Homework 7
Hidden Markov Models

CMU 10-601: Machine Learning (Fall 2018)
https://piazza.com/cmu/fall2018/10601bd
OUT: Nov 9, 2018
DUE: Nov 19, 2018
TAs: Aakanksha, Edgar, Sida, Varsha

Summary In this assignment you will implement a new named entity recognition system using Hidden Markov Models. You will begin by going through some multiple choice warm-up problems to build your intuition for these models and then use that intuition to build your own HMM models.

START HERE: Instructions[1]

- **Collaboration Policy**: Collaboration on solving the homework is allowed, after you have thought about the problems on your own. It is also OK to get clarification (but not solutions) from books or online resources, again after you have thought about the problems on your own. There are two requirements: first, cite your collaborators fully and completely (e.g., "Jane explained to me what is asked in Question 3.4"). Second, write your solution *independently*: close the book and all of your notes, and send collaborators out of the room, so that the solution comes from you only. See the collaboration policy on the website for more information: http://www.cs.cmu.edu/~mgormley/courses/10601bd-f18/about.html
- **Late Submission Policy:** See the late submission policy here: http://www.cs.cmu.edu/~mgormley/courses/10601bd-f18/about.html
- **Submitting your work:** You will use Gradescope to submit answers to all questions, and Autolab to submit your code. Please follow instructions at the end of this PDF to correctly submit all your code to Autolab.
  - **Gradescope:** For written problems such as derivations, proofs, or plots we will be using Gradescope (https://gradescope.com/). Submissions can be handwritten, but should be labeled and clearly legible. If your writing is not legible, you will not be awarded marks. Alternatively, submissions can be written in LaTeX. Upon submission, label each question using the template provided. Regrade requests can be made, however this gives the TA the opportunity to regrade your entire paper, meaning if additional mistakes are found then points will be deducted. Each derivation/proof should be completed on a separate page.
  - **Autolab:** You will submit your code for programming questions on the homework to Autolab (https://autolab.andrew.cmu.edu/). After uploading your code, our grading scripts will autograde your assignment by running your program on a virtual machine (VM). The software installed on the VM is identical to that on `linux.andrew.cmu.edu`, so you should check that your code runs correctly there. If developing locally, check that the version number of the programming language environment (e.g. Python 2.7, Octave 3.8.2, OpenJDK 1.8.0, g++ 4.8.5) and versions of permitted libraries (e.g. `numpy` 1.7.1) match those on `linux.andrew.cmu.edu`. Octave users: Please make sure you do not use any Matlab-specific libraries in your code that might make it fail against our tests. Python3 users: Please include a blank file called python3.txt (case-sensitive) in your tar submission. You have a **total of 10 Autolab submissions**. Use them wisely. In

[1]Compiled on Monday 19th November, 2018 at 21:53

order to not waste Autolab submissions, we recommend debugging your implementation on your local machine (or the linux servers) and making sure your code is running correctly first before any Autolab submission.

- **Materials:** Download from autolab the tar file ("Download handout"). The tar file will contain all the data that you will need in order to complete this assignment.

For multiple choice or select all that apply questions, shade in the box or circle in the template document corresponding to the correct answer(s) for each of the questions. For LaTeXusers, use ■ and ● for shaded boxes and circles, and don't change anything else.

Instructions for Specific Problem Types

For "Select One" questions, please fill in the appropriate bubble completely:

Select One: Who taught this course?

● Matt Gormley

○ Marie Curie

○ Noam Chomsky

If you need to change your answer, you may cross out the previous answer and bubble in the new answer:

Select One: Who taught this course?

● Matt Gormley

○ Marie Curie

● Noam Chomsky

For "Select all that apply" questions, please fill in all appropriate squares completely:

Select all that apply: Which are scientists?

■ Stephen Hawking

■ Albert Einstein

■ Isaac Newton

□ I don't know

Again, if you need to change your answer, you may cross out the previous answer(s) and bubble in the new answer(s):

Select all that apply: Which are scientists?

■ Stephen Hawking

■ Albert Einstein

■ Isaac Newton

■ I don't know

For questions where you must fill in a blank, please make sure your final answer is fully included in the given space. You may cross out answers or parts of answers, but the final answer must still be within the given space.

Fill in the blank: What is the course number?

10-601

10-7601

1 Written Questions [20 pts]

1.1 Multiple Choice [10 pts]

In this section we will test your understanding of several aspects of HMMs. Shade in the box or circle in the template document corresponding to the correct answer(s) for each of the questions below.

