Exam 1 Review

CS440/ECE448, Spring 2021

Exam date: Friday, March 5, 1:00pm

Question 1

Perceptrons can learn all Boolean functions of two inputs.

◯ True

√ **False**

Explain:

Solution: False they can only learn linearly separable functions. In particular, they cannot learn the XOR function.

Question 2

A particular perceptron is initialized with the weights $w = [1, 1, 1]$ and the bias $b = 1$.

(a) The perceptron undergoes one iteration of training, with learning rate $\eta = 1$, with the training token $x = [0.1, 0.6, 0.5]$, $y = -1$. After this one iteration of training, what are w and b?

Solution:

$$w = [0.9, 0.4, 0.5], \quad b = 0$$

(b) The perceptron undergoes one more iteration of training, with learning rate $\eta = 1$, with the training token $x = [0.1, 0.1, 0.4]$, $y = 1$. After this second iteration of training, what are w and b?

Solution:

$$w = [0.9, 0.4, 0.5], \quad b = 0$$

Question 3

A particular perceptron is initialized with the weights $w = [1, 1, 1]$ and the bias $b = 1$.

(a) The perceptron undergoes one iteration of training, with learning rate $\eta = 1$, with the training token $x = [0.1, 0.1, 0.8]$, $y = -1$. After this one iteration of training, what are w and b?

Solution:

$$w = [0.9, 0.9, 0.2], \quad b = 0$$

(b) The perceptron undergoes one more iteration of training, with learning rate $\eta = 1$, with the training token $x = [0.1, 0.3, 0.9]$, $y = 1$. After this second iteration of training, what are w and b?

Solution:

$$w = [0.9, 0.9, 0.2], \quad b = 0$$

Question 4

A particular perceptron is initialized with the weights $w = [1, 1, 1]$ and the bias $b = 1$.

(a) The perceptron undergoes one iteration of training, with learning rate $\eta = 1$, with the training token $x = [0.5, 0.9, 0.6]$, $y = 1$. After this one iteration of training, what are w and b?

Solution:

$$w = [1, 1, 1], \quad b = 1$$

(b) The perceptron undergoes one more iteration of training, with learning rate $\eta = 1$, with the training token $x = [0.4, 0.9, 0.3]$, $y = -1$. After this second iteration of training, what are w and b?

Solution:

$$w = [0.6, 0.1, 0.7], \quad b = 0$$

Question 5

Gradient descent is guaranteed to find a set of model parameters that has the smallest possible loss function on the training corpus.

○ True

√ **False**

Explain:

Solution: Gradient descent finds a local optimum of the loss function, but it is not guaranteed to find a global optimum.

Question 6

Cross-entropy is

$$\mathcal{L} = -\frac{1}{n}\sum_{i=1}^{n} \ln P(Y = y_i|x_i)$$

Suppose you have $n = 2$ training samples, $x_1 = [0.2, 0.6]^T$, $y_1 = 1$, $x_2 = [-1.2, 0.3]^T$, $y_2 = 0$. Suppose $P(Y = 1|x)$ is defined as $P(Y = 1|x) = \sigma(w^T x)$, where $\sigma(\cdot)$ is the logistic sigmoid function. This computation has already been performed, therefore you already know that $P(Y = 1|x_1) = 0.2$, $P(Y = 1|x_2) = 0.9$. Find the gradient of $\mathcal{L}$ with respect to the vector w.

Solution:

$$\nabla_w \mathcal{L} = -\frac{1}{2}\left[(0.8)(0.2) - (0.9)(-1.2), (0.8)(0.6) - (0.9)(0.3)\right]^T$$

Question 7

A particular neural net has scalar input x_i, and scalar training target y_i, where i is the training token number. There is just one hidden node, with activation $h_i = \text{ReLU}(x_i + b)$. The output is $\hat{y}_i = \text{ReLU}(wh_i + c)$, and the loss is

$$\mathcal{L} = \frac{1}{2n}\sum_i (\hat{y}_i - y_i)^2$$

Suppose the coefficients are initialized to $w = 1$, $b = c = 0$, and then the loss is computed with respect to the following n=4 training tokens: $x_1 = -2, y_1 = 4, x_2 = -1, y_2 = 1, x_3 = 1, y_3 = 1, x_4 = 2, y_4 = 4$. What is $d\mathcal{L}/db$?

