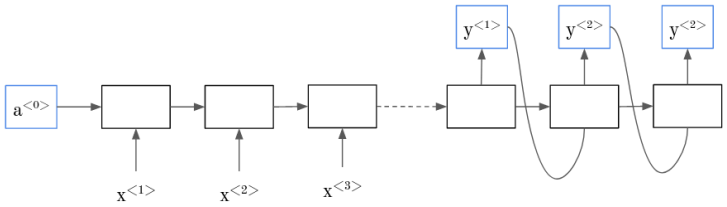
Week 3: Sequence models and Attention Mechanism

1. This model is a *conditional language model* in the sense that the encoder portion is modeling the probability of the input sentence x.



Ans: False

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

Ans:

- Beam search will run more slowly
- Beam search will take up more memory
- Beam search will find better solutions

3. In machine translation, if we carry out beam search without using sentence normalization, the algorithm will tend to output overly short translations.

Ans: True

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 $P(y|x)$. 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." We find by investigating, $P(\hat{y}|x) > P(y*|x)$

Ans: The error is expected in the RNN algorithm because the machine prediction is higher than human baseline. Beam search is not the expected cause of this anomaly.

5. 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 suggest you should focus your attention on improving the search algorithm

   Ans: True

6. Consider the attention model, Which of the following statements about $\alpha^{<t,t'>}$ are true?

   Ans:

   - We expect $\alpha < t, t' >$ to be generally larger for values of $a^{<t'>}$ that are highly relevant to the value the network should predict for $y^{<t>}$.
   - $\sum_{t'} \alpha_{t,t'} = 1$

7. The network learns where to pay attention by learning the values $e^{<t,t'>}$, which are computed using a small neural network. We can't replace $s^{<t-1>}$ with $s^{<t>}$ as an input because $s^{<t>}$ depends on $\alpha^{<t,t'>}$ which in turn depends on $e^{<t,t'>}$; so at the time we need to evaluate this network, we haven't computed $s^{<t>}$ yet.

   Ans: True

8. We expect the attention model to have the greatest advantage when:

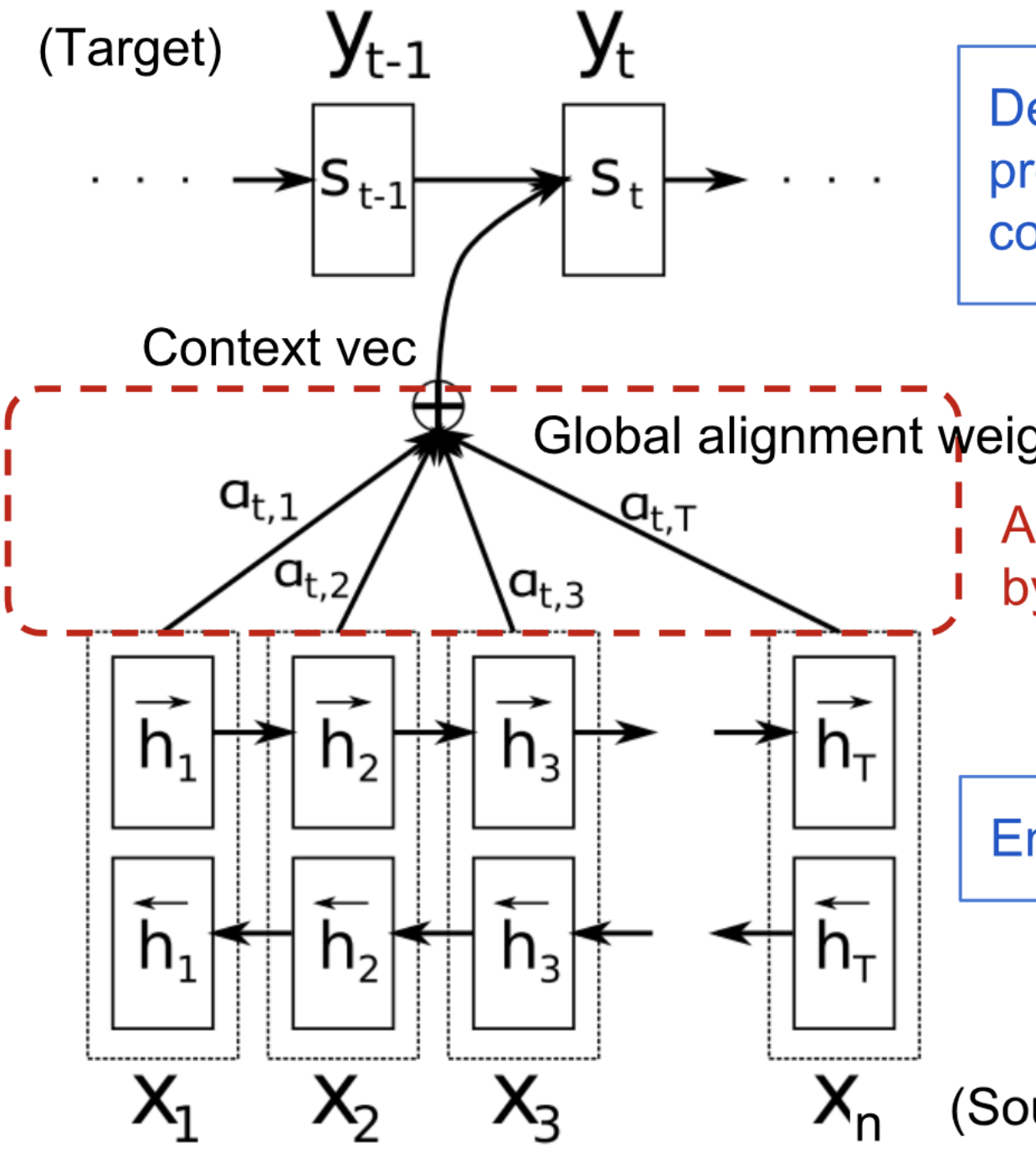
   Ans: Input sequence is long.

9. Under the CTC model, identical repeated characters not separated by the blank character are collapsed. What does "__c_oo_o_kk___b_ooooo__oo__kkk" collapse to?

   Ans: Cookbook

10. In trigger word detection, $x^{<t>}$ is,

    Ans: Features of audio, such as spectrogram


(Target)
y_{t-1}
y_t
s_{t-1}
s_t
Context vec
Global alignment weig
$a_{t,1}$
$a_{t,2}$
$a_{t,3}$
$a_{t,T}$
$\overrightarrow{h}_1$
$\overrightarrow{h}_2$
$\overrightarrow{h}_3$
$\overrightarrow{h}_T$
$\overleftarrow{h}_1$
$\overleftarrow{h}_2$
$\overleftarrow{h}_3$
$\overleftarrow{h}_T$
x_1
x_2
x_3
x_n
(Sou