1. (2 points. **Select all that apply**) Which of the following are true under the (first-order) Markov assumption in an HMM:

   ☐ The states are independent

   ☐ The observations are independent

   ☐ $y_t \perp y_{t-1} | y_{t-2}$

   ■ $y_t \perp y_{t-2} | y_{t-1}$

   ☐ None of the above

2. (2 points. **Select all that apply**) Which of the following independence assumptions hold in an HMM:

   ■ The current observation x_t is conditionally independent of all other observations given the current state y_t

   ■ The current observation x_t is conditionally independent of all other states given the current state y_t

   ■ The current state y_t is conditionally independent of all states given the previous state y_{t-1}

   ☐ The current observation x_t is conditionally independent of x_{t-2} given the previous observation x_{t-1}.

   ☐ None of the above

In the remaining questions you will always see two quantities and decide what is the strongest relation between them. (? means it's not possible to assign any true relation). As such there is **only one correct answer**.

3. (2 points. Select one) What is the relation between $\sum_{i=0}^{N-1}(\alpha_5(i) * \beta_5(i))$ and $P(\mathbf{x})$? Select only the **strongest** relation that necessarily holds.

   ● =
   ◯ >
   ◯ <
   ◯ ≤
   ◯ ≥
   ◯ ?

4. (2 points. Select one) What is the relation between $P(y_4 = s_1, y_5 = s_2, \mathbf{x})$ and $\alpha_4(s_1) \cdot \beta_5(s_2)$? Select only the **strongest** relation that necessarily holds.

   ◯ =
   ◯ >
   ◯ <
   ● ≤
   ◯ ≥

◯ ?

5. (2 points. Select one) What is the relation between $\alpha_5(i)$ and $\beta_5(i)$? Select only the **strongest** relation that necessarily holds.

◯ =
◯ >
◯ <
◯ ≤
◯ ≥
● ?

1.2 Warm-up Exercise: Forward-Backward Algorithm [4 pts]

To help you prepare to implement the HMM forward-backward algorithm (see Section 2.3 for a detailed explanation), we have provided a small example for you to work through by hand. This toy data set consists of a training set of three sequences with three unique words and two tags and a test set with a single sequence composed of the same unique words used in the training set. Before going through this example, please carefully read the algorithm description in Sections 2.2 and 2.3.

Training set:

```
you_B eat_A fish_B
you_B fish_B eat_A
eat_A fish_B
```

Where the training word sequences are:

$$x = \begin{bmatrix} you & eat & fish \\ you & fish & eat \\ eat & fish & \end{bmatrix}$$

And the corresponding tags are:

$$y = \begin{bmatrix} B & A & B \\ B & B & A \\ A & B & \end{bmatrix}$$

Test set:

```
fish eat you
```

or

$$x = \begin{bmatrix} fish & eat & you \end{bmatrix}$$

The following four questions are meant to encourage you to work through the forward backward algorithm by hand using this test example. Feel free to use a calculator, being careful to carry enough significant figures through your computations to avoid rounding errors. For each question below, please report the requested value in the text box next to the question (these boxes are only visible in the template document). When a number is requested, only write the number in the box. When a word/tag is requested, only write that word or tag. **DO NOT** include explanations or derivation work in the text box. Points will be deducted if anything else is included in the box.

1. (1 point) Compute $\alpha_2(A)$, the α value associated with the tag "A" for the second word in the test sequence. Please round your answer to **THREE** decimal places.

   0.131

2. (1 point) Compute $\beta_2(B)$, the β value associated with the tag "B" for the second word in the test sequence. Please round your answer to **THREE** decimal places.

   0.250

3. (1 point) Predict the tag for the third word in the test sequence.

   B

4. (1 point) Compute the log-likelihood for the entire test sequence, "`fish eat you`". Please round your answer to **THREE** decimal places.