Solution:

$$\frac{d\mathcal{L}}{db} = \sum_i \left(\frac{d\mathcal{L}}{d\hat{y}_i}\right)\left(\frac{d\hat{y}_i}{dh_i}\right)\left(\frac{dh_i}{db}\right)$$

$dh_i/db = 0$ for $i = 1$ and $i = 2$. $d\mathcal{L}/d\hat{y}_i = (1/4)(\hat{y}_i - y_i)$, which is zero for $i = 3$. Therefore, the only nonzero term is for $i = 4$, where we have

$$\frac{d\mathcal{L}}{db} = \left(\left(\frac{1}{8}\right)(2)(2-4)\right)\left(\left.\frac{d\text{ReLU}(wh_i)}{dh_i}\right|_{w=1,h_1=1}\right)\left(\left.\frac{d\text{ReLU}(x_i)}{dx_i}\right|_{x_1=2}\right) = -\frac{1}{2}$$

Question 8

Suppose

$$h_i = \frac{\exp(e_i)}{\sqrt{\sum_j(\exp(2e_j))}}$$

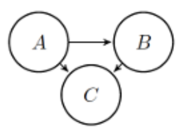
Find $d\ln(h_1)/de_2$ in terms of h_1 and/or h_2.

Solution: Let's define $NUM_1 = \exp(e_1)$, $NUM_2 = \exp(e_2)$, and $DEN = \sqrt{\sum_j (\exp(2e_j))}$, so that $h_1 = NUM/DEN$. Then

$$\begin{aligned}
\frac{d\ln(h_1)}{de_2} &= \left(\frac{1}{h_1}\right)\left(\frac{dh_1}{de_2}\right) && (1)\\
&= \left(\frac{1}{h_1}\right)\left(\frac{-NUM_1}{DEN^2}\right)\frac{dDEN}{de_2} && (2)\\
&= \left(\frac{DEN}{NUM_1}\right)\left(\frac{-NUM_1}{DEN^2}\right)0.5\left(\sum_j(\exp(e_j))\right)^{-0.5} 2\exp(2e_2) && (3)\\
&= -\left(\frac{1}{DEN}\right)0.5\left(\frac{1}{DEN}\right)2NUM_2^2 && (4)\\
&= -(h_2)^2 && (5)
\end{aligned}$$

Question 9

A, B, and C are binary random variables, whose dependencies are shown in the Bayes net below:



Variable A has a probability $P(A{=}1) = 0.3$. Conditional probabilities of the other two variables are given in the table below:

a	$P(B{=}1\|A{=}a)$	$P(C{=}1\|A{=}a, B{=}0)$	$P(C{=}1\|A{=}a, B{=}1)$
0	0.2	0.4	0.9
1	0.8	0.3	0.7

(a) What is $P(A = 1, B = 0, C = 1)$? Leave your answer in the form of a product of real numbers; do not simplify.

Solution: (0.3)(0.2)(0.3)

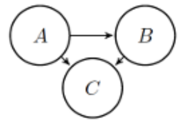
(b) What is $P(A{=}1|C{=}1)$? Leave your answer in the form of a ratio of sums of products; do not simplify.

Solution:

$$P(A{=}1|C{=}1) = \frac{(0.3)(0.2)(0.3) + (0.3)(0.8)(0.7)}{(0.3)(0.2)(0.3) + (0.3)(0.8)(0.7) + (0.7)(0.8)(0.4) + (0.7)(0.2)(0.9)}$$

Question 10

A, B, and C are binary random variables, whose dependencies are shown in the Bayes net below:



Variable A has a probability $P(A=1) = 0.8$. Conditional probabilities of the other two variables are given in the table below:

a	$P(B=1\|A=a)$	$P(C=1\|A=a, B=0)$	$P(C=1\|A=a, B=1)$
0	0.8	0.3	0.1
1	0.4	0.9	0.9

(a) What is $P(A = 0, B = 1, C = 0)$? Leave your answer in the form of a product of real numbers; do not simplify.

Solution: (0.2)(0.8)(0.9)

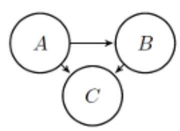
(b) What is $P(A=0|C=0)$? Leave your answer in the form of a ratio of sums of products; do not simplify.

