Recurrent Neural Networks

测验, 10 questions

1 point

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?



 $\bigcirc \quad x^{< i > (j)}$

 $\bigcirc \quad x^{(j) < i >}$

 $\bigcirc \quad x^{< j > (i)}$

1 point

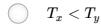
2.

Consider this RNN:

This secific type of architecture is appropriate when:



$$T_x = T_y$$



$$\bigcap T_x > T_y$$

$$T_x = 1$$

1 point

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)

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

Recurrent Neural Networks

测验, 10 questions

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)

1 point

4

You are training this RNN language model.

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

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

1 point

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?

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

Recurrent Neural Networks

\I- A		
测验.	10	auestions

point

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.

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.

1 point

7.

Suppose you are training a 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?



300

10000

1 point

8.

Here're the update equations for the GRU.

Alice proposes to simplify the GRU by always removing the Γ_u . I.e., setting Γ_u = 1. Betty proposes to simplify the GRU by removing the Γ_r . I. e., setting Γ_r = 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 Γ_u), because if $\Gamma_r \approx 0$ for a timestep, the gradient can propagate back through that timestep without much decay.

Alice's model (removing Γ_u), because if $\Gamma_r \approx 1$ for a timestep, the gradient can propagate back through that timestep without much

Recurrent Neural Networks

测验, 10 questions



Betty's model (removing Γ_r), because if $\Gamma_u \approx 0$ for a timestep, the gradient can propagate back through that timestep without much decay.

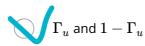
Betty's model (removing Γ_r), because if $\Gamma_u \approx 1$ for a timestep, the gradient can propagate back through that timestep without much decay.

1 point

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?



 \bigcap Γ_u and Γ_r

 \bigcap $1-\Gamma_u$ and Γ_u

 \bigcap Γ_r and Γ_u

1 point

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 $x^{<1>},\dots,x^{<365>}$. You've also collected data on your dog's mood, which you represent as $y^{<1>},\dots,y^{<365>}$. You'd like to build a model to map from $x\to y$. 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 $y^{< t>}$ depends only on $x^{< 1>}, \dots, x^{< t>}$, but not on $x^{< t+1>}, \dots, x^{< 365>}$

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

Recurrent Neural Networks

测验, 10 questions

升级后提交

