Encoder-decoder architectures

LATEST SUBMISSION GRADE

100%

1.Question 1

In the lecture as well as in this test we will have lots of formulas. Let us first make sure that we remember the used notation.

Please, name the following objects: II, JJ, x_ix_i, y_jy_j, h_ih_i, v_jv_j, s_js_j

○ length of target, length of source, encoder state, decoder state, source word, target word, context vector

○ length of source, length of target, source word, target word, encoder state, decoder state, context vector

● length of source, length of target, source word, target word, encoder state, context vector, decoder state

Correct

1 / 1 point

2.Question 2

How do we compute the context (thought) vector vv for the decoder position jj in a seq2seq model without attention?

● h_Ih_I, where h_Ih_I is the last encoder state

○ h_jh_j, where h_jh_j is the jj-th state of the encoder

○ \sum_i \alpha_i^j h_i$\sum_i \alpha_i^j h_i$, where \alpha_i^jα_i^j are some weights

Correct

Correct!

1 / 1 point

3.Question 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

○ No new parameters

Correct

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

1 / 1 point

4.Question 4

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

○ h_i^T W s_j$h_i^T W s_j$, where W is a matrix of parameters

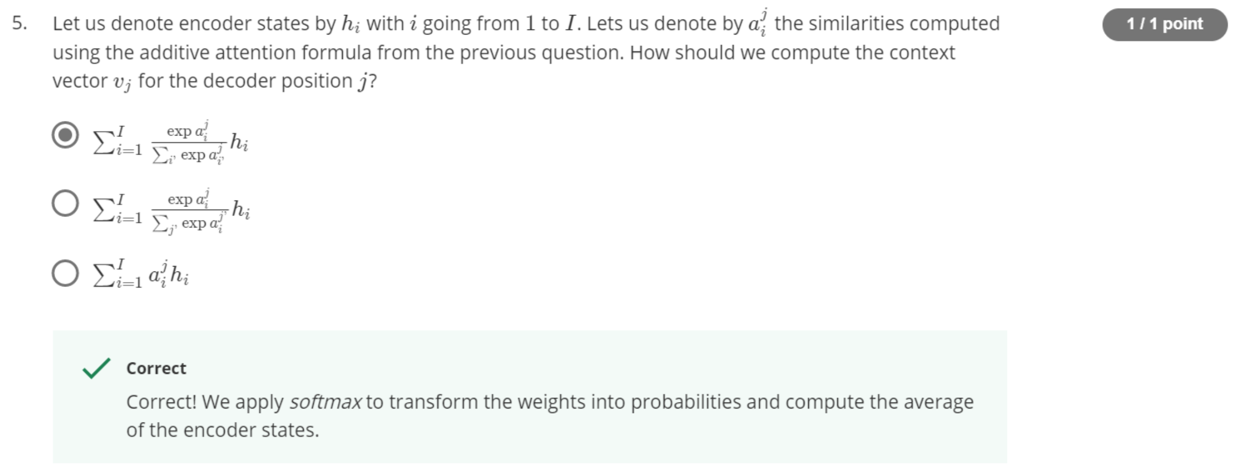
○ h_i^T s_j$h_i^T s_j$

● w^T tanh (W [h_i, s_j])$w^T tanh(W[h_i,s_j])$, where the brackets denote concatenation of the vectors, and ww and WW 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.

1 / 1 point



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 using the additive attention formula from the previous question. How should we compute the context vector v_j for the decoder position j?

1 / 1 point

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

○ $\sum_{i=1}^{I} \frac{\exp a_i^j}{\sum_{j'} \exp a_i^{j'}} h_i$

○ $\sum_{i=1}^{I} a_i^j h_i$

✓ **Correct**

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

1 / 1 point

6.Question 6

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

□ h_{i-1}h_{i-1} - the previous encoder state

□ h_ih_i - the ii-th encoder state

□ y_jy_j - the jj-th word in the target sequence

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

☑ y_{j-1}y_{j-1} - the previous word in the target sequence

Correct

☑ s_{j-1}s_{j-1} - the previous decoder state

Correct

☐ s_js_j - the jj-th decoder state

☑ v_jv_j - the context vector for position jj, calculated using attention

Correct

☐ x_ix_i - the ii-th word in the source sequence

1 / 1 point

7.Question 7

Which techniques would help if the data has rich morphology, informal spelling, and other sources of OOV tokens?

☐ Negative sampling

☑ Copy mechanism

Correct

☑ Byte-pair encoding

Correct

☑ Sub-word modeling

Correct

☐ Hierarchical softmax

1 / 1 point

8.Question 8

Let us imagine we have trained a conversational chat-bot as a seq2seq model on Harry Potter movies subtitles. What problems could we expect?

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

Correct

No personality problem!

☑ The bot doesn't remember what has already been decided in your dialogue

Correct

No memory of the context / history!

☑ 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!

☐ The bot makes lots of spelling mistakes