Inference and Representation, Fall 2017

Problem Set 1: Bayesian networks

Selected solutions

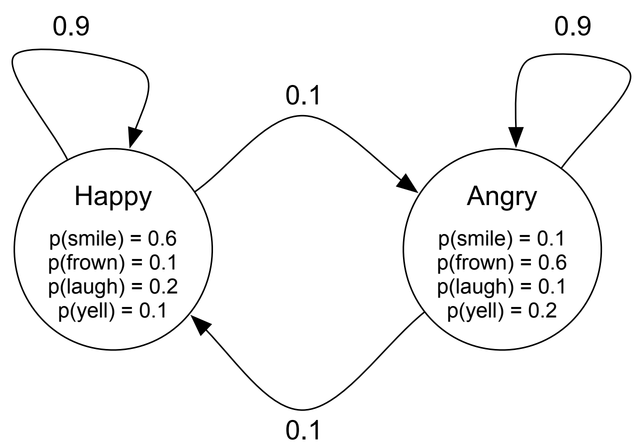
---

1. **Hidden Markov models** Harry lives a simple life. Some days he is Angry and some days he is Happy. But he hides his emotional state, and so all we can observe is whether he smiles, frowns, laughs, or yells. Harry's best friend is utterly confused about whether Harry is actually happy or angry and decides to model his emotional state using a hidden Markov model.

   Let $X_d \in \{\text{Happy, Angry}\}$ denote Harry's emotional state on day d, and let $Y_d \in \{\text{smile, frown, laugh, yell}\}$ denote the observation made about Harry on day d. **Assume that on day 1 Harry is in the Happy state**, i.e. $X_1 = \text{Happy}$. Furthermore, assume that Harry transitions between states exactly once per day (staying in the same state is an option) according to the following distribution: $p(X_{d+1} = \text{Happy} \mid X_d = \text{Angry}) = 0.1$, $p(X_{d+1} = \text{Angry} \mid X_d = \text{Happy}) = 0.1$, $p(X_{d+1} = \text{Angry} \mid X_d = \text{Angry}) = 0.9$, and $p(X_{d+1} = \text{Happy} \mid X_d = \text{Happy}) = 0.9$.

   The observation distribution for Harry's Happy state is given by $p(Y_d = \text{smile} \mid X_d = \text{Happy}) = 0.6, p(Y_d = \text{frown} \mid X_d = \text{Happy}) = 0.1, p(Y_d = \text{laugh} \mid X_d = \text{Happy}) = 0.2$, and $p(Y_d = \text{yell} \mid X_d = \text{Happy}) = 0.1$. The observation distribution for Harry's Angry state is $p(Y_d = \text{smile} \mid X_d = \text{Angry}) = 0.1, p(Y_d = \text{frown} \mid X_d = \text{Angry}) = 0.6, p(Y_d = \text{laugh} \mid X_d = \text{Angry}) = 0.1$, and $p(Y_d = \text{yell} \mid X_d = \text{Angry}) = 0.2$.

   All of this is summarized in the following figure:

   Be sure to show all of your work for the below questions. Note, the goal of this question is to get you to start thinking deeply about probabilistic inference. Thus, although you could look at Chapter 17 for an overview of HMMs, try to solve this question based on first principles (also: no programming needed!).

   (a) What is $p(X_2 = \text{Happy})$?

(b) What is $p(Y_2 = \text{frown})$?

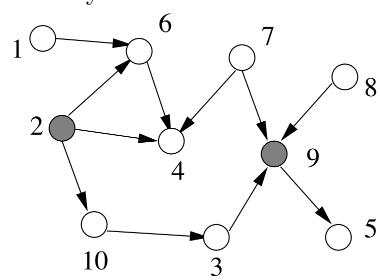
(c) What is $p(X_2 = \text{Happy} \mid Y_2 = \text{frown})$?

(d) What is $p(Y_{80} = \text{yell})$?

(e) Assume that $Y_1 = Y_2 = Y_3 = Y_4 = Y_5 = \text{frown}$. What is the most likely sequence of the states? That is, compute the MAP assignment $\arg\max_{x_1,\ldots,x_5} p(X_1 = x_1, \ldots, X_5 = x_5 \mid Y_1 = Y_2 = Y_3 = Y_4 = Y_5 = \text{frown})$.

2. **Bayesian networks must be acyclic.**

3. **D-separation.** Consider the Bayesian network shown in the below figure:



(a) For what pairs (i, j) does the statement $X_i \perp X_j$ hold? (Do not assume any conditioning in this part.)

Answer: The goal is to find all pairs (i, j) such that $X_i \perp X_j$. For node 1, an application of the d-separation algorithm shows that there is an active path from node 1 to nodes 6 and 4. For node 2, there are active paths to nodes 6, 4, 10, 3, 9, and 5. 3 and 10 can reach the sames nodes as node 2. Node 4 can reach every node but node 8. Node 5 can reach every node but node 1. Node 6 can reach any node but nodes 7 and 8. Node 7 can reach node 9, 4, and 5. Node 8 can only reach nodes 9 and 5. Node 9 can't reach node 1. Finally, node 10 can't reach nodes 1, 7, and 8. $(1, 2)$, $(1, 3)$, $(1, 5)$, $(1, 7)$, $(1, 8)$, $(1, 9)$, $(1, 10)$, $(2, 7)$, $(2, 8)$, $(3, 7)$, $(3, 8)$, $(4, 8)$, $(6, 7)$, $(6, 8)$, $(7, 8)$, $(7, 10)$, and $(8, 10)$. In all there are 17 distinct pairs.

(b) Suppose that we condition on $\{X_2, X_9\}$, shown shaded in the graph. What is the largest set A for which the statement $X_1 \perp X_A \mid \{X_2, X_9\}$ holds? The Bayes ball algorithm for d-separation may be helpful.

Answer: Conditioned on $\{X_2, X_9\}$ there is no active path starting at node 1 and reaching nodes 3, 10, 7, 8 and 5. Hence, $A = \{3, 5, 7, 8, 10\}$. Note that nodes 2 and 9 are not elements of the set A because we are conditioning on them.

4. Consider the following distribution over 3 binary variables X, Y, Z:

$$p(x, y, z) \quad = \quad \begin{cases} 1/12 & x \oplus y \oplus z = 0 \\ 1/6 & x \oplus y \oplus z = 1 \end{cases}$$

where $\oplus$ denotes the XOR function. Show that there is no directed acyclic graph G such that $I_{d-sep}(G) = I(p)$.

Answer: For the given distribution note that $I(p) \ni (X \perp Y)$, $I(p) \not\ni (X \perp Y|Z)$. Consequently

$$X \to Y \to Z,$$

$$X \to Y \leftarrow Z,$$

$$X \leftarrow Y \to Z$$

cannot satisfy the given condition since $I_{d-sep}(G) \not\ni (X \perp Y)$. Also

$$X, Y \to Z,$$

$$X, Y, Z$$

cannot satisfy the given condition since $I_{d-sep}(G) \ni (X \perp Y|Z)$. Finally note that these statements hold true however we permute X, Y and Z since p is invariant to such operations. Therefore no BN on variables X, Y, Z can satisfy $I_{d-sep}(G) = I(p)$, since it can always be reduced to one of the above five cases by permutation.

5. **Naive Bayes classifier** (Solution in a separate jupyter notebook)