techniques to HMMs. Consequently, a clear understanding of the material in this chapter is crucial before proceeding with the remainder of the book. The homework problem will help the dedicated reader clear any remaining questions regarding HMMs.

## 2.10 Problems

*When faced with a problem you do not understand,*
*do any part of it you do understand, then look at it again.*
— Robert Heinlein

1. Suppose that we train an HMM and obtain the model $\lambda = (A, B, \pi)$ where

$$A = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}, \quad B = \begin{pmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{pmatrix}, \quad \pi = \begin{pmatrix} 0.0 & 1.0 \end{pmatrix}.$$

Furthermore, suppose the hidden states correspond to $H$ and $C$, respectively, while the observations are $S$, $M$, and $L$, which are mapped to 0, 1, and 2, respectively. In this problem, we consider the observations $\mathcal{O} = (\mathcal{O}_0, \mathcal{O}_1, \mathcal{O}_2) = (M, S, L) = (1, 0, 2)$.

a) Directly compute $P(\mathcal{O} \mid \lambda)$. Since

$$P(\mathcal{O} \mid \lambda) = \sum_X P(\mathcal{O}, X \mid \lambda)$$

we use the probabilities in $\lambda = (A, B, \pi)$ to compute each of the following for given observation sequence:

$$P(\mathcal{O}, X = HHH) = \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ = \_\_\_\_\_$$
$$P(\mathcal{O}, X = HHC) = \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ = \_\_\_\_\_$$
$$P(\mathcal{O}, X = HCH) = \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ = \_\_\_\_\_$$
$$P(\mathcal{O}, X = HCC) = \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ = \_\_\_\_\_$$
$$P(\mathcal{O}, X = CHH) = \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ = \_\_\_\_\_$$
$$P(\mathcal{O}, X = CHC) = \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ = \_\_\_\_\_$$
$$P(\mathcal{O}, X = CCH) = \underline{\ 1.0\ } \cdot \underline{\ 0.2\ } \cdot \underline{\ 0.6\ } \cdot \underline{\ 0.7\ } \cdot \underline{\ 0.4\ } \cdot \underline{\ 0.5\ } = \_\_\_\_\_$$
$$P(\mathcal{O}, X = CCC) = \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ \cdot \_\_\_ = \_\_\_\_\_$$

The desired probability is the sum of these 8 probabilities.

b) Compute $P(\mathcal{O}|\lambda)$ using the $\alpha$ pass. That is, compute

$$\alpha_0(0) = \_\_\_ \cdot \_\_\_ = \_\_\_\_\_$$

$$\alpha_0(1) = \underline{\ 1.0\ } \cdot \underline{\ 0.2\ } = \_\_\_\_\_$$
$$\alpha_1(0) = (\_\_\_ \cdot \_\_\_ + \_\_\_ \cdot \_\_\_) \cdot \_\_\_ = \_\_\_\_\_$$
$$\alpha_1(1) = (\_\_\_ \cdot \_\_\_ + \_\_\_ \cdot \_\_\_) \cdot \_\_\_ = \_\_\_\_\_$$
$$\alpha_2(0) = (\_\_\_ \cdot \_\_\_ + \_\_\_ \cdot \_\_\_) \cdot \_\_\_ = \_\_\_\_\_$$
$$\alpha_2(1) = (\_\_\_ \cdot \_\_\_ + \_\_\_ \cdot \_\_\_) \cdot \_\_\_ = \_\_\_\_\_$$

where the recurrence for $\alpha_t(i)$ is

$$\alpha_0(i) = \pi_i b_i(\mathcal{O}_0), \text{ for } i = 0, 1, \ldots, N-1$$

and

$$\alpha_t(i) = \left( \sum_{j=0}^{N-1} \alpha_{t-1}(j) a_{ji} \right) b_i(\mathcal{O}_t)$$

for $t = 1, 2, \ldots, T-1$ and $i = 0, 1, \ldots, N-1$. The desired probability is given by

$$P(\mathcal{O} \mid \lambda) = \sum_{i=0}^{N-1} \alpha_{T-1}(i).$$

c) Explain the results you obtained for parts a) and b). Be sure to explain why you obtained the results you did.

d) In terms of $N$ and $T$, and counting only multiplications, what is the work factor for the method in part a)? The method in part b)?

