UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering

CS 440/ECE 448 Artificial Intelligence
Spring 2020

PRACTICE EXAM 2

Actual Exam will be held on Compass, Monday, March 30, 2020

- This is will be an OPEN BOOK exam. You will be allowed to use any textbook, notes, calculator, and/or internet search available to you.
- The actual exam will be held on Compass, on Monday, March 30, 2020.
- This practice exam is almost four times as long as the actual exam will be.

Name: ____

netid: ____

Problem 1 (4 points)

Use the axioms of probability to prove that $P(\neg A) = 1 - P(A)$.
Solution:

- From the third axiom, $P(A \vee \neg A) = P(A) + P(\neg A) - P(A \wedge \neg A)$.
- The event $(A \vee \neg A)$ is always true, so from the second axiom, $P(A \vee \neg A) = 1$. The event $(A \wedge \neg A)$ is always false, so from the second axiom, $P(A \wedge \neg A) = 0$.
- Combining the two statements above, $1 = P(A) + P(\neg A)$. Q.E.D.

Problem 2 (4 points)

Consider the following joint probability distribution:

$$\begin{aligned} P(A, B) &= 0.12 \\ P(A, \neg B) &= 0.18 \\ P(\neg A, B) &= 0.28 \\ P(\neg A, \neg B) &= 0.42 \end{aligned}$$

What are the marginal distributions of A and B? Are A and B independent and why?

Solution: $P(A) = 0.3, P(\neg A) = 0.7, P(B) = 0.4, P(\neg B) = 0.6$. They are independent, because $P(A)P(B) = P(A, B) = 0.12, P(A)P(\neg B) = P(A, \neg B) = 0.18$, and so on.

Problem 3 (4 points)

A couple has two children, and one of them is a boy. What is the probability that they're both boys?

(You may assume that, for this couple, the *a priori* probability of any child being male is exactly 50%).

Solution: Let A be the event "at least one boy," and let B be the event "two boys." $P(A) = 3/4$, $P(A \wedge B) = 1/4$, so $P(B|A) = P(A \wedge B)/P(A) = 1/3$.

Problem 4 (4 points)

A friend who works in a big city owns two cars, one small and one large. Three-quarters of the time he drives the small car to work, and one-quarter of the time he drives the large car. If he takes the small car, he usually has little trouble parking, and so is at work on time with probability 0.9. If he takes the large car, he is at work on time with probability 0.6. Given that he was on time on a particular morning, what is the probability that he drove the small car?

Solution: Let S be the event "takes the small car," and T is the event "arrives on time." Then

$$P(S|T) = \frac{P(T|S)P(S)}{P(T)} = \frac{P(T|S)P(S)}{P(T|S)P(S) + P(T|\neg S)P(\neg S)} = \frac{0.9(3/4)}{0.9(3/4) + 0.6(1/4)} = \frac{27}{33}$$

Problem 5 (8 points)

Let A and B be independent binary random variables with $p(A = 1) = 0.1$, $p(B = 1) = 0.4$. Let C denote the event that at least one of them is 1, and let D denote the event that exactly one of them is 1.

(a) What is $P(C)$?

Solution:

$$\begin{aligned} P(C) &= p(A = 1, B = 1) + p(A = 1, B = 0) + p(A = 0, B = 1) \\ &= (0.1)(0.4) + (0.1)(0.6) + (0.9)(0.4) = 0.46 \end{aligned}$$

where the last line follows from the independence of A and B.

(b) What is $P(D)$?

Solution:

$$\begin{aligned} P(D) &= p(A = 1, B = 0) + p(A = 0, B = 1) \\ &= (0.1)(0.6) + (0.9)(0.4) = 0.42 \end{aligned}$$

(c) What is $P(D|A = 1)$?

Solution:

$$\begin{aligned} P(D|A = 1) &= P(D, A = 1)/p(A = 1) \\ &= p(A = 1, B = 0)/p(A = 1) \\ &= \frac{0.06}{0.1} = 0.6 \end{aligned}$$

(d) Are A and D independent? Why?

Solution: No. $P(D|A = 1) \neq P(D)$.

