


Feedback — Lecture 7 Quiz

[Help Center](#)

You submitted this quiz on **Thu 30 Jun 2016 1:52 PM CEST**. You got a score of **3.75** out of **9.00**. However, you will not get credit for it, since it was submitted past the deadline.




Question 1

How many bits of information can be modeled by the vector of hidden activities (at a specific time) of a Recurrent Neural Network (RNN) with 16 logistic hidden units?

Your Answer	Score	Explanation
<input type="radio"/> >16		
<input type="radio"/> 16		
<input checked="" type="radio"/> 4	 0.00	
<input type="radio"/> 2		
Total	0.00 / 1.00	

Question 2

This question is about speech recognition. To accurately recognize what phoneme is being spoken at a particular time, one needs to know the sound data from 100ms before that time to 100ms after that time, i.e. a total of 200ms of sound data. Which of the following setups have access to enough sound data to recognize what phoneme was being spoken 100ms into the past?

Your Answer	Score	Explanation
<input type="checkbox"/> A Recurrent Neural Network (RNN) with 200ms of input	 0.00	
<input type="checkbox"/> A feed forward Neural Network with 30ms of input	 0.25	A feed forward Neural Network can not be used given 30ms because we need 200ms to perform the recognition task.
<input checked="" type="checkbox"/> A feed forward Neural Network with 200ms of input	 0.25	A feed forward Neural Network can be used only in case it is given all the input it needs for the recognition task because it doesn't store the input history.

☒ A Recurrent Neural Network (RNN) with 30ms of input

✓ 0.25

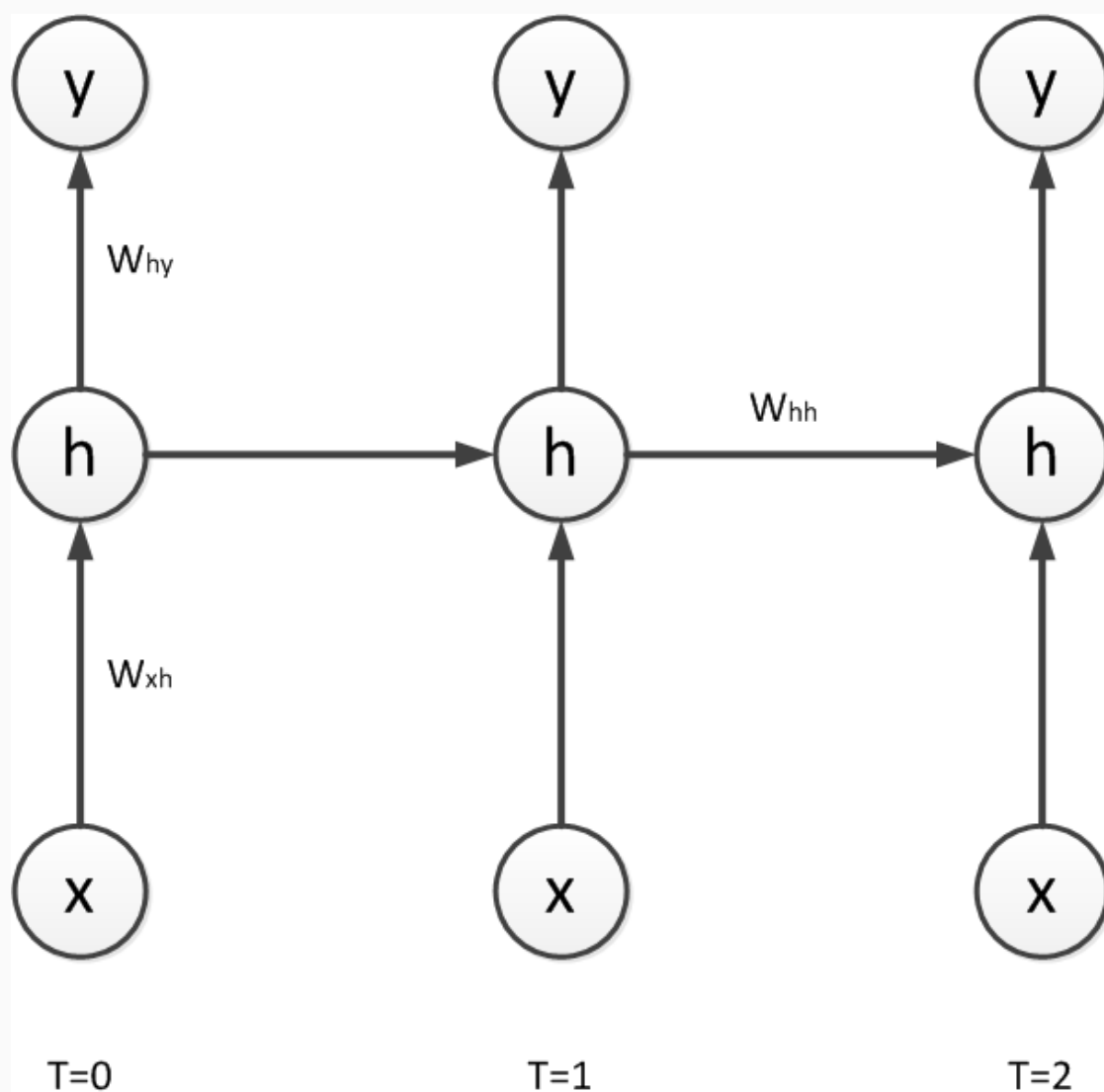
RNN can reliably memorize long term history due to its temporal connections between hidden states.

