축하합니다! 통과하셨습니다!

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

다음 항목으로 이동

1. Suppose your training examples are sentences (sequences of words). Which of the following refers to the s^{th}word in the r^{th}training example?

1/1점

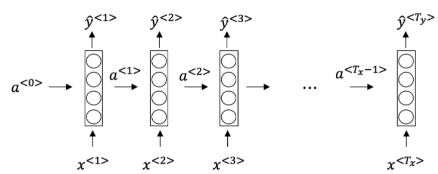
- ○ $x^{<r>(s)}$
- ○ $x^{<s>(r)}$
- ◉ $x^{(r)<s>}$
- ○ $x^{(s)<r>}$

더 보기

맞습니다
We index into the r^{th} row first to get to the r^{th} training example (represented by parentheses), then the s^{th} column to get to the s^{th} word (represented by the brackets).

2. Consider this RNN:

1/1점



True/False: This specific type of architecture is appropriate when Tx=Ty

- ○ False
- ◉ True

더 보기

맞습니다
It is appropriate when the input sequence and the output sequence have the same length or size.

3. Select the two tasks combination that could be addressed by a many-to-one RNN model architecture from the following:

1/1점

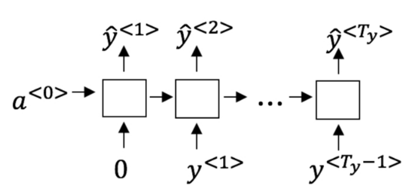
- ○ **Task 1:** Image classification. **Task 2:** Sentiment classification.
- ○ **Task 1:** Speech recognition. **Task 2:** Gender recognition from audio.
- ◉ **Task 1:** Gender recognition from audio. **Task 2:** Movie review (positive/negative) classification.
- ○ **Task 1:** Gender recognition from audio. **Task 2:** Image classification.

더 보기

맞습니다
Gender recognition from audio and movie review classification are two examples of many-to-one RNN architecture

4. Using this as the training model below, answer the following:

1/1점



True/False: At the t^{th} time step the RNN is estimating $P(y^{<t>} \mid y^{<1>}, y^{<2>}, \ldots, y^{<t-1>})$

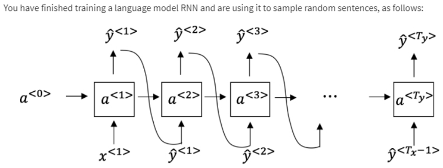
- ◉ True
- ○ False

더 보기

맞습니다
Yes, in a training model we try to predict the next step based on knowledge of all prior steps.

5. You have finished training a language model RNN and are using it to sample random sentences, as follows:

1/1점



What are you doing at each time step t?

- ○ (i) Use the probabilities output by the RNN to pick the highest probability word for that time-step as $\hat{y}^{<t>}$. (ii) Then pass the ground-truth word from the training set to the next time-step.
- ○ (i) Use the probabilities output by the RNN to randomly sample a chosen word for that time-step as $\hat{y}^{<t>}$.(ii) Then pass the ground-truth word from the training set to the next time-step.
- ○ (i) Use the probabilities output by the RNN to pick the highest probability word for that

time-step as $\hat{y}^{<t>}$.(ii) Then pass this selected word to the next time-step.

- (i) Use the probabilities output by the RNN to randomly sample a chosen word for that time-step as $\hat{y}^{<t>}$.(ii) Then pass this selected word to the next time-step.

더 보기

맞습니다

6. True/False: If you are training an RNN model, and find that your weights and activations are all taking on the value of NaN ("Not a Number") then you have a vanishing gradient problem. 1 / 1점

- False
- True

더 보기

맞습니다
Vanishing and exploding gradients are common problems in training RNNs, but in this case, your weights and activations taking on the value of NaN implies you have an exploding gradient problem.

7. Suppose you are training an LSTM. You have a 10000 word vocabulary, and are using an LSTM with 100-dimensional activations $a^{<t>}$. What is the dimension of Γ_u at each time step? 1 / 1점

- 1
- 100
- 300
- 10000

더 보기

맞습니다
Correct, Γ_u is a vector of dimension equal to the number of hidden units in the LSTM.

8. True/False: In order to simplify the GRU without vanishing gradient problems even when training on very long sequences you should remove the Γ_r i.e., setting $\Gamma_r = 1$ always. 1 / 1점

- False
- True

더 보기

맞습니다
If Γu≈0 for a timestep, the gradient can propagate back through that timestep without much decay. For the signal to backpropagate without vanishing, we need $c^{<t>}$ to be highly dependent on $c^{<t-1>}$.

9. True/False: Using the equations for the GRU and LSTM below the Update Gate and Forget Gate in the LSTM play a role similar to 1- Γu and Γu. 1 / 1점

GRU

$$\tilde{c}^{<t>} = \tanh(W_c[\Gamma_r * c^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[\, c^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_r = \sigma(W_r[\, c^{<t-1>}, x^{<t>}] + b_r)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$$

$$a^{<t>} = c^{<t>}$$

LSTM

$$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[\, a^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_f = \sigma(W_f[\, a^{<t-1>}, x^{<t>}] + b_f)$$

$$\Gamma_o = \sigma(W_o[\, a^{<t-1>}, x^{<t>}] + b_o)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$$

$$a^{<t>} = \Gamma_o * c^{<t>}$$

- False
- True

더 보기

맞습니다
Instead of using Γu to compute 1 - Γu, LSTM uses 2 gates (Γu and Γf) to compute the final value of the hidden state. So, Γf is used instead of 1 - Γu.

10. True/False: You would use unidirectional RNN if you were building a model map to show how your mood is heavily dependent on the current and past few days' weather. 0 / 1점

- True
- False

더 보기

틀립니다
Your mood is contingent on the current and past few days' weather, not on the current, past, AND future days' weather.