Back **Sequence Models & Attention Mechanism**
Graded Quiz • 30 min

coursera

Due Apr 25, 2:59 PM CST

Congratulations! You passed!

Item Navigation

Grade received 80% **To pass** 80% or higher

Go to next item

Sequence Models & Attention Mechanism

Quiz • 30 min

Sequence Models & Attention Mechanism

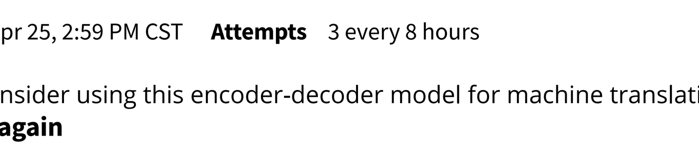
Submit your assignment

Latest Submission Grade 80%

Due Apr 25, 2:59 PM CST **Attempts** 3 every 8 hours

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

Try again



To P

You

80%

True/False: This model is a "conditional language model" in the sense that the decoder portion (shown in green) is modeling the probability of the input sentence x.

View Feedback

We keep your highest score

- ( ) True
- (•) False

Like Dislike Report an issue

Correct
The encoder-decoder model for machine translation models the probability of the output sentence y conditioned on the input sentence x. The encoder portion is shown in green, while the decoder portion is shown in purple.

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**

- [ ] Beam search will converge after fewer steps.
- [x] Beam search will run more slowly.

  Correct
- [x] Beam search will use up more memory.

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

  Correct

3. True/False: In machine translation, if we carry out beam search using sentence normalization, the algorithm will tend to output overly short translations. **0 / 1 point**

- (•) True
- ( ) False

Incorrect
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 $P(y \mid x)$. **0 / 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?

- ( ) 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.
- ( ) 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.
- (•) 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.
- ( ) 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.

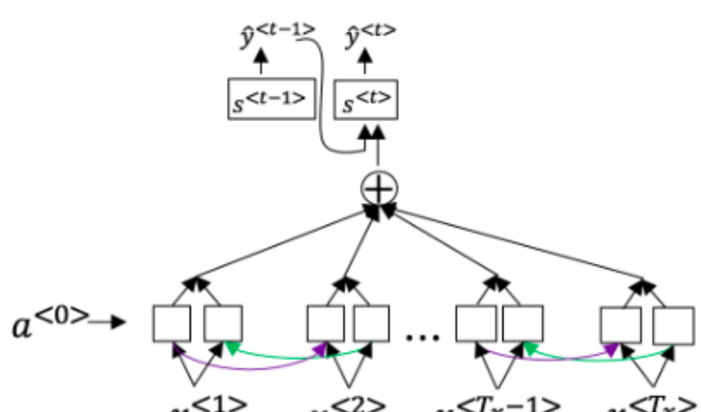
Incorrect

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 / 1 point**

- ( ) False.
- (•) True.

Correct

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.

- [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.)

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

  Correct
- [ ] $\sum_{t} \alpha^{<t,t'>} = 1$ (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 / 1 point**

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

- [x] $e^{<t,t'>}$

  Correct
  $s^{<t>}$ depends on $\alpha^{<t,t'>}$ which in turn depends on $e^{<t,t'>}$.
- [x] $\alpha^{<t,t'>}$

  Correct
  $s^{<t>}$ depends on $\alpha^{<t,t'>}$ which in turn depends on $e^{<t,t'>}$.
- [ ] $s^{<t+1>}$
- [ ] s^t is independent of $\alpha^{<t,t'>}$ and $e^{<t,t'>}$.

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.

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

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

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, $x^{<t>}$ represents the trigger word x being stated for the t-th time **1 / 1 point**

- ( ) True
- (•) False

Correct
$x^{<t>}$ represents the features of the audio at time t.