Congratulations! You passed!

Grade received 90% **Latest Submission Grade** 90% **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 j^{th} word in the i^{th} training example?

1 / 1 point

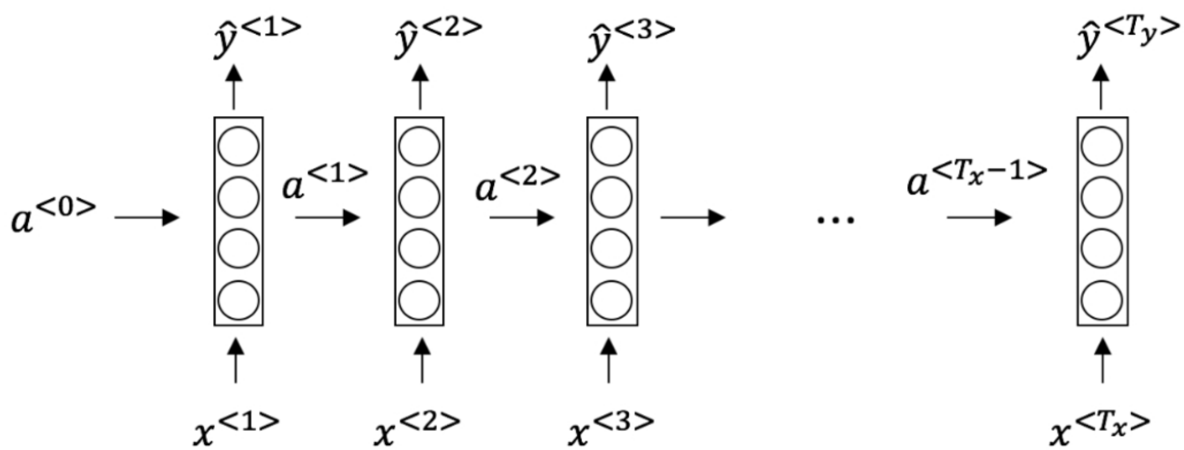
- ◉ $x^{(i)<j>}$
- ○ $x^{<i>(j)}$
- ○ $x^{(j)<i>}$
- ○ $x^{<j>(i)}$

Expand

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

2. Consider this RNN:

1 / 1 point



This specific type of architecture is appropriate when:

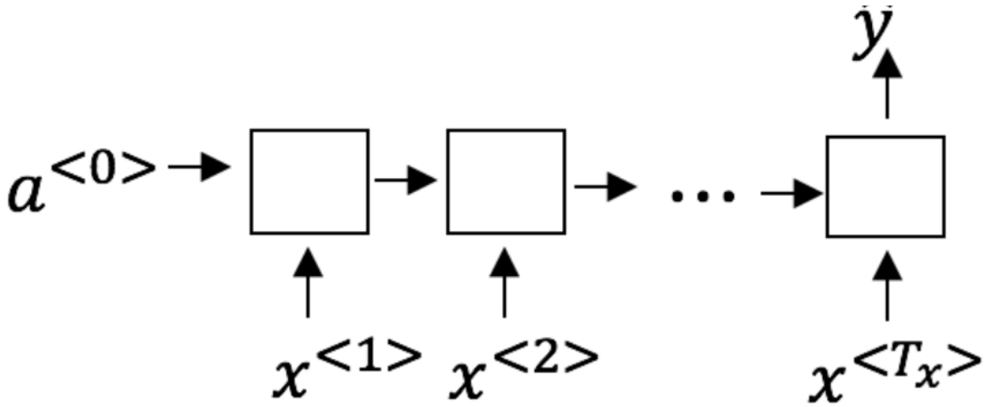
- ◉ $T_x = T_y$
- ○ $T_x < T_y$
- ○ $T_x > T_y$
- ○ $T_x = 1$

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? (Check all that apply).

0 / 1 point



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

  ! This should not be selected
  This is an example of many-to-many architecture.

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

  ✓ Correct
  Correct!

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

  ! This should not be selected
  This is an example of one-to-one architecture.

