**[Skip to Main Content](https://www.coursera.org/learn/nlp-sequence-models/exam/e4bJR/recurrent-neural-networks/view-attempt" \l "main)**

[[](https://www.coursera.org/)](https://www.coursera.org/)

Explore

Top of Form



Bottom of Form

* **1**
* Vijay Misra

1. [Sequence Models](https://www.coursera.org/learn/nlp-sequence-models/home/welcome)
2. [Week 1](https://www.coursera.org/learn/nlp-sequence-models/home/week/1)
3. Recurrent Neural Networks

[**Previous**](https://www.coursera.org/learn/nlp-sequence-models/supplement/TsCLx/lectures-in-pdf)[**Next**](https://www.coursera.org/learn/nlp-sequence-models/supplement/X39s5/how-to-download-your-notebook)

* **Recurrent Neural Networks**
* **Lecture Notes (Optional)**
* **Quiz**

**[Quiz:](https://www.coursera.org/learn/nlp-sequence-models/exam/e4bJR/recurrent-neural-networks)**[Recurrent Neural Networks](https://www.coursera.org/learn/nlp-sequence-models/exam/e4bJR/recurrent-neural-networks)

[10 questions](https://www.coursera.org/learn/nlp-sequence-models/exam/e4bJR/recurrent-neural-networks)

* **Programming Assignments**

**QUIZQuiz • 30 MIN30 minutes**

**Recurrent Neural Networks**

**Submit your assignment**

**DUE DATE**Jun 20, 11:59 PM PDTJune 20, 11:59 PM PDT

**ATTEMPTS**3 every 8 hours

Try again

**Receive grade**

**TO PASS**80% or higher

**Grade**

100%

View Feedback

We keep your highest score

Recurrent Neural Networks

Graded Quiz • 30 min

**Due** Jun 20, 11:59 PM PDT

**Congratulations! You passed!**

**TO PASS**80% or higher

Keep Learning

**GRADE**

100%

**Recurrent Neural Networks**

**LATEST SUBMISSION GRADE**

100%

1.

Question 1

Suppose your training examples are sentences (sequences of words). Which of the following refers to the j^{th}*jth* word in the i^{th}*ith* training example?

