Back

Recurrent Neural Networks

Quiz, 10 questions

8/10 points (80%)

**Congratulations! You passed!**

Next Item

loading   https://d3njjcbhbojbot.cloudfront.net/web/images/icons/loading.gif

Load Error!

Question 1

*Correct*

1 / 1 points

1. Question 1

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



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

**Correct**

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



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



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



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

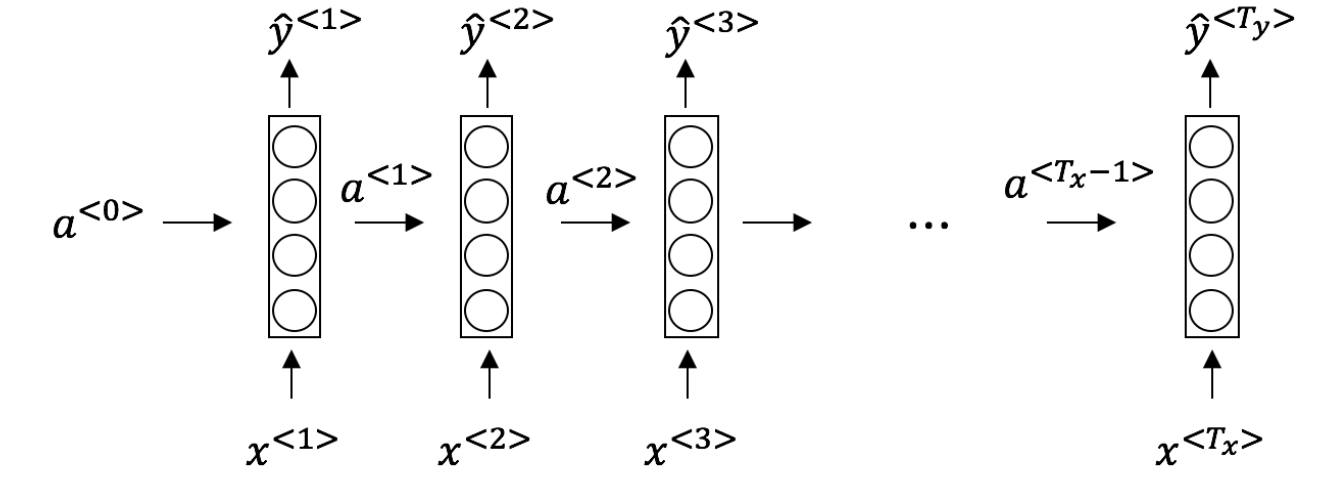
Question 2

*Correct*

1 / 1 points

2. Question 2

Consider this RNN:



This specific type of architecture is appropriate when:



**Tx=TyT\_x = T\_yTx​=Ty​**

**Correct**

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



**Tx<TyT\_x < T\_yTx​<Ty​**



**Tx>TyT\_x > T\_yTx​>Ty​**



**Tx=1T\_x = 1Tx​=1**

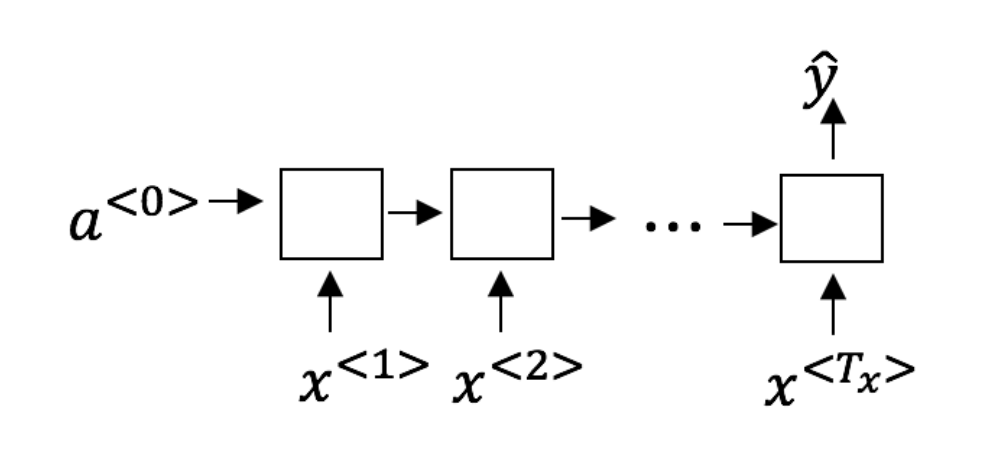
Question 3

*Correct*

1 / 1 points

3. Question 3

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





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

**Un-selected is correct**



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)

**Un-selected is correct**



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

**Correct**

Correct!

