CMSI 485 – Classwork 4
SOLUTION

Instructions:

This worksheet will not only provide you with practice problems whose concepts will help you understand HW3 and your final exam, and think about the types of applications in which some of our Markov models can be deployed in practice.
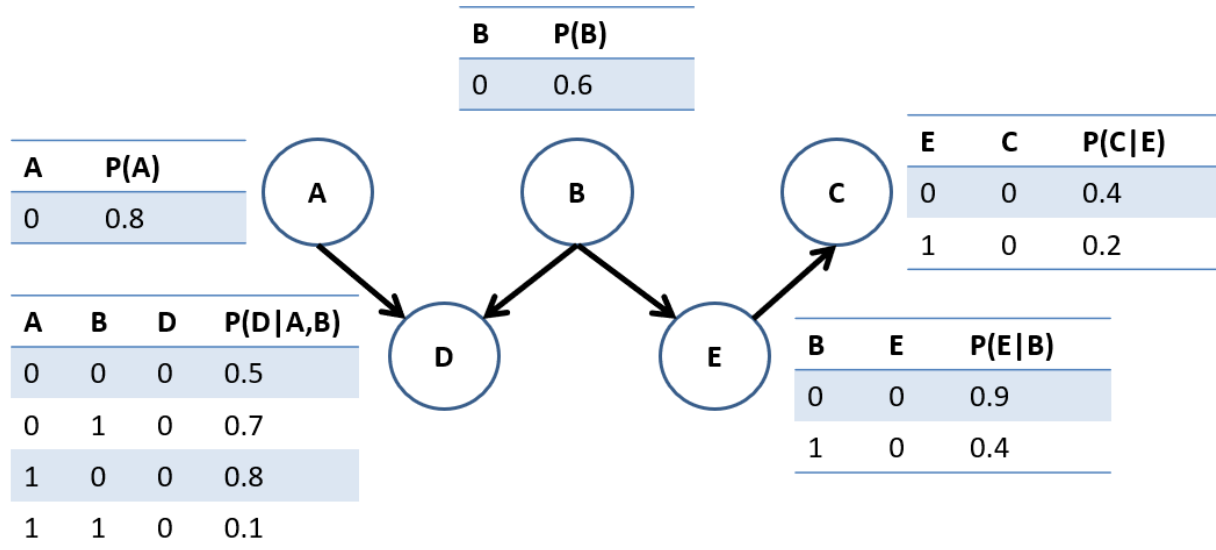
- Provide answers to each of the following questions and write your responses in the blanks. If you are expected to show your work in arriving at a particular solution, space will be provided for you.
- Place the names of your group members below:

Group Members:

1. _____
2. _____
3. _____

Problem 1 – Bayesian Network Approximate Inference

Consider the following Bayesian Network, and use it to answer the questions that follow:

B	P(B)
0	0.6

A	P(A)
0	0.8

E	C	P(C|E)
0	0	0.4
1	0	0.2

A	B	D	P(D|A,B)
0	0	0	0.5
0	1	0	0.7
1	0	0	0.8
1	1	0	0.1

B	E	P(E|B)
0	0	0.9
1	0	0.4

1.1. Provide a *possible* order in which variables would be sampled by *Prior Sampling:*

[A, B, D, E, C]
(among other possibilities, as long as parents are always sampled before their children)

Suppose now that Prior Sampling is used to generate the following set of 10 samples.

Sample	A	B	C	D	E
0	0	1	1	0	0
1	0	0	1	1	0
2	1	0	0	1	1
3	0	1	1	1	1
4	1	1	1	0	0
5	0	0	0	1	1
6	0	1	1	1	0
7	1	0	0	1	0
8	0	0	1	1	0
9	0	0	1	1	0

1.2. Using these samples, estimate $P(C = 1|D = 1)$

Find the proportion of C = 1 samples among the total of those for which D = 1:
$$\frac{count(C = 1, D = 1)}{count(D = 1)} = \frac{5}{8}$$

1.3. Using the same model, let's trace the generation of Gibbs Sampling in order to estimate $\hat{P}(B|C = 1, D = 1, E = 0)$. Suppose for the current sample we are sampling variables in the order of: $[A, B]$ and begin with the following random state. <u>For each variable sampled thereafter, assume (for this exercise) that the most likely value of the distribution sampled from is the one actually picked.</u>

A	B	C	D	E
0	1	1	1	0

Sampling A (Provided as an Example)

