CS 188 Spring 2015 Introduction to Artificial Intelligence Midterm 2 V1

- You have approximately 2 hours and 50 minutes.
- The exam is closed book, closed calculator, and closed notes except your one-page crib sheet.
- Mark your answers ON THE EXAM ITSELF. If you are not sure of your answer you may wish to provide a *brief* explanation or show your work.
- For multiple choice questions,
  - ☐ means mark **all options** that apply
  - ◯ means mark a **single choice**
- There are multiple versions of the exam. For fairness, this does not impact the questions asked, only the ordering of options within a given question.

First name	
Last name	
SID	
edX username	

First and last name of student to your left	
First and last name of student to your right	

For staff use only:

Q1.	Probability	/14
Q2.	Bayes' Nets: Representation	/8
Q3.	Bayes' Nets: Independence	/8
Q4.	Bayes' Nets: Inference	/6
Q5.	Bayes' Nets: Sampling	/10
Q6.	VPI	/13
Q7.	HMM: Where is the Car?	/13
Q8.	Particle Filtering: Where are the Two Cars?	/11
Q9.	Naive Bayes MLE	/7
Q10.	Perceptron	/10
	Total	/100

THIS PAGE IS INTENTIONALLY LEFT BLANK

Q1. [14 pts] Probability

(a) For the following questions, you will be given a set of probability tables and a set of conditional independence assumptions. Given these tables and independence assumptions, write an expression for the requested probability tables. Keep in mind that your expressions cannot contain any probabilities other than the given probability tables. If it is not possible, mark "Not possible."

(i) [1 pt] Using probability tables $\mathbf{P}(\mathbf{A}), \mathbf{P}(\mathbf{A} \mid \mathbf{C}), \mathbf{P}(\mathbf{B} \mid \mathbf{C}), \mathbf{P}(\mathbf{C} \mid \mathbf{A}, \mathbf{B})$ and no conditional independence assumptions, write an expression to calculate the table $\mathbf{P}(\mathbf{A}, \mathbf{B} \mid \mathbf{C})$.

$\mathbf{P}(\mathbf{A}, \mathbf{B} \mid \mathbf{C}) =$ _____ ○ Not possible.

(ii) [1 pt] Using probability tables $\mathbf{P}(\mathbf{A}), \mathbf{P}(\mathbf{A} \mid \mathbf{C}), \mathbf{P}(\mathbf{B} \mid \mathbf{A}), \mathbf{P}(\mathbf{C} \mid \mathbf{A}, \mathbf{B})$ and no conditional independence assumptions, write an expression to calculate the table $\mathbf{P}(\mathbf{B} \mid \mathbf{A}, \mathbf{C})$.

$\mathbf{P}(\mathbf{B} \mid \mathbf{A}, \mathbf{C}) =$ _____ ○ Not possible.

(iii) [1 pt] Using probability tables $\mathbf{P}(\mathbf{A} \mid \mathbf{B}), \mathbf{P}(\mathbf{B}), \mathbf{P}(\mathbf{B} \mid \mathbf{A}, \mathbf{C}), \mathbf{P}(\mathbf{C} \mid \mathbf{A})$ and conditional independence assumption $\mathbf{A} \perp\!\!\!\perp \mathbf{B}$, write an expression to calculate the table $\mathbf{P}(\mathbf{C})$.

$\mathbf{P}(\mathbf{C}) =$ _____ ○ Not possible.

(iv) [1 pt] Using probability tables $\mathbf{P}(\mathbf{A} \mid \mathbf{B}, \mathbf{C}), \mathbf{P}(\mathbf{B}), \mathbf{P}(\mathbf{B} \mid \mathbf{A}, \mathbf{C}), \mathbf{P}(\mathbf{C} \mid \mathbf{B}, \mathbf{A})$ and conditional independence assumption $\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C}$, write an expression for $\mathbf{P}(\mathbf{A}, \mathbf{B}, \mathbf{C})$.

$\mathbf{P}(\mathbf{A}, \mathbf{B}, \mathbf{C}) =$ _____ ○ Not possible.

(b) For each of the following equations, select the *minimal set* of conditional independence assumptions necessary for the equation to be true.

(i) [1 pt] $\mathbf{P}(\mathbf{A}, \mathbf{C}) = \mathbf{P}(\mathbf{A} \mid \mathbf{B})\ \mathbf{P}(\mathbf{C})$

- ☐ $A \perp\!\!\!\perp B \mid C$
- ☐ $B \perp\!\!\!\perp C \mid A$
- ☐ $A \perp\!\!\!\perp C$
- ☐ $A \perp\!\!\!\perp C \mid B$
- ☐ $B \perp\!\!\!\perp C$
- ☐ $A \perp\!\!\!\perp B$
- ☐ No independence assumptions needed.

(ii) [1 pt] $\mathbf{P}(\mathbf{A} \mid \mathbf{B}, \mathbf{C}) = \frac{\mathbf{P}(\mathbf{A})\ \mathbf{P}(\mathbf{B}|\mathbf{A})\ \mathbf{P}(\mathbf{C}|\mathbf{A})}{\mathbf{P}(\mathbf{B}|\mathbf{C})\ \mathbf{P}(\mathbf{C})}$

- ☐ $A \perp\!\!\!\perp C$
- ☐ $A \perp\!\!\!\perp B \mid C$
- ☐ $B \perp\!\!\!\perp C \mid A$
- ☐ $A \perp\!\!\!\perp B$
- ☐ $A \perp\!\!\!\perp C \mid B$
- ☐ $B \perp\!\!\!\perp C$
- ☐ No independence assumptions needed.

(iii) [1 pt] $\mathbf{P}(\mathbf{A}, \mathbf{B}) = \sum_{\mathbf{c}} \mathbf{P}(\mathbf{A} \mid \mathbf{B}, \mathbf{c})\ \mathbf{P}(\mathbf{B} \mid \mathbf{c})\ \mathbf{P}(\mathbf{c})$

- ☐ $B \perp\!\!\!\perp C \mid A$
- ☐ $B \perp\!\!\!\perp C$
- ☐ $A \perp\!\!\!\perp C \mid B$
- ☐ $A \perp\!\!\!\perp C$
- ☐ $A \perp\!\!\!\perp B$
- ☐ $A \perp\!\!\!\perp B \mid C$
- ☐ No independence assumptions needed.

(iv) [1 pt] $\mathbf{P}(\mathbf{A}, \mathbf{B} \mid \mathbf{C}, \mathbf{D}) = \mathbf{P}(\mathbf{A} \mid \mathbf{C}, \mathbf{D})\ \mathbf{P}(\mathbf{B} \mid \mathbf{A}, \mathbf{C}, \mathbf{D})$

- ☐ $A \perp\!\!\!\perp B \mid D$
- ☐ $C \perp\!\!\!\perp D \mid A$
- ☐ $C \perp\!\!\!\perp D \mid B$
- ☐ $C \perp\!\!\!\perp D$
- ☐ $A \perp\!\!\!\perp B$
- ☐ $A \perp\!\!\!\perp B \mid C$
- ☐ No independence assumptions needed.

