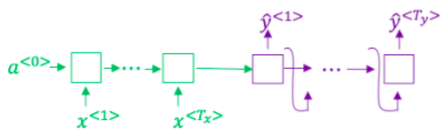
Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

Next item →

1. Consider using this encoder-decoder model for machine translation. **1 / 1 point**



True/False: This model is a “conditional language model” in the sense that the encoder portion (shown in green) is modeling the probability of the input sentence x.

- (●) False
- ( ) True

Correct
The encoder-decoder model for machine translation models the probability of the output sentence y conditioned on the input sentence x. The encoder portion is shown in green, while the decoder portion is shown in purple.

2. In beam search, if you increase the beam width B, which of the following would you expect to be true? Check all that apply. **1 / 1 point**

- [x] Beam search will generally find better solutions (i.e. do a better job maximizing $P(y \mid x)$)

Correct

- [x] Beam search will run more slowly.

Correct

- [ ] Beam search will converge after fewer steps.
- [x] Beam search will use up more memory.

Correct

3. True/False: In machine translation, if we carry out beam search without using sentence normalization, the algorithm will tend to output overly long translations. **1 / 1 point**

- (●) False
- ( ) True

Correct
In machine translation, if we carry out beam search without using sentence normalization, the algorithm will tend to output overly short translations.

4. Suppose you are building a speech recognition system, which uses an RNN model to map from audio clip x to a text transcript y. Your algorithm uses beam search to try to find the value of y that maximizes $P(y \mid x)$. **1 / 1 point**

On a dev set example, given an input audio clip, your algorithm outputs the transcript $\hat{y} =$ “I'm building an A Eye system in Silly con Valley.”, whereas a human gives a much superior transcript $y^* =$ “I'm building an AI system in Silicon Valley.”

According to your model,

$P(\hat{y} \mid x) = 1.09 * 10^-7$

$P(y^* \mid x) = 7.21 * 10^-8$

Would you expect increasing the beam width B to help correct this example?

- ( ) Yes, because $P(y^* \mid x) \leq P(\hat{y} \mid x)$ indicates the error should be attributed to the search algorithm rather than to the RNN.
- ( ) Yes, because $P(y^* \mid x) \leq P(\hat{y} \mid x)$ indicates the error should be attributed to the RNN rather than to the search algorithm.

◯ Yes, because $P(y^* \mid x) \leq P(\hat{y} \mid x)$ indicates the error should be attributed to the RNN rather than to the search algorithm.

◯ No, because $P(y^* \mid x) \leq P(\hat{y} \mid x)$ indicates the error should be attributed to the search algorithm rather than to the RNN.

◉ No, because $P(y^* \mid x) \leq P(\hat{y} \mid x)$ indicates the error should be attributed to the RNN rather than to the search algorithm.

Correct

5. Continuing the example from Q4, suppose you work on your algorithm for a few more weeks, and now find that for the vast majority of examples on which your algorithm makes a mistake, $P(y^* \mid x) > P(\hat{y} \mid x)$. This suggests you should focus your attention on improving the search algorithm.

1 / 1 point

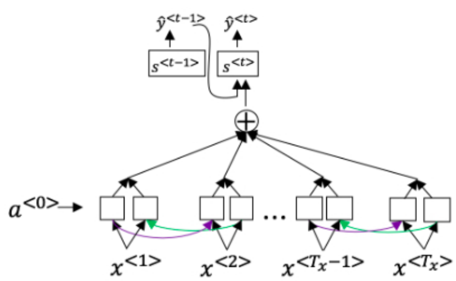
◉ True.

◯ False.

Correct

6. Consider the attention model for machine translation.

1 / 1 point



Further, here is the formula for $\alpha^{<t,t'>}$.

$$\alpha^{<t,t'>} = \frac{\exp(e^{<t,t'>})}{\sum_{t'=1}^{T_x} \exp(e^{<t,t'>})}$$

Which of the following statements about $\alpha^{<t,t'>}$ are true? Check all that apply.

☐ $\sum_{t'} \alpha^{<t,t'>} = 0$

☐ $\sum_{t'} \alpha^{<t,t'>} = -1$

☑ $\alpha^{<t,t'>}$ is equal to the amount of attention $y^{<t>}$ should pay to $a^{<t'>}$

Correct
Correct! $\alpha^{<t,t'>}$ = amount of attention $y^{<t>}$ should pay to $a^{<t'>}$

☐ We expect $\alpha^{<t,t'>}$ to be generally larger for values of $a^{<t'>}$ that are highly relevant to the value the network should output for $y^{<t'>}$. (**Note the indices in the superscripts.**)

7. The network learns where to "pay attention" by learning the values $e^{<t,t'>}$, which are computed using a small neural network:

1 / 1 point

We can replace $s^{<t-1>}$ with $s^{<t>}$ as an input to this neural network because $s^{<t>}$ is independent of $\alpha^{<t,t'>}$ and $e^{<t,t'>}$.

◯ True

◉ False

Correct
We can't replace $s^{<t-1>}$ with $s^{<t>}$ as an input to this neural network. This is because $s^{<t>}$ depends on $\alpha^{<t,t'>}$ which in turn depends on and $e^{<t,t'>}$; so at the time we need to evaluate this network, we haven't computed $s^{<t>}$.

8. The attention model performs the same as the encoder-decoder model, no matter the sentence length.

1 / 1 point

◯ True

◉ False

Correct

The performance of the encoder-decoder model declines as the amount of words increases. The attention model has the greatest advantage when the input sequence length T_x is large.

9. Under the CTC model, identical repeated characters not separated by the “blank” character (_) are collapsed. Under the CTC model, what does the following string collapse to? **1 / 1 point**

__c_oo_o_kk___b_ooooo__oo__kkk

- ○ cokbok
- ◉ cookbook
- ○ cook book
- ○ coookkbooooooookkk

Correct

10. In trigger word detection, $x^{<t>}$ represents the trigger word x being stated for the t-th time **1 / 1 point**

- ○ True
- ◉ False

Correct

$x^{<t>}$ represents the features of the audio at time t.