Solution:

$$P(A=0|C=0) = \frac{(0.2)(0.2)(0.7) + (0.2)(0.8)(0.9)}{(0.2)(0.2)(0.7) + (0.2)(0.8)(0.9) + (0.8)(0.6)(0.1) + (0.8)(0.4)(0.1)}$$

Question 11

A, B, and C are binary random variables, whose dependencies are shown in the Bayes net below:



Variable A has a probability $P(A=1) = 0.1$. Conditional probabilities of the other two variables are given in the table below:

a	$P(B=1\|A=a)$	$P(C=1\|A=a, B=0)$	$P(C=1\|A=a, B=1)$
0	0.8	0.1	0.2
1	0.1	0.6	0.4

(a) What is $P(A = 1, B = 1, C = 1)$? Leave your answer in the form of a product of real numbers; do not simplify.

Solution: (0.1)(0.1)(0.4)

(b) What is $P(A=1|C=1)$? Leave your answer in the form of a ratio of sums of products; do not simplify.

Solution:

$$P(A=0|C=0) = \frac{(0.1)(0.9)(0.6) + (0.1)(0.1)(0.4)}{(0.1)(0.9)(0.6) + (0.1)(0.1)(0.4) + (0.9)(0.2)(0.1) + (0.9)(0.8)(0.2)}$$

Question 12

We have a bag of three biased coins, a, b, and c, with probabilities of coming up heads of 20%, 60%, and 80%, respectively. One coin is drawn randomly from the bag (with equal likelihood of drawing each of the three coins), and then the coin is flipped three times to generate the outcomes X1, X2, and X3.

(a) Draw the Bayesian network corresponding to this setup and define the necessary conditional probability tables (CPTs).

Solution: You need an intermediate variable, $C \in \{a, b, c\}$, to specify which coin is drawn, then the graph is

$$\begin{tikzcd} & C \arrow[dl] \arrow[d] \arrow[dr] & \\ X_1 & X_2 & X_3 \end{tikzcd}$$

and the CPTs are

C	$P(C)$	$P(X_1 = H\|C)$	$P(X_2 = H\|C)$	$P(X_3 = H\|C)$
a	1/3	0.2	0.2	0.2
b	1/3	0.6	0.6	0.6
c	1/3	0.8	0.8	0.8

(b) Calculate which coin was most likely to have been drawn from the bag if the observed flips come out heads twice and tails once.

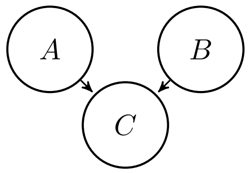
Solution:

$$P(C = a, HHT) = (0.2)(0.2)(0.8)/3 = 32/3000$$
$$P(C = b, HHT) = (0.6)(0.6)(0.4)/3 = 144/3000$$
$$P(C = c, HHT) = (0.8)(0.8)(0.2)/3 = 128/3000$$

The maximum-posterior-probability event is also the maximum-joint-probability event, which is the event $C = b$.

Question 13

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



$P(A) = 0.4, P(B) = 0.1$

A, B	$P(C \mid A, B)$
False,False	0.7
False,True	0.7
True,False	0.1
True,True	0.9

(a) What is $P(C)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(C) &= P(\neg A, \neg B, C) + P(\neg A, B, C) + P(A, \neg B, C) + P(A, B, C) \\ &= (0.6)(0.9)(0.7) + (0.6)(0.1)(0.7) + (0.4)(0.9)(0.1) + (0.4)(0.1)(0.9) \end{aligned}$$

(b) What is $P(A \mid B = True, C = True)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(A \mid B, C) &= \frac{P(A, B, C)}{P(A, B, C) + P(\neg A, B, C)} \\ &= \frac{(0.4)(0.1)(0.9)}{(0.4)(0.1)(0.9) + (0.6)(0.1)(0.7)} \end{aligned}$$

(c) You've been asked to re-estimate the parameters of the network based on the following observations:

Observation	A	B	C
1	True	False	False
2	False	False	True
3	True	True	False
4	False	False	False

Given the data in the table, what are the maximum likelihood estimates of the model parameters? If there is a model parameter that cannot be estimated from these data, mark it "UNKNOWN."

Solution: $P(A) = 2/4$, $P(B) = 1/4$, and

A, B	$P(C \mid A, B)$
F, F	1/2
F, T	UNKNOWN
T, F	0/1
T, T	0/1

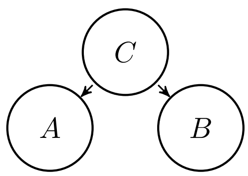
(d) Use the table of data given in part (c), but this time, estimate the data using Laplace smoothing, with a smoothing parameter of $k = 1$.