(c) (i) [2 pts] Mark **all** expressions that are equal to $\mathbf{P}(\mathbf{A} \mid \mathbf{B})$, **given no independence assumptions**.

- ☐ $\sum_c P(A, c \mid B)$
- ☐ $\frac{P(A,C|B)}{P(C|B)}$
- ☐ $\sum_c P(A \mid B, c)$
- ☐ $\frac{\sum_c P(A,B,c)}{\sum_c P(B,c)}$
- ☐ $\frac{P(B|A)\ P(A|C)}{\sum_c P(B,c)}$
- ☐ $\frac{P(A|C,B)\ P(C|A,B)}{P(C|B)}$
- ☐ None of the provided options.

(ii) [2 pts] Mark **all** expressions that are equal to $\mathbf{P}(\mathbf{A}, \mathbf{B}, \mathbf{C})$, given that $\mathbf{A} \perp\!\!\!\perp \mathbf{B}$.

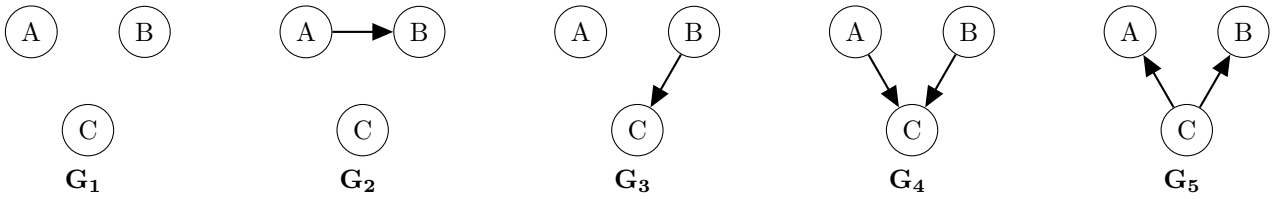
- ☐ $P(A)\ P(B)\ P(C \mid A, B)$
- ☐ $P(C)\ P(A \mid C)\ P(B \mid C)$
- ☐ $P(A)\ P(B \mid A)\ P(C \mid A, B)$
- ☐ $P(A \mid C)\ P(C \mid B)\ P(B)$
- ☐ $P(A)\ P(C \mid A)\ P(B \mid C)$
- ☐ $P(A, C)\ P(B \mid A, C)$
- ☐ None of the provided options.

(iii) [2 pts] Mark **all** expressions that are equal to $\mathbf{P}(\mathbf{A}, \mathbf{B} \mid \mathbf{C})$, given that $\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C}$.

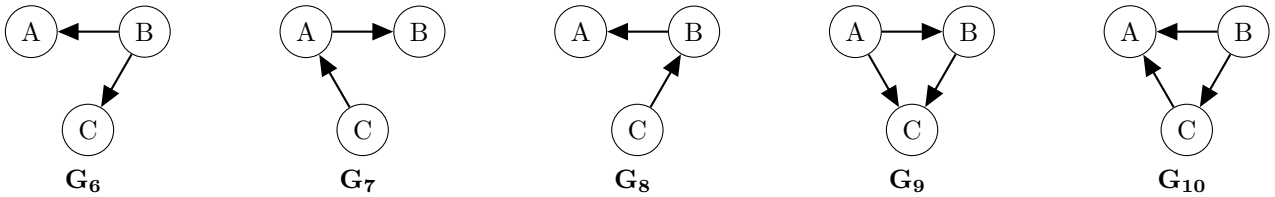
- ☐ $P(A \mid C)\ P(B \mid C)$
- ☐ $\frac{\sum_c P(A,B,c)}{P(C)}$
- ☐ $\frac{P(C)\ P(B|C)\ P(A|C)}{P(C|A,B)}$
- ☐ $\frac{P(A)\ P(B|A)\ P(C|A,B)}{\sum_c P(A,B,c)}$
- ☐ $\frac{P(C,A|B)\ P(B)}{P(C)}$
- ☐ $P(A \mid B)\ P(B \mid C)$
- ☐ None of the provided options.

Q2. [8 pts] Bayes' Nets: Representation

Assume we are given the following ten Bayes' nets, labeled $\mathbf{G_1}$ to $\mathbf{G_{10}}$:

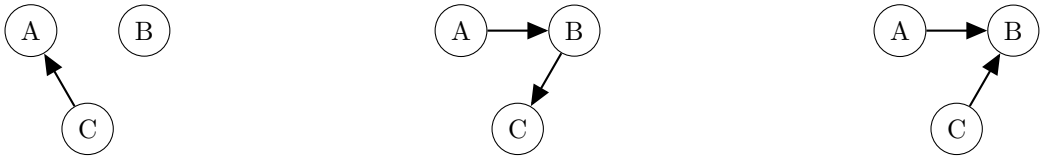


$\mathbf{G_1}$ $\mathbf{G_2}$ $\mathbf{G_3}$ $\mathbf{G_4}$ $\mathbf{G_5}$



$\mathbf{G_6}$ $\mathbf{G_7}$ $\mathbf{G_8}$ $\mathbf{G_9}$ $\mathbf{G_{10}}$

Assume we are also given the following three Bayes' nets, labeled $\mathbf{B_1}$ to $\mathbf{B_3}$:



$\mathbf{B_1}$ $\mathbf{B_2}$ $\mathbf{B_3}$

(a) [2 pts] Assume we know that a joint distribution $\mathbf{d_1}$ (over $\mathbf{A}, \mathbf{B}, \mathbf{C}$) can be represented by Bayes' net $\mathbf{B_1}$. Mark all of the following Bayes' nets that are guaranteed to be able to represent $\mathbf{d_1}$.

☐ $\mathbf{G_1}$ ☐ $\mathbf{G_2}$ ☐ $\mathbf{G_3}$ ☐ $\mathbf{G_4}$ ☐ $\mathbf{G_5}$

☐ $\mathbf{G_6}$ ☐ $\mathbf{G_7}$ ☐ $\mathbf{G_8}$ ☐ $\mathbf{G_9}$ ☐ $\mathbf{G_{10}}$

☐ None of the above.

(b) [2 pts] Assume we know that a joint distribution $\mathbf{d_2}$ (over $\mathbf{A}, \mathbf{B}, \mathbf{C}$) can be represented by Bayes' net $\mathbf{B_2}$. Mark all of the following Bayes' nets that are guaranteed to be able to represent $\mathbf{d_2}$.

☐ $\mathbf{G_1}$ ☐ $\mathbf{G_2}$ ☐ $\mathbf{G_3}$ ☐ $\mathbf{G_4}$ ☐ $\mathbf{G_5}$

☐ $\mathbf{G_6}$ ☐ $\mathbf{G_7}$ ☐ $\mathbf{G_8}$ ☐ $\mathbf{G_9}$ ☐ $\mathbf{G_{10}}$

☐ None of the above.

(c) [2 pts] Assume we know that a joint distribution $\mathbf{d_3}$ (over $\mathbf{A}, \mathbf{B}, \mathbf{C}$) ***cannot*** be represented by Bayes' net $\mathbf{B_3}$. Mark all of the following Bayes' nets that are guaranteed to be able to represent $\mathbf{d_3}$.