Problem 6 (4 points)

Consider a Nave Bayes classifier with 100 feature dimensions. The label Y is binary with $P(Y=0) = P(Y=1) = 0.5$. All features are binary, and have the same conditional probabilities: $P(X_i = 1|Y = 0) = a$ and $P(X_i = 1|Y = 1) = b$ for $i = 1, \dots, 100$. Given an item X with alternating feature values $(X_1 = 1, X_2 = 0, X_3 = 1, ..., X_{100} = 0)$, compute $P(Y = 1|X)$.

Solution:

$$
\begin{aligned}
P(Y=1|X) &= \frac{P(Y=1)\prod_{i=1}^{100} P(X_i|Y=1)}{P(Y=1)\prod_{i=1}^{100} P(X_i|Y=1) + P(Y=0)\prod_{i=1}^{100} P(X_i|Y=0)} \\
&= \frac{0.5b^{50}(1-b)^{50}}{0.5b^{50}(1-b)^{50} + 0.5a^{50}(1-a)^{50}} \\
&= \frac{b^{50}(1-b)^{50}}{b^{50}(1-b)^{50} + a^{50}(1-a)^{50}}
\end{aligned}
$$

Problem 7 (8 points)

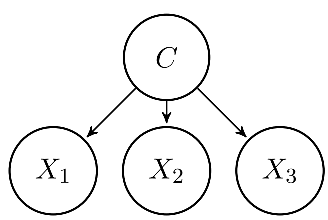
Consider the data points in Table 1, representing a set of seven patients with up to three different symptoms. We want to use the Naïve Bayes assumption to diagnose whether a person has the flu based on the symptoms.

Sore Throat	Stomachache	Fever	Flu
No	No	No	No
No	No	Yes	Yes
No	Yes	No	No
Yes	No	No	No
Yes	No	Yes	Yes
Yes	Yes	No	Yes
Yes	Yes	Yes	No

Table 1: Symptoms of seven patients, three of whom had the flu.

(a) Define random variables, and show the structure of the Bayes network representing a Naïve Bayes classifier for the flu, using the variables shown in Table 1.

Solution: The binary variables could be called F, T, S, and E, representing the presence of flu, sore throat, stomach ache, and fever, respectively. The Bayes net is then



(b) Calculate the maximum likelihood conditional probability tables.

Solution:

F	$P(F)$	$P(T\|F)$	$P(S\|F)$	$P(E\|F)$
0	4/7	1/2	1/2	1/4
1	3/7	2/3	1/3	2/3

(c) If a person has stomachache and fever, but no sore throat, what is the probability of him or her having the flu (according to the conditional probability tables you calculated in part (b))?

Solution:

$$\begin{aligned}
P(F|\neg T, S, E) &= \frac{P(\neg T, S, E, F)}{P(\neg T, S, E)} \\
&= \frac{P(F, \neg T, S, E)}{P(F, \neg T, S, E) + P(\neg F, \neg T, S, E)} \\
&= \frac{(3/7)(1/3)(1/3)(2/3)}{(3/7)(1/3)(1/3)(2/3) + (4/7)(1/2)(1/2)(1/4)} \\
&= \frac{8}{17}
\end{aligned}$$

Problem 8 (8 points)

You're creating sentiment analysis. You have a training corpus with four movie reviews:

Review #	Sentiment	Review
1	+	what a great movie
2	+	I love this film
3	-	what a horrible movie
4	-	I hate this film

Let $Y = 1$ for positive sentiment, $Y = 0$ for negative sentiment.

(a) What's the maximum likelihood estimate of $P(Y = 1)$?

Solution: Maximum likelihood estimate is

$$P(Y=1) = \frac{\#\text{ times } Y = 1}{\#\text{ training tokens}} = \frac{2}{4} = \frac{1}{2}$$

(b) Find maximum likelihood estimates $P(W|Y = 1)$ and $P(W|Y = 0)$ for the ten words $W \in \{\text{what,a,movie,I,this,film,great,love,horrible,hate}\}$.

Solution: There are three cases. For the words $W \in \{\text{what,a,movie,I,this,film}\}$, $P(W|Y = 0) = P(W|Y = 1) = 1/8$. For the words $W \in \{\text{great,love}\}$, $P(W|Y = 0) = 0$, and $P(W|Y = 1) = 1/8$. For the words $W \in \{\text{horrible,hate}\}$, $P(W|Y = 1) = 0$, and $P(W|Y = 0) = 1/8$.