**1 / 1 point**



**x^{(i)<j>}*x*(*i*)<*j*>**



**x^{<i>(j)}*x*<*i*>(*j*)**



**x^{(j)<i>}*x*(*j*)<*i*>**



**x^{<j>(i)}*x*<*j*>(*i*)**

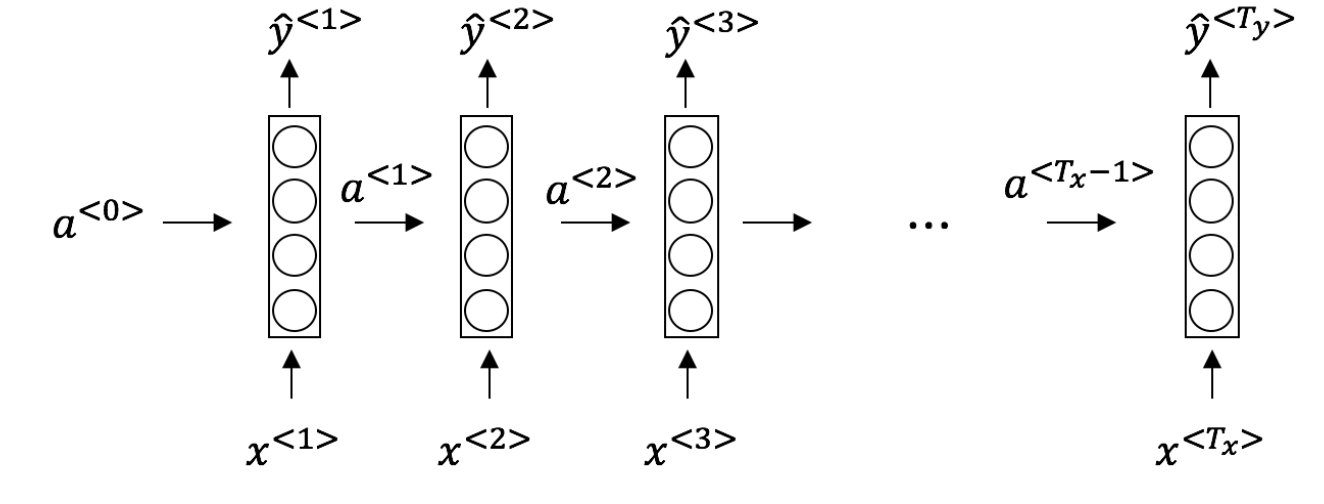
**Correct**

We index into the i^{th}*ith* row first to get the i^{th}*ith* training example (represented by parentheses), then the j^{th}*jth* column to get the j^{th}*jth* word (represented by the brackets).

2.

Question 2

Consider this RNN:



This specific type of architecture is appropriate when:

**1 / 1 point**



T\_x = T\_y*Tx*​=*Ty*​



T\_x < T\_y*Tx*​<*Ty*​



T\_x > T\_y*Tx*​>*Ty*​



T\_x = 1*Tx*​=1

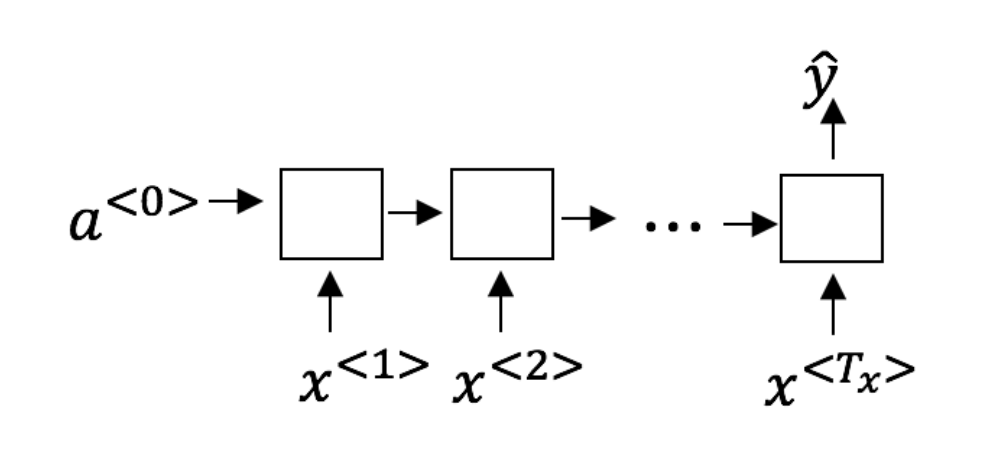
**Correct**

It is appropriate when every input should be matched to an output.

3.

Question 3

To which of these tasks would you apply a many-to-one RNN architecture? (Check all that apply).



**1 / 1 point**



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



Sentiment classification (input a piece of text and output a 0/1 to denote positive or negative sentiment)

**Correct**

Correct!



Image classification (input an image and output a label)



Gender recognition from speech (input an audio clip and output a label indicating the speaker’s gender)

