

# CS 177, Homework 4

Applications of Probability in Computer Science: Winter 2007

Due Date: Tuesday February 6th, in class

## Suggested Reading

- Chapter 7, pages 408 to 417.
- Class notes on Markov chains.

In all of your solutions **clearly** show step-by-step how you arrived at your answer.

## Problem 1

Your newly started company has a telephone hotline that PC owners can call for technical support. Mr. John Doe calls your hotline to report a problem with his PC running Windows. Let  $E$  be the event that his machine has a virus. Let  $F$  be the probability that the registry file on his PC is corrupted. The following table represents the joint probabilities of various combinations of these two events, i.e., each entry represents the probability of the "AND" of the corresponding row and column.

	$E$	$\text{NOT}(E)$
$F$	0.01	0.19
$\text{NOT}(F)$	0.02	0.78

1. What is  $P(E)$ ?
2. What is  $P(F)$ ?
3. What is  $P(E|F)$ ?
4. What is  $P(F|E)$ ?
5. If John Doe first reports that his hard drive has a virus, the probability that his registry file is corrupted has changed: has it increased or decreased? and by what multiplicative factor?
6. If John Doe first reports that his registry file is corrupted, the probability that his hard drive has a virus has changed: increased or decreased? and by what multiplicative factor?

## Problem 2

Consider building a model for the following fault diagnosis problem. The class variable  $C$  represents the health of a disk drive:  $C = 1$  means it is operating normally, and  $C = 0$  means it is in a failed state. There are 2 binary features that we can measure,  $X$  and  $Y$ , where each takes values 0 or 1.

The following is the joint pmf of all three variables:

$x$	$y$	$c$	$p(x, y, c)$
0	0	0	0.1
0	1	0	0.15
1	0	0	0.15
1	1	0	0.1
0	0	1	0.35
0	1	1	0.1
1	0	1	0.05
1	1	1	0.0

In the following problems for the cases where a numerical value is required, provide an equation (or multiple equations) to show how you calculated the answer and the numerical value of the answer.

1. What is the probability  $p(C = 1)$ ?
2. What is the probability  $p(C = 0|X = 1, Y = 0)$ ?
3. What is the probability  $p(X = 0, Y = 0)$ ?
4. What is the probability  $p(C = 0|X = 0)$ ?
5. Are  $X$  and  $Y$  independent? justify your answer
6. Are  $X$  and  $Y$  conditionally independent given  $C$ ? justify your answer

In the next few problems the phrase “Markov chain” should be interpreted as a discrete-time, finite-state, homogeneous Markov chain, as discussed in class and in the textbook. In the transition matrices of size  $m \times m$ , states are numbered from 1 to  $m$  in correspondence with the rows and columns, and row  $i$  is the set of transition probabilities out of state  $i$ ,  $1 \leq i \leq m$ .

### Problem 3

Consider a very simple model for the weather each day in Irvine, where if it rains the weather is recorded as “rainy” and otherwise the weather is recorded as “sunny” for that day. If it is rainy on a given day, then the probability is 0.5 that it is rainy again the next day. If it is sunny on a given day then the probability is 0.9 that it is sunny the next day.

1. Define the transition matrix for this Markov chain.
2. If it rains on Friday, what is the probability of it being sunny on Saturday.
3. If it rains on Friday, what is the probability of Saturday, Sunday, and Monday all being rainy?
4. If it rains on Friday, what is the probability of rain on Sunday?
5. If it rains on Friday, what is the probability of rain on Monday?

### Problem 4

Consider a very simple Markov model for letters, digits, and other symbols (punctuation, blank space, etc) in text. This is the transition matrix:

$$\begin{array}{ccc} 0.7 & 0.1 & 0.2 \\ 0.5 & 0.3 & 0.2 \\ 0.5 & 0.1 & 0.4 \end{array}$$

Assume that state 1 is a letter, state 2 is a digit, and state 3 is “other.” A simple model like this could be used by systems that try to automatically scan and read documents (e.g., faxes, addresses on envelopes, etc). Assume in the questions below that the current symbol is a letter.

1. What is the probability that the next 3 symbols are digits?
2. What is the probability that the next 2 symbols are digits and the 2 symbols after that are letters?
3. What is the probability that the next 3 symbols are the same?
4. Mention in 1 sentence one important disadvantage of using a Markov model like this in a real problem involving text.