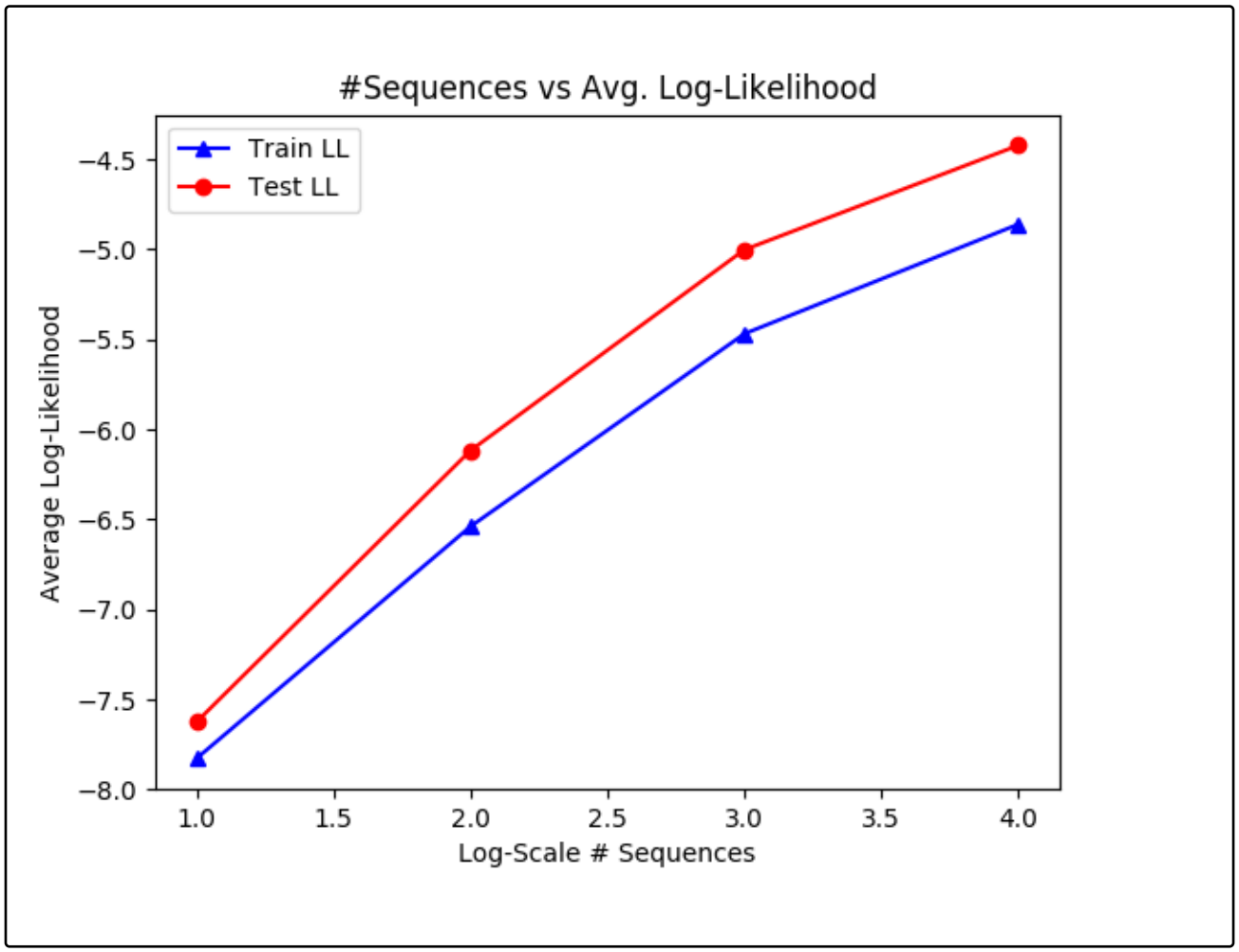
   -3.044

1.3 Empirical Questions [6 pts]

[Return to these questions after implementing your `learnhmm.{py|java|cpp|m}` and `forwardbackward.{py|java|cpp|m}` functions]

Using the fulldata set **trainwords.txt** in the handout using your implementation of `learnhmm.{py|java|cpp|m}` to learn parameters for an hmm model using the first 10, 100, 1000, and 10000 sequences in the file. Use these learned parameters perform prediction on the **trainwords.txt** and the **testwords.txt** files using your `forwardbackward.{py|java|cpp|m}`. Construct a plot with number of sequences used for training on the x-axis (log-scale) and average log likelihood across all sequences from the **trainwords.txt** or the **testwords.txt** on the y-axis (see Section 2.3 for details on computing the log data likelihood for a sequence). Each table entry is worth 0.5 points. Write the resulting log likelihood values in the table in the template. Include your plot in the large box in the template (2 points). To receive credit for your plot, you must submit a computer generated plot. **DO NOT** hand draw your plot.

#sequences	Train average log-likelihood	Test average log-likelihood
10	-7.827	-7.625
100	-6.539	-6.117
1000	-5.469	-5.002
10000	-4.861	-4.423



1.4 Collaboration Policy

After you have completed all other components of this assignment, report your answers to the collaboration policy questions detailed in the Academic Integrity Policies found here.

1. Did you receive any help whatsoever from anyone in solving this assignment? Is so, include full details.
2. Did you give any help whatsoever to anyone in solving this assignment? Is so, include full details.
3. Did you find or come across code that implements any part of this assignment ? If so, include full details.