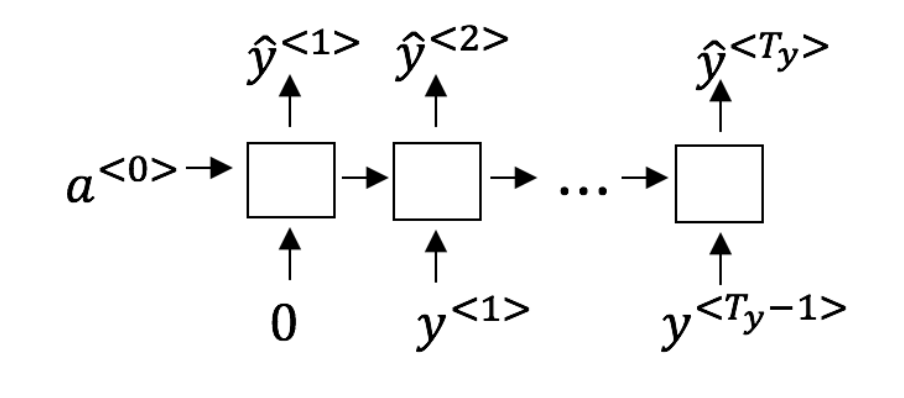
**Correct**

Correct!

4.

Question 4

You are training this RNN language model.



At the t^{th}*tth* time step, what is the RNN doing? Choose the best answer.

**1 / 1 point**



Estimating P(y^{<1>}, y^{<2>}, …, y^{<t-1>})*P*(*y*<1>,*y*<2>,…,*y*<*t*−1>)



Estimating P(y^{<t>})*P*(*y*<*t*>)



Estimating P(y^{<t>} \mid y^{<1>}, y^{<2>}, …, y^{<t-1>})*P*(*y*<*t*>∣*y*<1>,*y*<2>,…,*y*<*t*−1>)



Estimating P(y^{<t>} \mid y^{<1>}, y^{<2>}, …, y^{<t>})*P*(*y*<*t*>∣*y*<1>,*y*<2>,…,*y*<*t*>)

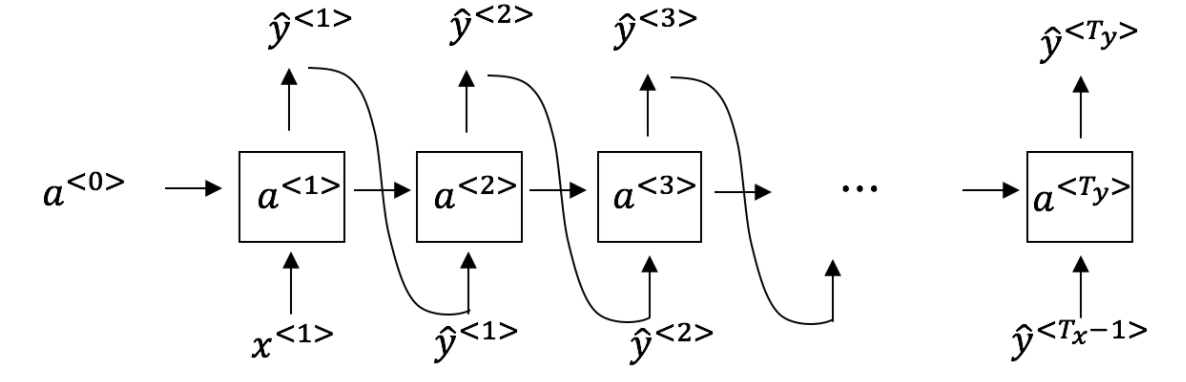
**Correct**

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

5.

Question 5

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



What are you doing at each time step t*t*?

**1 / 1 point**



(i) Use the probabilities output by the RNN to pick the highest probability word for that time-step as \hat{y}^{<t>}*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>}*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>}*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>}*y*^​<*t*>.(ii) Then pass this selected word to the next time-step.

**Correct**

6.

Question 6

You are training an RNN, 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.



ReLU activation function g(.) used to compute g(z), where z is too large.



Sigmoid activation function g(.) used to compute g(z), where z is too large.

**Correct**

7.

Question 7

Suppose you are training a LSTM. You have a 10000 word vocabulary, and are using an LSTM with 100-dimensional activations a^{<t>}*a*<*t*>. What is the dimension of \Gamma\_uΓ*u*​ at each time step?

