CS 188 Fall 2019

Introduction to Artificial Intelligence

Written HW 3

Due: Friday 11/22/2019 at 11:59pm (Submit via Gradescope).

Policy: Can be solved in groups (acknowledge collaborators) but must be written up individually

Last Updated 11/21. Fixed typo in Q2 c(i)

First name	
Last name	
SID	
Collaborators	

Q1. Single Eyelid or Double Eyelid

Single eyelid or double eyelid are determined by genes. Typically, the eyelid gene is located on a pair of chromosomes, and we use A/a to represent the dominant / recessive gene on the chromosomes, respectively. If a person's gene is AA, Aa or aA, they will have double eyelids, whereas aa will result in single eyelids. According to a recent study, the proportion of recessive eyelid genes (a) is p among all the people in the world ($0 < p < 1$), and the dominance/recesiveness is **independent** between each chromosome.

(a) Based on this study, what is the probability of randomly selecting a person with single eyelids if we sample from the world?

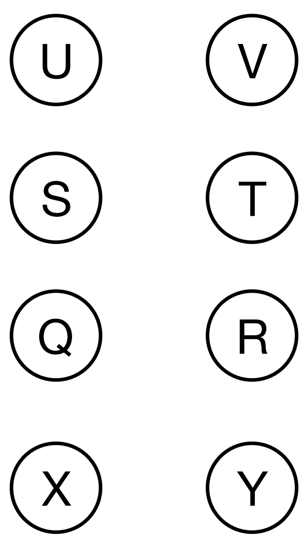
Bob is Alice's friend and Alice knows the attribute of Bob's eyelid. Although Alice has never met Bob's family, she knows that Bob's family contains 4 members: Bob, Bob's father, Bob's mother, and Bob's sister Anna. Alice has never met Anna and is very curious about whether Anna is has single eyelids or double eyelids. Let us use Bayes Nets to help her solve this puzzle.

We assume that a child will obtain the two chromosomes from the parents through this manner: picking one of the two chromosomes from their father with **equal** probability, and picking the other chromosome from their mother with **equal** probability. We assume that no further change will occur beyond this. We assume the genes between the parents are **independent**.

(b) Construct the Bayes Net by using the nodes from below.

To make the solution to this question unique, your *unconditional* probability tables may contain expressions of p, but your *conditional* probability tables may only contain the 0, 0.5, or 1 (or expressions involving those constants). You do not need to write down the probability tables, but they should satisfy the above conditions.

- Q: Bob's gene (AA or Aa or aA or aa).
- R: Anna's gene (AA or Aa or aA or aa).
- S: Bob and Anna's father's gene (AA or Aa or aA or aa).
- T: Bob and Anna's mother's gene (AA or Aa or aA or aa).
- U: Bob and Anna's father's eyelid attribute (single or double).
- V: Bob and Anna's mother's eyelid attribute (single or double).
- X: Bob's eyelid attribute (single or double).
- Y: Anna's eyelid attribute (single or double).



(c) Write down the expression for $P(Q, R, S, T, U, V, X, Y)$ in terms of the Bayes Net probability tables.

(d) Suppose that we observe the eyelid attribute of Bob (X), list all pairs of nodes (among the remaining 7 nodes) that have changed their dependency property (i.e. gone from possibly dependent to independent, or independent to possibly dependent)? Note we treat Q-R and R-Q as one pair, rather than two.

(e) Alice claims that we should also take Bob's grandparents into account by augmenting the Bayes Net. Suppose we do not know anything about Bob's grandparents, does this augmentation of the Bayes Net change $P(Y|X)$?

(f) To get $P(Y|X = double)$, we use variable elimination. If we eliminate the nodes with the optimal ordering so that the largest table of factors generated is minimized, what is the size of this minimized factor table?

(g) Calculate $P(Y = single|X = single)$. Assume that the prior $P(S), P(T)$ are according to the study mentioned in the first paragraph. You can include p in your provided answer.

Q2. HMMs and Dynamic Bayesian Networks

Self driving cars are finally here! Before we put them on the market, you've been asked to investigate the behavior of the cars, to check that the cars' programming complies with ethical and legal standards.

On a high level, the car's behavior can be modeled using 4 variables: if the passenger prefers a fast jerky car or not (**P**), whether the destination of the car is far or not (**F**), whether the car is aggressive toward other cars or not (**A**), and whether the car drives quickly (**Q**). Unfortunately, the engineers at the self-driving car company don't know exactly how these variables interact (they found the algorithm using a combination of deep learning and evolutionary strategies), so your first task is to fit a Bayesian network to model the behavior of their car.

