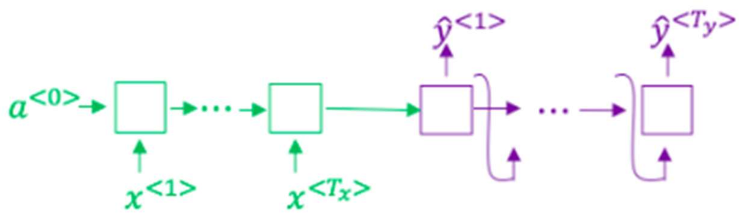
Question 1

Consider using this encoder-decoder model for machine translation.



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 xx.

1 / 1 point

False

True

Correct

2.

Question 2

In beam search, if you increase the beam width BB, which of the following would you expect to be true? Check all that apply.

1 / 1 point

Beam search will run more slowly.

Correct

Beam search will use up more memory.

Correct

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

Correct

Beam search will converge after fewer steps.

3.

Question 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.

Question 4

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

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

According to your model,

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

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

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

1 / 1 point

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

Correct

5.

Question 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)$P(y^*|x) > P(\hat{y}|x)$. This suggests you should focus your attention on improving the search algorithm.

1 / 1 point

False.

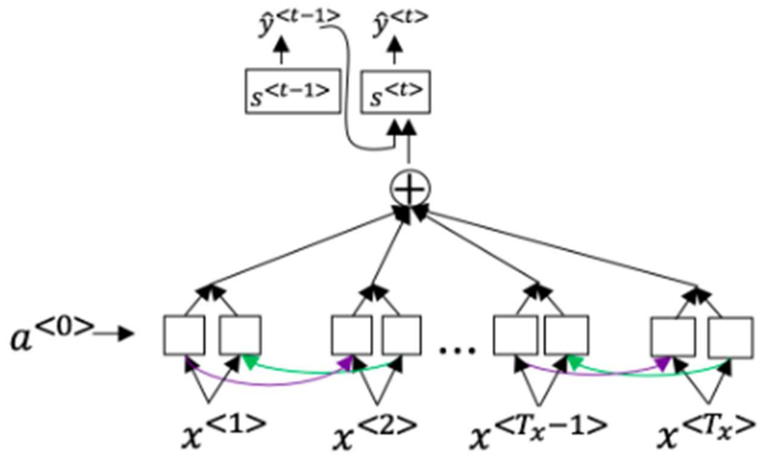
True.

Correct

6.

Question 6

Consider the attention model for machine translation.



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

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

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

1 / 1 point

☐

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

☐

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

Correct

☐

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

Correct

\sum_{t} \alpha^{<t,t’>} = 1$\sum_t \alpha^{<t,t'>}=1$ (Note the summation is over tt.)

7.

Question 7

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

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

1 / 1 point

False

True

Correct

8.

Question 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_xT_x is small.

The input sequence length T_xT_x is large.

Correct

9.

Question 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?

__c_oo_o_kk___b_ooooo__oo__kkk

1 / 1 point

cokbok

cookbook

cooookkbooooooookkk

cook book

Correct

10.

Question 10

In trigger word detection, x^{<t>}$x^{<t>}$ is:

1 / 1 point

Features of the audio (such as spectrogram features) at time tt.

Whether the trigger word is being said at time tt.

Whether someone has just finished saying the trigger word at time tt.

The tt-th input word, represented as either a one-hot vector or a word embedding.

Correct