CS471- Homework 6

Due date: April 23 at 11:59pm EST

Introduction

The objective of this assignment is to get familiar with Dynamic Bayesian Networks, Utility Theory, and (machine learning topics here).

Submission

Your homework must be typed and must contain your name and Purdue ID. To submit your assignment, log into `data.cs.purdue.edu` (physically go to the lab or use ssh remotely) and follow these steps:

1. To ssh use the command: `ssh username@data.cs.purdue.edu`
2. Make a directory named `username-hw6` (all letters in lower case)
3. Copy your PDF and code inside it. To do it remotely use the comand from your computer:
   `scp ./path/to/your-file.pdf username@data.cs.purdue.edu:./remote/path/from-home-dir/`
4. Go to the directory containing `username-hw6` (e.g., if the files are in /homes/aporco/aporco-hw6, go to /homes/aporco), and execute the following command:

   ```
   turnin -c cs471 -p hw6 username-hw6
   ```

   (e.g. Aldo would use: `turnin -c cs471 -p hw6 aporco-hw6` to submit his work)
5. To overwrite an old submission, simply execute this command again.
6. To verify the contents of your submission, execute the following command:

   ```
   turnin -v -c cs471 -p hw6
   ```

Required files

You will need to submit 1 file:

- The PDF containing your typed answers.

Problem Set

Problem 1: Dynamic Bayesian Networks 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.

Problem 2: Utility Theory 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:

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

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.
c. What should Sam do?

Problem 3: Learning Theory 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 $a_i \leqslant b_i$ are considered for learning.

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%.

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?

Problem 4: Bayesian Learning 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?