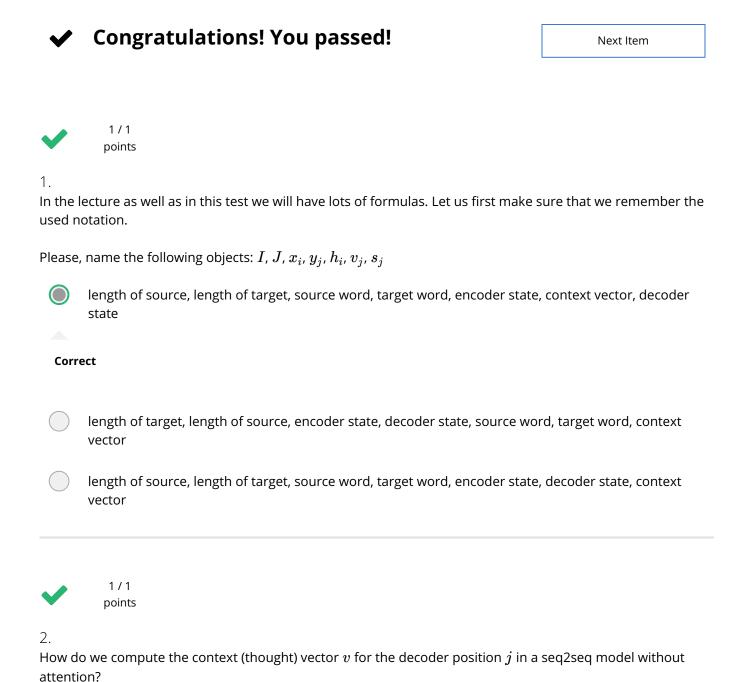
Encoder-decoder architectures

Quiz, 8 questions



 h_I , where h_I is the last encoder state

 h_j , where h_j is the j-th state of the encoder

Correct

Correct!

Encoder-decoder architectures

Quiz, 8 questions



1/1 points

3

How many new parameters for the network are introduced to calculate **multiplicative attention** weights? (Just to calculate, we are not yet looking into how we use them afterwards).

- The length of the source, multiplied by the length of the target
- The dimension of an encoder state, multiplied by the dimension of a decoder state

Correct

Exactly! This is the number of parameters for **multiplicative** attention.

No new parameters



1/1 points

4.

Which of the following formulas stand for the **additive attention**? Note that h_i is the i-th encoder state, s_j is the j-th decoder state, and we are interested in the similarity between them.



 $w^T tanh(W[h_i,s_j])$, where the brackets denote concatenation of the vectors, and w and W are a vector and a matrix of parameters respectively.



Correct

Exactly! Take a moment to see, that this is just a different form of the same additive attention formula that was introduced at the lecture.

- $h_i^TWs_j$, where W is a matrix of parameters
- $igcup h_i^T s_j$



1/1 points

5.

Let us denote encoder states by h_i with i going from 1 to I. Lets us denote by a_i^j the similarities computed Encode the decide i the context vector i and i the context vector i and i the context vector i and i are decoder position i?

$$\sum_{i=1}^{I}a_{i}^{j}h_{i}$$

$$\sum_{i=1}^{I} rac{\exp a_i^j}{\sum_{i'} \exp a_{i'}^j} \, h_i$$

Correct

Correct! We apply *softmax* to transform the weights into probabilities and compute the average of the encoder states.

$$\sum_{i=1}^{I}rac{\exp a_{i}^{j}}{\sum_{j'}\exp a_{i}^{j'}}\,h_{i}$$



1/1 points

6.

Which three vectors should be passed to a decoder state s_j in a seq2seq with attention model from the lecture?

 h_{i-1} - the previous encoder state

Un-selected is correct

 $oxed{oxed} v_j$ - the context vector for position j, calculated using attention

Correct

 s_j - the j-th decoder state

Un-selected is correct

 x_{i-1} - the previous word in the source sequence

Un-selected is correct

 y_j - the j-th word in the target sequence

Encoder-decoder-architectures

Quiz, 8 d	questions
-----------	-----------

	x_i - the i -th word in the source sequence
Un-selected is correct	
Corr	y_{j-1} - the previous word in the target sequence
Corr	s_{j-1} - the previous decoder state $oldsymbol{e}$
	h_i - the i -th encoder state
Un-s	selected is correct
~	1/1 points
7. Which tokens	techniques would help if the data has rich morphology, informal spelling, and other sources of OOV 5?
	Hierarchical softmax
Un-s	selected is correct
	Byte-pair encoding
	Byte-pair encoding
Un-s Corr	Byte-pair encoding

10/13/2018 Natural Language Processing - Home | Coursera Encoder-decoder architectures Quiz, 8 question sub-word modeling Correct Copy mechanism Correct 1/1 points Let us imagine we have trained a conversational chat-bot as a seq2seq model on Harry Potter movies subtitles. What problems could we expect? The bot suggests to use a time-turner or probably some spell if you say you do not have enough time for your Coursera studies Correct Dramatic/unrealistic topics, influenced by the training corpus! If asked in English, the bot replies in French or some other language **Un-selected** is correct The bot makes lots of spelling mistakes **Un-selected** is correct The bot doesn't remember what has already been decided in your dialogue Correct No memory of the context / history!

When asked "What's your name?", the bot is not sure and says Harry, or Ron, or Hermione from

time to time.

Encoder architectures Quiz, 8 quest personality problem!