(c) Use Laplace smoothing, with a smoothing parameter of $k = 1$, to estimate $P(W|Y = 1)$ and $P(W|Y = 0)$ for the ten words $W \in \{\text{what,a,movie,I,this,film,great,love,horrible,hate}\}$.

Solution: There are three cases. For the words $W \in \{\text{what,a,movie,I,this,film}\}$, $P(W|Y = 0) = P(W|Y = 1) = 2/18$. For the words $W \in \{\text{great,love}\}$, $P(W|Y = 0) = 1/18$, and $P(W|Y = 1) = 2/18$. For the words $W \in \{\text{horrible,hate}\}$, $P(W|Y = 1) = 1/18$, and $P(W|Y = 0) = 2/18$.

(d) Using some other method (unknown to you), your professor has estimated the following conditional probability table:

Y	$P(\text{great}\|Y)$	$P(\text{love}\|Y)$	$P(\text{horrible}\|Y)$	$P(\text{hate}\|Y)$
1	0.01	0.01	0.005	0.005
0	0.005	0.005	0.01	0.01

and $P(Y = 1) = 0.5$. All other words (except great, love, horrible, and hate) can be considered out-of-vocabulary, and you can assume that $P(W|Y) = 1$ for all out-of-vocabulary words. Under these assumptions, what is the probability $P(Y = 1|R)$ that the following 14-word review is a positive review?

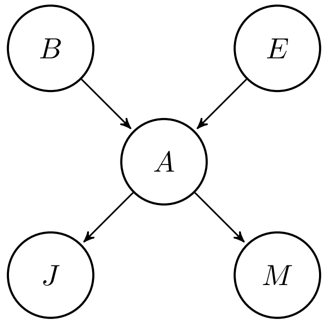
$$R = \{\text{“I'm horrible fond of this movie, and I hate anyone who insults it.”}\}$$

Solution:

$$P(Y=1|R) = \frac{P(Y=1,R)}{P(Y=1,R)+P(Y=0,R)} = \frac{(0.5)(0.005)(0.005)}{(0.5)(0.005)(0.005)+(0.5)(0.01)(0.01)} = \frac{1}{5}$$

Problem 9 (4 points)

Consider the "Burglary" Bayesian network:



(a) How many independent parameters does this network have? How many entries does the full joint distribution table have?

Solution: There are five binary variables: two with no parents (B and E, one parameter each), one with two parents (A, four parameters), and two with one parent each (J and M, two parameters each), for a total of ten independent parameters. The full joint distribution table has $2^5 - 1 = 31$ parameters.

(b) If no evidence is observed, are B and E independent?

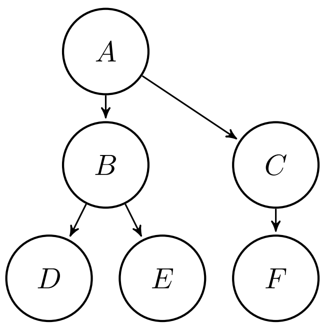
Solution: Yes, because they have no common ancestors.

(c) Are B and E conditionally independent given the observation that $A =$True?

Solution: No. Knowing that Earthquake=True makes Burglary less probable.

Problem 10 (8 points)

Consider the following Bayes network (all variables are binary):



$P(A) = 0.8$

A	$P(B\|A)$	$P(C\|A)$
0	0.2	0.6
1	0.5	0.8

B	$P(D\|B)$	$P(E\|B)$
0	0.5	0.8
1	0.5	0.8

C	$P(F\|C)$	
0	0.01	
1	0.2	

(a) Are D and E independent?

Solution: Yes. This is a trick question. The structure of the Bayes net shows them to be conditionally independent given B, but not independent. However, in the probability table, notice that $P(D|B) = P(D|\neg B)$, therefore D is independent of B, despite the arrow shown in the Bayes net. Similarly, $P(E|B) = P(E|\neg B)$, therefore E is independent of B, despite the arrow shown in the Bayes net. Since there is no other path connecting D to E except the one going through B, they are independent.

(b) Are D and E conditionally independent given B?

Solution: Yes. This is not a trick question. The structure of the Bayes net shows that they are conditionally independent given B.

(c) If you did not know the Bayesian network, how many numbers would you need to represent the full joint probability table?

Solution: There are 2^6 possible combinations of 6 binary variables, so you'd need $2^6 - 1 = 63$ numbers.

