1. The Transition matrix A defined in lecture allows you to: 1 / 1 point

- ○ Compute the probability of going from a word to a part of speech tag.
- ○ Compute the probability of going from a part of speech tag to a word.
- ◉ Compute the probability of going from a part of speech tag to another part of speech tag.
- ○ Compute the probability of going from a word to another word.

✓ **Correct**
Correct.

2. The Emission matrix B defined in lecture allows you to: 1 / 1 point

- ○ Compute the probability of going from a part of speech tag to another part of speech tag.
- ○ Compute the probability of going from a word to another word.
- ◉ Compute the probability of going from a part of speech tag to a word.
- ○ Compute the probability of going from a word to a part of speech tag.

✓ **Correct**
Correct.

3. The column sum of the emission matrix has to be equal to 1. 1 / 1 point

- ◉ False.
- ○ True.

✓ **Correct**
It is the row sum that has to be 1.

4. The row sum of the transition matrix has to be 1. 1 / 1 point

- ○ False, it has to be the column sum.
- ◉ True

✓ **Correct**
Correct.

5. Why is smoothing usually applied? Select all that apply. 1 / 1 point

- [x] Applying smoothing, for the majority of cases, allows us to decrease the probabilities in the transition and emission matrices and this allows us to have non zero probabilities.

  Correct
  Correct.

- [ ] Applying smoothing, for the majority of cases, allows us to increase the probabilities in the transition and emission matrices and this allows us to have non zero probabilities.
- [ ] Applying smoothing is a bad idea and we should not use it.
- [x] Applying smoothing, for the minority of cases, allows us to increase the probabilities in the transition and emission matrices and this allows us to have non zero probabilities.

  Correct
  Correct.

6. Given the following D matrix, what would be the sequence of tags for the words on the right? 1 / 1 point

$D =$

	w_1	w_2	w_3	w_4	w_5
t_1	0	1	3	2	3
t_2	0	2	4	1	3
t_3	0	2	4	1	4
t_4	0	4	4	3	1

<s> w1 w2 w3 w4 w5

$$s = \underset{i}{\operatorname{argmax}}\, c_{i,K} = 1$$

- t_3, t_4, t_2, t_2, t_1
- t_3, t_4, t_2, t_3, t_1
- t_2, t_3, t_1, t_3, t_1
- t_1, t_3, t_1, t_2, t_1

Correct
Correct

- ○ t_3, t_4, t_2, t_2, t_1
- ○ t_3, t_4, t_2, t_3, t_1
- ◉ t_2, t_3, t_1, t_3, t_1
- ○ t_1, t_3, t_1, t_2, t_1

✓ Correct
Correct

7. Previously, we have been multiplying the raw probabilities, but in reality we take the log of those probabilities. Why might that be the case? 1 / 1 point

- ○ Because the log probabilities force the numbers to be between 0 and 1 and hence, we want to take a probability.
- ○ The log probabilities should not be used because they introduce noise to our original computed scores.
- ◉ We take the log probabilities because probabilities are bounded between 0 and 1 and as a result, the numbers could be too small and will go towards 0.
- ○ The log probabilities help us with the inference as they bound the numbers between -1 and 1.

✓ Correct
Correct.

8. Which of the following are useful for applications for parts of speech tagging? 1 / 1 point

- ☑ Speech recognition

✓ Correct
Correct.

- ☑ Coreference Resolution

✓ Correct
Correct.

- ☐ Sentiment Analysis
- ☑ Named Entity Recognition

✓ Correct
Correct.