CS 271 Spring 2020
Assignment 2

Student ID: 014498887
Name: Yash Sahasrabuddhe

Q.14.

The question directs us to train multiple HMMs so as to find the local maximum with the local minimum initial values of the A, B and Pi matrices.

In this case though, we are going to use multiple random restarts using the expectation maximization (EM) approach.

To train an HMM for a substitution cryptanalysis, we need high number of random restarts. In this question, we will train 4 HMMs with 1 model, 10 models, 100 models and 1000 models respectively.

We will read the text from the Brown corpus data set. Then we will store this observation sequence of the plain text and also calculate a cipher text based on some random shift of the alphabets. According to the question, we will remember this key using a global variable. We will store this as well as the alpha, beta, gammas and digammas using structures.

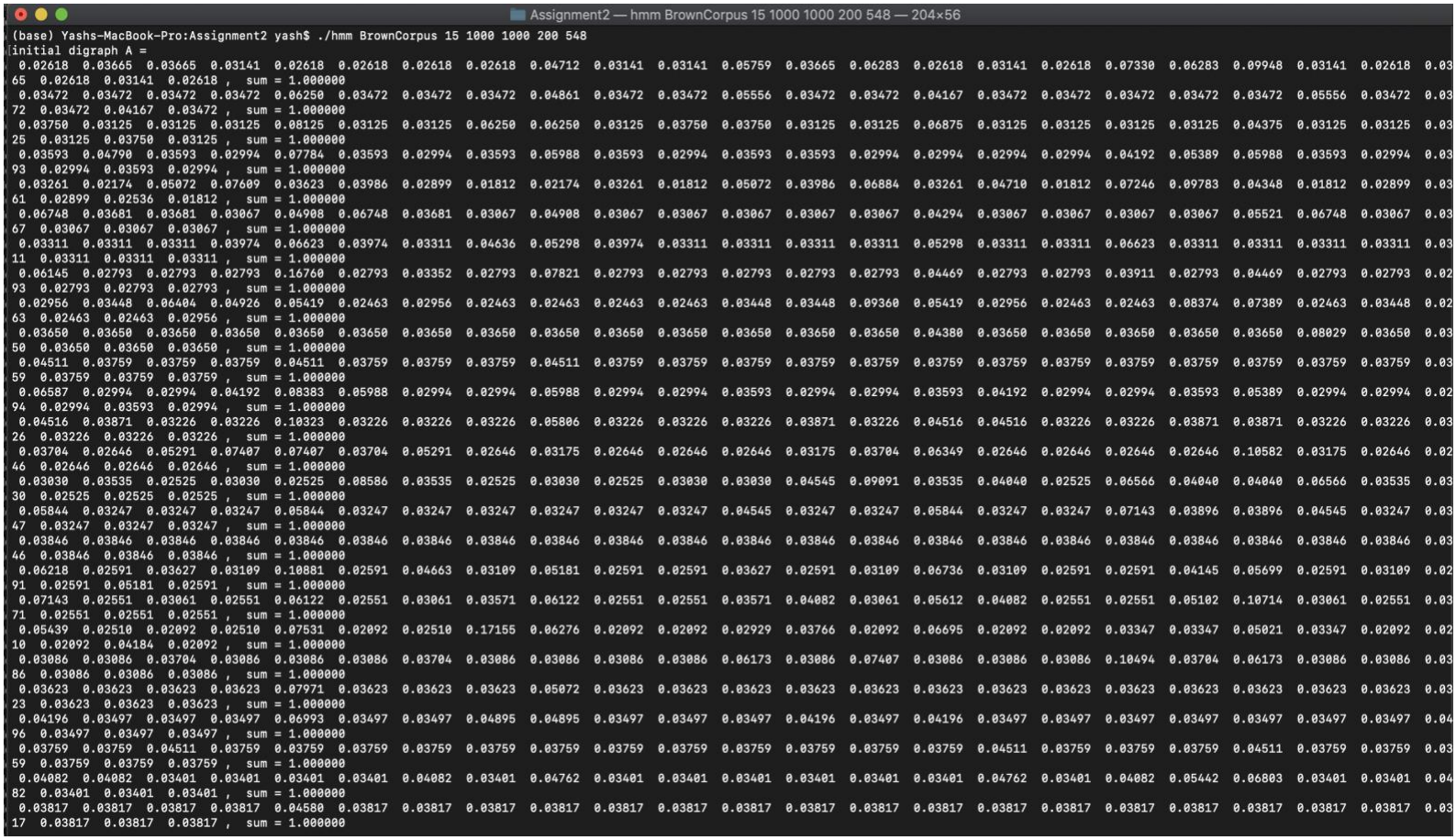
Firstly, we will read 1,000,000 characters from the data set to generate a digraph frequency matrix. We will not res-estimate this matrix. This forms the first sub-question. The digraph matrix's screenshot is added in the subsequent sub-points.