☐ $\mathbf{G_1}$ ☐ $\mathbf{G_2}$ ☐ $\mathbf{G_3}$ ☐ $\mathbf{G_4}$ ☐ $\mathbf{G_5}$

☐ $\mathbf{G_6}$ ☐ $\mathbf{G_7}$ ☐ $\mathbf{G_8}$ ☐ $\mathbf{G_9}$ ☐ $\mathbf{G_{10}}$

☐ None of the above.

(d) [2 pts] Assume we know that a joint distribution $\mathbf{d_4}$ (over $\mathbf{A}, \mathbf{B}, \mathbf{C}$) can be represented by Bayes' nets $\mathbf{B_1}$, $\mathbf{B_2}$, and $\mathbf{B_3}$. Mark all of the following Bayes' nets that are guaranteed to be able to represent $\mathbf{d_4}$.

☐ $\mathbf{G_1}$ ☐ $\mathbf{G_2}$ ☐ $\mathbf{G_3}$ ☐ $\mathbf{G_4}$ ☐ $\mathbf{G_5}$

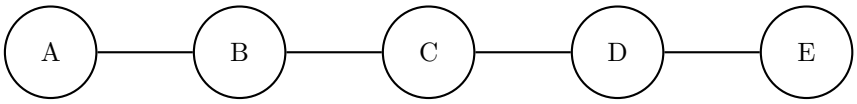
☐ $\mathbf{G_6}$ ☐ $\mathbf{G_7}$ ☐ $\mathbf{G_8}$ ☐ $\mathbf{G_9}$ ☐ $\mathbf{G_{10}}$

☐ None of the above.

Q3. [8 pts] Bayes' Nets: Independence

For the following questions, each edge shown in the Bayes' nets below does not have a direction. For each of the edges shown, assign a direction (by adding an arrowhead at one end of each edge) to ensure that the Bayes' Net structure implies the assumptions provided. You cannot add new edges. The Bayes' nets can imply more assumptions than listed, but they *must* imply the ones listed. If there does not exist an assignment of directions that satisfies all the assumptions listed, clearly mark the *Not Possible* choice. *If you mark the Not Possible choice, the directions that you draw in the Bayes' net will not be looked at.* Keep in mind that Bayes' Nets cannot have directed cycles. You may find it useful to use the front of the next page to work on this problem.

(a) [2 pts]

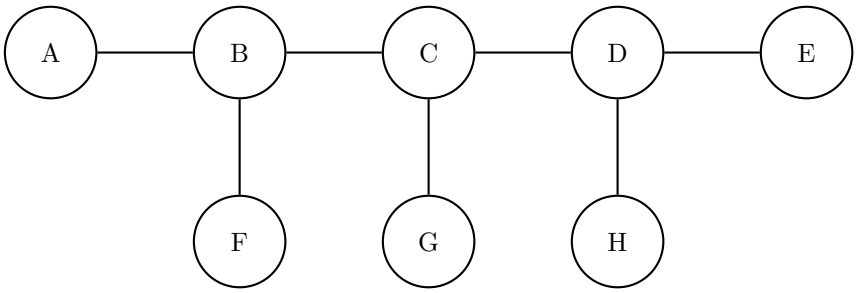


Assumptions:

- $A \perp\!\!\!\perp E$
- $B \perp\!\!\!\perp E \mid D$
- $A \perp\!\!\!\perp E \mid C$

◯ Not Possible

(b) [3 pts]

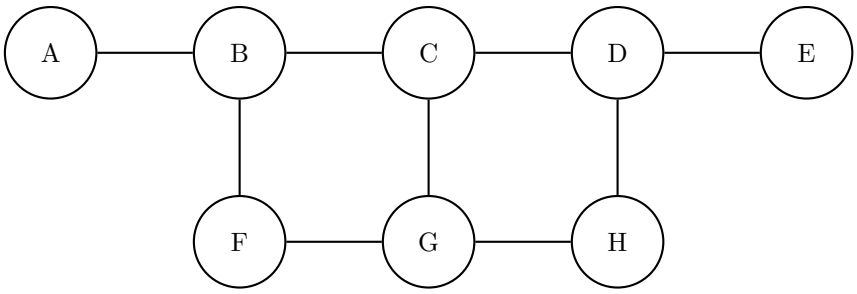


Assumptions:

- $C \perp\!\!\!\perp E \mid H$
- $A \perp\!\!\!\perp H \mid B, D$
- $F \perp\!\!\!\perp G \mid A, D$

◯ Not Possible

(c) [3 pts]



Assumptions:

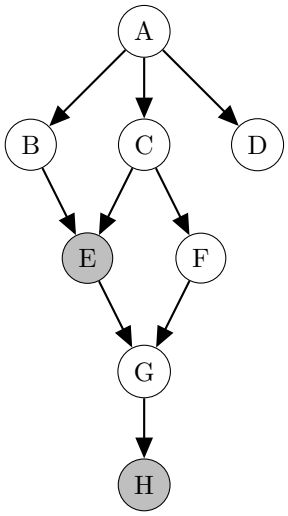
- $B \perp\!\!\!\perp H$
- $B \perp\!\!\!\perp G \mid E, F$
- $A \perp\!\!\!\perp E \mid C, H$

◯ Not Possible

This page is intentionally left blank for scratch work.
Nothing on this page will be graded.

Q4. [6 pts] Bayes' Nets: Inference

Assume we are given the following Bayes' net, and would like to perform inference to obtain $P(B, D \mid E = e, H = h)$.



(a) [1 pt] What is the number of rows in the largest factor generated by *inference by enumeration*, for this query $P(B, D \mid E = e, H = h)$? Assume all the variables are binary.

◯ 2^8 ◯ 2^6 ◯ 2^2 ◯ 2^3

◯ None of the above.

(b) [2 pts] Mark all of the following variable elimination orderings that are optimal for calculating the answer for the query $P(B, D \mid E = e, H = h)$. Optimality is measured by the sum of the sizes of the factors that are generated. Assume all the variables are binary.

☐ C, A, F, G ☐ G, F, C, A ☐ F, G, C, A ☐ A, C, F, G

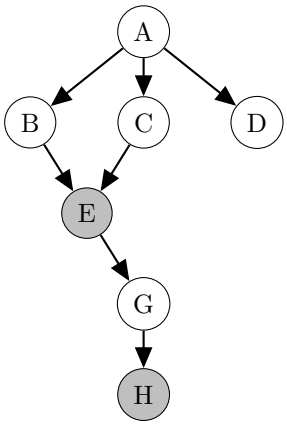
☐ None of the above.

(c) Suppose we decide to perform variable elimination to calculate the query $P(B, D \mid E = e, H = h)$, and choose to eliminate F first.

(i) [2 pts] When F is eliminated, what intermediate factor is generated and how is it calculated? Make sure it is clear which variable(s) come before the conditioning bar and which variable(s) come after.