1.3.1. From what distribution will A be sampled? Show the likelihoods of sampling each value.

Distribution Sampled: $P(A|B = 1, C = 1, D = 1, E = 0)$

$$P(A = 0|B = 1, C = 1, D = 1, E = 0) = \frac{P(A = 0)P(D = 1|A = 0, B = 1)}{\sum_a P(A = a)P(D = 1|A = a, B = 1)}$$

$$= \frac{0.8 * 0.3}{0.8 * 0.3 + 0.2 * 0.9} \approx 0.57$$

1.3.2. Assuming our random number generator selects the value for A that is most likely, what will be the result of the above sample, and thus, the next state? (Fill in the value for A below):

A	B	C	D	E
0	1	1	1	0

Sampling B (Your Turn)

1.3.3. From what distribution will B be sampled? Show the likelihoods of sampling each value.

Distribution Sampled: $P(B|A = 0, C = 1, D = 1, E = 0)$

$$P(B = 0|A = 0, C = 1, D = 1, E = 0)$$
$$= \frac{P(B = 0)P(D = 1|A = 0, B = 0)P(E = 0|B = 0)}{\sum_b P(B = b)P(D = 1|A = 0, B = b)P(E = 0|B = b)}$$
$$= \frac{0.6 * 0.5 * 0.9}{0.6 * 0.5 * 0.9 + 0.4 * 0.3 * 0.4}$$
$$\approx 0.85$$

1.3.4. Assuming our random number generator selects the value for B that is most likely, what will be the result of the above sample, and thus, the next state? (Fill in the value for B below):

A	B	C	D	E
0	0	1	1	0

Problem 2 – Hidden Markov Models

In bioinformatics, analyses of nucleotide sequences in DNA processing is a central, but noisy, process; determining the identity of each nucleotide (from the 4 possible: ACGT) is not only imperfect, but many nucleotides in the sequence are "non-coding," meaning they do not encode any protein sequences (and are, for many applications, ignorable). It turns out the likelihood of producing certain nucleotides (viz., GC) increases during "coding" ($X = 1$) sections, while the likelihoods of these two to express *decrease* during non-coding ($X = 0$). From lab data, we have parameterized the following HMM describing this process:

Initial			Transitions			Emissions		
X	$P(X_0)$	X_t	X_{t+1}	$P(X_{t+1}|X_t)$		E_t	$P(E_t|X_t = 0)$	$P(E_t|X_t = 1)$
0	0.50	0	0	0.60		A	0.30	0.20
		1	0	0.50		C	0.20	0.30
						G	0.20	0.30
						T	0.30	0.20

2.1. Suppose we observe a (small) sequence of nucleotides: GGC. Treating our time step t as position within the sequence, use filtering to compute $B(X_3)$, our belief of the current state after observing the entire sequence, starting with the following as given:

	$t = 0$ *Init*	$t = 1$ $E_1 = G$	$t = 2$ $E_2 = G$	$t = 3$ $E_3 = C$
$B(X_t = 0)$	0.5	0.4	0.44	0.443
$B(X_t = 1)$	0.5	0.6	0.56	0.557

Forward Algorithm (Combining both Updates for Time and Observation):

$$B(X_3 = 0) = \alpha\, P(E_3 = C|X_3 = 0) \sum_{x_2} P(X_3 = 0|X_2 = x_2)B(x_2)$$

$$= \alpha\, P(E_3 = C|X_3 = 0)[P(X_3 = 0|X_2 = 0)B(X_2 = 0) + P(X_3 = 0|X_2 = 1)B(X_2 = 1)]$$
$$= \alpha\, 0.2 * [0.6 * 0.44 + 0.5 * 0.56] = \alpha\, 0.1088$$

$$B(X_3 = 1) = \alpha\, P(E_3 = C|X_3 = 1)[P(X_3 = 1|X_2 = 0)B(X_2 = 0) + P(X_3 = 1|X_2 = 1)B(X_2 = 1)]$$
$$= \alpha\, 0.3 * [0.4 * 0.44 + 0.5 * 0.56] = \alpha\, 0.1368$$

$$\alpha = \frac{1}{0.1088 + 0.1368} = \frac{1}{0.2456}$$

$$B(X_3 = 0) = \frac{0.1088}{0.2456} \approx 0.443; \therefore B(X_3 = 1) = 1 - B(X_3 = 0) \approx 0.557$$

In this setting, since we really care about the *most likely state* at any given part of the sequence, it's often useful to seek an **explanation** of what led us to witness the evidence at any given time. However, to compute this *most likely explanation*, we need to amend our Forward Algorithm slightly; instead of summing over all *possible* previous states in the transitions, we focus on only the most likely, labeled $M(X_t)$, in what is known as the **Viterbi Algorithm**.