Then, we will train the various models (which are nothing but random restarts) and calculate a putative key. This key is generated using the final B matrix's probabilities. We match these probabilities with the alphabet's probabilities and try and match each alphabet from this cipher text to the original plain text. We will count the number of correct matches and compute a score for each model. First, we will read 1000 observations of plain text and see the results for each model as well as the best score for each of those 4 HMMs. This is the solution for Sub-question (b).

We will then repeat the process for 400 and 300 observations and compute the results.

The final outputs are shown here, and the output file has been added as an attachment in the submission.

a)



b)

```
(base) Yashs-MacBook-Pro:Assignment2 yash$ ./hmm BrownCorpus 15 1000 1000 200 548

HMM 1 with model 1 has score  0.07692
The best score for HHM 1 was with model 1 with score  0.07692


HMM 2 with model 1 has score  0.00000
HMM 2 with model 2 has score  0.03846
HMM 2 with model 3 has score  0.73077
HMM 2 with model 4 has score  0.03846
HMM 2 with model 5 has score  0.73077
HMM 2 with model 6 has score  0.73077
HMM 2 with model 7 has score  0.00000
HMM 2 with model 8 has score  0.73077
HMM 2 with model 9 has score  0.00000
HMM 2 with model 10 has score  0.00000
The best score for HHM 2 was with model 3 with score  0.73077
```

```
HMM 3 with model 98 has score  0.73077
HMM 3 with model 99 has score  0.73077
HMM 3 with model 100 has score  0.00000
The best score for HHM 3 was with model 3 with score  0.73077


HMM 4 with model 1 has score  0.73077
HMM 4 with model 2 has score  0.11538
HMM 4 with model 3 has score  0.73077
HMM 4 with model 4 has score  0.00000
HMM 4 with model 5 has score  0.73077
HMM 4 with model 6 has score  0.00000
HMM 4 with model 7 has score  0.03846
HMM 4 with model 8 has score  0.03846
HMM 4 with model 9 has score  0.03846
HMM 4 with model 10 has score  0.00000
HMM 4 with model 11 has score  0.03846
HMM 4 with model 12 has score  0.03846
HMM 4 with model 13 has score  0.00000
```

```
HMM 4 with model 986 has score  0.73077
HMM 4 with model 987 has score  0.73077
HMM 4 with model 988 has score  0.15385
HMM 4 with model 989 has score  0.00000
HMM 4 with model 990 has score  0.00000
HMM 4 with model 991 has score  0.73077
HMM 4 with model 992 has score  0.73077
HMM 4 with model 993 has score  0.03846
HMM 4 with model 994 has score  0.73077
HMM 4 with model 995 has score  0.00000
HMM 4 with model 996 has score  0.03846
HMM 4 with model 997 has score  0.03846
HMM 4 with model 998 has score  0.03846
HMM 4 with model 999 has score  0.03846
HMM 4 with model 1000 has score  0.73077
The best score for HHM 4 was with model 51 with score  0.80769
```

The highest accuracy which the model was able to achieve is about 80%.

c)

```
(base) Yashs-MacBook-Pro:Assignment2 yash$ ./hmm BrownCorpus 15 1000 400 200 548

HMM 1 with model 1 has score  0.00000
The best score for HHM 1 was with model 1 with score  0.00000


HMM 2 with model 1 has score  0.00000
HMM 2 with model 2 has score  0.00000
HMM 2 with model 3 has score  0.00000
HMM 2 with model 4 has score  0.15385
HMM 2 with model 5 has score  0.00000
HMM 2 with model 6 has score  0.00000
HMM 2 with model 7 has score  0.73077
HMM 2 with model 8 has score  0.73077
HMM 2 with model 9 has score  0.00000
HMM 2 with model 10 has score  0.00000
The best score for HHM 2 was with model 7 with score  0.73077


HMM 3 with model 1 has score  0.00000
HMM 3 with model 2 has score  0.00000
HMM 3 with model 3 has score  0.38462
HMM 3 with model 4 has score  0.73077
HMM 3 with model 5 has score  0.00000
```

```
HMM 3 with model 94 has score  0.69231
HMM 3 with model 95 has score  0.00000
HMM 3 with model 96 has score  0.00000
HMM 3 with model 97 has score  0.73077
HMM 3 with model 98 has score  0.00000
HMM 3 with model 99 has score  0.73077
HMM 3 with model 100 has score  0.73077
The best score for HHM 3 was with model 7 with score  0.73077


HMM 4 with model 1 has score  0.00000
HMM 4 with model 2 has score  0.11538
HMM 4 with model 3 has score  0.00000
HMM 4 with model 4 has score  0.73077
HMM 4 with model 5 has score  0.00000
HMM 4 with model 6 has score  0.38462
HMM 4 with model 7 has score  0.00000
HMM 4 with model 8 has score  0.00000
HMM 4 with model 9 has score  0.73077
HMM 4 with model 10 has score  0.15385
HMM 4 with model 11 has score  0.00000
```

```
HMM 4 with model 993 has score  0.00000
HMM 4 with model 994 has score  0.00000
HMM 4 with model 995 has score  0.73077
HMM 4 with model 996 has score  0.73077
HMM 4 with model 997 has score  0.03846
HMM 4 with model 998 has score  0.73077
HMM 4 with model 999 has score  0.00000
HMM 4 with model 1000 has score  0.73077
The best score for HHM 4 was with model 830 with score  0.76923
```

