받은 학점 90% **최신 제출물 학점** 90% **통과 점수**: 80% 이상

다음 향목으로 이동

1/1점

1/1점

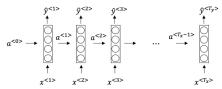
1/1점

1/1점

- $\textbf{1.} \ \ \, \text{Suppose your training examples are sentences (sequences of words)}. \ \, \text{Which of the following refers to the } s^{th} \text{word in the } r^{th} \text{training example?}$
 - $\bigcirc \ x^{< r > (s)}$
 - (r) x < r>(r)
 - ○ x^{(s)<r>}

∠ ^ 더보기

 \bigcirc **9:044** We index into the r^{th} row first to get to the r^{th} training example (represented by parentheses), then the s^{th} column to get to the s^{th} word (represented by the brackets).



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

- False
- ① True

∠^ 터보기

② 契合니다 It is appropriate when the input sequence and the output sequence have the same length or size.

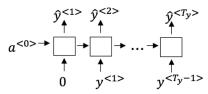
- 3. Select the two tasks combination that could be addressed by a many-to-one RNN model architecture from the

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

₹ 대보기

② ₹âUU 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:

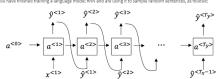


True/False: At the t^{th} time step the RNN is estimating $P(y^{< t>} \mid y^{< 1>}, y^{< 2>}, \ldots, y^{< t-1>})$

- True
- False

∠ 전보기

受験合니다Yes, in a training model we try to predict the next step based on knowledge of all prior steps.



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}^{GD} . (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