(d) If you knew the Bayes network as shown above, but the variables were ternary instead of binary, how many values would you need to represent the full joint probability table and the conditional probability tables, respectively?

Solution: Conditional probability tables: For each variable, the number of trainable parameters is (# possible values of the variable, minus 1)$\times$(# possible values of its parents). $P(A)$ would need 2 trainable parameters, each of the other five variables would need $2 \times 3 = 6$ trainable parameters, for a total of $2 + 5 \times 2 \times 3 = 32$ trainable parameters.

Full joint probability table: there are 3^6 possible combinations of the variables, so you would need to store $3^6 - 1$ parameters.

(e) Write down the expression for the joint probability of all the variables in the network, in terms of the model parameters given above.

Solution:

$$P(A, B, C, D, E, F) = P(A)P(B|A)P(C|A)P(D|B)P(E|B)P(F|C)$$

(f) Find $P(A = 0, B = 1, C = 1, D = 0)$.

Solution:

$$P(A = 0, B = 1, C = 1, D = 0) = (0.2)(0.2)(0.6)(0.5) = \frac{3}{250}$$

(g) Find $P(B|A = 1, D = 0)$.

Solution:

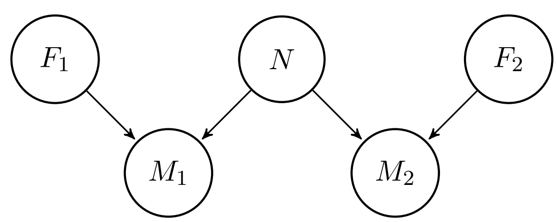
$$\begin{aligned} P(B|A = 1, D = 0) &= \frac{P(A = 1, B = 1, D = 0)}{P(A = 1, B = 1, D = 0) + P(A = 1, B = 0, D = 0)} \\ &= \frac{(0.8)(0.5)(0.5)}{(0.8)(0.5)(0.5) + (0.8)(0.5)(0.5)} \\ &= \frac{1}{2} \end{aligned}$$

Problem 11 (8 points)

Two astronomers in different parts of the world make measurements M_1 and M_2 of the number of stars N in some small region of the sky, using their telescopes. Under normal circumstances, this experiment has three possible outcomes: either the measurement is correct, or the measurement overcounts the stars by one (one star too high), or the measurement undercounts the stars by one (one star too low). There is also the possibility, however, of a large measurement error in either telescope (events F_1 and F_2, respectively), in which case the measured number will be **at least** three stars too low (regardless of whether the scientist makes a small error or not), or, if N is less than 3, fail to detect any stars at all.

(a) Draw a Bayesian network for this problem.

Solution: A solution must include the variables N, M_1, M_2 with the dependencies shown below. The variables F_1, F_2 are optional:



(b) Write out a conditional distribution for $P(M_1|N)$ for the case where $N \in \{1,2,3\}$ and $M_1 \in \{0,1,2,3,4\}$. Each entry in the conditional distribution table should be expressed as a function of the parameters e and/or f.

Solution:

	M_1				
N	0	1	2	3	4
1	$e+f$	$1-2e-f$	e	0	0
2	f	e	$1-2e-f$	e	0
3	f	0	e	$1-2e-f$	e

(c) Suppose $M_1 = 1$ and $M_2 = 3$. What are the possible numbers of stars if you assume no prior constraint on the values of N?

Solution: $N = 2$ is possible, if both made small mistakes. $N = 4$ is possible, if M_2 made a small and M_1 a big mistake. $N \geq 6$ is possible, if both M_1 and M_2 made big mistakes.

(d) What is the most likely number of stars, given the observations $M_1 = 1, M_2 = 3$? Explain how to compute this, or if it is not possible to compute, explain what additional information is needed and how it would affect the result.

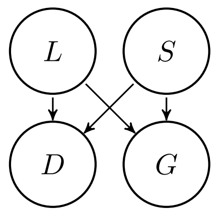
Solution: We need to find the value of N that maximizes $P(N, M_1 = 1, M_2 = 3)$. We have that $P(N = 2, M_1 = 1, M_2 = 3) = P(N = 2)e^2$. We know that $P(N =$

$4, M_1 = 1, M_2 = 3) \leq P(N = 4)fe$; we don't know exactly how much it is, because we don't know $P(M_1 = 1|N = 4)$, but we know that $P(M_1 = 1|N = 4) \leq f$. So if $P(N = 2)e > P(N = 4)f$, $N = 2$ is the most probable value. If $P(N = 2)e \leq P(N = 4)f$, then it depends on the way in which big errors are distributed among the various values that are "at least three stars" too small.

