HW7 Solution Template

1 Written Questions [20 pts]

1.1 Multiple Choices [10 pts]

In this section we will test your understanding of several aspects of HMMs. Shade in the box or circle corresponding to the correct answer(s) for each of the questions below. For LaTeXusers, use ■ and ●for shaded boxes and circles, and don't change anything else.

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

   ☐ $y_t \perp y_{t-1} | y_{t-2}$

   ☑ $y_t \perp y_{t-2} | y_{t-1}$

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}

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.

   ○ $=$ A

   ○ $>$

   ○ $<$

   ○ $\leq$

   ○ $\geq$

   ○ ?

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.

   ◯ =

   ◯ >

   ◯ <

   ◯ ≤ D

   ◯ ≥

   ◯ ?

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.

   ◯ =

   ◯ >

   ◯ <

   ◯ ≤ F

   ◯ ≥

   ◯ ?

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 five unique words and three 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, and write or type **only the number** in the box.

   $\alpha_2(A)$ = 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, and write or type **only the number** in the box.

   $\beta_2(B)$ = 0.250

3. (1 point) Predict the tag for the third word in the test sequence, and write or type **only the tag** in the box.

   The third tag is: B

4. (1 point) Compute the log-likelihood for the entire test sequence, "`fish eat you`". Please round your answer to **THREE** decimal places , and write or type **only the number** in the box.

   $l(test)$ = -3.044

1.3 Empirical Question [6 pts]

[Return to these questions after implementing your `learnhmm.{py|java|cpp|m}` and `forwardbackward.{py|ja`
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. 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, please round you answers to **THREE** decimal places. 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	-122.378	-130.798
100	-110.171	-116.001
1000	-100.936	-104.832
10000	-95.279	-97.989