Forward Algorithm	*Viterbi Algorithm*
$B(X_{t+1})$ $= \alpha\, P(E_{t+1}\|X_{t+1}) \sum_{x_t} P(X_{t+1}\|x_t) B(x_t)$	$M(X_{t+1})$ $= \alpha\, P(E_{t+1}\|X_{t+1}) \max_{x_t} P(X_{t+1}\|x_t) M(x_t)$

2.2. Suppose we observe the sequence $GGCA$ and are using the Viterbi Algorithm to find the most likely explanation of the state at each point in its generation. Using the following partially-completed computations, find $M(X_4)$: (hint: it *is* as simple as it seems, that's why it's nice).

	$t = 0$ Init	$t = 1$ $E_1 = G$	$t = 2$ $E_2 = G$	$t = 3$ $E_3 = C$	$t = 4$ $E_4 = A$
$M(X_t = 0)$	0.5	0.4	0.4	0.4	0.6
$M(X_t = 1)$	0.5	0.6	0.6	0.6	0.4

$$M(X_4 = 0) = \alpha\, P(E_4 = A|X_4 = 0) \max_{x_3} P(X_4 = 0|X_3 = x_3)\, M(X_3 = x_3)$$
$$= \alpha\, 0.3 \max(0.6 * 0.4, 0.5 * 0.6)$$
$$= \alpha\, 0.09$$

$$M(X_4 = 1) = \alpha\, P(E_4 = A|X_4 = 1) \max_{x_3} P(X_4 = 1|X_3 = x_3)\, M(X_3 = x_3)$$
$$= \alpha\, 0.2 \max(0.4 * 0.4, 0.5 * 0.6)$$
$$= \alpha\, 0.06$$

$$\therefore \alpha = \frac{1}{0.15}$$
$$\therefore M(X_4 = 0) = 0.6; \quad M(X_4 = 1) = 0.4$$

2.3. Examining the completed table above, determine the *explanation* (i.e., the most likely value of the state) for the observed sequence in all steps $t = 1 \rightarrow 4$:

	$t = 1$ $E_1 = G$	$t = 2$ $E_2 = G$	$t = 3$ $E_3 = C$	$t = 4$ $E_3 = A$
X_t^*	$X = 1$	$X = 1$	$X = 1$	$X = 0$

Consider instead that we decided to use HMM Approximate Inference with a Particle Filter (doesn't make a whole lot of sense because the state is so small here but meh, practice!) with $N = 5$ particles (again, very small). Consider that our approximate belief state $\hat{B}(X_t)$ is represented by the current allocation of particles as drawn below:
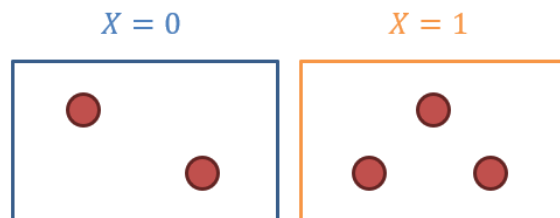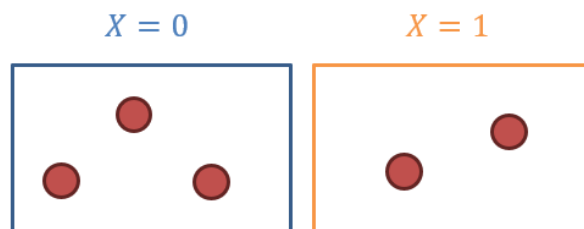

2.4. Suppose we now observe $E_{t+1} = T$ as the next nucleotide in the sequence, and we want to update our approximate belief state.

2.4.1. *Elapse Time:* What's the chance that each particle in the $X = 0$ state remain in this state?

Just check the proper transition distribution: $P(X_{t+1} = 0 | X_t = 0) = 0.6$

Suppose after elapsing time, we are left with the following configuration of particles:

2.4.2. *Observe Evidence:* Labeling the particles in the above diagram from left-to-right as $S^{(0)}, S^{(1)}, ...,$ how are each weighted after observing $E_{t+1} = T$?

	$S^{(0)}$	$S^{(1)}$	$S^{(2)}$	$S^{(3)}$	$S^{(4)}$
$w(S^{(i)})$	0.3	0.3	0.3	0.2	0.2

2.4.3. *Resample:* What's the likelihood that each particle is resampled at $X = 0$ and $X = 1$?

$P\left(S_{t+1}^{(i)} = 0 | E_{t+1} = T, S_t^{(i)} = s_t^{(i)}\right) =$

$$\frac{0.9}{1.3}$$

$P\left(S_{t+1}^{(i)} = 1 | E_{t+1} = T, S_t^{(i)} = s_t^{(i)}\right) =$

$$\frac{0.4}{1.3}$$