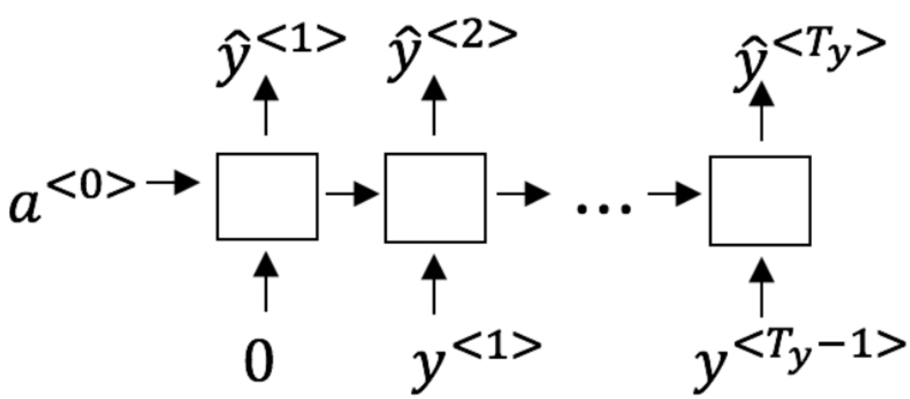
- [x] Gender recognition from speech (input an audio clip and output a label indicating the speaker's gender)

  ✓ Correct
  Correct!

Expand

Incorrect
You chose the extra incorrect answers.

4. You are training this RNN language model. **1 / 1 point**



At the t^{th} time step, what is the RNN doing?

- ○ Estimating $P(y^{<1>}, y^{<2>}, \ldots, y^{<t-1>})$

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

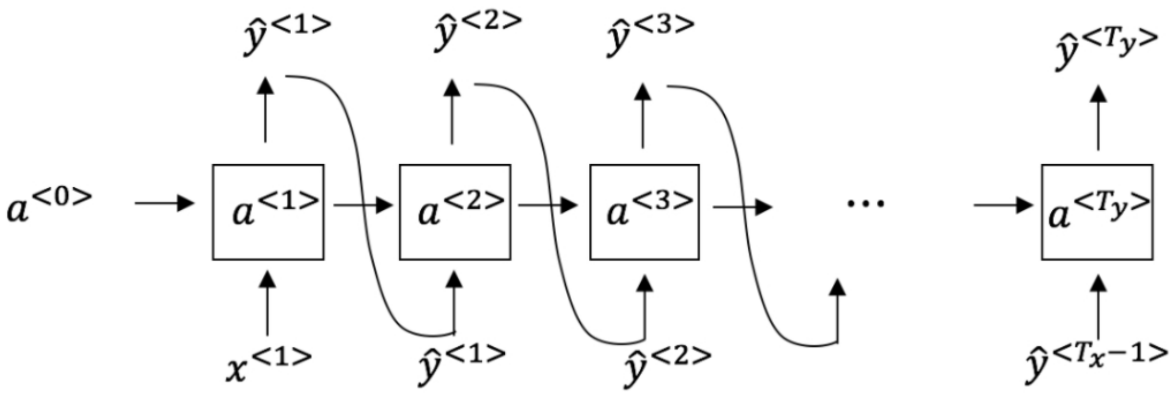
Estimating $P(y^{<t>} \mid y^{<1>}, y^{<2>}, \ldots, y^{<t-1>})$

Expand

Correct
Yes, in a language model we try to predict the next step 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 pick the highest probability word for that time-step. Then it passes the ground-truth word from the training set to the next time-step.

True

False

Expand

Correct
The probabilities output by the RNN are not used to pick the highest probability word and the ground-truth word from the training set is not the input to the next time-step.

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 / 1 point

False

True

Expand

Correct
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 / 1 point

- 1
- 100
- 300
- 10000

Expand

Correct
Correct, $$\Gamma_u$$ 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 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. Here are the equations for the GRU and the LSTM: **1 / 1 point**

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

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

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

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

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

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

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

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?

- Γ_u and $1 - \Gamma_u$
- Γ_u and Γ_r
- $1 - \Gamma_u$ and Γ_u
- Γ_r 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

- ○ False
- ◉ True

Expand

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