축하합니다! 통과하셨습니다!

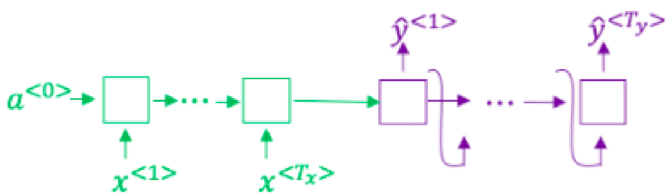
받은 학점 90% **최신 제출물 학점** 90% **통과 점수:** 80% 이상

다음 항목으로 이동

7 시간 57 분 후에 과제를 다시 풀어보세요.

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

1 / 1점



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.

- (•) False
- ( ) True

더 보기

맞습니다

2. In beam search, if you increase the beam width B, which of the following would you expect to be true?

1 / 1점

- ( ) Beam search will run more quickly.
- (•) Beam search will generally find better solutions (i.e. do a better job maximizing $P(y \mid x)$).
- ( ) Beam search will use up less memory.

Beam search will converge after fewer steps.

더 보기

맞습니다
As the beam width increases, beam search runs more slowly, uses up more memory, and converges after more steps, but generally finds better solutions.

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 / 1점

False

True

더 보기

맞습니다
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 / 1점

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?

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 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.
- ◉ 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.

더 보기

맞습니다

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 / 1점

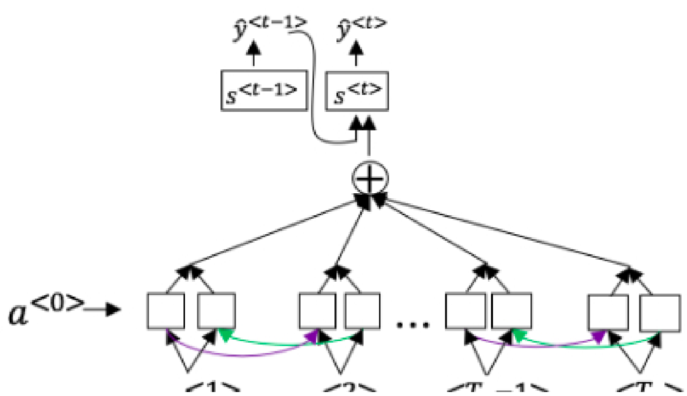
- ◉ True.
- ○ False.

더 보기

맞습니다

6. Consider the attention model for machine translation.

1 / 1점



$x^{<1>}$ $x^{<2>}$ $x^{<T_x-1>}$ $x^{<T_x>}$

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

Correct

Correct! If we sum over $\alpha^{<t,t'>}$ for all t' (the formulation can be seen in the image), the numerator will be equal to the denominator, therefore, $\sum_{t'} \alpha^{<t,t'>} = 1$.

- [ ] $\sum_{t'} \alpha^{<t,t'>} = 0$. (**Note the summation is over t'.**)
- [ ] 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] $\alpha^{<t,t'>}$ is equal to the amount of attention $y^{<t>}$ should pay to

더 보기

맞습니다

Great, you got all the right answers.

7. The network learns where to "pay attention" by learning the values $e^{<t,t'>}$, which are computed using a small neural network: 0 / 1점

We can replace $s^{<t-1>}$ with $s^{<t>}$ as an input to this neural network because $s^{<t>}$ is independent of $\alpha^{<t,t'>}$ and $e^{<t,t'>}$.

True

False

더 보기

틀립니다
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 and $e^{<t,t'>}$; so at the time we need to evaluate this network, we haven't computed $s^{<t>}$.

8. The attention model performs the same as the encoder-decoder model, no matter the sentence length.

1 / 1점

True

False

더 보기

맞습니다
The performance of the encoder-decoder model declines as the amount of words increases. The attention model has the greatest advantage when the input sequence length T_x is large.

9.

1 / 1점

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

- [x] keeper
- [ ] kkeeeeepeeeeeeeeerrrrr
- [ ] keper
- [ ] ke epe r

더 보기

맞습니다
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, $x^{<t>}$ is:

1 / 1점

- [x] Features of the audio (such as spectrogram features) at time t.
- [ ] Whether someone has just finished saying the trigger word at time t.
- [ ] The 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.

더 보기

맞습니다