### Problem 5

Consider the following Markov transition matrix:

$$\begin{array}{ccc} 0.7 & 0.3 & 0.0 \\ 0.5 & 0.4 & 0.1 \\ 0.2 & 0.0 & 0.8 \end{array}$$

Answer the following questions:

1. Draw the state-diagram for this Markov chain.
2. Is this Markov chain irreducible?
3. If the system is in state 1 at time  $t$  what is the probability of the following events:
  - (a) the system is in state 1 at time  $t + 1$
  - (b) the system is in state 3 at time  $t + 2$
  - (c) the system is in state 2 at time  $t + 2$
  - (d) the system is in state 1 at time  $t + 4$

## Problem 6

Consider an artificial agent in a game whose behavior is determined by a Markov chain with the following Markov transition matrix:

$$\begin{array}{ccccc} 0.8 & 0.1 & 0.0 & 0.0 & 0.1 \\ 0.1 & 0.8 & 0.1 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.8 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.1 & 0.8 & 0.1 \\ 0.1 & 0.0 & 0.0 & 0.1 & 0.8 \end{array}$$

Answer the following questions:

1. Draw the state-diagram for this Markov chain.
2. Is this Markov chain irreducible?
3. If the agent is in state 1 at time  $t$  what is the probability of the following events:
  - (a) the agent is in state 2 at time  $t + 1$
  - (b) the agent is in state 2 at time  $t + 2$
  - (c) the agent is in state 2 at time  $t + 3$
  - (d) the agent is in state 4 at time  $t + 4$
  - (e) the agent is in state 1 at time  $t + 4$

## Problem 7

Consider a Markov chain with 4 states, named 1, 2, 3, 4, with the following transition probability matrix:

$$P = \begin{pmatrix} 0.3 & 0.1 & 0.2 & 0.4 \\ 0.0 & 0.6 & 0.2 & 0.2 \\ 0.0 & 0.0 & 0.5 & 0.5 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

The initial state probability vector is  $(1, 0, 0, 0)$ , i.e., the system always starts in state 1.

This could be a very simple model to represent how a user navigates a simple ecommerce Web site, where each state represents a set of similar pages on the site. For example state 1 could be the set of introductory pages, state 2 represents specific product information pages, state 3 could be the product purchase and checkout pages, and state 4 could represent the state where the person leaves the Web site. State 4 is known as an “absorbing” state: once the system enters this state it never leaves. Here we use state 4 to represent the end of an individual session on the Web site.

(For simplicity in this problem to keep the math relatively simple there are no transitions “backwards”, e.g., from state 3 to state 2 — in a real problem we would certainly allow such transitions since real users typically go back and forth between different parts of a Web site).

Answer the following questions about this Markov chain. In each case clearly explain you arrived at the solution.

1. What is the probability that a person will leave the site exactly at time step 2? (i.e., what is the probability that the 2nd state in a sequence is state 4). Note that the state at time step 1 is state 1.
2. What is the probability that a person will leave the site exactly at time step 3? (i.e., what is the probability that the 3rd state in a sequence is state 4). Hint: there is more than one way for this to happen.
3. What is the probability that a person exits the site without visiting the product purchase page?, i.e., the probability that they get to state 4 without visiting state 3. (Be careful in answering this one: make sure you consider the various possible ways that this can happen).

## Problem 8

Consider a 2-state Markov chain with states 1 and 2 and transition probabilities  $P_{ij}$  indicating transitions from  $i$  to  $j$ ,  $i, j \in \{1, 2\}$ .

1. Show that the log of the probability of a particular sequence  $\{s_1, s_2, s_3, \dots, s_T\}$ , under a 2-state Markov model, can be written as a sum of a finite number of different terms where the number of terms in this expression does not depend on  $T$ . (Hint: let  $n_{ij}$  be the number of transitions in the sequence from state  $i$  to state  $j$  and consider writing your solution using such terms).
2. Let the self-transition probabilities  $P_{11}$  and  $P_{22}$  both be 0.9. The initial state probability vector is  $(0.5, 0.5)$ , i.e., the system is equally likely to start in either state 1 or state 2. Compute the log of the probability (base 10) of each of the following sequences:

Sequence 1 is  $\{2, 2, 2, 2, 1, 1, 1, 1, 2, 2\}$

Sequence 2 is  $\{1, 2, 2, 2, 1, 2, 1, 2, 1, 2, \}$

Which sequence is more likely according to the model?

3. Repeat part 2 above, for the same 2 sequences, but now where the model is different: specifically, the self-transition probabilities  $P_{11}$  and  $P_{22}$  are now both 0.5, and the initial state probability vector is still  $(0.5, 0.5)$ .