CS 8750: Artificial Intelligence II

Spring 2015

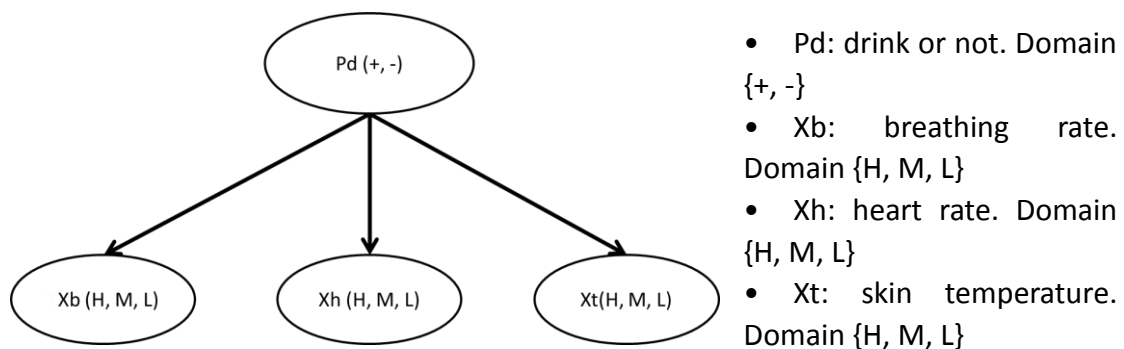## Programming assignment #2 (12 points)

1. Problem description

For this assignment, you may form teams with up to 3 members. Only one submission is required for each team.

In this programming assignment, you are required to implement four Bayesian networks to answer questions of a person had a drink or not, given evidence of breathing rate, heart rate, skin temperature, and ambulation status. You may use any programming language.
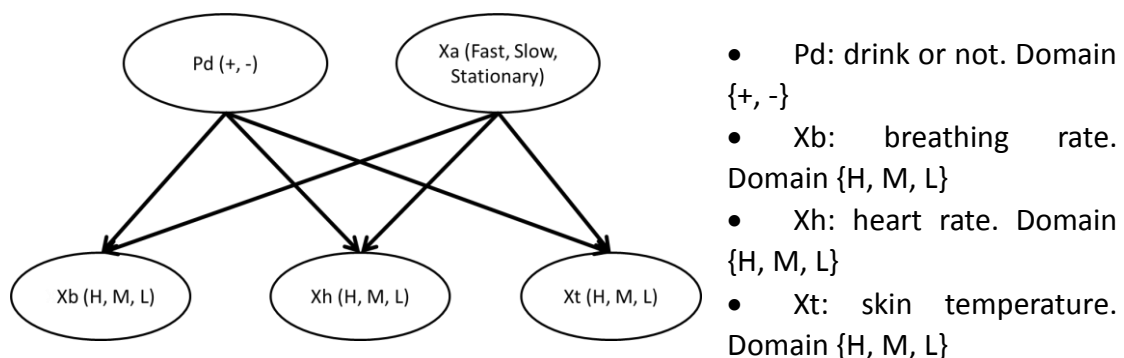
The structures and variables of the BNs are as follows. One set of parameters (CPDs) are to be learned from each of the two datasets, 1001 and 1004, in which the 1st column is Xb, 2nd Xh, 3rd Xt, 4th Xa, and 5th Pd. For Xa, stationary is 1, slow is 2, and fast is 3. For Pd, + is 1 and - is 0.

a) BN#1:

- Pd: drink or not. Domain {+, -}
- Xb: breathing rate. Domain {H, M, L}
- Xh: heart rate. Domain {H, M, L}
- Xt: skin temperature. Domain {H, M, L}