Problem 12 (8 points)

Maria likes ducks and geese. She notices that when she leaves the heat lamp on (in her back yard), she is likely to see ducks and geese. When the heat lamp is off, she sees ducks and geese in the summer, but not in the winter.

(a) The following Bayes net summarizes Maria's model, where the binary variables D,G,L, and S denote the presence of ducks, geese, heat lamp, and summer, respectively:



On eight randomly selected days throughout the year, Maria makes the observations shown in Table 1.

day	D	G	L	S	day	D	G	L	S
1	0	1	1	0	5	1	0	0	1
2	1	0	1	0	6	1	0	1	1
3	0	0	0	0	7	0	1	1	1
4	0	0	0	0	8	0	1	0	1

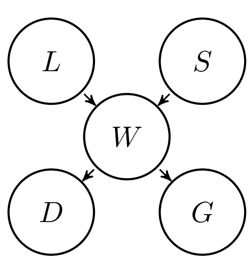
Table 1: Observations of the presence of ducks and geese, as a function of season (S) and heat lamp (L).

Write the maximum-likelihood conditional probability tables for D, G, L and S.

Solution: We have that $P(S) = 0.5$, $P(L) = 0.5$, and

S	L	$P(D\|S,L)$	$P(G\|S,L)$
0	0	0	0
0	1	0.5	0.5
1	0	0.5	0.5
1	1	0.5	0.5

(b) Maria speculates that ducks and geese don't really care whether the lamp is lit or not, they only care whether or not the temperature in her yard is warm. She defines a binary random variable, W, which is 1 when her back yard is warm, and she proposes the following revised Bayes net:



She forgot to measure the temperature in her back yard, so W is a hidden variable. Her initial guess is that $P(D|W) = \frac{2}{3}$, $P(D|\neg W) = \frac{1}{3}$, $P(G|W) = \frac{2}{3}$, $P(G|\neg W) = \frac{1}{3}$, $P(W|L \wedge S) = \frac{2}{3}$, $P(W|\neg(L \wedge S)) = \frac{1}{3}$. Find the posterior probability $P(W|\text{day})$ for each of the 8 days, day $\in \{1, \ldots, 8\}$, whose observations are shown in Table 1.

Solution:

day	1	2	3	4	5	6	7	8
$P(W\|\text{day})$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$

Problem 13 (8 points)

Suppose you have a Bayes net with two binary variables, Jahangir (J) and Shahjahan (S):

$$J \longrightarrow S$$

This network has three trainable parameters: $P(J) = a$, $P(S|J) = b$, and $P(S|\neg J) = c$. Suppose you have a training dataset in which S is observed, but J is hidden. Specifically, there are N training tokens for which $S =$ True, and M training tokens for which $S =$ False. Given current estimates of a, b, and c, you want to use the EM algorithm to find improved estimates $\hat{a}$, $\hat{b}$, and $\hat{c}$.

(a) Find the following expected counts, in terms of M, N, a, b, and c:

$$E[\#\text{ times } J \text{ True}] =$$
$$E[\#\text{ times } J \text{ and } S \text{ True}] =$$
$$E[\#\text{ times } J \text{ True and } S \text{ False}] =$$

Solution:

$$E[\#\text{ times } J \text{ True}] = \frac{abN}{ab + (1-a)c} + \frac{a(1-b)M}{a(1-b) + (1-a)(1-c)}$$
$$E[\#\text{ times } J \text{ and } S \text{ True}] = \frac{abN}{ab + (1-a)c}$$
$$E[\#\text{ times } J \text{ True and } S \text{ False}] = \frac{a(1-b)M}{a(1-b) + (1-a)(1-c)}$$

(b) Find re-estimated values $\hat{a}$, $\hat{b}$, and $\hat{c}$ in terms of M, N, $E[\#\text{ times } J \text{ True}]$, $E[\#\text{ times } J \text{ and } S \text{ True}]$, and $E[\#\text{ times } J \text{ True and } S \text{ False}]$.