$f_1(\_\_\_\_\_ \mid \_\_\_\_\_) = \sum_f$ _____

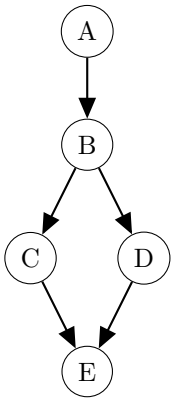
(ii) [1 pt] Now consider the set of distributions that can be represented by the remaining factors *after F is eliminated*. Draw the minimal number of directed edges on the following Bayes' Net structure, so that it can represent any distribution in this set. If no additional directed edges are needed, please fill in that option below.



◯ No additional directed edges needed

Q5. [10 pts] Bayes' Nets: Sampling

Assume we are given the following Bayes' net, with the associated conditional probability tables (CPTs).



A	P(A)
$+a$	0.5
$-a$	0.5

A	B	P(B | A)
$+a$	$+b$	0.2
$+a$	$-b$	0.8
$-a$	$+b$	0.5
$-a$	$-b$	0.5

B	C	P(C | B)
$+b$	$+c$	0.4
$+b$	$-c$	0.6
$-b$	$+c$	0.8
$-b$	$-c$	0.2

B	D	P(D | B)
$+b$	$+d$	0.2
$+b$	$-d$	0.8
$-b$	$+d$	0.2
$-b$	$-d$	0.8

C	D	E	P(E | C, D)
$+c$	$+d$	$+e$	0.6
$+c$	$+d$	$-e$	0.4
$+c$	$-d$	$+e$	0.2
$+c$	$-d$	$-e$	0.8
$-c$	$+d$	$+e$	0.4
$-c$	$+d$	$-e$	0.6
$-c$	$-d$	$+e$	0.8
$-c$	$-d$	$-e$	0.2

You are given a set of the following samples, but are not told whether they were collected with rejection sampling or likelihood weighting.

$$\begin{array}{lllll} -a & -b & +c & +d & +e \\ -a & +b & +c & -d & +e \\ -a & -b & -c & -d & +e \\ -a & -b & +c & -d & +e \\ -a & +b & +c & +d & +e \end{array}$$

Throughout this problem, you may answer as either numeric expressions (e.g. $0.1*0.5$) or numeric values (e.g. 0.05).

(a) [2 pts] Assuming these samples were generated from *rejection sampling*, what is the sample based estimate of $P(+b \mid -a, +e)$?

Answer: _____

(b) [2 pts] Assuming these samples were generated from *likelihood weighting*, what is the sample-based estimate of $P(+b \mid -a, +e)$?

Answer: _____

(c) [2 pts] Again, assume these samples were generated from *likelihood weighting*. However, you are not sure about the original CPT for $P(E \mid C, D)$ given above being the CPT associated with the Bayes' Net: With 50% chance, the CPT associated with the Bayes' Net is the original one. With the other 50% chance, the CPT is actually the CPT below.

C	D	E	P(E | C, D)
$+c$	$+d$	$+e$	0.8
$+c$	$+d$	$-e$	0.2
$+c$	$-d$	$+e$	0.4
$+c$	$-d$	$-e$	0.6
$-c$	$+d$	$+e$	0.2
$-c$	$+d$	$-e$	0.8
$-c$	$-d$	$+e$	0.6
$-c$	$-d$	$-e$	0.4

Samples from previous page copied below for convenience:

$$\begin{array}{lllll} -a & -b & +c & +d & +e \\ -a & +b & +c & -d & +e \\ -a & -b & -c & -d & +e \\ -a & -b & +c & -d & +e \\ -a & +b & +c & +d & +e \end{array}$$

Given this uncertainty, what is the sample-based estimate of $P(+b \mid -a, +e)$?

Answer: _____

(d) [1 pt] Now assume you can only sample a *small, limited number of samples*, and you want to estimate $P(+b, +d \mid -a)$ and $P(+b, +d \mid +e)$. You are allowed to estimate the answer to one query with likelihood weighting, and the other answer with rejection sampling. In order to obtain the best estimates for both queries, *which query should you estimate with likelihood weighting*? (The other query will have to be estimated with rejection sampling.)

- ◯ $P(+b, +d \mid -a)$
- ◯ $P(+b, +d \mid +e)$
- ◯ Either – both choices allow you to obtain the best estimates for both queries.

(e) Suppose you choose to use Gibbs sampling to estimate $P(B, E \mid +c, -d)$. Assume the CPTs are the same as the ones for parts (a) and (b). Currently your assignments are the following:

$$-a \quad -b \quad +c \quad -d \quad +e$$

(i) [1 pt] Suppose the next step is to resample E.
What is the probability that the new assignment to E will be +e?

Answer: _____

(ii) [1 pt] Instead, suppose the next step is to resample A.
What is the probability that the new assignment to A will be $+a$?

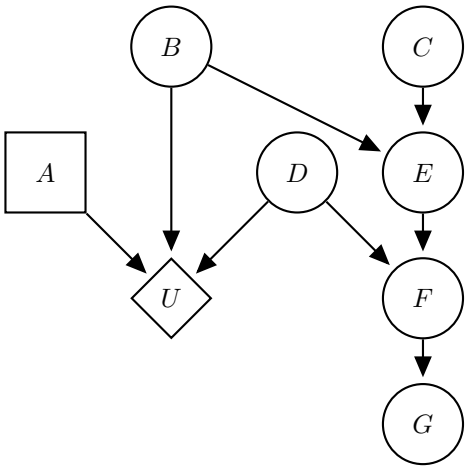
Answer: _____

(iii) [1 pt] Instead, suppose the next step is to resample B.
What is the probability that the new assignment to B will be +b?

Answer: _____

Q6. [13 pts] VPI

Consider a decision network with the following structure. Node A is an action, and node U is the utility:



(a) [1 pt] Choose the option which is *guaranteed* to be true, or "Neither guaranteed" if no option is guaranteed to be true.

◯ $VPI(C) = 0$ ◯ $VPI(C) > 0$ ◯ Neither guaranteed

(b) [2 pts] Mark *all* of the following that are *guaranteed* to be true.

☐ $VPI(E) \leq VPI(B)$ ☐ $VPI(E) \geq VPI(B)$

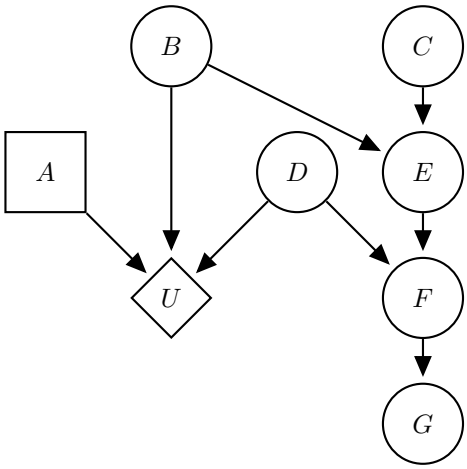
☐ $VPI(E) = VPI(B)$ ☐ None of the above

(c) [2 pts] Mark *all* of the following that are *guaranteed* to be true.

☐ $VPI(E \mid F) \leq VPI(B \mid F)$ ☐ $VPI(E \mid F) \geq VPI(B \mid F)$

☐ $VPI(E \mid F) = VPI(B \mid F)$ ☐ None of the above

The decision network on the previous page has been reproduced below:



(d) [2 pts] Noting that $E \perp\!\!\!\perp G \mid F$, mark *all* of the following that are *guaranteed* to be true.