 (i) Use the probabilities output by the RNN to randomly time-step as ŷ^{ct>}.(ii) Then pass this selected word to th 	e next time-step.	
∠^ 터보기		
⊙ 맞습니다		
6. True/False: If you are training an RNN model, and find that y	your weights and activations are all taking on the value	1/1점
of NaN ("Not a Number") then you have a vanishing gradier		*/**
False		
○ True		
√ 전보기		
	tome in training DNNs but in this case your weights	
and activations taking on the value of NaN implies yo		
7. Suppose you are training an LSTM. You have a 10000 word v dimensional activations $a^{< t>}$. What is the dimension of Γ_0	ocabulary, and are using an LSTM with 100- at each time step?	1/1점
O 1		
100		
○ 300		
O 10000		
₹ 대보기		
\bigcirc 맞습니다 Correct, Γ_u is a vector of dimension equal to the num	ber of hidden units in the LSTM.	
8. True/False: In order to simplify the GRU without vanishing g sequences you should remove the Γ_r i.e., setting $\Gamma_r=1$	gradient problems even when training on very long	1/1점
False	into)s.	
True		
9 100		
2 ct wat		
√² 더보기 △ 막스니다		
고 대보기 ③ 맛습니다 If 「마~0 for a timestep, the gradient can propagate baths signal to backpropagate without vanishing, we ne	ick through that timestep without much decay. For $=$ ed $\sigma^{$.	
○ 맛습니다 If Tu≈0 for a timestep, the gradient can propagate bac	isk through that timestep without much decay. For $\stackrel{\wedge}{=}$ eed $c^{<\!C\!>}$ to be highly dependent on $c^{<\!C\!-1\!>}$.	
 ⊙ %℃↓□ If Fu=0 for a timestep, the gradient can propagate bas the signal to backpropagate without vanishing, we ne True/False: Using the equations for the GRU and LSTM belor 	ed a to be highly dependent on a to to be highly dependen	1/1점
	ed a to be highly dependent on a to to be highly dependen	1/1점
 ⊕ %4LG If Fu=0 for a timestep, the gradient can propagate bas the signal to backpropagate without vanishing, we ne True/False: Using the equations for the GRU and LSTM belor role similar to 1- Fu and Fu. 	with Eupdate Gate and Forget Gate in the LSTM play a	1/1점
True/False: Using the equations for the GRU and LSTM belor role similar to 1- Fu and Fu. GRU	w the Update Gate and Forget Gate in the LSTM play a	1/1정
 ⑦ 代金니다 If Turb for a timestep, the gradient can propagate bat the signal to backpropagate without vanishing, we no	with Euphanian Grant and Forget Gate in the LSTM play a $ \text{LSTM} $ $ \mathcal{E}^{<\text{LS}} = \tanh(W_c[a^{<\text{L}-1}, x^{<\text{L}>}] + b_c) $	1/1점
② RéLIQ If Fund for a timestep, the gradient can propagate bat the signal to backpropagate without vanishing, we not signal to backpropagate without vanishing, we not see that the signal to backpropagate without vanishing, we not see that the signal to 1. Fund fund for the GRU and LSTM belonger signal to 1. Fund fund fund fund fund fund fund fund f	with Eupdate Gate and Forget Gate in the LSTM play a $ \text{LSTM} $ $\mathcal{E}^{<\text{LS}} = \tanh(W_c[a^{<\text{L}-1>}, x^{<\text{LS}}] + b_c) $ $\Gamma_{\text{W}} = \sigma(W_{\text{W}}[a^{<\text{L}-1>}, x^{<\text{LS}}] + b_{\text{W}}) $	1/18
② RéLICI If Fu=0 for a timestep, the gradient can propagate bat the signal to backpropagate without vanishing, we not 9. True/False: Using the equations for the GRU and LSTM belor role similar to 1- Fu and Fu. GRU $\mathcal{E}^{}, x^{}] + b_c)$ $\Gamma_u = \sigma(W_u[c^{}, x^{}] + b_u)$ $\Gamma_r = \sigma(W_r[c^{}, x^{}] + b_r)$	with Update Gate and Forget Gate in the LSTM play a	1/18
