테스트테스트 • 30 min30 minutes

Recurrent Neural Networks

과제 제출
기한년 8월 23일 오후 3:59 KST년 8월 23일 오후 3:59 KST
시도하기8 hours당 3회

계속하기

성적 받기
통과 점수:80% 이상
성적
100%

피드백 보기

최고 점수가 유지됩니다.

탐색 확인

이 페이지에서 나가시겠습니까?

이 페이지에 머물기 이 페이지에서 나가기

Recurrent Neural Networks
성적 평가 퀴즈 • 30 min

만료 년 8월 23일 오후 3:59 KST

축하합니다! 통과하셨습니다!

통과 점수: 80% 이상

학습 계속하기

성적

100%

Recurrent Neural Networks

최신 제출물 성적

100%

1.

질문 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 / 1점

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

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

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

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

맞습니다

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.

질문 2

Consider this RNN:

This specific type of architecture is appropriate when:

1 / 1점

$T_x = T_y$Tx = Ty

$T_x < T_y$Tx < Ty

$T_x > T_y$Tx > Ty

$T_x = 1$Tx = 1

맞습니다

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

3.

질문 3

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

1 / 1점

☐ ☐

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)

✓

맞습니다

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)

✓

맞습니다

Correct!

4.
질문 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 / 1점

◯ ◯

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>)

✓

맞습니다

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

5.
질문 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 tt?

1 / 1점

◯ ◯

(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.

✓
맞습니다
6.
질문 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 / 1점
○○

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.

✓
맞습니다
7.
질문 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 Γ_uΓu at each time step?

1 / 1점
○○

1

◉◉

100

○○

300

○○

10000

✓
맞습니다

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

8.

질문 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?

1 / 1점

○

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 decay.

◉

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.

맞습니다

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

9.
질문 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 / 1점

◉

Γ_u and $1-\Gamma_u$

○

Γ_u and Γ_r

○

$1-\Gamma_u$ and Γ_u

○

Γ_r and Γ_u

맞습니다

Yes, correct!

10.
질문 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 $x^{<1>}, \ldots, x^{<365>}$. You've also collected data on your dog's mood, which you represent as $y^{<1>}, \ldots, y^{<365>}$. 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?

1 / 1점

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 $x^{<1>}, \ldots, x^{<t>}$, but not on $x^{<t+1>}, \ldots, x^{<365>}$

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

맞습니다

Yes!