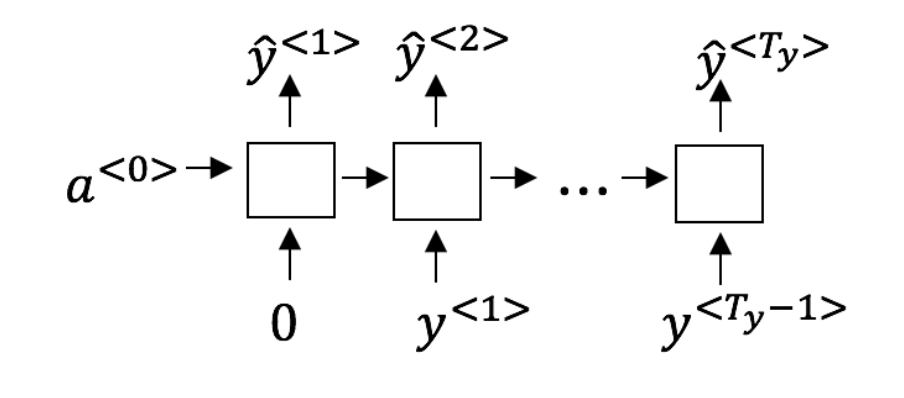
Question 4

*Correct*

1 / 1 points

4. Question 4

You are training this RNN language model.



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



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



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



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

**Correct**

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



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

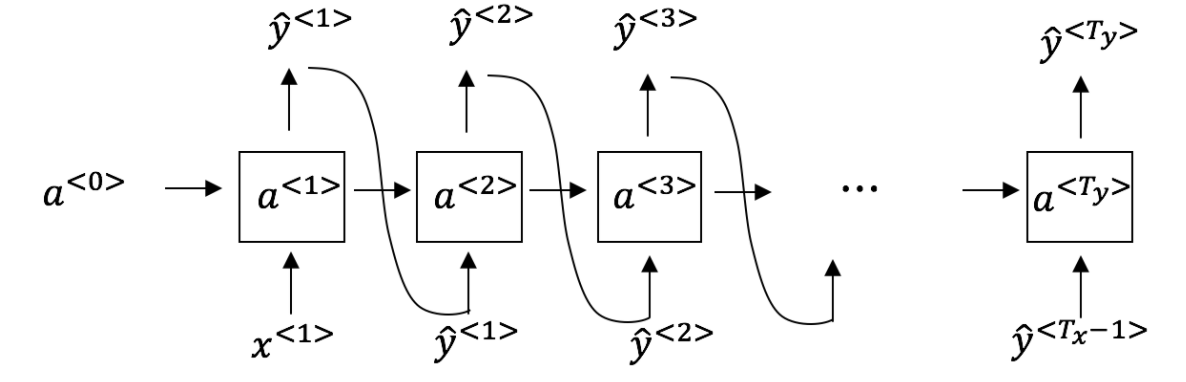
Question 5

*Incorrect*

0 / 1 points

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 ttt?



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

**This should not be selected**

The probabilities output by the RNN are not used to pick the highest probability word.



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

Question 6

*Correct*

1 / 1 points

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?



Vanishing gradient problem.



Exploding gradient problem.

**Correct**



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.

Question 7

*Correct*

1 / 1 points

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>}a<t>. What is the dimension of Γu\Gamma\_uΓu​ at each time step?



1



100

**Correct**

Correct, Γu\Gamma\_uΓu​ is a vector of dimension equal to the number of hidden units in the LSTM.



300



10000

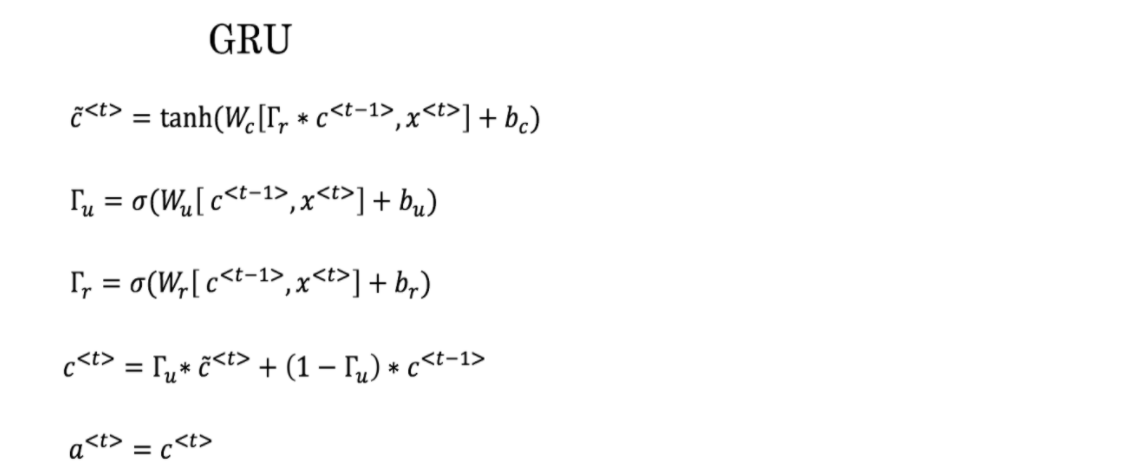
Question 8

*Incorrect*

0 / 1 points

8. Question 8

Here’re the update equations for the GRU.



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



Alice’s model (removing Γu\Gamma\_uΓu​), because if Γr≈0\Gamma\_r \approx 0Γr​≈0 for a timestep, the gradient can propagate back through that timestep without much decay.



