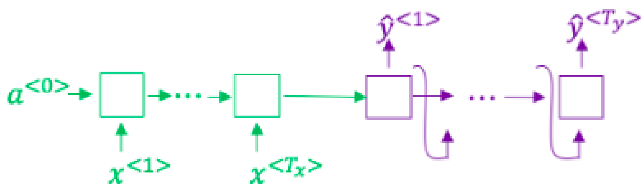
1.Consider using this encoder-decoder model for machine translation. 1 point



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.

- ◯ True
- ◯ False

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

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 point

- ◯ True
- ◯ False

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 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,

$$P(\hat{y} \mid x) = 1.95*10^{-7}$$

$$P(y^* \mid x) = 3.42*10^{-9}$$

True/False: Trying a different network architecture could help correct this example.

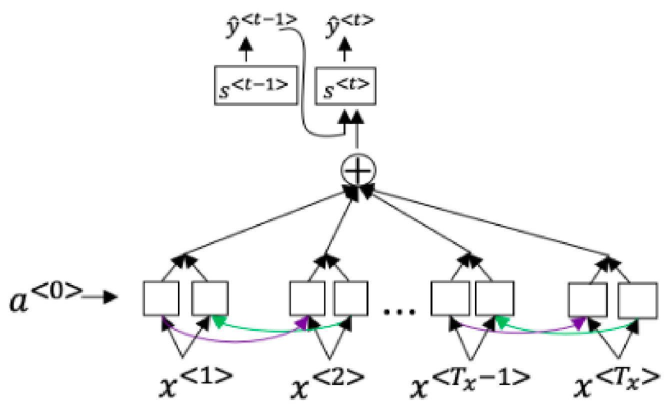
- ◯ True
- ◯ False

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 point

- ◯ True.

○ False.

6.Consider the attention model for machine translation. 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.

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

7.The network learns where to "pay attention" by learning the values $e^{<t,t'>}$, which are computed using a small neural network: 1 point

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 $e^{<t,t'>}$; so at the time we need to evaluate this network, we haven't computed $s^{<t>}$ yet.

○ False

○ True

8.The attention model performs the same as the encoder-decoder model, no matter the sentence length. 1 point

○ True

○ False

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_ddddddddd________v_aaaaaa_rrrr_________kk

- aaaaaaaaarrddddddddvaaaaaarrrrkk
- aa rd var k
- ardvark
- aardvark

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

- True
- False