d)

```
(base) Yashs-MacBook-Pro:Assignment2 yash$ ./hmm BrownCorpus 15 1000 300 200 548

HMM 1 with model 1 has score  0.07692
The best score for HHM 1 was with model 1 with score  0.07692


HMM 2 with model 1 has score  0.03846
HMM 2 with model 2 has score  0.00000
HMM 2 with model 3 has score  0.07692
HMM 2 with model 4 has score  0.00000
HMM 2 with model 5 has score  0.34615
HMM 2 with model 6 has score  0.00000
HMM 2 with model 7 has score  0.69231
HMM 2 with model 8 has score  0.00000
HMM 2 with model 9 has score  0.23077
HMM 2 with model 10 has score  0.69231
The best score for HHM 2 was with model 7 with score  0.69231
```

```
HMM 3 with model 84 has score  0.15385
HMM 3 with model 85 has score  0.00000
HMM 3 with model 86 has score  0.03846
HMM 3 with model 87 has score  0.00000
HMM 3 with model 88 has score  0.30769
HMM 3 with model 89 has score  0.00000
HMM 3 with model 90 has score  0.03846
HMM 3 with model 91 has score  0.00000
HMM 3 with model 92 has score  0.00000
HMM 3 with model 93 has score  0.03846
HMM 3 with model 94 has score  0.07692
HMM 3 with model 95 has score  0.00000
HMM 3 with model 96 has score  0.30769
HMM 3 with model 97 has score  0.07692
HMM 3 with model 98 has score  0.00000
HMM 3 with model 99 has score  0.57692
HMM 3 with model 100 has score  0.00000
The best score for HHM 3 was with model 21 with score  0.73077
```

```
HMM 4 with model 992 has score  0.00000
HMM 4 with model 993 has score  0.11538
HMM 4 with model 994 has score  0.00000
HMM 4 with model 995 has score  0.00000
HMM 4 with model 996 has score  0.69231
HMM 4 with model 997 has score  0.26923
HMM 4 with model 998 has score  0.11538
HMM 4 with model 999 has score  0.00000
HMM 4 with model 1000 has score  0.00000
The best score for HHM 4 was with model 21 with score  0.73077

(base) Yashs-MacBook-Pro:Assignment2 yash$
```

Note: - A detailed output file has been attached with the submission.

Q.15.

This question is about the Zodiac Cipher 408. To find and score key of the Zodiac cipher, we will first create a digraph matrix same as the one in the previous question to find probabilities of state changes between alphabets.

We will then store the Zodiac Cipher 408 in our observation sequence in the structure.

Following this, we will train the HMM models on multiple random restarts starting with 1000 iterations for the first sub-question to 1,000,000 random restarts for the last sub-question.

We will try again to map the alphabets again with the B matrix which has converged after the iterations and check how many alphabets have actually mapped to the letters. We will compute a score based on this. The hindsight here is that, one alphabet may be mapped to more than one letter. For this, we will have to compute the matching in consideration of this.

