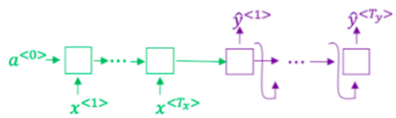
1. Consider using this encoder-decoder model for machine translation.

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.

- (●) False
- ( ) True

44:29

Expand

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 point

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

44:28

Expand

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 point

- (●) False
- ( ) True

44:28

Expand

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

44:27

Expand

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 point

( ) True

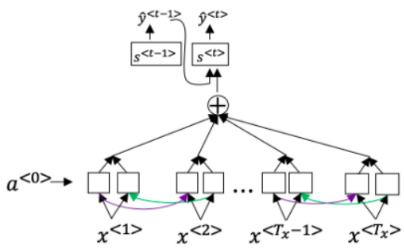
(•) False

44:27

Expand

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.

- [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.**)
- [ ] $\alpha^{<t,t'>}$ is equal to the amount of attention $y^{<t>}$ should pay to $a^{<t'>}$
- [x] $\sum_{t'} \alpha^{<t,t'>} = -1$
- [ ] $\sum_{t'} \alpha^{<t,t'>} = 0$

44:26

Expand

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 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

44:25

Expand

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 point

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

44:25

Expand

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 point

__c_oo_o_kk___b_ooooo__oo__kkk

- (•) cookbook
- ( ) coookkboooooookkk
- ( ) cokbok
- ( ) cook book

44:25

Expand

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

1 point

- ( ) Whether someone has just finished saying the trigger word at time t.
- ( ) Whether the trigger word is being said at time t.
- ( ) The t-th input word, represented as either a one-hot vector or a word embedding.
- (•) Features of the audio (such as spectrogram features) at time t.

44:24

Expand