Solution: $P(A) = 3/6$, $P(B) = 1/3$, and

A, B	$P(C\|A, B)$
F, F	2/4
F, T	1/2
T, F	1/3
T, T	1/3

Question 14

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



$P(C) = 0.1$

C	$P(A\|C)$	$P(B\|C)$
False	0.8	0.7
True	0.4	0.7

(a) What is $P(A)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(A) &= P(\neg C, A) + P(C, A) \\ &= (0.9)(0.8) + (0.1)(0.4) \end{aligned}$$

(b) What is $P(C|A = True, B = True)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(C|A, B) &= \frac{P(A, B, C)}{P(A, B, C) + P(A, B, \neg C)} \\ &= \frac{(0.1)(0.4)(0.7)}{(0.1)(0.4)(0.7) + (0.9)(0.8)(0.7)} \end{aligned}$$

(c) You've been asked to re-estimate the parameters of the network based on the following observations:

Observation	A	B	C
1	False	True	False
2	True	True	False
3	False	False	True
4	False	False	True

Given the data in the table, what are the maximum likelihood estimates of the model parameters? If there is a model parameter that cannot be estimated from these data, mark it "UNKNOWN."

Solution: $P(C) = 2/4$, and

C	$P(A\|C)$	$P(B\|C)$
F	1/2	2/2
T	0/2	0/2

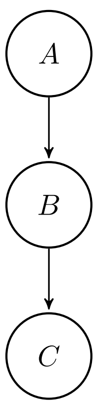
(d) Use the table of data given in part (c), but this time, estimate the data using Laplace smoothing, with a smoothing parameter of $k = 1$.

Solution: $P(C) = 3/6$, and

C	$P(A\|C)$	$P(B\|C)$
F	2/4	3/4
T	1/4	1/4

Question 15

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



$P(A) = 0.8$

A	$P(B\|A)$
False	0.7
True	0.3

B	$P(C\|B)$
False	0.5
True	0.7

(a) What is $P(C)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(C) &= P(\neg A, \neg B, C) + P(\neg A, B, C) + P(A, \neg B, C) + P(A, B, C) \\ &= (0.2)(0.3)(0.5) + (0.2)(0.7)(0.7) + (0.8)(0.7)(0.5) + (0.8)(0.3)(0.7) \end{aligned}$$

(b) What is $P(A|B = True, C = True)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$P(A|B,C) = \frac{P(A,B,C)}{P(A,B,C) + P(\neg A,B,C)}$$
$$= \frac{(0.8)(0.3)(0.7)}{(0.8)(0.3)(0.7) + (0.2)(0.7)(0.7)}$$

(c) You've been asked to re-estimate the parameters of the network based on the following observations:

Observation	A	B	C
1	True	False	False
2	False	False	True
3	True	True	False
4	False	False	False

Given the data in the table, what are the maximum likelihood estimates of the model parameters? If there is a model parameter that cannot be estimated from these data, mark it "UNKNOWN."

Solution:

$$P(A) = 2/4$$

A	$P(B\|A)$
False	0/2
True	1/2

B	$P(C\|B)$
False	1/3
True	0/1

(d) Use the table of data given in part (c), but this time, estimate the data using Laplace smoothing, with a smoothing parameter of $k = 1$.

Solution: $P(A) = 3/6$, $P(B) = 1/3$, and

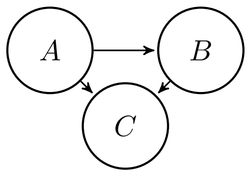
$$P(A) = 3/6$$

A	$P(B\|A)$
False	1/4
True	2/4

B	$P(C\|B)$
False	2/5
True	1/3

Question 16

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



$P(A) = 0.4$

A	$P(B\|A)$
False	0.1
True	0.2

A, B	$P(C\|A,B)$
False,False	0.9
False,True	0.3
True,False	0.7
True,True	0.5

(a) What is $P(C)$? Write your answer in numerical form, but you don't need to simplify.

> **Solution:**
>
> $$\begin{aligned} P(C) &= P(\neg A, \neg B, C) + P(\neg A, B, C) + P(A, \neg B, C) + P(A, B, C) \\ &= (0.6)(0.9)(0.9) + (0.6)(0.1)(0.3) + (0.4)(0.8)(0.7) + (0.4)(0.2)(0.5) \end{aligned}$$

(b) What is $P(A|B = True, C = True)$? Write your answer in numerical form, but you don't need to simplify.