2. For this problem, use the same model $\lambda$ and observation sequence $\mathcal{O}$ given in Problem 1.

   a) Determine the "best" hidden state sequence $(X_0, X_1, X_2)$ in the dynamic programming sense.

   b) Determine the "best" hidden state sequence $(X_0, X_1, X_2)$ in the HMM sense.

3. Summing the numbers in the "probability" column of Table 2.2, we find $P(\mathcal{O} = (0, 1, 0, 2)) = 0.009629$.

   a) By a similar direct calculation, compute $P(\mathcal{O} = (\mathcal{O}_0, \mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_3))$, where each $\mathcal{O}_i \in \{0, 1, 2\}$, and verify that $\sum P(\mathcal{O}) = 1$. You will use the probabilities for $A$, $B$ and $\pi$ given in equations (2.3), (2.4) and (2.5) in Section 2.2, respectively.

   b) Use the forward algorithm to compute each $P(\mathcal{O})$ and verify that you obtain the same results as in part a).

4. Write the re-estimation formulae (3), (7) and (12) directly in terms of $\alpha$ and $\beta$.

5. In the re-estimation formulae obtained in the previous problem, substitute $\widehat{\alpha}$ and $\widehat{\beta}$ for $\alpha$ and $\beta$, respectively, and show that the resulting re-estimation formulae are exact.

6. Instead of using $c_t$ to scale the $\beta_t(i)$, scale each $\beta_t(i)$ by

$$d_t = 1 \Big/ \sum_{j=0}^{N-1} \widetilde{\beta}_t(j)$$

where the definition of $\widetilde{\beta}$ is similar to that of $\widetilde{\alpha}$.

   a) Using the scaling factors $c_t$ and $d_t$ show that the re-estimation formulae obtained in Exercise 1 are exact with $\widehat{\alpha}$ and $\widehat{\beta}$ in place of $\alpha$ and $\beta$.

   b) Write $\log\big(P(\mathcal{O} \mid \lambda)\big)$ in terms of $c_t$ and $d_t$.

7. When training a model, the elements of $\lambda$ are typically initialized to approximately uniform. That is, we initialize $\pi_i \approx 1/N$ and $a_{ij} \approx 1/N$ and $b_j(k) \approx 1/M$, subject to the row stochastic conditions. In Section 2.5.3, it is stated that it is a bad idea to initialize the values to exactly uniform, since this would cause the model to be stuck at a local maximum, and hence the HMM could not climb to an improved solution. Suppose that $\pi_i = 1/N$ and $a_{ij} = 1/N$ and $b_j(k) = 1/M$. Show that the re-estimation process leaves all of these values unchanged.

8. Consider an HMM where for each $t$ the state transition matrix is time dependent. Then for each $t$, there is an $N \times N$ row-stochastic $A_t = \{a_{ij}^t\}$ that is used in place of $A$ in the HMM computations. For such an HMM,

   a) Give pseudo-code to solve Problem 1.

   b) Give pseudo-code to solve Problem 2.

9. Consider an HMM of order two, that is, an HMM where the underlying Markov process is of order two. Then the the state at time $t$ depends on the states at time $t-1$ and $t-2$ and a fixed set of probabilities. For an HMM of order two,

   a) Give pseudo-code to solve Problem 1.

   b) Give pseudo-code to solve Problem 2.

   c) Give pseudo-code to solve Problem 3.

10. Write an HMM program to solve the English language problem discussed in Section 9.2 for each of the following cases.

   a) There are $N = 2$ hidden states. Explain your results.

b) There are $N = 3$ hidden states. Explain your results.

c) There are $N = 4$ hidden states. Explain your results.

11. In this problem, you will use an HMM to break a simple substitution ciphertext message. For each HMM, train using 200 iterations of the Baum-Welch re-estimation algorithm.

    a) Obtain an English plaintext message consisting of 50,000 plaintext characters, where the characters consist only of lower case `a` through `z` (i.e., remove all punctuation and spaces, and convert all upper case to lower case). Encrypt this plaintext using a Caesar's cipher.

    b) Train an HMM with $N = 2$ and $M = 26$ on your ciphetext from part a). From the resulting $B$ matrix, determine the ciphertext letters that correspond to consonants and those that correspond to vowels.