Total

0.75 /
1.00

Question 3

The figure below shows a Recurrent Neural Network (RNN) with one input unit x , one logistic hidden unit h , and one linear output unit y . The RNN is unrolled in time for $T=0, 1$, and 2 .



The network parameters are: $W_{xh} = 0.5$, $W_{hh} = -1.0$, $W_{hy} = -0.7$, $h_{bias} = -1.0$, and $y_{bias} = 0.0$. Remember, $\sigma(k) = \frac{1}{1+\exp(-k)}$.

If the input x takes the values 9, 4, -2 at time steps 0, 1, 2 respectively, what is the value of the output y at $T = 1$? Give your answer to at least two digits after the decimal point.

You entered:

0.5206

Your Answer

Score

Explanation

0.5206

✖

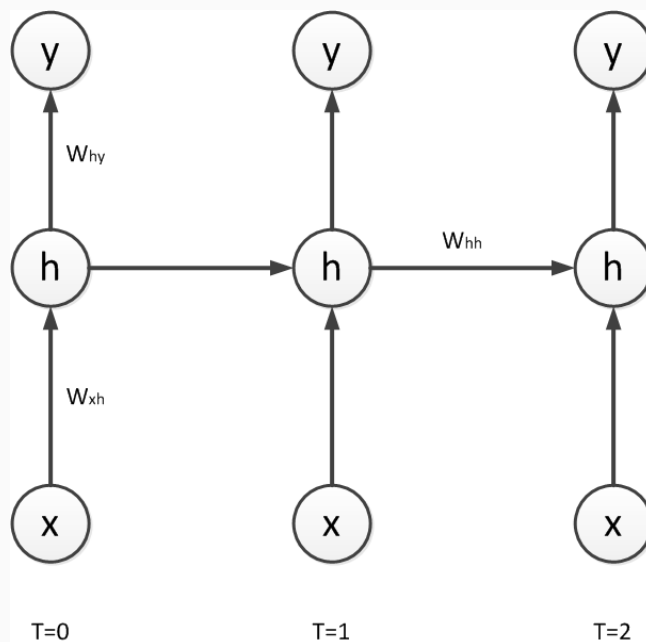
0.00

Total

0.00 / 2.00

Question 4

The figure below shows a Recurrent Neural Network (RNN) with one input unit x , one logistic hidden unit h , and one linear output unit y . The RNN is unrolled in time for $T=0, 1$, and 2 .



The network parameters are: $W_{xh} = -0.1$, $W_{hh} = 0.5$, $W_{hy} = 0.25$, $h_{\text{bias}} = 0.4$, and $y_{\text{bias}} = 0.0$.

If the input x takes the values $18, 9, -8$ at time steps $0, 1, 2$ respectively, the hidden unit values will be $0.2, 0.4, 0.8$ and the output unit values will be $0.05, 0.1, 0.2$ (you can check these values as an exercise). A variable z is defined as the total input to the hidden unit before the logistic nonlinearity.

