Lecture 7 Quiz

4/6 points earned (66%)

You haven't passed yet. You need at least 80% to pass.
Review the material and try again! You have 3 attempts every 8 hours.

Review Related Lesson

---

1 / 1
points

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?

- 2
- >16

Correct

- 16
- 4

---

1 / 1
points

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?

- A feed forward Neural Network with 200ms of input

Correct

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 200ms of input

Correct

- [ ] A Recurrent Neural Network (RNN) with 30ms of input

Correct

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

- [ ] A feed forward Neural Network with 30ms of input

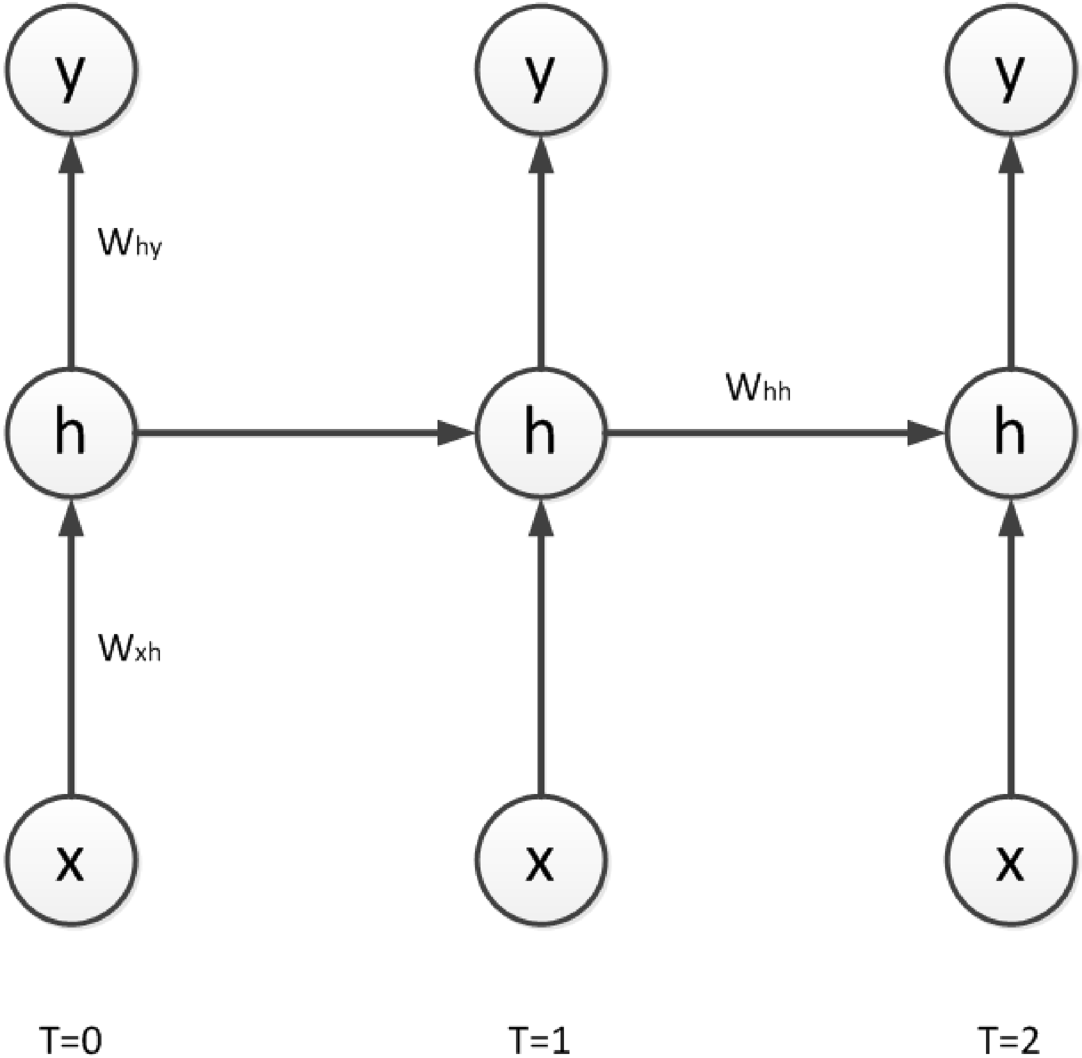
Un-selected is correct

✖ 0 / 1 points

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.