② Stàtiq If fu=0 for a timestep, the gradient can propagate bas the signal to backpropagate without vanishing, we not signal to backpropagate without vanishing, we not signal to the signal to backpropagate without vanishing, we not signal to 1- fu and fu. GRU $\mathcal{E}^{ \Gamma_u = \sigma(W_u[e^{-Ct-1}, x^{ \Gamma_r = \sigma(W_r[e^{-Ct-1}, x^{ e^{-Ct} = \Gamma_u * \mathcal{E}^{$	with Update Gate and Forget Gate in the LSTM play a $ \begin{aligned} \mathbf{LSTM} \\ & \mathcal{E}^{<\mathrm{CP}} = \mathrm{tanh}(W_c[a^{<\mathrm{CP}-1}, x^{<\mathrm{CP}}] + b_c) \\ & \Gamma_u = \sigma(W_u[a^{<\mathrm{CP}-1}, x^{<\mathrm{CP}}] + b_u) \\ & \Gamma_f = \sigma(W_f[a^{<\mathrm{CP}-1}, x^{<\mathrm{CP}}] + b_f) \end{aligned} $ $ \Gamma_0 = \sigma(W_0[a^{<\mathrm{CP}-1}, x^{<\mathrm{CP}}] + b_f) $	1/1%
② Stàtiq If fu=0 for a timestep, the gradient can propagate bas the signal to backpropagate without vanishing, we not signal to backpropagate without vanishing, we not signal to the signal to backpropagate without vanishing, we not signal to 1- fu and fu. GRU $\mathcal{E}^{ \Gamma_u = \sigma(W_u[e^{-Ct-1}, x^{ \Gamma_r = \sigma(W_r[e^{-Ct-1}, x^{ e^{-Ct} = \Gamma_u * \mathcal{E}^{$	with Update Gate and Forget Gate in the LSTM play a	1/18
② Stàtiq If fu=0 for a timestep, the gradient can propagate bas the signal to backpropagate without vanishing, we not signal to backpropagate without vanishing, we not signal to the signal to backpropagate without vanishing, we not signal to 1- fu and fu. GRU $\mathcal{E}^{ \Gamma_u = \sigma(W_u[e^{-Ct-1}, x^{ \Gamma_r = \sigma(W_r[e^{-Ct-1}, x^{ e^{-Ct} = \Gamma_u * \mathcal{E}^{$	with Update Gate and Forget Gate in the LSTM play a	1/1%
 ③ YéLIQ If Funo for a timestep, the gradient can propagate bas the signal to backpropagate without vanishing, we not the signal to backpropagate without vanishing, we not similar to 1- Fu and Fu. GRU ē^{<1>} = tanh(W_c[Γ_x * c^{<1-1>}, x^{<1>}] + b_c) Γ_u = σ(W_u[c^{<1-1>}, x^{<1>}] + b_u) Γ_r = σ(W_r[c^{<1-1>}, x^{<1>}] + b_r) c^{<1>} = Γ_u * ξ^{<1} + (1 - Γ_u) * c^{<1-1>} a^{<1>} = c^{<1>} 	with Update Gate and Forget Gate in the LSTM play a	1/18
③ Stàtiq if fu=0 for a timestep, the gradient can propagate bas the signal to backpropagate without vanishing, we not see that the signal to backpropagate without vanishing, we not see that the signal to backpropagate without vanishing, we not see that the signal to backpropagate without vanishing, we not see that the signal to 1. True/False: Using the equations for the GRU and LSTM below role similar to 1. Fu and fu. GRU	with Update Gate and Forget Gate in the LSTM play a	1/1%
③ Stàtiq if fu=0 for a timestep, the gradient can propagate bas the signal to backpropagate without vanishing, we not see that the signal to backpropagate without vanishing, we not see that the signal to backpropagate without vanishing, we not see that the signal to backpropagate without vanishing, we not see that the signal to 1. True/False: Using the equations for the GRU and LSTM below role similar to 1. Fu and fu. GRU	with Update Gate and Forget Gate in the LSTM play a	1/12
③ % £\(\text{LIQ}\) If \(\text{Ind for a timestep, the gradient can propagate base the signal to backpropagate without vanishing, we not the signal to backpropagate without vanishing, we not color similar to 1-\text{Tu and }\text{Fu.} \[\text{GRU}\] \[\text{\$\delta^{<\text{C}\text{-1}} = \text{tanh}(W_c[\Gamma_r \in \delta^{<\text{C}\text{-1}}, \text{\$\delta^{<\text{-1}}\right)\$ } + \text{\$b_c\$}\) \[\Gamma_u = \sigma(W_u[\Gamma_c^{<\text{C}\text{-1}}, \text{\$\delta^{<\text{-1}}\right)\$ } + \text{\$b_c\$}\) \[\Gamma_v = \sigma(W_v[\Gamma_c^{<\text{C}\text{-1}}, \text{\$\delta^{<\text{-1}}\right)\$ } + \text{\$b_c\$}\) \[\text{\$\delta^{<\text{C}\text{-1}} = \Gamma_u^{<\text{C}\text{-1}} \right)\$ $\text{$\delta^{<\text{C}\text{-1}} \right)$ } \text{$\delta^{<\text{C}\text{-1}} \right)$} \[\text{$\delta^{<\text{C}\text{-1}} \right]$ \[\text{$\delta^{\text{C}\text{-1}} \right]$ \[\delta^{\text{C}\text{-1}} \right]$ \[\text{$\delta^{\text{C}\text{-1}} \right]$ \[\text{$\delta^{\text{C}\text{-1}} \right]$ \[\delta^{\text{C}\text{-1}} \right]$ \[\delta^{\text{C}\text{-1}} \right]$ $	with Update Gate and Forget Gate in the LSTM play a	បរន