**1 / 1 point**



1



100



300



10000

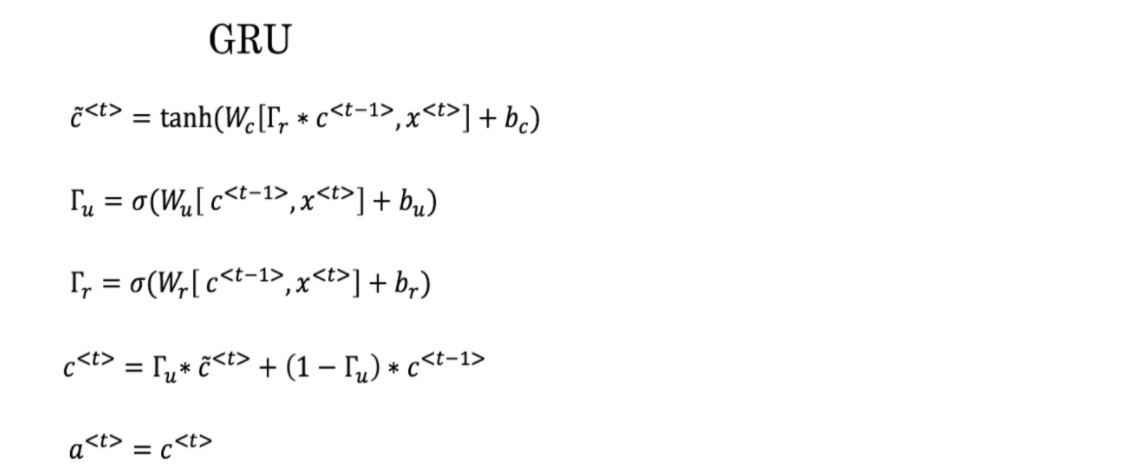
**Correct**

Correct, \Gamma\_uΓ*u*​ is a vector of dimension equal to the number of hidden units in the LSTM.

8.

Question 8

Here’re the update equations for the GRU.



Alice proposes to simplify the GRU by always removing the \Gamma\_uΓ*u*​. I.e., setting \Gamma\_uΓ*u*​ = 1. Betty proposes to simplify the GRU by removing the \Gamma\_rΓ*r*​. I. e., setting \Gamma\_rΓ*r*​ = 1 always. Which of these models is more likely to work without vanishing gradient problems even when trained on very long input sequences?

**1 / 1 point**



Alice’s model (removing \Gamma\_uΓ*u*​), because if \Gamma\_r \approx 0Γ*r*​≈0 for a timestep, the gradient can propagate back through that timestep without much decay.



Alice’s model (removing \Gamma\_uΓ*u*​), because if \Gamma\_r \approx 1Γ*r*​≈1 for a timestep, the gradient can propagate back through that timestep without much decay.



Betty’s model (removing \Gamma\_rΓ*r*​), because if \Gamma\_u \approx 0Γ*u*​≈0 for a timestep, the gradient can propagate back through that timestep without much decay.



Betty’s model (removing \Gamma\_rΓ*r*​), because if \Gamma\_u \approx 1Γ*u*​≈1 for a timestep, the gradient can propagate back through that timestep without much decay.