☐ $VPI(E, G \mid F) = VPI(E \mid F) + VPI(G \mid E, F)$

☐ $VPI(E, G \mid F) = VPI(E \mid F)VPI(G \mid E, F)$

☐ $VPI(E, G \mid F) = VPI(E \mid F) + VPI(G \mid F)$

☐ $VPI(E, G \mid F) = VPI(E \mid F)VPI(G \mid F)$

☐ None of the above

(e) [3 pts] Suppose we have two actions, a_1 and a_2. In addition, you are given $P(B)$ and $P(D)$ below. Fill in the empty entries in the utility table below **with either** 1 **or** -1 such that $VPI(B) = VPI(D) = 0$, but $VPI(B, D) > 0$.

Note on grading: You will get 1 point for each condition you enforce correctly, that is, you get 1 point each for ensuring $VPI(B) = 0$, $VPI(D) = 0$, and $VPI(B, D) > 0$. If you cannot enforce all three conditions, try to enforce two for partial credit.

$P(B)$	
+b	0.5
-b	0.5

$P(D)$	
+d	0.5
-d	0.5

Action a_1: $U(a_1, B, D)$		
$+b$	$+d$	1
$-b$	$+d$	
$+b$	$-d$	
$-b$	$-d$	

Action a_2: $U(a_2, B, D)$		
$+b$	$+d$	
$-b$	$+d$	
$+b$	$-d$	
$-b$	$-d$	

(f) For this question, assume you did the previous part correctly - that is, you should assume that $VPI(B) = VPI(D) = 0$, and $VPI(B, D) > 0$.

Now taking into account your answer from the previous part, choose the option which is *guaranteed* to be true, or "Neither guaranteed" if no option is guaranteed to be true. **This means that you should fill in one circle for each row.**

(i) [1 pt] ◯ $VPI(E) = 0$ ◯ $VPI(E) > 0$ ◯ Neither guaranteed

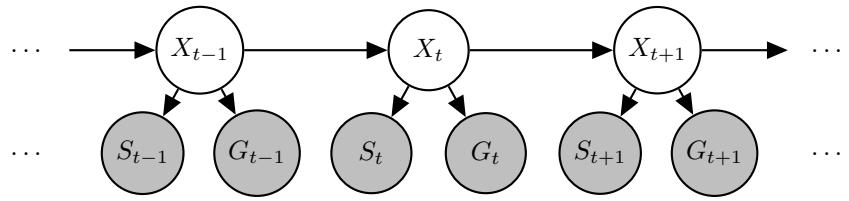
(ii) [1 pt] ◯ $VPI(G) = 0$ ◯ $VPI(G) > 0$ ◯ Neither guaranteed

(iii) [1 pt] ◯ $VPI(D \mid B) = 0$ ◯ $VPI(D \mid B) > 0$ ◯ Neither guaranteed

Q7. [13 pts] HMM: Where is the Car?

Transportation researchers are trying to improve traffic in the city but, in order to do that, they first need to estimate the location of each of the cars in the city. They need our help to model this problem as an inference problem of an HMM. For this question, assume that only *one* car is being modeled.

(a) The structure of this modified HMM is given below, which includes X, the location of the car; S, the noisy location of the car from the signal strength at a nearby cell phone tower; and G, the noisy location of the car from GPS.



We want to perform filtering with this HMM. That is, we want to compute the belief $P(x_t|s_{1:t}, g_{1:t})$, the probability of a state x_t given all past and current observations.

The **dynamics update** expression has the following form:

$$P(x_t|s_{1:t-1}, g_{1:t-1}) = \underline{\quad\textbf{(i)}\quad}\ \underline{\textbf{(ii)}}\ \underline{\quad\textbf{(iii)}\quad}\ P(x_{t-1}|s_{1:t-1}, g_{1:t-1}).$$

Complete the expression by choosing the option that fills in each blank.

(i) [1 pt] ○ $P(s_{1:t}, g_{1:t})$ ○ $P(s_{1:t-1}, g_{1:t-1})$ ○ $P(s_{1:t})P(g_{1:t})$ ○ $P(s_{1:t-1})P(g_{1:t-1})$ ○ 1

(ii) [1 pt] ○ $\sum_{x_t}$ ○ $\sum_{x_{t-1}}$ ○ $\max_{x_{t-1}}$ ○ $\max_{x_t}$ ○ 1

(iii) [1 pt] ○ $P(x_{t-2}, x_{t-1})$ ○ $P(x_{t-1} \mid x_{t-2})$ ○ $P(x_{t-1}, x_t)$ ○ $P(x_t|x_{t-1})$ ○ 1

The **observation update** expression has the following form:

$$P(x_t|s_{1:t}, g_{1:t}) = \underline{\quad\textbf{(iv)}\quad}\ \underline{\textbf{(v)}}\ \underline{\quad\textbf{(vi)}\quad}\ P(x_t|s_{1:t-1}, g_{1:t-1}).$$

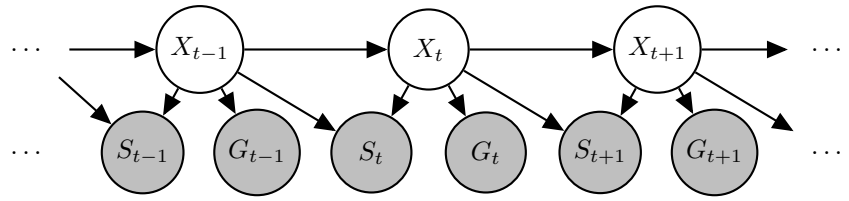
Complete the expression by choosing the option that fills in each blank.

(iv) [1 pt] ○ $P(s_t, g_t|s_{1:t-1}, g_{1:t-1})$ ○ $P(s_{1:t-1}, g_{1:t-1}|s_t, g_t)$ ○ $P(s_t|s_{1:t-1})P(g_t|g_{1:t-1})$
○ $P(s_{1:t-1}|s_t)P(g_{1:t-1}|g_t)$ ○ $\dfrac{1}{P(s_t, g_t|s_{1:t-1}, g_{1:t-1})}$ ○ $\dfrac{1}{P(s_{1:t-1}, g_{1:t-1}|s_t, g_t)}$
○ $\dfrac{1}{P(s_t|s_{1:t-1})P(g_t|g_{1:t-1})}$ ○ $\dfrac{1}{P(s_{1:t-1}|s_t)P(g_{1:t-1}|g_t)}$ ○ 1

(v) [1 pt] ○ $\sum_{x_{t-1}}$ ○ $\sum_{x_t}$ ○ $\max_{x_{t-1}}$ ○ $\max_{x_t}$ ○ 1

