Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

Go to next item

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

1 / 1 point

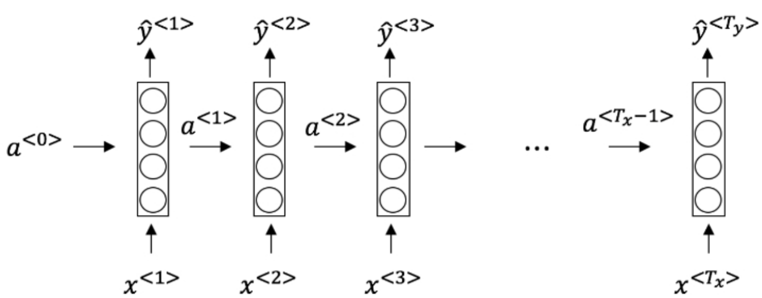
- ○ $x^{<k>(l)}$
- ◉ $x^{(k)<l>}$
- ○ $x^{(l)<k>}$
- ○ $x^{<l>(k)}$

Expand

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

2. Consider this RNN:

1 / 1 point



This specific type of architecture is appropriate when:

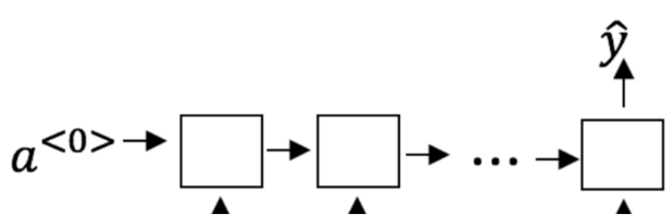
- ◉ $T_x = T_y$
- ○ $T_x < T_y$
- ○ $T_x > T_y$
- ○ $T_x = 1$

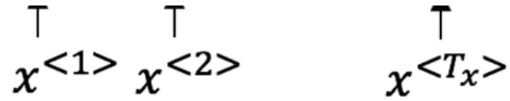
Expand

Correct
It is appropriate when every input should have an output.

3. To which of these tasks would you apply a many-to-one RNN architecture?

1 / 1 point





- [ ] Image classification (input an image and output a label)
- [x] Music genre recognition

> ✓ **Correct**
> This is an example of many-to-one architecture.

- [x] Language recognition from speech (input an audio clip and output a label indicating the language being spoken)

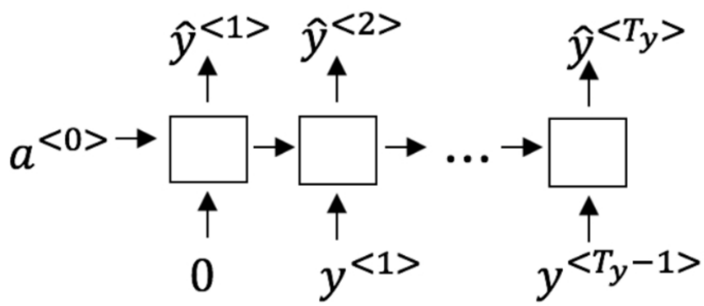
> ✓ **Correct**
> This is an example of many-to-one architecture.

- [ ] Speech recognition (input an audio clip and output a transcript)

Expand

Correct
Great, you got all the right answers.

4. Using this as the training model below, answer the following: **1 / 1 point**



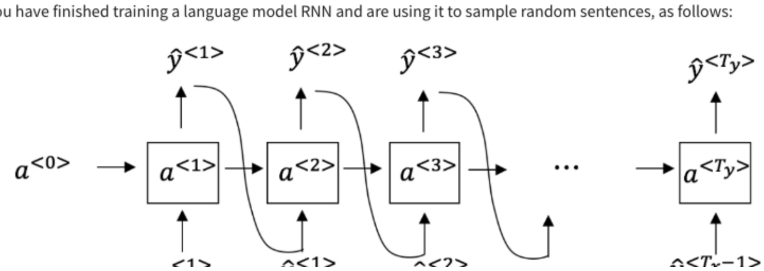
True/False: At the t^{th} time step the RNN is estimating $P(y^{<t>})$

- ( ) True
- (•) False

Expand

Correct
No, in a training model we try to predict the next steps based on the 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 / 1 point**



True/False: In this sample sentence, step t uses the probabilities output by the RNN to randomly sample a chosen word for that time-step. Then it passes this selected word to the next time-step.

- (•) True
- ( ) False

Expand

Correct
Step t uses the probabilities output by the RNN to randomly sample a chosen word for that time-step. Then it passes this selected word to the next time-step.

6. You are training an RNN model, and find that your weights and activations are all taking on the value of NaN ("Not a Number"). Which of these is the most likely cause of this problem?

1 / 1 point

- ( ) Vanishing gradient problem.
- (•) Exploding gradient problem.
- ( ) The model used the ReLU activation function to compute g(z), where z is too large.
- ( ) The model used the Sigmoid activation function to compute g(z), where z is too large.

Expand

Correct

7. Suppose you are training an LSTM. You have a 50000 word vocabulary, and are using an LSTM with 500-dimensional activations $a^{<t>}$. What is the dimension of Γ_u at each time step?

1 / 1 point

- (•) 500
- ( ) 50000
- ( ) 200
- ( ) 5

Expand

Correct
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 / 1 point

- (•) True
- ( ) False

Expand

Correct

If $\Gamma_u \approx 0$ 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. Here are the equations for the GRU and the LSTM:

1 / 1 point

GRU	LSTM
$\tilde{c}^{<t>} = \tanh(W_c[\Gamma_r * c^{<t-1>}, x^{<t>}] + b_c)$	$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c)$
$\Gamma_u = \sigma(W_u[\, c^{<t-1>}, x^{<t>}] + b_u)$	$\Gamma_u = \sigma(W_u[\, a^{<t-1>}, x^{<t>}] + b_u)$
$\Gamma_r = \sigma(W_r[\, c^{<t-1>}, x^{<t>}] + b_r)$	$\Gamma_f = \sigma(W_f[\, a^{<t-1>}, x^{<t>}] + b_f)$
$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$	$\Gamma_o = \sigma(W_o[\, a^{<t-1>}, x^{<t>}] + b_o)$
$a^{<t>} = c^{<t>}$	$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$
	$a^{<t>} = \Gamma_o * c^{<t>}$

From these, we can see that the Update Gate and Forget Gate in the LSTM play a role similar to _______ and _______ in the GRU. What should go in the blanks?

- (selected) Γ_u and $1 - \Gamma_u$
- Γ_u and Γ_r
- $1 - \Gamma_u$

  and

  Γ_u

Expand

Correct

Yes, correct!

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.

1 / 1 point

- (selected) True
- False

Expand

Correct

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