Solution:

$$\hat{a} = \frac{E[\#\text{ times } J \text{ True}]}{M+N}$$
$$\hat{b} = \frac{E[\#\text{ times } J \text{ and } S \text{ True}]}{E[\#\text{ times } J \text{ True}]}$$
$$\hat{c} = \frac{E[\#\text{ times } J \text{ False and } S \text{ True}]}{M + N - E[\#\text{ times } J \text{ False}]}$$

Problem 14 (4 points)

In a context-free grammar (CFG), every production rule can be written in the form

$$N_1 \rightarrow \pumpkin$$

where N_1 is a non-terminal, and $\pumpkin$ is some output. In a normal-form CFG, what are the possible values of $\pumpkin$?

Solution: $\pumpkin$ can be either the sequence of two non-terminals ($N_2 N_3$), or a terminal (T).

Problem 15 (4 points)

In a context-free grammar, what is a terminal symbol? What is a non-terminal symbol?

Solution:A terminal symbol is one that is observed in the data, for example, a word. A non-terminal is a generalization, for example, a phrase or a part of speech, that is never observed in the data, but that can generate one or more terminals.

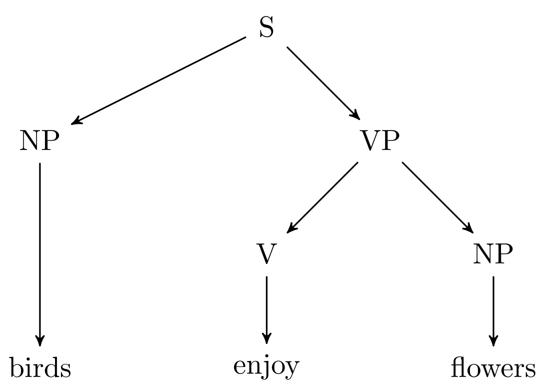
Problem 16 (8 points)

Consider the following probabilistic context-free grammar:

$$
\begin{array}{rll}
\text{S}\rightarrow & \text{NP VP} & P = 1.0 \\
\text{NP}\rightarrow & \text{birds} & P = 0.5 \\
\text{NP}\rightarrow & \text{flower} & P = 0.5 \\
\text{VP}\rightarrow & \text{V} & P = 0.5 \\
\text{VP}\rightarrow & \text{V NP} & P = 0.5 \\
\text{V}\rightarrow & \text{enjoy} & P = 0.5 \\
\text{V}\rightarrow & \text{grow} & P = 0.5
\end{array}
$$

(a) Draw a tree showing how the S nonterminal can produce the sentence "birds enjoy flowers".

Solution:



(b) What is the probability, according to this model, of the sentence "birds enjoy flowers"?

Solution:

$$
\begin{array}{rll}
\text{S}\rightarrow & \text{NP VP} & P = 1.0 \\
\text{NP}\rightarrow & \text{birds} & P = 0.5 \\
\text{NP}\rightarrow & \text{V NP} & P = 0.5 \\
\text{V}\rightarrow & \text{enjoy} & P = 0.5 \\
\text{NP}\rightarrow & \text{flower} & P = 0.5
\end{array}
$$

Multiplying the probabilities of the five production rules, we get $P = 1/16$.

Problem 17 (8 points)

The University of Illinois Vaccavolatology Department has four professors, named Aya, Bob, Cho, and Dale. The building has only one key, so we take special care to protect it. Every day Aya goes to the gym, and on the days she has the key, 60% of the time she forgets it next to the bench press. When that happens one of the other three TAs, equally likely, always finds it since they work out right after. Bob likes to hang out at Einstein Bagels and 50% of the time he is there with the key, he forgets the key at the shop. Luckily Cho always shows up there and finds the key whenever Bob forgets it. Cho has a hole in her pocket and ends up losing the key 80% of the time somewhere on Goodwin street. However, Dale takes the same path to campus and always finds the key. Dale has a 10% chance to lose the key somewhere in the Vaccavolatology classroom, but then Cho picks it up. The professors lose the key at most once per day, around noon (after losing it they become extra careful for the rest of the day), and they always find it the same day in the early afternoon.

(a) Let X_t = the first letter of the name of the person who has the key ($X_t \in \{A, B, C, D\}$). Find the maximum likelihood estimates of the Markov transition probabilities $P(X_t|X_{t-1})$.