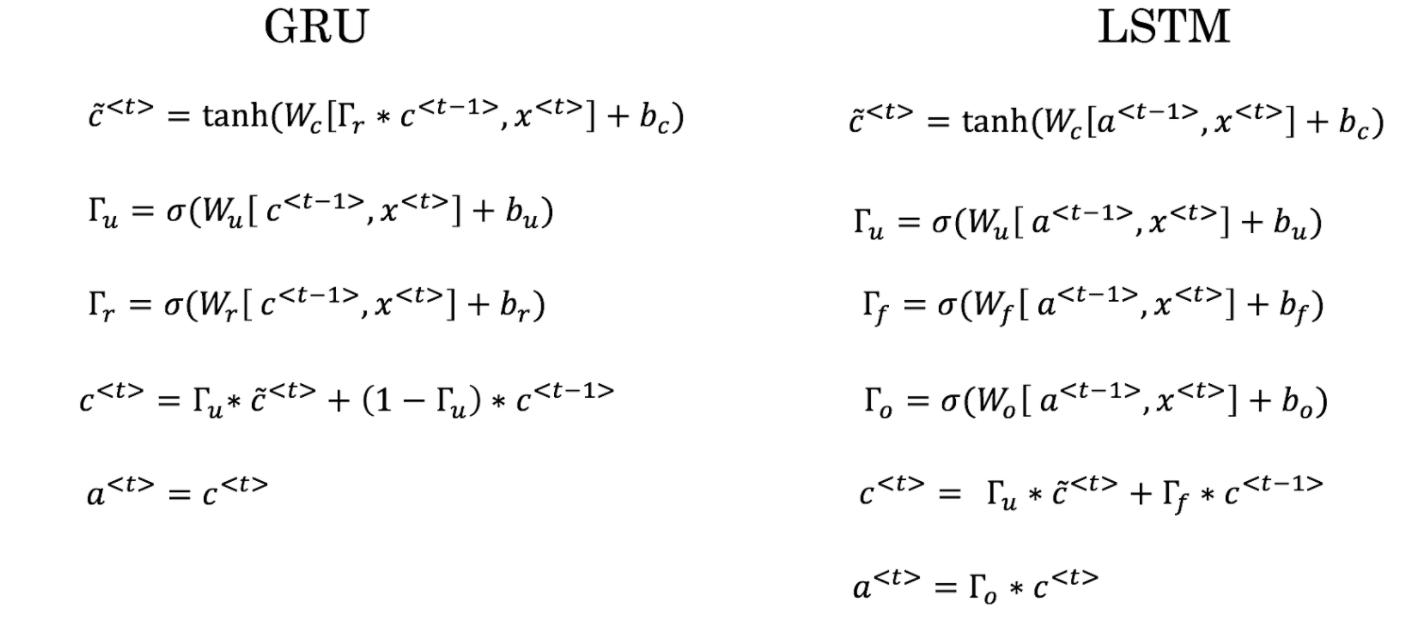
**Correct**

Yes. For the signal to backpropagate without vanishing, we need c^{<t>}*c*<*t*> to be highly dependent on c^{<t-1>}*c*<*t*−1>.

9.

Question 9

Here are the equations for the GRU and the LSTM:



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?

**1 / 1 point**



\Gamma\_uΓ*u*​ and 1-\Gamma\_u1−Γ*u*​



\Gamma\_uΓ*u*​ and \Gamma\_rΓ*r*​



1-\Gamma\_u1−Γ*u*​ and \Gamma\_uΓ*u*​



\Gamma\_rΓ*r*​ and \Gamma\_uΓ*u*​

**Correct**

Yes, correct!

10.

Question 10

You have a pet dog whose mood is heavily dependent on the current and past few days’ weather. You’ve collected data for the past 365 days on the weather, which you represent as a sequence as x^{<1>}, …, x^{<365>}*x*<1>,…,*x*<365>. You’ve also collected data on your dog’s mood, which you represent as y^{<1>}, …, y^{<365>}*y*<1>,…,*y*<365>. You’d like to build a model to map from x \rightarrow y*x*→*y*. Should you use a Unidirectional RNN or Bidirectional RNN for this problem?

**1 / 1 point**



Bidirectional RNN, because this allows the prediction of mood on day t to take into account more information.



Bidirectional RNN, because this allows backpropagation to compute more accurate gradients.



Unidirectional RNN, because the value of y^{<t>}*y*<*t*> depends only on x^{<1>}, …, x^{<t>}*x*<1>,…,*x*<*t*>, but not on x^{<t+1>}, …, x^{<365>}*x*<*t*+1>,…,*x*<365>



Unidirectional RNN, because the value of y^{<t>}*y*<*t*> depends only on x^{<t>}*x*<*t*>, and not other days’ weather.

**Correct**

Yes!