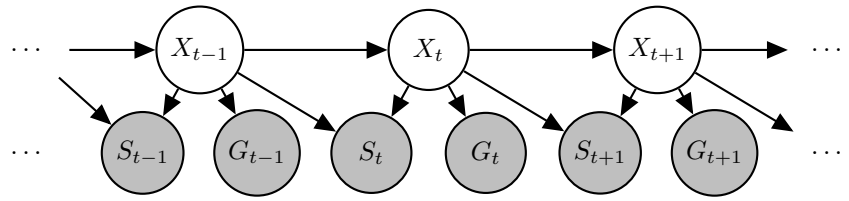
(vi) [1 pt] ○ $P(s_{t-1}|x_{t-1})P(g_{t-1}|x_{t-1})$ ○ $P(x_t, s_t)P(x_t, g_t)$ ○ $P(x_t, s_t, g_t)$
○ $P(x_{t-1}, s_{t-1})P(x_{t-1}, g_{t-1})$ ○ $P(x_{t-1}, s_{t-1}, g_{t-1})$ ○ $P(x_t|s_t)P(x_t|g_t)$
○ $P(x_{t-1}|s_{t-1})P(x_{t-1}|g_{t-1})$ ○ $P(s_t|x_t)P(g_t|x_t)$ ○ 1

(b) It turns out that if the car moves too fast, the quality of the cell phone signal decreases. Thus, the signal-dependent location S_t not only depends on the current state X_t but it also depends on the previous state X_{t-1}. Thus, we modify our original HMM for a new more accurate one, which is given below.



Again, we want to compute the belief $P(x_t|s_{1:t}, g_{1:t})$. In this part we consider an update that combines the dynamics and observation update in a *single* update.

For convenience, the HMM from part **(b)** from the last page is redrawn below.



$P(x_t|s_{1:t}, g_{1:t}) =$ **(i)** **(ii)** **(iii)** **(iv)** $P(x_{t-1}|s_{1:t-1}, g_{1:t-1})$.

Complete the **forward update** expression by choosing the option that fills in each blank.

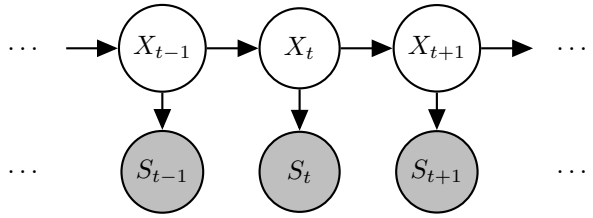
(i) [1 pt]
- ○ $P(s_t, g_t|s_{1:t-1}, g_{1:t-1})$
- ○ $P(s_{1:t-1}, g_{1:t-1}|s_t, g_t)$
- ○ $P(s_t|s_{1:t-1})P(g_t|g_{1:t-1})$
- ○ $\frac{1}{P(s_t, g_t|s_{1:t-1}, g_{1:t-1})}$
- ○ $\frac{1}{P(s_{1:t-1}, g_{1:t-1}|s_t, g_t)}$
- ○ $P(s_{1:t-1}|s_t)P(g_{1:t-1}|g_t)$
- ○ $\frac{1}{P(s_t|s_{1:t-1})P(g_t|g_{1:t-1})}$
- ○ $\frac{1}{P(s_{1:t-1}|s_t)P(g_{1:t-1}|g_t)}$
- ○ 1

(ii) [1 pt)
- ○ $\sum_{x_{t-1}}$
- ○ $\sum_{x_t}$
- ○ $\max_{x_{t-1}}$
- ○ $\max_{x_t}$
- ○ 1

(iii) [1 pt]
- ○ $P(x_{t-2}, x_{t-1}, s_{t-1})P(x_{t-1}, g_{t-1})$
- ○ $P(x_{t-1}, x_t, s_t)P(x_t, g_t)$
- ○ $P(s_{t-1}, g_{t-1}|x_{t-1})$
- ○ $P(s_{t-1}|x_{t-2}, x_{t-1})P(g_{t-1}|x_{t-1})$
- ○ $P(s_t|x_{t-1}, x_t)P(g_t|x_t)$
- ○ $P(s_t, g_t|x_t)$
- ○ $P(x_{t-2}, x_{t-1}|s_{t-1})P(x_{t-1}|g_{t-1})$
- ○ $P(x_{t-1}, x_t|s_t)P(x_t|g_t)$
- ○ 1
- ○ $P(x_{t-2}, x_{t-1}, s_{t-1}, g_{t-1})$
- ○ $P(x_{t-1}, x_t, s_t, g_t)$

(iv) [1 pt]
- ○ $P(x_{t-1}, x_t)$
- ○ $P(x_t|x_{t-1})$
- ○ $P(x_{t-2}, x_{t-1})$
- ○ $P(x_{t-1}|x_{t-2})$
- ○ 1

(c) The Viterbi algorithm finds the most probable sequence of hidden states $X_{1:T}$, given a sequence of observations $s_{1:T}$, for some time $t = T$. Recall the canonical HMM structure, which is shown below.



For this canonical HMM, the Viterbi algorithm performs the following dynamic programming computations:

$$m_t[x_t] = P(s_t|x_t) \max_{x_{t-1}} P(x_t|x_{t-1}) m_{t-1}[x_{t-1}].$$

We consider extending the Viterbi algorithm for the modified HMM from part **(b)**. We want to find the most likely sequence of states $X_{1:T}$ given the sequence of observations $s_{1:T}$ and $g_{1:T}$. The dynamic programming update for $t > 1$ for the modified HMM has the following form:

$m_t[x_t] =$ **(i)** **(ii)** **(iii)** $m_{t-1}[x_{t-1}]$.

Complete the expression by choosing the option that fills in each blank.

(i) [1 pt]
- ○ $\sum_{x_t}$
- ○ $\sum_{x_{t-1}}$
- ○ $\max_{x_t}$
- ○ $\max_{x_{t-1}}$
- ○ 1

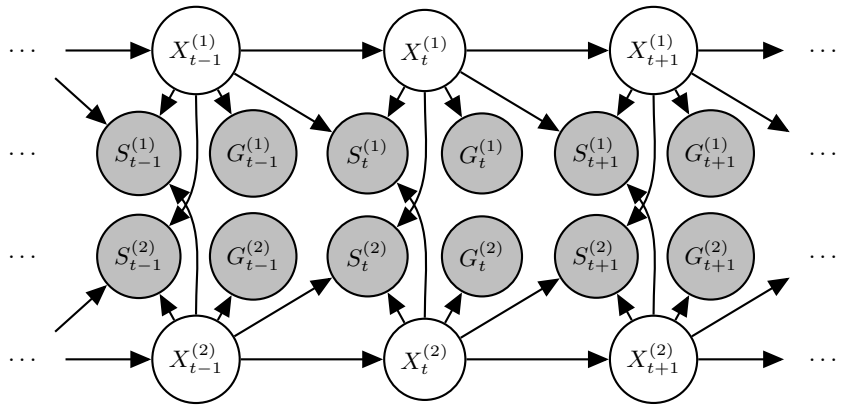
(ii) [1 pt]
- ○ $P(x_{t-2}, x_{t-1}, s_{t-1})P(x_{t-1}, g_{t-1})$
- ○ $P(x_{t-1}, x_t, s_t)P(x_t, g_t)$
- ○ $P(s_{t-1}, g_{t-1}|x_{t-1})$
- ○ $P(s_{t-1}|x_{t-2}, x_{t-1})P(g_{t-1}|x_{t-1})$
- ○ $P(s_t|x_{t-1}, x_t)P(g_t|x_t)$
- ○ $P(s_t, g_t|x_t)$
- ○ $P(x_{t-2}, x_{t-1}|s_{t-1})P(x_{t-1}|g_{t-1})$
- ○ $P(x_{t-1}, x_t|s_t)P(x_t|g_t)$
- ○ 1
- ○ $P(x_{t-2}, x_{t-1}, s_{t-1}, g_{t-1})$
- ○ $P(x_{t-1}, x_t, s_t, g_t)$

(iii) [1 pt]
- ○ $P(x_{t-1}, x_t)$
- ○ $P(x_t|x_{t-1})$
- ○ $P(x_{t-2}, x_{t-1})$
- ○ $P(x_{t-1}|x_{t-2})$
- ○ 1

Q8. [11 pts] Particle Filtering: Where are the Two Cars?