Solution:

	X_t			
X_{t-1}	A	B	C	D
A	0.4	0.2	0.2	0.2
B	0	0.5	0.5	0
C	0	0	0.2	0.8
D	0	0	0.1	0.9

(b) Sunday night Bob had the key (the initial state distribution assigns probability 1 to $X_0 = B$ and probability 0 to all other states). The first lecture of the week is Tuesday at 4:30pm, so one of the professors needs to open the building at that time. What is the probability for each professor to have the key at that time? Let X_0, X_{Mon} and X_{Tue} be random variables corresponding to who has the key Sunday, Monday, and Tuesday evenings, respectively. Fill in the probabilities in the table below.

Professor	$P(X_0)$	$P(X_{Mon})$	$P(X_{Tue})$
A	0		
B	1		
C	0		
D	0		

Solution:

Professor	$P(X_0)$	$P(X_{Mon})$	$P(X_{Tue})$
A	0	0	0
B	1	0.5	0.25
C	0	0.5	0.35
D	0	0	0.4

Problem 18 (8 points)

Consider a hidden Markov model (HMM) whose hidden variable denotes part of speech (POS), $X_t \in \{N, V\}$ where N =noun, V =verb, the initial state probability is $P(X_1 = N) = 0.8$, and the transition probabilities are $P(X_t = N|X_{t-1} = N) = 0.1$ and $P(X_t = V|X_{t-1} = V) = 0.1$. Suppose we have the observation probability matrix given in Table 1.

E_t	rose	bill	likes
$P(E_t\|X_t = N)$	0.4	0.4	0.2
$P(E_t\|X_t = V)$	0.2	0.2	0.6

Table 1: Observation probabilities for a simple POS HMM.

You are given the sentence "bill rose." You want to figure out whether each of these two words, "bill" and "rose", is being used as a noun or a verb.

(a) List the four possible combinations of (X_1, X_2). For each possible combination, give $P(X_1, E_1, X_2, E_2)$.

Solution:

$P(X_1, E_1, X_2, E_2)$	$X_2 = N$	$X_2 = V$
$X_1 = N$	(0.8)(0.4)(0.1)(0.4)	(0.8)(0.4)(0.9)(0.2)
$X_1 = V$	(0.2)(0.2)(0.9)(0.4)	(0.2)(0.2)(0.1)(0.2)

(b) Find $P(X_2 = V|E_1 = \text{bill}, E_2 = \text{rose})$.

Solution: Using the forward algorithm, we can compute the joint probabilities as

$$\begin{aligned} P(E, X_2 = V) &= P(X_1 = N, E_1, X_2 = V, E_2) + P(X_1 = V, E_1, X_2 = V, E_2) \\ &= (0.8)(0.4)(0.9)(0.2) + (0.2)(0.2)(0.1)(0.2) \\ P(E, X_2 = N) &= P(X_1 = N, E_1, X_2 = N, E_2) + P(X_1 = V, E_1, X_2 = N, E_2) \\ &= (0.8)(0.4)(0.1)(0.4) + (0.2)(0.2)(0.9)(0.4) \end{aligned}$$

Dividing the first row by the sum of the two rows, we get

$$P(X_2 = V|E) = \frac{(0.8)(0.4)(0.9)(0.2) + (0.2)(0.2)(0.1)(0.2)}{(0.8)(0.4)(0.9)(0.2) + (0.2)(0.2)(0.1)(0.2) + (0.8)(0.4)(0.1)(0.4) + (0.2)(0.2)(0.9)(0.4)}$$

(c) Use the Viterbi algorithm to find the most likely state sequence for this sentence.

Solution:

- To find the backpointer from $X_2 = N$, we find the maximum among the two possibilities $P(X_1 = N, E_1, X_2 = N, E_2)$ and $P(X_1 = V, E_1, X_2 = N, E_2)$. The larger of the two is $P(X_1 = V, E_1, X_2 = N, E_2) = (0.2)(0.2)(0.9)(0.4)$, so the backpointer from $X_2 = N$ points to $X_1 = V$.
- To find the backpointer from $X_2 = V$, we find the maximum among the two possibilities $P(X_1 = N, E_1, X_2 = V, E_2)$ and $P(X_1 = V, E_1, X_2 = V, E_2)$. The larger of the two is $P(X_1 = N, E_1, X_2 = V, E_2) = (0.8)(0.4)(0.9)(0.2)$, so the backpointer from $X_2 = V$ points to $X_1 = N$.

