification, give an upper bound on the size of the training set required to be sure with 99% probability that the learned function will have true error rate of at most 5%.

$$R \geq \frac{0.69}{\epsilon}(\log(9150625)+\log(\frac{1}{\delta}))=410.82$$

c. Compare |H| (the answer to question a) and the required training dataset size R (the answer to question b). Why does R not seem to be very affected by the number of possible hypotheses? What parameter makes R increase quickly and why?

The parameter ϵ effects R the most in comparison to the number of hypotheses because the hypotheses are logarithmic. This makes sense because the smaller

Max O'Cull
0028687211
CS 471
Homework 6
04/22/19

the error you desire, the much larger your dataset must become. Separately from how many hypotheses we have, a large training set which has been learned will likely classify independent test points consistently.

4. Two statisticians go to the doctor and are both given the same prognosis: A 40% chance that the problem is the deadly disease A, and a 60% chance of the fatal disease B. Fortunately, there are anti-A and anti-B drugs that are inexpensive, 100% effective, and free of side-effects. The statisticians have the choice of taking one drug, both, or neither. What will the first statistician (an avid Bayesian) do? How about the second statistician, who always uses the maximum likelihood hypothesis? The doctor does some research and discovers that disease B actually comes in two versions, dextro-B and levo-B, which are equally likely and equally treatable by the anti-B drug. Now that there are three hypotheses, what will the two statisticians do?

    The statistician which is an avid Bayesian would opt to take both anti-A and anti-B drugs. The statistician who always uses maxmimum likelihood hypothesis would only take the anti-B drug.

    If there are two strains of disease B, the Bayesian would still opt to take both the anti-A and anti-B drugs, while the other statistician who prefers maximum likelihood hypothesis would change to only taking the anti-A drug.