As before, we are trying to estimate the location of cars in a city, but now, we model two cars jointly, i.e. car i for $i \in \{1, 2\}$. The modified HMM model is as follows:

- $X^{(i)}$ – the location of car i
- $S^{(i)}$ – the noisy location of the car i from the signal strength at a nearby cell phone tower
- $G^{(i)}$ – the noisy location of car i from GPS



d	$D(d)$	$E_L(d)$	$E_N(d)$	$E_G(d)$
-4	0.05	0	0.02	0
-3	0.10	0	0.04	0.03
-2	0.25	0.05	0.09	0.07
-1	0.10	0.10	0.20	0.15
0	0	0.70	0.30	0.50
1	0.10	0.10	0.20	0.15
2	0.25	0.05	0.09	0.07
3	0.10	0	0.04	0.03
4	0.05	0	0.02	0

The signal strength from one car gets noisier if the other car is at the same location. Thus, the observation $S_t^{(i)}$ also depends on the current state of the other car $X_t^{(j)}$, $j \neq i$.

The transition is modeled using a drift model D, the GPS observation $G_t^{(i)}$ using the error model E_G, and the observation $S_t^{(i)}$ using one of the error models E_L or E_N, depending on the car's speed and the relative location of both cars. These drift and error models are in the table above. **The transition and observation models are:**

$$P(X_t^{(i)}|X_{t-1}^{(i)}) = D(X_t^{(i)} - X_{t-1}^{(i)})$$

$$P(S_t^{(i)}|X_{t-1}^{(i)}, X_t^{(i)}, X_t^{(j)}) = \begin{cases} E_N(X_t^{(i)} - S_t^{(i)}), & \text{if } |X_t^{(i)} - X_{t-1}^{(i)}| \geq 2 \text{ or } X_t^{(i)} = X_t^{(j)} \\ E_L(X_t^{(i)} - S_t^{(i)}), & \text{otherwise} \end{cases}$$

$$P(G_t^{(i)}|X_t^{(i)}) = E_G(X_t^{(i)} - G_t^{(i)}).$$

Throughout this problem you may give answers either as unevaluated numeric expressions (e.g. $0.1 \cdot 0.5$) or as numeric values (e.g. 0.05). The questions are decoupled.

(a) Assume that at $t = 3$, we have the single particle $(X_3^{(1)} = -1, X_3^{(2)} = 2)$.

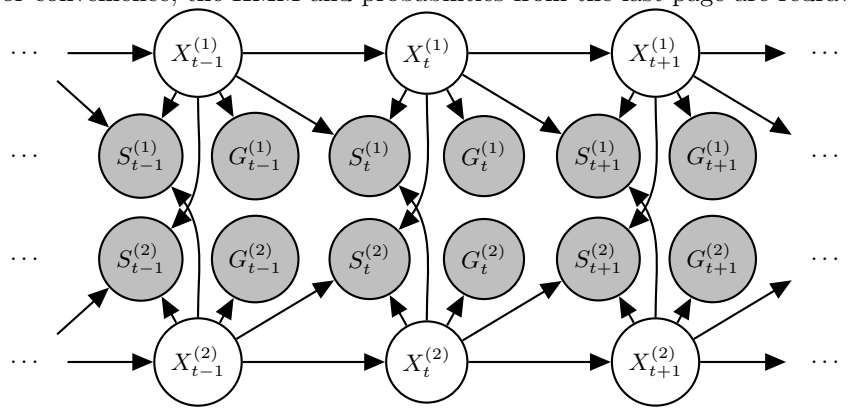
(i) [2 pts] What is the probability that this particle becomes $(X_4^{(1)} = -3, X_4^{(2)} = 3)$ after passing it through the dynamics model?

Answer: _____

(ii) [2 pts] Assume that there are no sensor readings at $t = 4$. What is the joint probability that the *original* single particle (from $t = 3$) becomes $(X_4^{(1)} = -3, X_4^{(2)} = 3)$ and then becomes $(X_5^{(1)} = -4, X_5^{(2)} = 4)$?

Answer: _____

For convenience, the HMM and probabilities from the last page are redrawn and rewritten below.



d	$D(d)$	$E_L(d)$	$E_N(d)$	$E_G(d)$
-4	0.05	0	0.02	0
-3	0.10	0	0.04	0.03
-2	0.25	0.05	0.09	0.07
-1	0.10	0.10	0.20	0.15
0	0	0.70	0.30	0.50
1	0.10	0.10	0.20	0.15
2	0.25	0.05	0.09	0.07
3	0.10	0	0.04	0.03
4	0.05	0	0.02	0

$$P(X_t^{(i)}|X_{t-1}^{(i)}) = D(X_t^{(i)} - X_{t-1}^{(i)})$$

$$P(S_t^{(i)}|X_{t-1}^{(i)}, X_t^{(i)}, X_t^{(j)}) = \begin{cases} E_N(X_t^{(i)} - S_t^{(i)}), & \text{if } |X_t^{(i)} - X_{t-1}^{(i)}| \geq 2 \text{ or } X_t^{(i)} = X_t^{(j)} \\ E_L(X_t^{(i)} - S_t^{(i)}), & \text{otherwise} \end{cases}$$

$$P(G_t^{(i)}|X_t^{(i)}) = E_G(X_t^{(i)} - G_t^{(i)}).$$

For the remaining of this problem, we will be using 2 particles at each time step.

(b) At $t = 6$, we have particles $[(X_6^{(1)} = 3, X_6^{(2)} = 0), (X_6^{(1)} = 3, X_6^{(2)} = 5)]$. Suppose that after weighting, resampling, and transitioning from $t = 6$ to $t = 7$, the particles become $[(X_7^{(1)} = 2, X_7^{(2)} = 2), (X_7^{(1)} = 4, X_7^{(2)} = 1)]$.

(i) [2 pts] At $t = 7$, you get the observations $S_7^{(1)} = 2, G_7^{(1)} = 2, S_7^{(2)} = 2, G_7^{(2)} = 2$. What is the weight of each particle?

Particle	Weight
$(X_7^{(1)} = 2, X_7^{(2)} = 2)$	
$(X_7^{(1)} = 4, X_7^{(2)} = 1)$	

(ii) [2 pts] Suppose both cars' cell phones died so you only get the observations $G_7^{(1)} = 2, G_7^{(2)} = 2$. What is the weight of each particle?

