[**Skip to Main Content**](https://www.coursera.org/learn/nlp-sequence-models/exam/4CCc4/sequence-models-attention-mechanism/view-attempt#main)

[[](https://www.coursera.org/)](https://www.coursera.org/)

Explore

Top of Form



Bottom of Form

* **1**
* Vijay Misra

1. [Sequence Models](https://www.coursera.org/learn/nlp-sequence-models/home/welcome)
2. [Week 3](https://www.coursera.org/learn/nlp-sequence-models/home/week/3)
3. Sequence Models & Attention Mechanism

[**Previous**](https://www.coursera.org/learn/nlp-sequence-models/supplement/Z2Xaf/lectures-in-pdf)[**Next**](https://www.coursera.org/learn/nlp-sequence-models/supplement/0kud6/instructions-if-you-are-unable-to-open-your-notebook)

* **Various Sequence To Sequence Architectures**
* **Speech Recognition - Audio Data**
* **Lecture Notes (Optional)**
* **Quiz**

**[Quiz:](https://www.coursera.org/learn/nlp-sequence-models/exam/4CCc4/sequence-models-attention-mechanism)**[Sequence Models & Attention Mechanism](https://www.coursera.org/learn/nlp-sequence-models/exam/4CCc4/sequence-models-attention-mechanism)

[10 questions](https://www.coursera.org/learn/nlp-sequence-models/exam/4CCc4/sequence-models-attention-mechanism)

* **Programming Assignments**

**QUIZQuiz • 30 MIN30 minutes**

**Sequence Models & Attention Mechanism**

**Submit your assignment**

**DUE DATE**Jul 4, 11:59 PM PDTJuly 4, 11:59 PM PDT

**ATTEMPTS**3 every 8 hours

Try again

**Receive grade**

**TO PASS**80% or higher

**Grade**

85%

View Feedback

We keep your highest score

Sequence Models & Attention Mechanism

Graded Quiz • 30 min

**Due** Jul 4, 11:59 PM PDT

**Congratulations! You passed!**

**TO PASS**80% or higher

Keep Learning

**GRADE**

85%

**Sequence Models & Attention Mechanism**

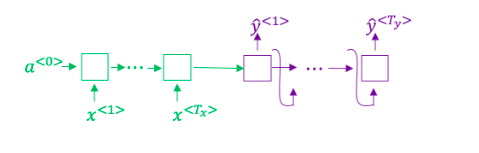
**LATEST SUBMISSION GRADE**

85%

1.

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 x*x*.

**1 / 1 point**



False



True

**Correct**

2.

Question 2

In beam search, if you increase the beam width B*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.



Beam search will use up more memory.

**Correct**



Beam search will run more slowly.

**Correct**



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

**Correct**

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**



False



True

**Correct**

4.

Question 4

Suppose you are building a speech recognition system, which uses an RNN model to map from audio clip x*x* to a text transcript y*y*. Your algorithm uses beam search to try to find the value of y*y* 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}=*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*(*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**



No, because P(y^\* \mid x) \leq P(\hat{y} \mid x)*P*(*y*∗∣*x*)≤*P*(*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*)≤*P*(*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*)≤*P*(*y*^​∣*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)*P*(*y*∗∣*x*)≤*P*(*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*(*y*^​∣*x*). This suggests you should focus your attention on improving the search algorithm.

**0 / 1 point**



True.



False.

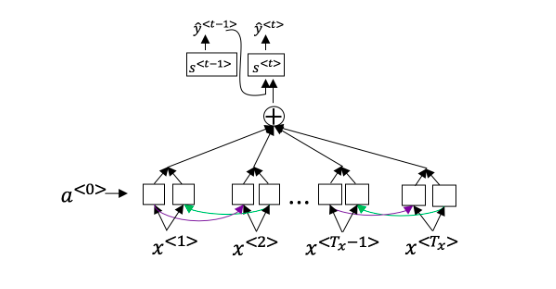
**Incorrect**

You didn’t select an answer.

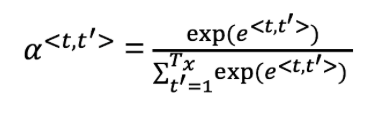
6.

Question 6

Consider the attention model for machine translation.



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



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

**0.5 / 1 point**



We expect \alpha^{<t,t’>}*α*<*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∑*t*​*α*<*t*,*t*’>=1 (Note the summation is over t*t*.)

**This should not be selected**



We expect \alpha^{<t,t’>}*α*<*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.)



\sum\_{t’} \alpha^{<t,t’>} = 1∑*t*’​*α*<*t*,*t*’>=1 (Note the summation is over t’*t*’.)

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’>}*α*<*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\_x*Tx*​ is large.



The input sequence length T\_x*Tx*​ is small.

**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**



cookbook



coookkboooooookkk



cook book



cokbok

**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 t*t*.



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



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



Whether the trigger word is being said at time t*t*.

**Correct**