The values of Xb, Xh, and Xt in the data files are real number. To discretize the real values of each attribute into H, M, and L, first calculate the mean, m, and std (standard deviation), d. Then, convert a value bigger than m+d to H, less than m-d to L, and the rest to M.
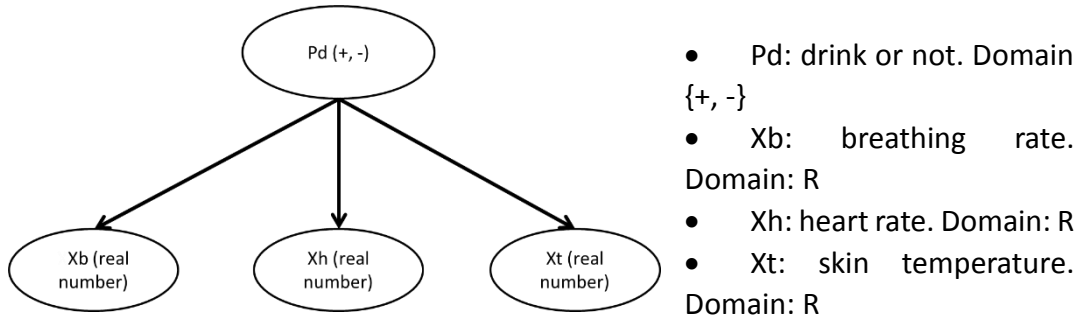
b) BN#2:

- Pd: drink or not. Domain {+, -}
- Xb: breathing rate. Domain {H, M, L}
- Xh: heart rate. Domain {H, M, L}
- Xt: skin temperature. Domain {H, M, L}

- Xa: ambulation status. Domain {Fast, Slow, Stationary}

Use the same method in BN#1 to discretize the real values of Xb, Xh, and Xt into H, M, and L.

c) **BN#3:**

- Pd: drink or not. Domain {+, -}
- Xb: breathing rate. Domain: R
- Xh: heart rate. Domain: R
- Xt: skin temperature. Domain: R

Use Gaussian density functions to represent the probabilities of Xb, Xh, and Xt, where each Gaussian density function is defined based on its mean and std.

d) **BN#4:**

- Pd: drink or not. Domain {+, -}
- Xb: breathing rate. Domain: R
- Xh: heart rate. Domain: R
- Xt: skin temperature. Domain: R
- Xa: ambulation status. Domain {Fast, Slow, Stationary}
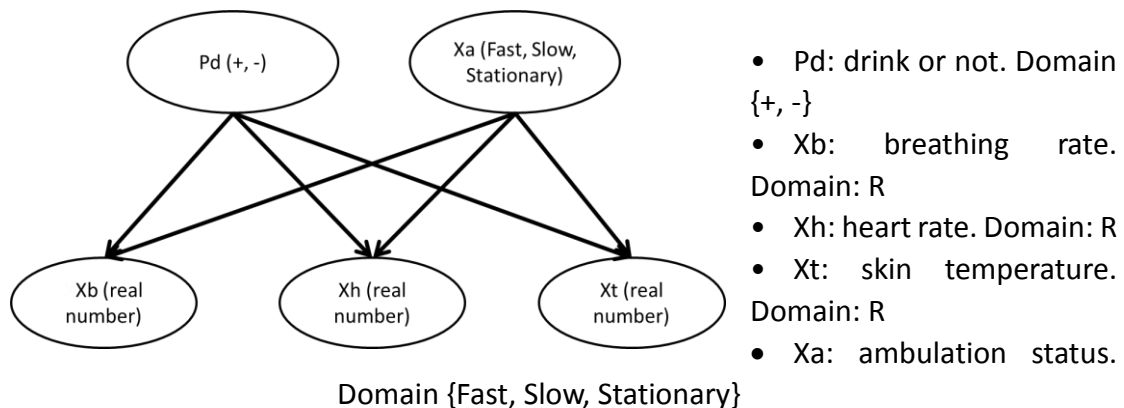
Use the same method in BN#3 to represent the probabilities of Xb, Xh, and Xt.

2. **Submission requirement**
   Your submission is required to have the following 3 parts:
   1) (4 points) For each BN, the formulas of Maximum Likelihood to learn the parameters from the given data, your learning program code, and the actual parameters learned from the two datasets (outputs of your program).
   2) (4 points) For each BN, the derivations of the formulas to predict Pd from Xb, Xh, Xt, and optionally Xa, your program code that implements the prediction.
   3) (4 points) Run each BN on each dataset to predict Pd from Xb, Xh, Xt, and optionally Xa. Report its confusion matrix and prediction accuracy. There should be a total of 8 confusion matrices and 8 accuracy numbers.