Congratulations! You passed!

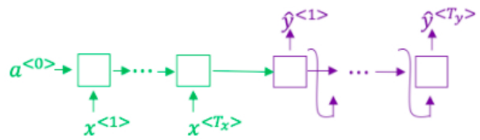
Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

Retake the assignment in **23h 53m**

Go to next item

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.

- ○ True
- ◉ False

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.

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 / 1 point

- ◉ False

- True

Expand

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

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,

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

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

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

- (selected) True
- False

Expand

Correct
$P(y^* \mid x) < P(\hat{y} \mid x)$ indicates the error should be attributed to the RNN rather than to the search algorithm. If the RNN model is at fault, then a deeper layer of analysis could help to figure out if you should add regularization, get more training data, or try a different network architecture.

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

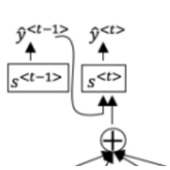
- True
- (selected) False

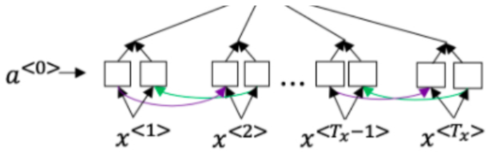
Expand

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.

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.

- ◯ We expect $\alpha^{<t,t'>}$ to be generally larger for values of $\alpha^{<t'>}$ 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 $a < t' >$
- ◯ $\sum_{t'} \alpha^{<t,t'>} = 0$
- ◯ $\sum_{t'} \alpha^{<t,t'>} = -1$

Expand

Correct

$\alpha^{<t,t'>}$ = amount of attention $y^{<t>}$ should pay to $a^{<t'>}$

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

- ◉ True
- ◯ False

Expand

Correct

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_x is small.
- ◉ The input sequence length T_x is large.

Expand

Correct

9. 1 / 1 point

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?

kk_eee____ee_p__eeeeeeee_____rrrrr

- ( ) keper
- ( ) kkeeeeepeeeeeeeerrrrr
- (•) keeper
- ( ) ke epe r

Expand

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

- ( ) 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.
- (•) Someone has just finished saying the trigger word at time <i>t</i>.

Expand

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