0.5268319

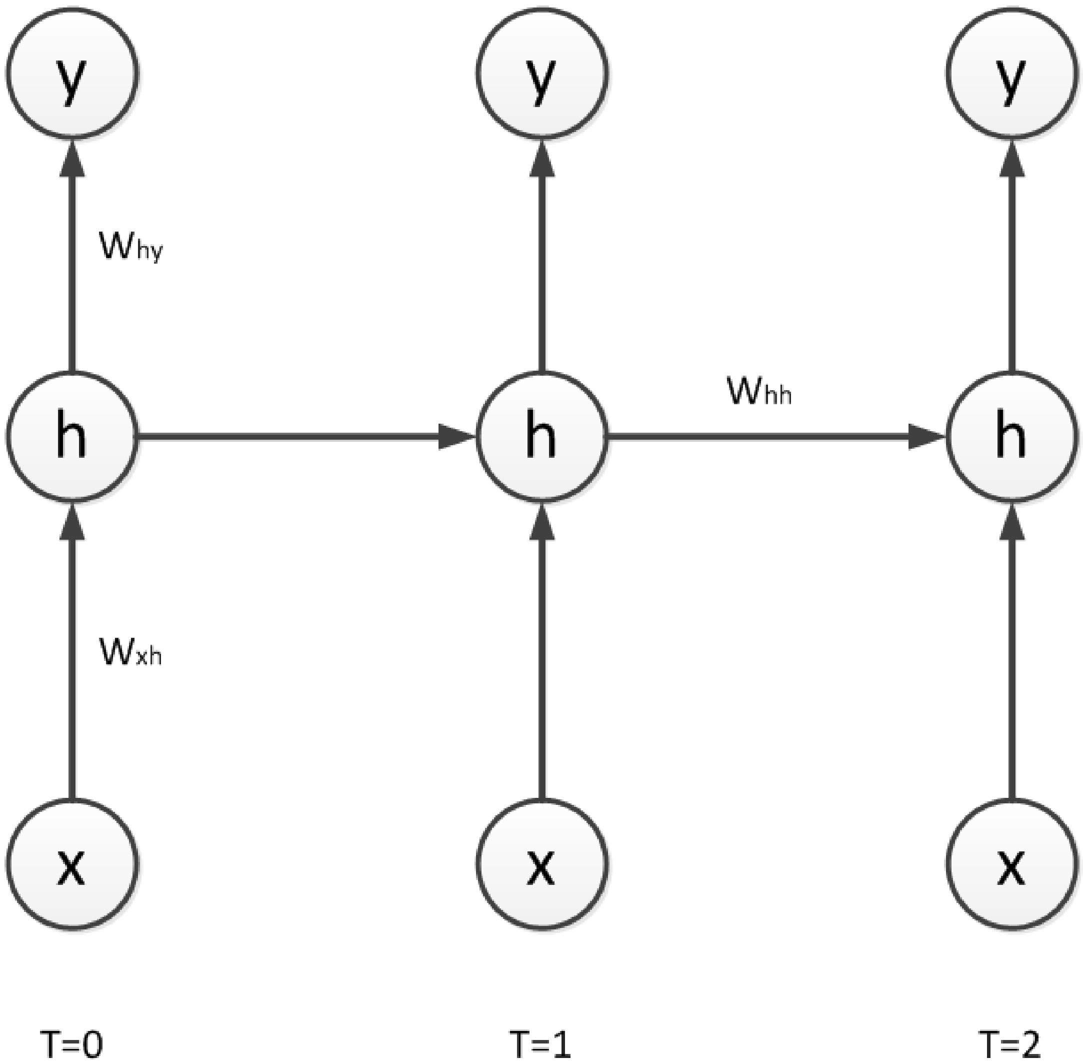
Incorrect Response

0 / 1 points

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:

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

$$E = E_0 + E_1 + E_2$$

If the target output values are $t_0 = 0.1, t_1 = -0.1, t_2 = -0.2$ and the squared error loss is used, what is the value of the error derivative just before the hidden unit nonlinearity at $T = 2$ (i.e. $\frac{\partial E}{\partial z_2}$)? Write your answer up to at least the fourth decimal place.