② 90 ± 0.1 Gr at immstep, the gradient can propagate bat the signal to backpropagate without vanishing, we net the signal to 0.1 True/False: Using the equations for the GRU and LSTM belor cole similar to 1.1 Fu and Γ_U . GRU $\mathcal{C}^{}, x^{<1>}] + b_u)$ $\Gamma_r = \sigma(W_r[c^{}, x^{<1>}] + b_r)$ $c^{} + (1 - \Gamma_u) * e^{}$ $\alpha^{$ © False True $\mathcal{C}^{}$ $\mathcal{C}^{}$ $\mathcal{C}^{}$ $\mathcal{C}^{}$ $\mathcal{C}^{}$	with Update Gate and Forget Gate in the LSTM play a	1/12
② 92 ± LEQ If Funds for a timestep, the gradient can propagate base the signal to backpropagate without vanishing, we need to see that the signal to backpropagate without vanishing, we need to see that the signal to backpropagate without vanishing, we need to see that the signal to 1-10 and Fu. 9. True/False: Using the equations for the GRU and LSTM belor role similar to 1-1-10, and Fu. GRU $\mathcal{E}^{<1} = \tan (W_{e} [\Gamma_{e} * c^{<1-1}, x^{<1-1}] + b_{u})$ $\Gamma_{r} = \sigma(W_{r} [c^{<1-1}, x^{<1-1}] + b_{u})$ $c^{<1-1} = \Gamma_{u} * c^{<1-1}, x^{<1-1}] + b_{u})$ $c^{<1-1} = \Gamma_{u} * c^{<1-1}, x^{<1-1}] + b_{u}$ $c^{<1-1} = \Gamma_{u} * c^{<1-1}, x^{<1-1}] + b_{u}$ 0 False 0 False 1 True $v^{2} \in M^{2}$ $v^{3} \in M^{2}$ $v^{4} \in M^{2}$ $v^{5} \in M^{2}$ $v^{5} \in M^{2}$ $v^{6} \in M^{2}$ $v^{6} \in M^{2}$ $v^{6} \in M^{2}$ $v^{6} \in M^{2}$ $v^{7} \in M^{2}$ $v^{7} \in M^{2}$ $v^{7} \in M^{2}$ $v^{7} \in M^{2}$ $v^{8} \in M^{2}$ $v^{9} \in M^{2}$ $v^{9} \in M^{2}$ $v^{9} \in M^{2}$ $v^{9} \in M^{2}$ v^{9	with Update Gate and Forget Gate in the LSTM play a	
③ % £\(\text{LIQ}\) If \(\text{Ind for a timestep, the gradient can propagate base the signal to backpropagate without vanishing, we not the signal to backpropagate without vanishing, we not color similar to 1-\text{Tu and }\text{Fu.} \[\text{GRU}\] \[\text{\$\delta^{<\text{C}\text{-1}} = \text{tanh}(W_c[\Gamma_r \in \delta^{<\text{C}\text{-1}}, \text{\$\delta^{<\text{-1}}\right)\$ } + \text{\$b_c\$}\) \[\Gamma_u = \sigma(W_u[\Gamma_c^{<\text{C}\text{-1}}, \text{\$\delta^{<\text{-1}}\right)\$ } + \text{\$b_c\$}\) \[\Gamma_v = \sigma(W_v[\Gamma_c^{<\text{C}\text{-1}}, \text{\$\delta^{<\text{-1}}\right)\$ } + \text{\$b_c\$}\) \[\text{\$\delta^{<\text{C}\text{-1}} = \Gamma_u^{<\text{C}\text{-1}} \right)\$ $\text{$\delta^{<\text{C}\text{-1}} \right)$ } \text{$\delta^{<\text{C}\text{-1}} \right)$} \[\text{$\delta^{<\text{C}\text{-1}} \right]$ \[\text{$\delta^{\text{C}\text{-1}} \right]$ \[\delta^{\text{C}\text{-1}} \right]$ \[\text{$\delta^{\text{C}\text{-1}} \right]$ \[\text{$\delta^{\text{C}\text{-1}} \right]$ \[\delta^{\text{C}\text{-1}} \right]$ \[\delta^{\text{C}\text{-1}} \right]$ $	with Update Gate and Forget Gate in the LSTM play a	0/1점
 StellG If U=0 for a timestep, the gradient can propagate bat the signal to backpropagate without vanishing, we net to learn the signal to 1- Fu and Fu. GRU C<!-- --> = tanh(W_c[Γ_r * c <1-1>, x <1>] + b_c) Γ_u = σ(W_u[c <1-1>, x <1>] + b_u) Γ_r = σ(W_r[c <1-1>, x <1>] + b_r) c <1-1> a <1-1	with Update Gate and Forget Gate in the LSTM play a	
 StellG If Γu=0 for a timestep, the gradient can propagate bat the signal to backpropagate without vanishing, we net to learn the signal to be signal to signal to	with Update Gate and Forget Gate in the LSTM play a	
	with Update Gate and Forget Gate in the LSTM play a	
	with Update Gate and Forget Gate in the LSTM play a	
SizeLIQ If Tuno for a timestep, the gradient can propagate bat the signal to backpropagate without vanishing, we net the signal to backpropagate without vanishing, we net the signal to backpropagate without vanishing, we net the signal to to 1- Fu and Fu. GRU	with Update Gate and Forget Gate in the LSTM play a	