> **Solution:**
>
> $$\begin{aligned} P(A|B,C) &= \frac{P(A,B,C)}{P(A,B,C) + P(\neg A, B, C)} \\ &= \frac{(0.4)(0.2)(0.5)}{(0.4)(0.2)(0.5) + (0.6)(0.1)(0.3)} \end{aligned}$$

(c) You've been asked to re-estimate the parameters of the network based on the following observations:

Observation	A	B	C
1	True	True	False
2	False	True	True
3	False	True	False
4	False	False	True

Given the data in the table, what are the maximum likelihood estimates of the model parameters? If there is a model parameter that cannot be estimated from these data, mark it "UNKNOWN."

> **Solution:**
>
> $$P(A) = 1/4$$

A	$P(B\|A)$
False	2/3
True	1

A, B	$P(C\|A, B)$
False,False	1/1
False,True	1/2
True,False	UNKNOWN
True,True	0/1

(d) Use the table of data given in part (c), but this time, estimate the data using Laplace smoothing, with a smoothing parameter of $k = 1$.

Solution:

$$P(A) = 2/6$$

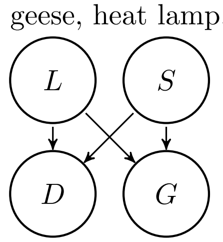
A	$P(B\|A)$
False	3/5
True	2/3

A, B	$P(C\|A, B)$
False,False	2/3
False,True	2/4
True,False	1/2
True,True	1/3

Question 17

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 the following table:

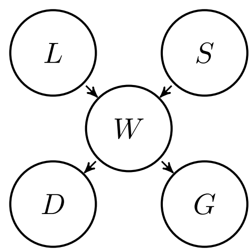
day	D	G	L	S	day	D	G	L	S
1	0	1	1	0	5	1	0	0	1
2	1	0	1	0	6	1	0	1	1
3	0	0	0	0	7	0	1	1	1
4	0	0	0	0	8	0	1	0	1

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:



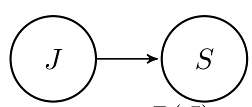
She forgot to measure the temperature in her back yard, so W is a hidden variable. Her initial guess is that $P(D|W) = \frac{2}{3}$, $P(D|\neg W) = \frac{1}{3}$, $P(G|W) = \frac{2}{3}$, $P(G|\neg W) = \frac{1}{3}$, $P(W|L \wedge S) = \frac{2}{3}$, $P(W|\neg(L \wedge S)) = \frac{1}{3}$. Find the posterior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

Solution:

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

Question 18

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

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

Question 19

There is a lion in a cage in the dungeons under Castle Rock.

- The zookeeper goes on vacation with a probability of P.
- If the zookeeper is on vacation, the lion doesn't get fed. If not, the lion gets fed with probability Q, and goes hungry with probability $1-Q$.
- If the lion has not been fed, and you try to pet it, then it will bite your hand with probability R. If it has been fed, it will only bite you with probability S.

(a) (2 points) Draw a Bayes network with three random variables: $Z=1$ if the zookeeper is on vacation, $F=1$ if the lion gets fed today, $B=1$ if it will bite the hand of the next person who tries to pet it. Draw edges to show the dependencies specified by the problem statement above.

Solution:

$$Z \longrightarrow F \longrightarrow B$$

(b) (3 points) Circe pets the lion, and it bites her hand. In terms of the unknown parameters P, Q, R, and S, what is the probability that the zookeper is on vacation?

Solution:

$$P(Z|B) = \frac{P(Z,F,B) + P(Z,\neg F,B)}{P(Z,F,B) + P(Z,\neg F,B) + P(\neg Z,F,B) + P(\neg Z,\neg F,B)}$$

$$= \frac{PR}{PR + (1-P)QS + (1-P)(1-Q)R}$$

(c) (2 points) Lord Lucky, the Lord of Castle Rock, hires a troupe of circus performers to pet the lion, once per day, in an attempt to learn the parameters P, Q, R, and S. Over the course of seven days, he collects the following observations. Based on these observations, estimate P, Q, R, and S.