- To find the best terminal state, then, we find the maximum among the two possibilities $P(X_1 = V, E_1, X_2 = N, E_2)$ and $P(X_1 = N, E_1, X_2 = V, E_2)$. The larger of the two is $P(X_1 = N, E_1, X_2 = V, E_2) = (0.8)(0.4)(0.9)(0.2)$, so the maximum likelihood state sequence is $(X_1, X_2) = (N, V)$.

Problem 19 (4 points)

In a pinhole camera, a light source at (x, y, z) is projected onto a pixel at $(x', y', -f)$ through a pinhole at $(0, 0, 0)$. Write $\sqrt{(x')^2 + (y')^2}$ in terms of x, y, z, and f.

Solution: The pinhole camera equations are

$$x' = \frac{-fx}{z}, \quad y' = \frac{-fy}{z}$$

from which we derive

$$\sqrt{(x')^2 + (y')^2} = \frac{f}{z}\sqrt{x^2 + y^2}$$

Problem 20 (4 points)

Under what circumstances is a difference-of-Gaussians filter more useful for edge detection than a simple pixel difference?

Solution:A difference-of-Gaussians filter first smooths the input image (using a Gaussian smoother), then computes a pixel difference. The smoothing step can reduce random noise. Therefore, this procedure is more useful if the input image has some random noise in it.

Problem 21 (4 points)

The real world contains two parallel infinite-length lines, whose equations, in terms of the coordinates (x, y, z), are parameterized as $ax + by + cz = d$ and $ax + by + cz = e$; in addition, both of these lines are on the ground plane, $y = g$, for some constants (a, b, c, d, e, g). Show that the images of these two lines, as imaged by a pinhole camera, converge to a vanishing point, and give the coordinates (x', y') of the vanishing point.

Solution: The pinhole camera equations are

$$x' = \frac{-fx}{z}, \quad y' = \frac{-fy}{z}$$

From which we derive

$$x = \frac{-zx'}{f}, \quad y = \frac{-zy'}{f}$$

So the equations of the two lines are

$$\begin{aligned} -\frac{ax'}{f} - \frac{by'}{f} + c &= \frac{d}{z} \\ -\frac{ax'}{f} - \frac{by'}{f} + c &= \frac{e}{z} \end{aligned}$$

As $z \to \infty$, the right-hand-sides of these two equations both go to zero, and the equations of both lines converge to

$$ax' + by' = cf$$

In addition, we have $y = g$, so $y' = -fg/z \to 0$, and therefore $x' = cf/a$. The coordinates are $(x', y') = (cf/a, 0)$.

Problem 22 (4 points)

Consider the convolution equation

$$Z(x', y') = \sum_m \sum_n h(m, n) Y(x' - m, y' - n)$$

Where $Y(x', y')$ is the original image, $Z(x', y')$ is the filtered image, and the filter $h(m, n)$ is given by

$$h(m, n) = \begin{cases} \frac{1}{21} & 1 \le m \le 3, \quad -3 \le n \le 3 \\ -\frac{1}{21} & -3 \le m \le -1, \quad -3 \le n \le 3 \end{cases}$$

Would this filter be more useful for smoothing, or for edge detection? Why?

Solution: The sum of $h(m, n)$, over all m and n, is 0. So if it is filtering a constant-color region, the output would always be zero, regardless of the input color. So it's not very useful for smoothing.

Any given pixel of $Z(x', y')$ is the difference between the pixels $Y(x', y')$ to its left, minus those to its right. Since it's computing a difference, it would be useful for edge detection.

Problem 23 (4 points)

The pinhole camera equations are

$$x' = \frac{-fx}{z}, \quad y' = \frac{-fy}{z}$$

Explain in words how these equations can be used to show that the image of any object gets smaller as the object gets farther from the camera.

Solution: Two points, on opposite sides of the object, project images onto the film at positions that are inversely proportional to the distance (z) from the object to the camera. Since the positions of these two points on the image are inversely proportional to z, the distance between them is also inversely proportional to z, therefore as the object gets farther from the camera, the distance between the opposite sides of the object (in the image) decreases.