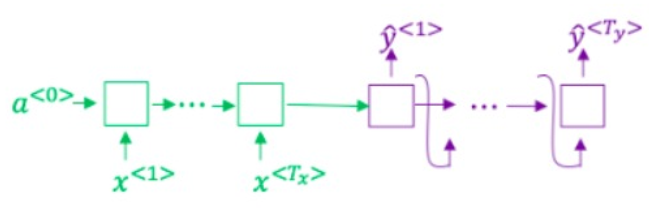
Your grade: 100%

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 decoder portion (shown in purple) is modeling the probability of the output sentence y given 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.

1 / 1 point

2. In beam search, if you increase the beam width B, which of the following would you expect to be true?

- ◉ Beam search will generally find better solutions (i.e. do a better job maximizing $P(y|x)$).
- ○ Beam search will converge after fewer steps.
- ○ Beam search will run more quickly.
- ○ Beam search will use up less memory.

Correct
As the beam width increases, beam search runs more slowly, uses up more memory, and converges after more steps, but generally finds better solutions.

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

- ◉ True
- ○ False

Correct

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 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 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.
- ○ 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.

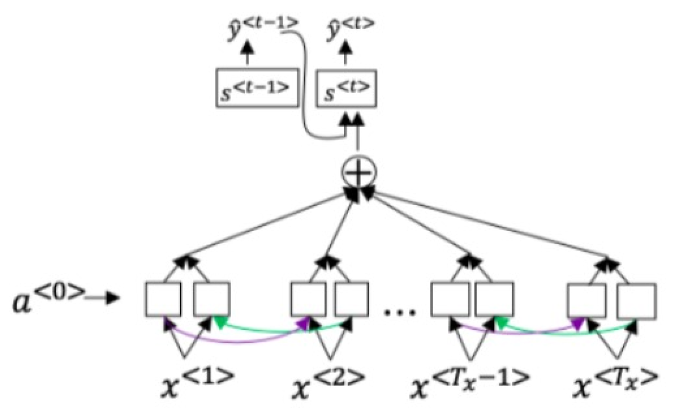
Correct

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 not focus your attention on improving the search algorithm. 1 / 1 point

- ◉ False
- ○ True

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.

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.

- ☑ $\sum_{t'} \alpha^{<t,t'>} = 1$ (**Note the summation is over** t'.)

Correct
Correct! If we sum over $\alpha^{<t,t'>}$ for all t' (the formulation can be seen in the image), the numerator will be equal to the denominator, therefore, $\sum_{t'} \alpha^{<t,t'>} = 1$.

- ☐ 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.**)
- ☐ $\sum_{t} \alpha^{<t,t'>} = 1$ (**Note the summation is over**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.**)

Correct
Correct! $\alpha^{<t,t'>}$is equal to the amount of attention $y^{<t>}$ should pay to $a^{<t'>}$. So, if a value of $a^{<t'>}$ is highly relevant to $y^{<t>}$, then the attention coefficient $\alpha^{<t,t'>}$should be larger. Note the difference between a (activation) and α (attention coefficient).

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

- ◉ False
- ○ True

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.

1 / 1 point

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?

aaa_aaaaaa________rr_dddddddd_______v_aaaaaa_rrrr________kk

- ○ aaaaaaaaarrddddddddvaaaaaarrrrkk
- ○ aa rd var k
- ○ ardvark
- ◉ aardvark

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, if the target label for $x^{<t>}$ is 1: 1 / 1 point

- ◉ Someone has just finished saying the trigger word at time *t*.
- ○ There is exactly one trigger word.
- ○ The total time that the trigger word detection algorithm has been running is 1.
- ○ Only one word has been stated.

Correct
Target labels indicate whether or not a trigger word has been said.