Particle	Weight
$(X_7^{(1)} = 2, X_7^{(2)} = 2)$	
$(X_7^{(1)} = 4, X_7^{(2)} = 1)$	

(c) [3 pts] To decouple this question, assume that you got the following weights for the two particles.

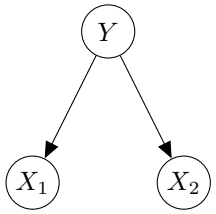
Particle	Weight
$(X_7^{(1)} = 2, X_7^{(2)} = 2)$	0.09
$(X_7^{(1)} = 4, X_7^{(2)} = 1)$	0.01

What is the belief for the location of car 1 and car 2 at $t = 7$?

Location	$P(X_7^{(1)})$	$P(X_7^{(2)})$
$X_7^{(i)} = 1$		
$X_7^{(i)} = 2$		
$X_7^{(i)} = 4$		

Q9. [7 pts] Naive Bayes MLE

Consider a naive Bayes classifier with two features, shown below. We have prior information that the probability model can be parameterized by λ and p, as shown below: Note that $P(X_1 = 0|Y = 0) = P(X_1 = 1|Y = 1) = p$ and $P(X_1|Y) = P(X_2|Y)$ (they share the parameter p). Call this model M1.



Y	$P(Y)$
0	λ
1	1 - λ

X_1	Y	$P(X_1\|Y)$
0	0	p
1	0	$1-p$
0	1	$1-p$
1	1	p

X_2	Y	$P(X_2\|Y)$
0	0	p
1	0	$1-p$
0	1	$1-p$
1	1	p

We have a training set that contains all of the following:

- n_{000} examples with $X_1 = 0$, $X_2 = 0$, $Y = 0$
- n_{010} examples with $X_1 = 0$, $X_2 = 1$, $Y = 0$
- n_{100} examples with $X_1 = 1$, $X_2 = 0$, $Y = 0$
- n_{110} examples with $X_1 = 1$, $X_2 = 1$, $Y = 0$
- n_{001} examples with $X_1 = 0$, $X_2 = 0$, $Y = 1$
- n_{011} examples with $X_1 = 0$, $X_2 = 1$, $Y = 1$
- n_{101} examples with $X_1 = 1$, $X_2 = 0$, $Y = 1$
- n_{111} examples with $X_1 = 1$, $X_2 = 1$, $Y = 1$

(a) [2 pts] Solve for the maximum likelihood estimate (MLE) of the parameter p with respect to $n_{000}, n_{100}, n_{010}, n_{110}, n_{001}, n_{101}, n_{011}$, and n_{111}.

$p =$

(b) [2 pts] For each of the following values of λ, p, X_1, and X_2, classify the value of Y. Hint: No detailed calculation should be necessary.

λ	p	X_1	X_2	Y
3/4	5/8	0	0	
2/5	4/7	1	0	

(c) [1 pt] For the following value of λ, p, X_1, and X_2, classify the value of Y. Detailed calculation may be necessary.

λ	p	X_1	X_2	Y
3/5	3/7	0	0	

(d) [2 pts] Now let's consider a new model M2, which has the same Bayes' Net structure as M1, but where we have a p_1 value for $P(X_1 = 0 \mid Y = 0) = P(X_1 = 1 \mid Y = 1) = p_1$ and a separate p_2 value for $P(X_2 = 0 \mid Y = 0) = P(X_2 = 1 \mid Y = 1) = p_2$, and we don't constrain $p_1 = p_2$. Let L_{M1} be the likelihood of the training data under model M1 with the maximum likelihood parameters for M1. Let L_{M2} be the likelihood of the training data under model M2 with the maximum likelihood parameters for M2. Which of the following properties are guaranteed to be true?

- ◯ $L_{M1} \geq L_{M2}$
- ◯ $L_{M1} \leq L_{M2}$
- ◯ Insufficient information, the above relationships rely on the particular training data.
- ◯ None of the above.

Q10. [10 pts] Perceptron

(a) [1 pt] Suppose you have a binary perceptron in 2D with weight vector $\mathbf{w} = r\ [w_1, w_2]^T$. You are given w_1 and w_2, and are given that $r > 0$, but otherwise not told what r is. *Assume that ties are broken as positive.*

Can you determine the perceptron's classification of a new example x with known feature vector $f(x)$?

◯ Always ◯ Sometimes ◯ Never

(b) Now you are learning a multi-class perceptron between 4 classes. The weight vectors are currently $[1, 0]^T$, $[0, 1]^T$, $[-1, 0]^T$, $[0, -1]^T$ for the classes A, B, C, and D. The next training example x has a **label of A** and feature vector $f(x)$.
For the following questions, *do not make any assumptions about tie-breaking.* (Do not write down a solution that creates a tie.)

(i) [1 pt] Write down a feature vector in which no weight vectors will be updated.

$f(x) = \begin{bmatrix} \\ \\ \end{bmatrix}$ ◯ Not possible

(ii) [1 pt] Write down a feature vector in which **only** $\mathbf{w}_A$ will be updated by the perceptron.

$f(x) = \begin{bmatrix} \\ \\ \end{bmatrix}$ ◯ Not possible

(iii) [1 pt] Write down a feature vector in which **only** $\mathbf{w}_A$ and $\mathbf{w}_B$ will be updated by the perceptron.

$f(x) = \begin{bmatrix} \\ \\ \end{bmatrix}$ ◯ Not possible

(iv) [1 pt] Write down a feature vector in which **only** $\mathbf{w}_A$ and $\mathbf{w}_C$ will be updated by the perceptron.

$f(x) = \begin{bmatrix} \\ \\ \end{bmatrix}$ ◯ Not possible

The weight vectors are the same as before, but now there is a bias feature with value of 1 for all x and the weight of this bias feature is $0, -2, 1, -1$ for classes A, B, C, and D respectively. As before, the next training example x has a **label of A** and a feature vector $f(x)$. The always "1" bias feature is the first entry in $f(x)$.

(v) [1 pt] Write down a feature vector in which **only** $\mathbf{w}_B$ and $\mathbf{w}_C$ will be updated by the perceptron.

$f(x) = \begin{bmatrix} 1 \\ \\ \end{bmatrix}$ ◯ Not possible

(vi) [1 pt] Write down a feature vector in which **only** $\mathbf{w}_A$ and $\mathbf{w}_C$ will be updated by the perceptron.

$f(x) = \begin{bmatrix} 1 \\ \\ \end{bmatrix}$ ◯ Not possible

(c) Suppose your training data is linearly separable and you are classifying between label Y and label Z. The mistake bound ($\frac{k}{\delta^2}$) is equal to 3, which is the maximum number of weight vector updates the perceptron might have to do before it is guaranteed to converge. There are 100 examples in your training set. Assume the perceptron cycles through the 100 examples in a fixed order.

(i) [1 pt] What is the maximum number of classifications the perceptron might make before it is guaranteed to have converged?

◯ 300 ◯ 103 ◯ 100^3 ◯ 3 ◯ 3^{100} ◯ None of the options

(ii) [2 pts] [*true* or *false*] After convergence, the learned perceptron will correctly classify *all* training examples.