CS161: FUNDAMENTALS OF ARTIFICIAL INTELLIGENCE

Fall 2018
Assignment 8. Due Thursday, November 29, 11:55pm

Please submit your solutions on CCLE. The submitted file should be plain text or a formatted PDF file (no scans or pictures). Text files should have lines no longer than 100 characters and should be well-aligned when viewed with a monospace font. In addition to your solutions file, you will need to submit one .net files as indicated below.

All queries for this homework should be done using SamIam. You can download SamIam from http://reasoning.cs.ucla.edu/samiam. Included with SamIam are video tutorials on its use.

1. (25 pts) Consider the following problem which was discussed in class:

> Suppose that we have a patient who was just tested for a particular disease and the test came out positive. We know that one in every thousand people has this disease. We also know that the test is not reliable: it has a false positive rate of 3% and a false negative rate of 6%. Our goal is then to assess our belief in the patient having the disease given that the test came out positive. If we let the propositional variable D stand for the patient has the disease, and the propositional variable T stand for the test came out positive, our goal is then to compute $Pr(D|T)$.

   We considered a version of this problem in class. You may also recall being surprised that $Pr(D|T)$ was somewhat lower than expected. The goal of this question is then to identify conditions under which this probability will be no less than .30. You will need to find the answer to this by constructing a Bayesian Network and using the sensitivity analysis engine of SamIam. You need to turn in:

   - Your complete Bayesian network (structure and CPTs) in test.net file.
   - A constraint on each of the following, which is sufficient to ensure that $Pr(D|T) \geq 0.3$: The prior probability of having the disease, the false positive for the test, and the false negative for the test.

2. (25 pts) Figure 1 shows a Bayesian network classifier that predicts if a movie will be a box-office success. It takes in 4 binary inputs: S (*Original Screenplay*), G (*Great Cinematography*), F (*Famous Cast*), and M (*Marketing*), and outputs a binary classification: $C = 0$ (*failure*) or $C = 1$ (*success*). For convenience, the parameters of the Bayesian network are not shown but the decision function of the classifier is given in Table 1. Construct the reduced decision graph for this decision function, using the variable order: S, G, F, M.

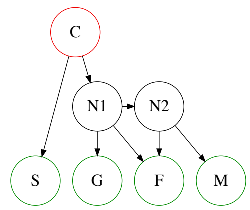

Figure 1: A Bayesian network classifier for movie success.

	S	G	F	M	C
1	-	-	-	-	-
2	-	-	-	+	-
3	-	-	+	-	+
4	-	-	+	+	+
5	-	+	-	-	-
6	-	+	-	+	-
7	-	+	+	-	+
8	-	+	+	+	+
9	+	-	-	-	-
10	+	-	-	+	-
11	+	-	+	-	-
12	+	-	+	+	+
13	+	+	-	-	-
14	+	+	-	+	-
15	+	+	+	-	+
16	+	+	+	+	+

Table 1: The decision function computed by the classifier in Figure 1.

3. (25 pts) Using the decision graph in Figure 2, answer the following explanation queries.

   (a) Consider a movie that is an original screenplay and has great cinematography, a famous cast, and poor marketing $\{S = 1, G = 1, F = 1, M = 0\}$. Identify a largest set of features that can be turned off (0) without changing the decision on this instance.

   (b) Consider a movie that is an original screenplay and has poor cinematography, a famous cast, and good marketing $\{S = 1, G = 0, F = 1, M = 1\}$. Identify a smallest set of features α that renders the remaining features β irrelevant to the decision on this instance. That is, if we fix features α to their current values, we can change the values of features β arbitrarily without changing the current decision.

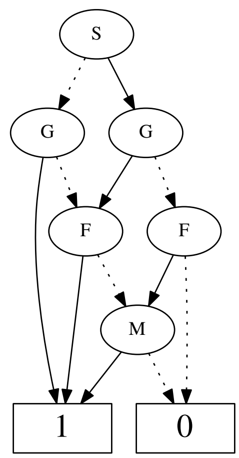

Figure 2: An ordered decision graph.

4. (25 pts) Consider the Bayesian network in Figure 3 (supplied as a genetics.net).

   > Our physical traits, also called phenotypes are known to be caused by genetic mutations such as single nucleotide polymorphisms (SNPs). It can be difficult to identify if it's a SNP that causes a phenotype or if it's caused by something else (which we call environmental effects). The following Bayesian network looks at four different possible effects: 2 independent SNPs, sex, and being over 60. Each has a probability of being present as well as a probability of causing the 2 phenotypes: high BMI and high blood pressure where BMI can also effect Blood Pressure.

   Use this network to answer the following queries:

   (a) *What are the prior marginal distributions for BMI and for blood pressure?*

   (b) *What is the posterior marginal distribution for 'over* 60*' assuming either high BMI or high blood pressure? (you will need to add an additional node to the network to answer this query)*

   (c) *If we know that SNP 2 is causal, what is the MAP for blood pressure and BMI?*

   (d) *If we know that both SNPs are present and are causal, what is the MPE?*

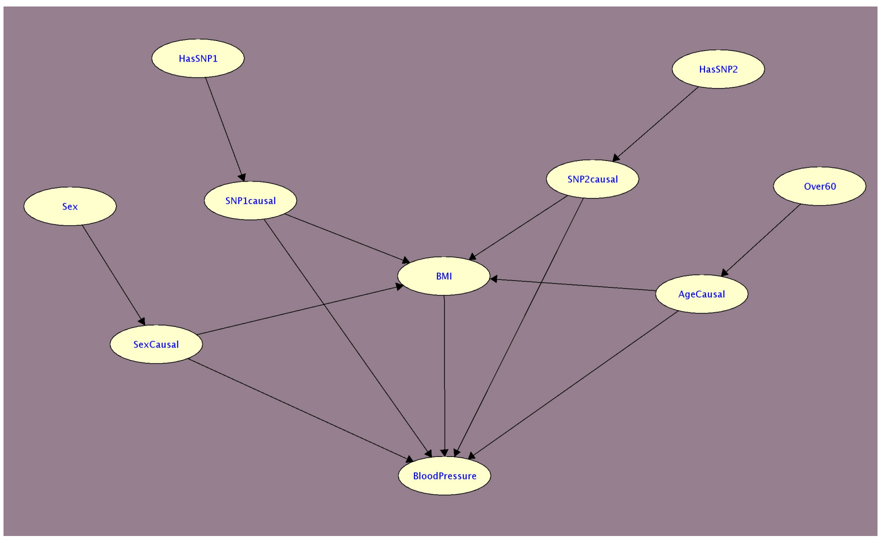

Figure 3: A Bayesian network.