2

Incorrect Response

We need to calculate $\frac{\partial E}{\partial z_2}$. We can do this by backpropagation:

$$\begin{aligned}\frac{\partial E}{\partial z_2} &= \frac{\partial E_2}{\partial z_2} \\ &= \frac{\partial E}{\partial y_2}\frac{\partial y_2}{\partial h_2}\frac{\partial h_2}{\partial z_2}\end{aligned}$$

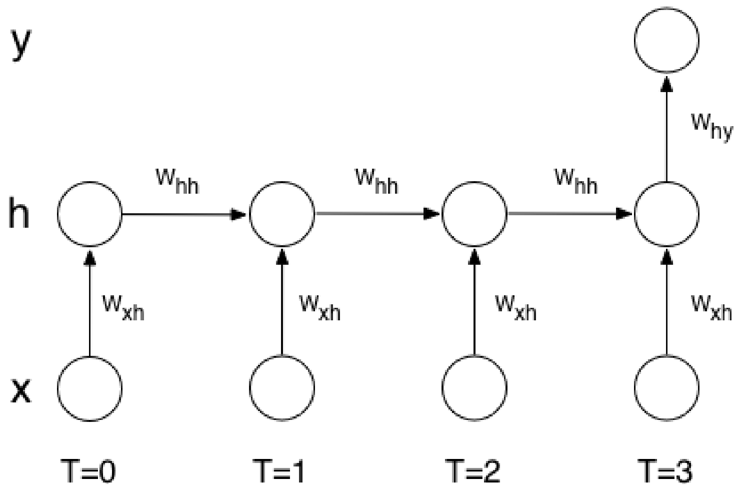
The first equality holds because only the error at the last time step depends on z_2.

1 / 1 points

5.

Consider a Recurrent Neural Network with one input unit, one logistic hidden unit, and one linear output unit. This RNN is for modeling sequences of

length 4 only, and the output unit exists only at the last time step, i.e. T=3. This diagram shows the RNN unrolled in time:



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

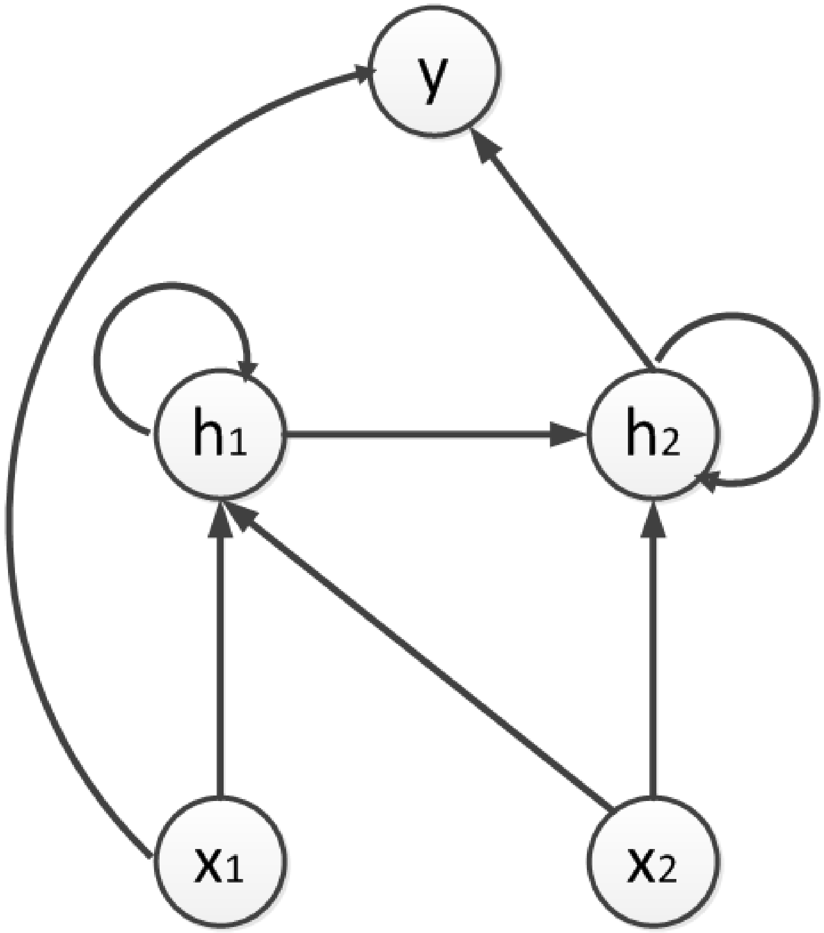
slope $\frac{\partial y}{\partial z}$ of the logistic function, and what role that plays in the backpropagation process.

- Exploding gradient
- Vanishing gradient

Correct

1 / 1
points

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?