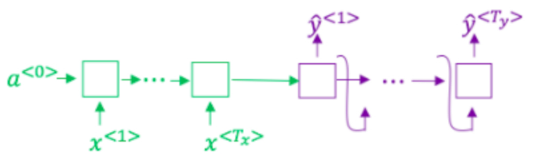
1. Consider using this encoder-decoder model for machine translation.

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

1 / 1 point

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

  Correct

- [x] Beam search will run more slowly.

  Correct

- [ ] Beam search will converge after fewer steps.
- [x] Beam search will use up more memory.

  Correct

Expand

Correct

Great, you got all the right answers.

Great, you got all the right answers.

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

- ○ False
- ◉ True

Expand

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 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 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.

Loading [MathJax]/jax/output/CommonHTML/jax.js

- ○ 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.

Expand

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 focus your attention on improving the RNN.

1 / 1 point

- ○ True
- ◉ False

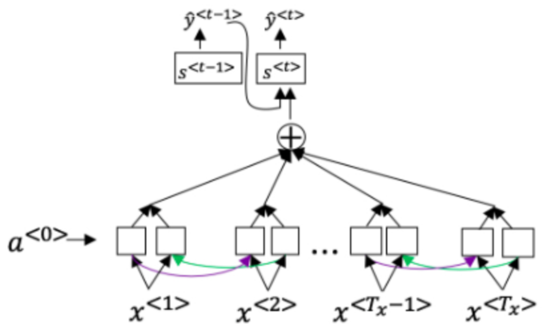
Expand

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.

6. Consider the attention model for machine translation.

0 / 1 point



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

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

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

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

Correct

We expect $\alpha^{<t,t'>}$ to be larger for activation values that are highly relevant to the value the network should output for $y^{<t>}$.

- [ ] $\sum_{t'} \alpha^{<t,t'>} = 0$

$\sum_{t'} \alpha^{<t,t'>} = 0$. (Note the summation is over t'.)

- [ ] $$\alpha^{<t,t’>}$$ is equal to the amount of attention $$y^{<t>}$$ should pay to $$\alpha^{<t’>}$$
- [x] We expect $$\alpha^{<t,t’>}$$ to be generally larger for values of $$\alpha^{<t’>}$$ that are highly relevant to the [illegible] the indices in the superscripts.)

$\alpha^{<t'>}$

Expand

Incorrect

You didn't select all the correct 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

Which of the following does $s^{<t>}$ depend on? Select all that apply.

- [ ] $s^{<t+1>}$
- [ ] $e^{<t,t'>}$
- [x] s^t is independent of $\alpha^{<t,t'>}$
- [ ] $\alpha^{<t,t'>}$

Expand

Incorrect
You didn't select all the correct answers

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

1 / 1 point

- (•) False
- ( ) True

Expand

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.

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 / 1 point

__c_oo_o_kk___b_ooooo__oo__kkk

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

Expand

Correct

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

1 / 1 point

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

Expand

Correct