The final output's screenshot is shown here and the full output file is attached in the submission folder.

a)

```
Model 989 has score  0.03846
Model 990 has score  0.00000
Model 991 has score  0.00000
Model 992 has score  0.00000
Model 993 has score  0.00000
Model 994 has score  0.03846
Model 995 has score  0.03846
Model 996 has score  0.03846
Model 997 has score  0.03846
Model 998 has score  0.00000
Model 999 has score  0.00000
Model 1000 has score  0.03846
The best score was for model 486 with score  0.15385
```

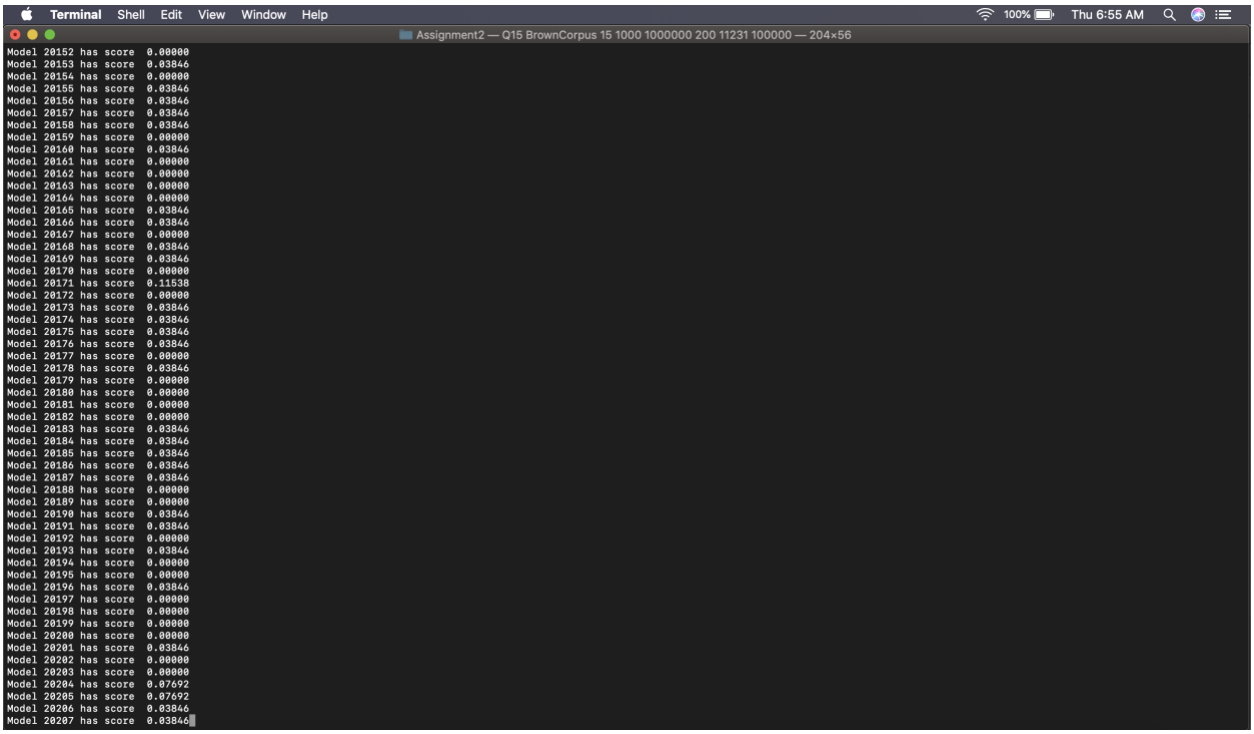
b)

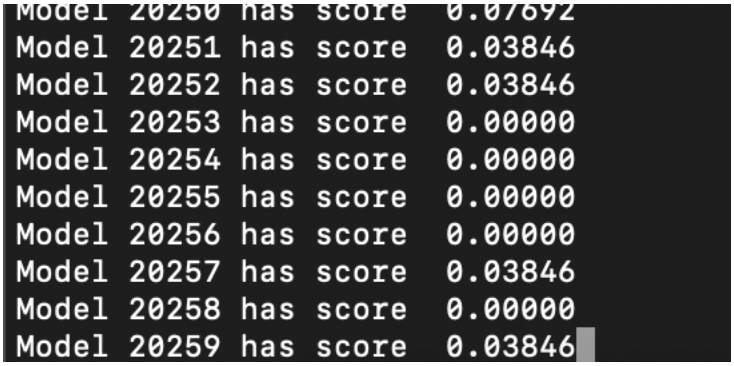
```
Model 9995 has score  0.00000
Model 9996 has score  0.00000
Model 9997 has score  0.07692
Model 9998 has score  0.07692
Model 9999 has score  0.00000
Model 10000 has score  0.03846
The best score was for model 486 with score  0.15385

(base) Yashs-MacBook-Pro:Assignment2 yash$
```

c)

For 100,000 iterations, about 24 hours were taking for the computations due to which I could not compute the results for these many random restarts. I tried to make the code more efficient and faster but that is the best time constraint for 100,000 iterations. I was able to complete only about 20000 iterations. The below screenshot illustrates that.





d)

Similarly, for 1,000,000 iterations it will take days to compute the results based on this code without any high-performance computing. Hence, the results for this also could not be computed.

Note: - A detailed output file has been attached with the submission.