    c) Generate a digraph frequency matrix $A$ for English text, where $a_{ij}$ is the count of the number of times that letter $i$ is followed by letter $j$. Note that `a` is letter 0, `b` is letter 1, `c` is letter 2, and so on. This matrix must be based on 1,000,000 characters where, as above, only the 26 letters of the alphabet are used. Next, add 5 to each element in your $26 \times 26$ matrix $A$. Finally, normalize your matrix $A$ by dividing each element by its row sum. The resulting matrix $A$ will be row stochastic, and it will not contain any 0 probabilities.

    d) Train an HMM with $N = M = 26$, using the first 1000 characters of ciphertext you generated in part a), where the $A$ matrix is initialized with the $A$ matrix from part c). Also, in your HMM, do not re-estimate $A$. Use the final $B$ matrix to determine a putative key and give the fraction of putative key elements that match the actual key (as a decimal, to four places). For example, if 22 of the 26 key positions are correct, then your answer would be $22/26 = 0.8462$.

12. Write an HMM program to solve the problem discussed in Section 9.2, replacing English text with

    a) French.

    b) Russian.

    c) Chinese.

13. Perform an HMM analysis similar to that discussed in Section 9.2, replacing English with "Hamptonese", the mysterious writing system developed by James Hampton. For information on Hamptonese, see

    http://www.cs.sjsu.edu/faculty/stamp/Hampton/hampton.html

14. Since HMM training is a hill climb technique, we are only assured of a local maximum. And, as with any hill climb, the specific local maximum we obtain depends on our choice of the initial values. Therefore, by training multiple times with different initial values, we might expect to obtain better results than when training only once. In [18], the authors use expectation maximization (EM) with multiple random restarts to attack homophonic substitution ciphers. An analogous HMM-based technique is considered in [169] and [170], where the effectiveness on simple substitution cryptanalysis is analyzed in detail. The advantage of such an approach is that attacks will succeed more often when limited data is available. However, the tradeoff is that the work factor can be high, since the number of restarts required may be large.

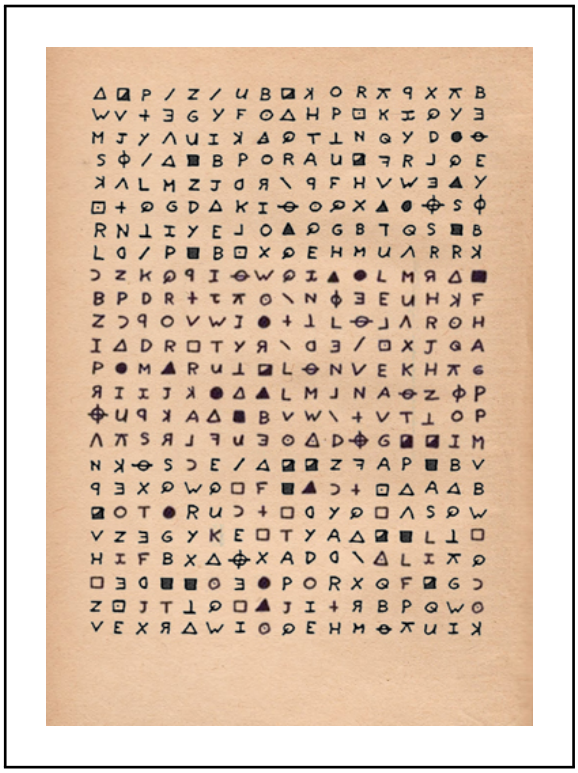
    a) Complete part a) of Problem 11, except that we only require a ciphertext of length 1000. Also, complete part c) of Problem 11.

    b) For each of $n = 1$, $n = 10$, and $n = 100$, train $n$ HMM models as in Problem 11, part d), except using $T = 1000$ observations. For a given $n$ select the best result based on the model scores and give the fraction of the putative key that is correct, calculated as in Problem 11, part d).

    c) Repeat part b), except only use $T = 400$ observations, and also include the $n = 1000$ case.

    d) Repeat part c), except only use $T = 300$ observations.

15. The Zodiac killer murdered at least five people in the San Francisco bay area in the late 1960s and early 1970s. Although police had a prime suspect, no arrest was ever made and the murders remain officially unsolved. The killer sent several messages to local newspapers, taunting police for their failure to catch him. One of these messages was a homophonic substitution consisting of 408 strange symbols which, appropriately, is known as the Zodiac 408 cipher. The ciphertext was sent in three parts to local newspapers. Combining the all three into one, the complete ciphertext is as follows.



