Congratulations! You passed!

Grade received 100% To pass 80% or higher

Go to next item

Recurrent Neural Networks

Latest Submission Grade 100%

1. Suppose your training examples are sentences (sequences of words). Which of the following refers to the $a^{[3]}$ word in the i^{th} training example?

1 / 1 point

- $x^{(i)<3>}$
- $x^{<i>(3)}$
- $x^{(3)<i>}$
- $x^{<3>(i)}$

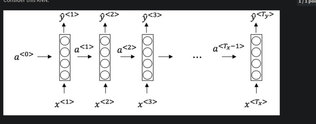
Expand

Correct

We index into the i^{th} row first to get to the i^{th} training example (represented by parentheses), then the 3^{th} column to get to the 3^{th} word (represented by the brackets).

2. Consider this RNN:

1 / 1 point



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

- False
- True

Expand

Correct

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 / 1 point

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

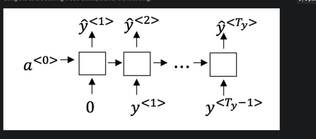
Expand

Correct

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 / 1 point



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

- False
- True

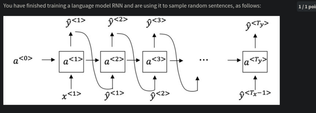
Expand

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 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. 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

- 200
- 5
- 500
- 50000

Expand

Correct

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

8. Sarah proposes to simplify the GRU by always removing the Γ_u i.e., setting $\Gamma_u = 0$. Ashely proposes to simplify the GRU by removing the Γ_r i.e., setting $\Gamma_r = 1$ 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

GRU

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

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

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

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

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

- Sarah's model (removing Γ_{r}), because if $\Gamma_{r} \approx 0$ for a timestep, the gradient can propagate back through that timestep without much decay.
- Ashely's model (removing Γ_{r}), because if $\Gamma_{r} \approx 0$ for a timestep, the gradient can propagate back through that timestep without much decay.
- Ashely's model (removing Γ_{r}), because if $\Gamma_{r} \approx 1$ for a timestep, the gradient can propagate back through that timestep without much decay.
- Sarah's model (removing Γ_{r}), because if $\Gamma_{r} \approx 1$ for a timestep, the gradient can propagate back through that timestep without much decay.

Expand

Correct

Yes. 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 $1-\Gamma_u$ and Γ_u.

1 / 1 point

GRU

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

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

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

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

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

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

Expand

Correct

Instead of using Γ_u to compute $1 - \Gamma_u$, LSTM uses 2 gates (Γ_u and Γ_f) to compute the final value of the hidden state. So, Γ_f is used instead of $1 - \Gamma_u$.

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.