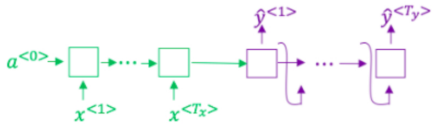
Congratulations! You passed!

Grade received 80% **Latest Submission Grade** 80% **To pass** 80% or higher

Go to next item

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



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

Expand

Correct

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

- [x] Beam search will run more quickly.

  Correct
  As the beam width decreases, beam search runs more quickly, uses up less memory, and converges after fewer steps, but will generally not find the maximum P(y|x).

- [x] Beam search will converge after fewer steps.

  Correct
  As the beam width decreases, beam search runs more quickly, uses up less memory, and converges after fewer steps, but will generally not find the maximum P(y|x).

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

Expand

Correct
Great, you got all the right answers.

3. True/False: In machine translation, if we carry out beam search using sentence normalization, the algorithm will tend to output overly short translations. **1 / 1 point**

- (•) False
- ( ) True

Expand

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,

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

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

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

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

Expand

Correct

$P(y^* \mid x) > P(\hat{y} \mid x)$ indicates the error should be attributed to the search algorithm rather than to the RNN. Increasing the beam width will generally allow beam search to find better solutions.

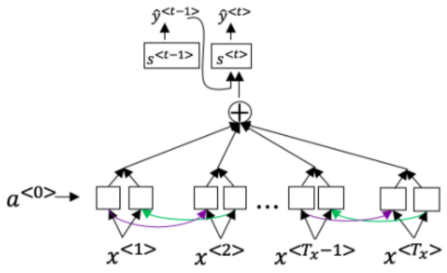
5. Continuing the example from Q4, suppose you work on your algorithm for a few more weeks, and now find that for the vast majority of examples on which your algorithm makes a mistake, $P(y^* \mid x) > P(\hat{y} \mid x)$. This suggests you should focus your attention on improving the search algorithm. **1 / 1 point**

- False.
- True.

Expand

Correct

6. Consider the attention model for machine translation. **0 / 1 point**



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

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

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

- [x] 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.)**

  This should not be selected
  Incorrect! 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 $y^{<t>}$, **not for** $y^{<t'>}$

- [ ] $\sum_{t'} \alpha^{<t,t'>} = -1$

- [x] $\alpha^{<t,t'>}$ 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 $a^{<t'>}$

- [ ] $\sum a^{<t,t'>} = 0$

Expand

Incorrect
You chose the extra incorrect answers.

7. The network learns where to "pay attention" by learning the values $e^{<t,t'>}$, which are computed using a small neural network: **0 / 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

Expand

Incorrect
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. Compared to the encoder-decoder model shown in Question 1 of this quiz (which does not use an attention mechanism), we expect the attention model to have the greatest advantage when: **1 / 1 point**

- ( ) The input sequence length T_x is large.
- ( ) The input sequence length T_x is small.

Expand

Correct

9. **1 / 1 point**

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?

aaa_aaaaaa_____rr_ddddddddd_____v_aaaaaa_rrrr_____kk

- ( ) aardvark
- ( ) aa rd var k
- ( ) ardvark

aaaaaaaaarrddddddddvaaaaaarrrrkk

Expand

Correct
The basic rule for the CTC cost function is to collapse repeated characters not separated by "blank". If a character is repeated, but separated by a “blank”, it is included in the string.

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

1 / 1 point

- (•) False
- ( ) True

Expand

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