Within days of its release, the Zodiac 408 was broken by Donald and Bettye Harden, who were school teachers in Salinas, California. The plaintext is given by

```
I L I K E K I L L I N G P E O P L
E B E C A U S E I T I S S O M U C
H F U N I T I S M O R E F U N T H
A N K I L L I N G W I L D G A M E
I N T H E F O R R E S T B E C A U
S E M A N I S T H E M O S T D A N
G E R O U E A N A M A L O F A L L
T O K I L L S O M E T H I N G G I
V E S M E T H E M O S T T H R I L
L I N G E X P E R E N C E I T I S
E V E N B E T T E R T H A N G E T
T I N G Y O U R R O C K S O F F W
I T H A G I R L T H E B E S T P A
R T O F I T I S T H A E W H E N I
D I E I W I L L B E R E B O R N I
N P A R A D I C E A N D A L L T H
E I H A V E K I L L E D W I L L B
E C O M E M Y S L A V E S I W I L
L N O T G I V E Y O U M Y N A M E
B E C A U S E Y O U W I L L T R Y
T O S L O I D O W N O R A T O P M
Y C O L L E C T I O G O F S L A V
E S F O R M Y A F T E R L I F E E
B E O R I E T E M E T H H P I T I
```

Note the (apparently intentional) misspellings in the plaintext, including "FORREST", "ANAMAL", and so on. Also, the final 18 characters (as highlighted above) appear to be random filler.

a) Solve the Zodiac 408 cipher using the HMM approach discussed in Section 9.4. Initialize the $A$ matrix based on the Brown Corpus, and do not re-estimate $A$. Use 1000 random restarts of the HMM, and give your answer as the percentage of characters of the actual plaintext that are recovered correctly.

b) Repeat part a), using 10,000 random restarts.

c) Repeat part b), using 100,000 random restarts.

d) Repeat part c), using 1,000,000 random restarts.

e) Repeat part a), except also re-estimate the $A$ matrix.

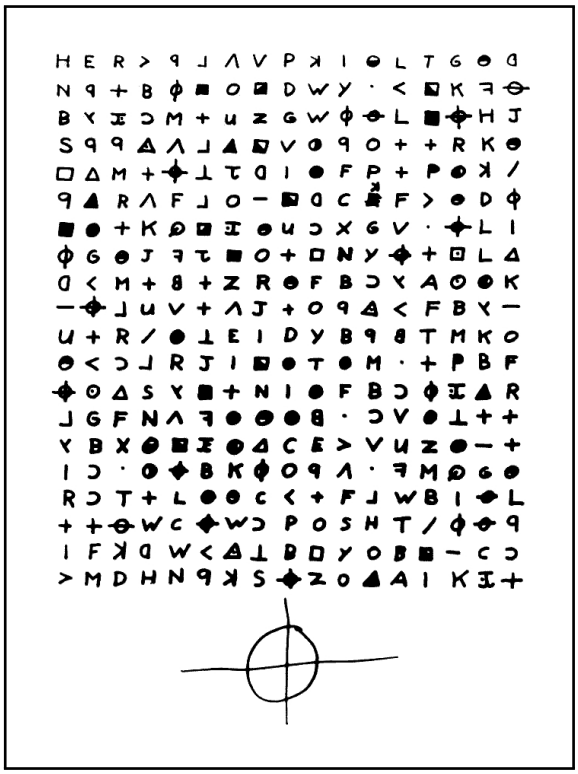
f) Repeat part b), except also re-estimate the $A$ matrix.

g) Repeat part c), except also re-estimate the $A$ matrix.

h) Repeat part d), except also re-estimate the $A$ matrix.

16. In addition to the Zodiac 408 cipher, the Zodiac killer (see Problem 15) released another similar-looking cipher with 340 symbols. This cipher

is known as the Zodiac 340 and remains unsolved to this day.[4] The ciphertext is



a) Repeat Problem 15, parts a) through d), but using the Zodiac 340 in place of the Zodiac 408. Since the plaintext is unknown, in each case, simply print the decryption obtained from your highest scoring model.

b) Repeat part a) of this problem, except use parts e) through h) of Problem 15.

[4]It is possible that the Zodiac 340 is not a cipher at all, but instead just a random collection of symbols designed to frustrate would-be cryptanalysts. If that's the case, the "cipher" has been wildly successful.