If we are using the squared loss, with targets t_0, t_1, t_2 , then the sequence of calculations required to compute the total error E is as follows:

$$\begin{aligned}
 z_0 &= W_{xh}x_0 + h_{\text{bias}} & z_1 &= W_{xh}x_1 + W_{hh}h_0 + h_{\text{bias}} & z_2 &= W_{xh}x_2 + W_{hh}h_1 + h_{\text{bias}} \\
 h_0 &= \sigma(z_0) & h_1 &= \sigma(z_1) & h_2 &= \sigma(z_2) \\
 y_0 &= W_{hy}h_0 + y_{\text{bias}} & y_1 &= W_{hy}h_1 + y_{\text{bias}} & y_2 &= W_{hy}h_2 + y_{\text{bias}} \\
 E_0 &= \frac{1}{2} (t_0 - y_0)^2 & E_1 &= \frac{1}{2} (t_1 - y_1)^2 & E_2 &= \frac{1}{2} (t_2 - y_2)^2 \\
 E &= E_0 + E_1 + E_2
 \end{aligned}$$

Suppose that the model has learned the following parameter values:

- $w_{xh} = 1$
- $w_{hh} = -2$
- $w_{hy} = 1$
- All biases are 0

For one specific training case, the input is 1 at $T=0$ and 0 at $T=1$, $T=2$, and $T=3$. The target output (at $T=3$) is 0.5, and we're using the squared error loss function.

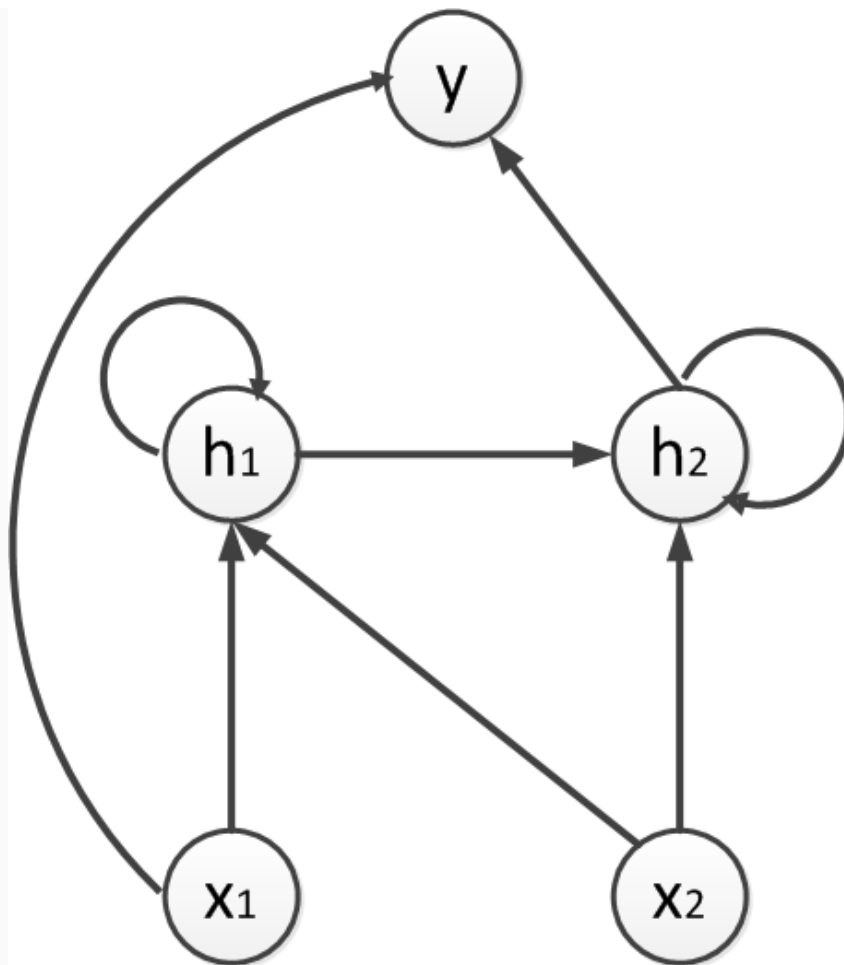
We're interested in the gradient for w_{xh} , i.e. $\frac{\partial E}{\partial w_{xh}}$. Because it's only at $T=0$ that the input is not zero, and it's only at $T=3$ that there's an output, the error needs to be backpropagated from $T=3$ to $T=0$, and that's the kind of situations where RNN's often get either vanishing gradients or exploding gradients. Which of those two occurs in this situation?

You can either do the calculations, and find the answer that way, or you can find the answer with more thinking and less math, by thinking about the slope $\frac{\partial y}{\partial z}$ of the logistic function, and what role that plays in the backpropagation process.

Your Answer	Score	Explanation
<input type="radio"/> Exploding gradient		
<input checked="" type="radio"/> Vanishing gradient	✓ 2.00	
Total	2.00 / 2.00	

Question 6

Consider the following Recurrent Neural Network (RNN):



As you can see, the RNN has two input units, two hidden units, and one output unit.