Here are the six candidate networks that might be used to explain the car's behavior:

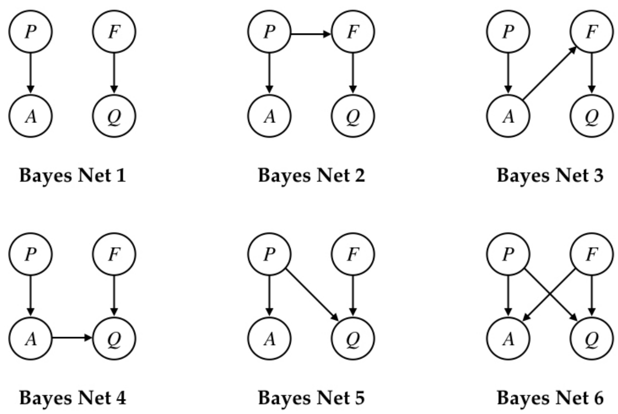


Figure 1: The six candidate Bayesian networks

(a) For each of the following joint probability tables, select the smallest Bayesian network(s) (i.e., one with least edges) that can represent the joint probability distribution. *Hint: A Bayes Net can represent a given joint distribution if the Bayes Net does not make more independence assumptions than the joint distribution. Work through these problems using process of elimination by checking independences.*

(i) First, the engineers show you the behavior of these four variables in simulation:

P	**F**	**A**	**Q**	$P(P, F, A, Q)$
-	-	-	-	1/4
-	-	-	+	0
-	-	+	-	0
-	-	+	+	0
-	+	-	-	0
-	+	-	+	1/4
-	+	+	-	0
-	+	+	+	0
+	-	-	-	0
+	-	-	+	0
+	-	+	-	1/4
+	-	+	+	0
+	+	-	-	0
+	+	-	+	0
+	+	+	-	0
+	+	+	+	1/4

What is the smallest Bayesian Network(s) that can represent the joint probability distribution?
☐ Bayes Net 1 ☐ Bayes Net 2 ☐ Bayes Net 3
☐ Bayes Net 4 ☐ Bayes Net 5 ☐ Bayes Net 6

(ii) Next, the engineers show you the behavior of this car on your company's test track:

P	**F**	**A**	**Q**	$P(P,F,A,Q)$
-	-	-	-	1/6
-	-	-	+	0
-	-	+	-	1/6
-	-	+	+	0
-	+	-	-	1/24
-	+	-	+	1/24
-	+	+	-	1/24
-	+	+	+	1/24
+	-	-	-	0
+	-	-	+	0
+	-	+	-	1/6
+	-	+	+	0
+	+	-	-	0
+	+	-	+	0
+	+	+	-	1/6
+	+	+	+	1/6

What is the smallest Bayesian Network(s) that can represent the joint probability distribution?

☐ Bayes Net 1 ☐ Bayes Net 2 ☐ Bayes Net 3
☐ Bayes Net 4 ☐ Bayes Net 5 ☐ Bayes Net 6

(iii) Finally, the engineers show you the behavior of this car on the road.

P	**F**	**A**	**Q**	$P(P,F,A,Q)$
-	-	-	-	1/16
-	-	-	+	1/16
-	-	+	-	1/16
-	-	+	+	1/16
-	+	-	-	0
-	+	-	+	1/8
-	+	+	-	0
-	+	+	+	1/8
+	-	-	-	0
+	-	-	+	0
+	-	+	-	1/4
+	-	+	+	0
+	+	-	-	0
+	+	-	+	0
+	+	+	-	1/8
+	+	+	+	1/8

What is the smallest Bayesian Network(s) that can represent the joint probability distribution?

☐ Bayes Net 1 ☐ Bayes Net 2 ☐ Bayes Net 3
☐ Bayes Net 4 ☐ Bayes Net 5 ☐ Bayes Net 6

Having (somewhat) figured out how our own self-driving car works, it's time to write an inference module to help us understand what our competitors' cars are doing. Unlike for our cars, however, we do not have access to all the variables! Suppose we observe two pieces of evidence A_t, Q_t at each timestep. Our hidden Markov Model (HMM) can be formulated as:

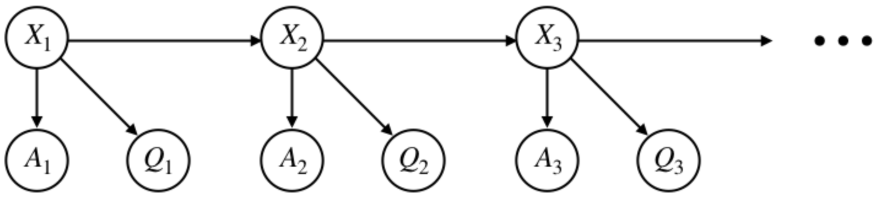


Figure 2: Our HMM.

with prior $P(X_1)$, transition probabilities $P(X_{t+1}|X_t)$, and emission probabilities $P(A_t|X_t), P(Q_t|X_t)$.

(b) Assume we have a Markov chain with X_t as the state at time t, and transition probabilities $P(X_t|X_{t-1})$. Show that the Markov property holds in the reverse direction: $P(X_t = x_t|X_{t+1} = x_{t+1}, \ldots, X_{t+n} = x_{t+n}) = P(X_t = x_t|X_{t+1} = x_{t+1})$ for all $n \in \mathbb{N}$. **Do not use D-separation. You may only use the fact that X_t is independent of $X_1, \ldots, X_{t-2}$ given X_{t-1} (the Markov Property).**

That is, for retrodicting the past behavior of a competitor's car, its current state is sufficient, and we don't need to know about the its future states.

(c) For the HMM in figure 2, what will our time-elapse and observe update functions look like (note the different indexing for each expression)?

(i) $P(X_t|a_{1:t-1}, q_{1:t-1}) =$ ____ $P(x_{t-1}|a_{1:t-1}q_{1:t-1})$

(ii) $P(X_t|a_{1:t}, q_{1:t}) \propto$ ____ $P(X_t|a_{1:t-1}q_{1:t-1})$