Day	Z	F	B
1	1	0	1
2	0	1	1
3	0	1	0
4	1	0	0
5	0	0	1
6	0	1	0
7	0	1	0

Solution: The probability estimates are

$$P = P(Z) = \frac{2}{7}$$

$$Q = P(F|\neg Z) = \frac{4}{5}$$

$$R = P(B|\neg F) = \frac{2}{3}$$

$$S = P(B|F) = \frac{1}{4}$$

Question 20

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 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

Question 21

A particular hidden Markov model (HMM) has state variable Y_t, and observation variables X_t, where t denotes time. Suppose that this HMM has two states, $Y_t \in \{0, 1\}$, and three possible observations, $X_t \in \{0, 1, 2\}$. The initial state probability is $P(Y_1 = 1) = 0.3$. The transition and observation probability matrices are

Y_{t-1}	$P(Y_t = 1 \vert Y_{t-1})$	Y_t	$P(X_t = 0 \vert Y_t)$	$P(X_t = 1 \vert Y_t)$
0	0.6	0	0.4	0.1
1	0.4	1	0.1	0.6

Suppose that, in a particular test of the HMM, the observation sequence is

$$\{X_1, X_2\} = \{2, 1\}$$

(a) What is the joint probability $P(Y_1 = 1, X_1 = 2, Y_2 = 0)$?

Solution:

$$\begin{aligned} P(Y_1 = 1, X_1 = 2, Y_2 = 0) &= P(Y_1 = 1)P(X_1 = 2|Y_1 = 1)P(Y_2 = 0|Y_1 = 1) \\ &= (0.3)(0.3)(0.6) \end{aligned}$$

(b) What is the probability of the most likely state sequence ending in $Y_2 = 0$? In other words, what is $\max_{Y_1} P(Y_1, X_1 = 2, Y_2 = 0, X_2 = 1)$?

Solution:

$$\begin{aligned} \max_{Y_1} P(Y_1, X_1 = 2, Y_2 = 0, X_2 = 1) &= \max_{Y_1} P(Y_1)P(X_1 = 2|Y_1)P(Y_2 = 0|Y_1)P(X_2 = 1|Y_2 = 0) \\ &= \max((0.7)(0.5)(0.4)(0.1), (0.3)(0.3)(0.6)(0.1)) \\ &= (0.7)(0.5)(0.4)(0.1) \end{aligned}$$

Question 22

A particular HMM has a binary state variable $Y_t \in \{0, 1\}$, and a binary observation variable $X_t \in \{0, 1\}$. Suppose the HMM starts at time $t = 1$ with initial probability $P(Y_1 = 1) = 0.5$. The transition probabilities and observation probabilities are given in the following table:

	$P(Y_{t+1} = 1\|Y_t)$	$P(X_t = 1\|Y_t)$
$Y_t = 0$	0.8	0.5
$Y_t = 1$	0.6	0.3

What is $P(Y_1 = 1, X_1 = 0, X_2 = 1)$? Express your answer as a sum of products; do not simplify.

Solution:

$$\begin{aligned} P(Y_1 = 1, X_1 = 0, X_2 = 1) &= P(Y_1 = 1)P(X_1 = 0|Y_1 = 1)P(Y_2 = 0|Y_1 = 1)P(X_2 = 1|Y_2 = 0) \\ &+ P(Y_1 = 1)P(X_1 = 0|Y_1 = 1)P(Y_2 = 1|Y_1 = 1)P(X_2 = 1|Y_2 = 1) \\ &= (0.5)(0.7)(0.4)(0.5) + (0.5)(0.7)(0.6)(0.3) \end{aligned}$$

Question 23

A particular HMM has a binary state variable $X_t \in \{0, 1\}$, and a binary observation variable $E_t \in \{0, 1\}$. Suppose the HMM starts at time $t = 1$ with initial probability $P(X_1 = 1) = 0.5$. The transition probabilities and observation probabilities are given in the following table:

		$P(X_{t+1} = 1\|X_t)$	$P(E_t = 1\|X_t)$
X_t	0	0.2	0.4
	1	0.8	0.3

What is $P(X_1 = 0, E_1 = 0, E_2 = 0)$? Express your answer as a sum of products; do not simplify.

Solution:

$$(0.5)(0.6)(0.8)(0.6) + (0.5)(0.6)(0.2)(0.7)$$

Question 24

A particular HMM has a binary state variable $X_t \in \{0, 1\}$, and a binary observation variable $E_t \in \{0, 1\}$. Suppose the HMM starts at time $t = 1$ with initial probability $P(X_1 = 1) = 0.2$. The transition probabilities and observation probabilities are given in the following table:

		$P(X_{t+1} = 1 \mid X_t)$	$P(E_t = 1 \mid X_t)$
X_t	0	0.3	0.8
	1	0.3	0.9

What is $P(X_1 = 0, E_1 = 0, E_2 = 0)$? Express your answer as a sum of products; do not simplify.

Solution:

$$(0.8)(0.2)(0.7)(0.2) + (0.8)(0.2)(0.3)(0.1)$$