For this question, every arrow denotes the effect of a variable at time t on a variable at time $t + 1$.

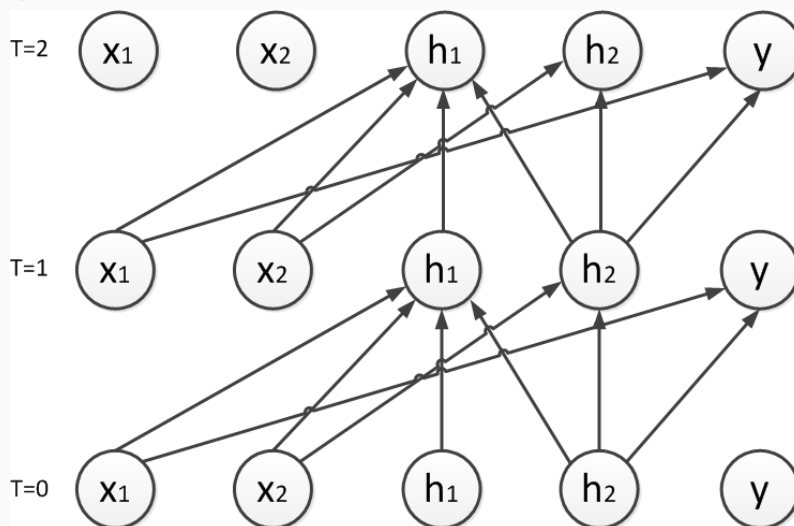
Which feed forward Neural Network is equivalent to this network unrolled in time?

Your Answer

Score

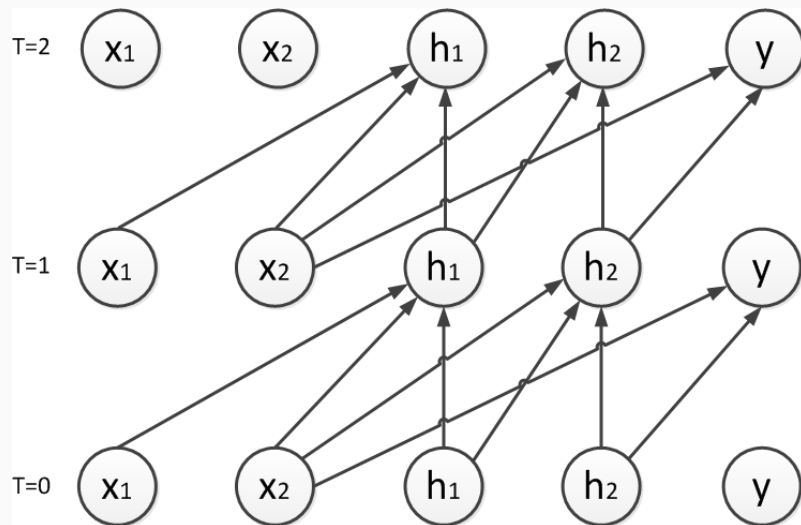
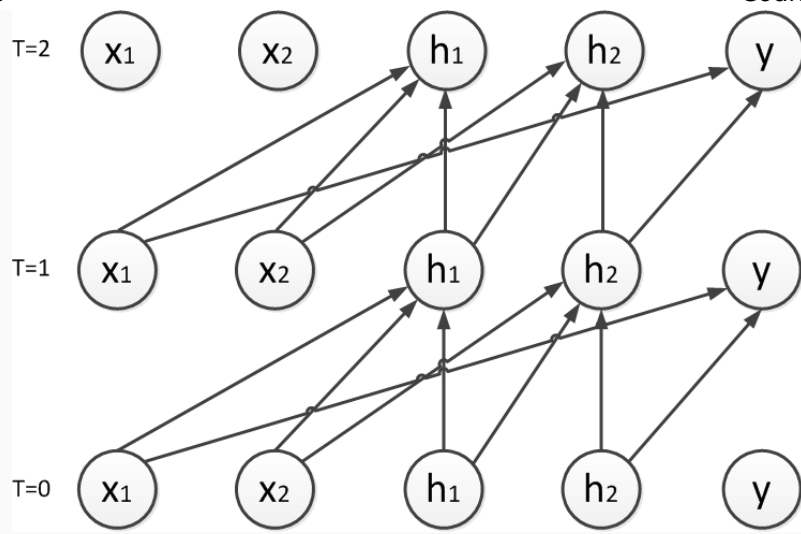
Explanation

☐



☒

✓ 1.00



Total

1.00 /
1.00