Alice’s model (removing Γu\Gamma\_uΓu​), because if Γr≈1 \Gamma\_r \approx 1Γr​≈1 for a timestep, the gradient can propagate back through that timestep without much decay.

**This should not be selected**

No. For the signal to backpropagate without vanishing, we need c<t>c^{<t>}c<t> to be highly dependant on c<t−1>c^{<t-1>}c<t−1>.



Betty’s model (removing Γr\Gamma\_rΓr​), because if Γu≈0\Gamma\_u \approx 0Γu​≈0 for a timestep, the gradient can propagate back through that timestep without much decay.



Betty’s model (removing Γr\Gamma\_rΓr​), because if Γu≈1\Gamma\_u \approx 1Γu​≈1 for a timestep, the gradient can propagate back through that timestep without much decay.

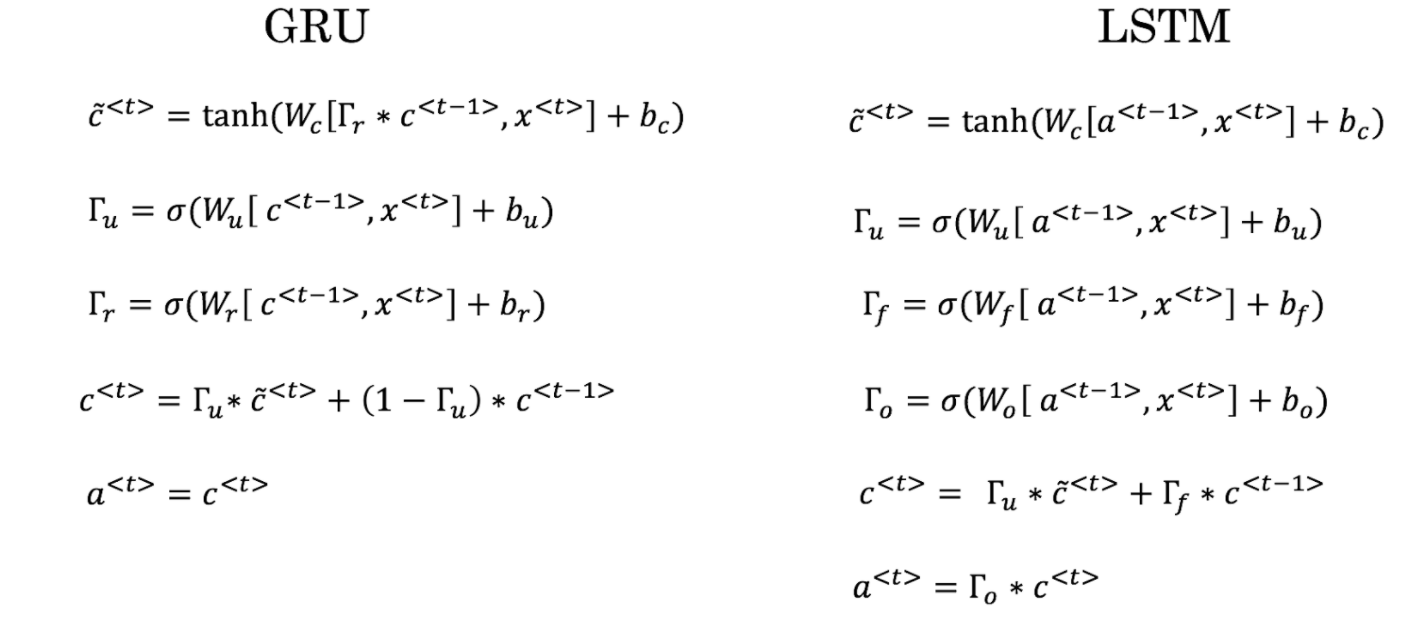
Question 9

*Correct*

1 / 1 points

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 the blanks?



Γu\Gamma\_uΓu​ and 1−Γu1-\Gamma\_u1−Γu​

**Correct**

Yes, correct!



Γu\Gamma\_uΓu​ and Γr\Gamma\_rΓr​



1−Γu1-\Gamma\_u1−Γu​ and Γu\Gamma\_uΓu​



Γr\Gamma\_rΓr​ and Γu\Gamma\_uΓu​

Question 10

*Correct*

1 / 1 points

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>}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>}y<1>,…,y<365>. You’d like to build a model to map from x→yx \rightarrow yx→y. Should you use a Unidirectional RNN or Bidirectional RNN for this problem?



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>}y<t> depends only on x<1>,…,x<t>x^{<1>}, …, x^{<t>}x<1>,…,x<t>, but not on x<t+1>,…,x<365>x^{<t+1>}, …, x^{<365>}x<t+1>,…,x<365>

**Correct**

Yes!



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



I, **James John Hendrik Piggott**, understand that submitting work that isn’t my own may result in permanent failure of this course or deactivation of my Coursera account. Learn more about Coursera’s Honor Code

Submitting...Submit error! Please try again.Submit QuizSubmit Quiz

-->

Confirm Navigation

Are you sure you want to leave this page?

Stay on this Page  Leave this Page

Confirm Navigation

Are you sure you want to leave this page?

Stay on this Page  Leave this Page