

Quiz-07

- Due Mar 2 at 11:59pm
- Points 10
- Questions 10
- Available Feb 28 at 6pm - Mar 2 at 11:59pm
- Time Limit None
- Allowed Attempts 3

Take the Quiz Again

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	139 minutes	7 out of 10

⚠ Correct answers are hidden.

Score for this attempt: 7 out of 10

Submitted Mar 2 at 1:26pm

This attempt took 139 minutes.



Question 1

1 / 1 pts

For the purpose of predicting stock values, a well-trained finite response system that looks 7-time steps into the past (where each time step represents a day) can be expected to be able to predict: [Select all that apply].

Hint: Lecture 13- slides 10 - 24

- ☐ Bi-Weekly trends (once every two weeks)
- ☒ Weekly trends
- ☐ Annual Trends
- ☒ Daily Trends



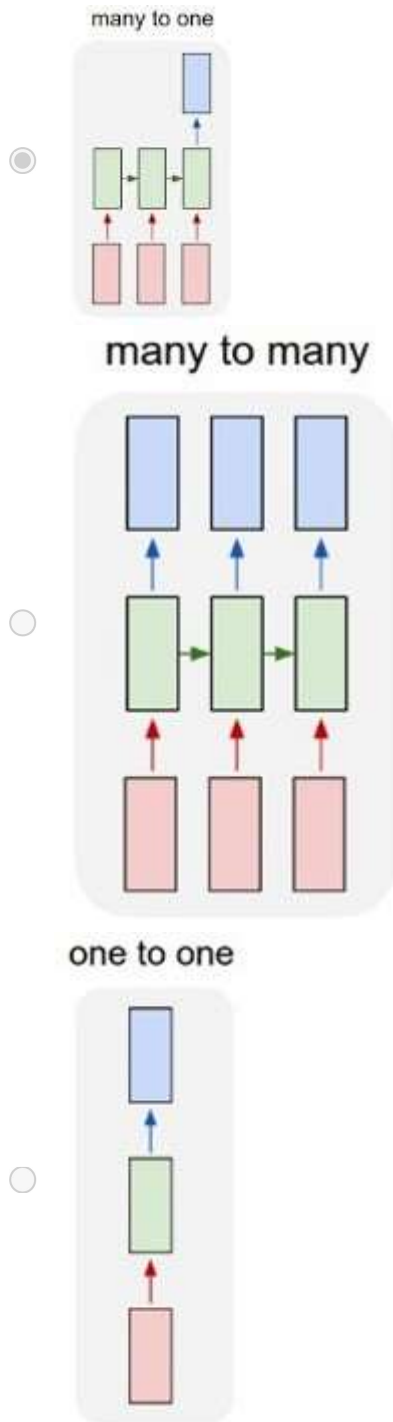
Question 2

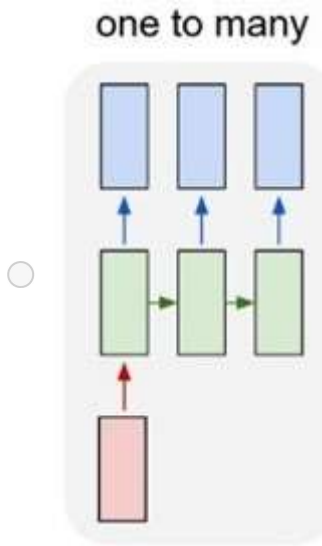
1 / 1 pts

In a problem where an RNN is used to analyze a sequence of words to predict the next word, what kind of model would best fit the problem. Please refer to the definition of different types of recurrent networks

mentioned in the lecture.

Hint: Lecture 13- slides 68, 70





IncorrectQuestion 3

0 / 1 pts

In Backpropagation through time, in principle the divergence is a

Hint: Lecture 13- slides 72 - 101

- ☒ joint function of all the outputs of the network
- ☐ average of the divergences at individual time instances
- ☐ divergence with respect to the output of only the last time instance
- ☒ sum of the divergences at individual time instances



PartialQuestion 4

0.5 / 1 pts

The following statements relate to backpropagation in a bi-directional RNN. Select all that are true. All statements must be assumed to be complete, and not partial facts.

Hint: Lecture 13- slides 115 - 123

- ☐ Backpropagation propagates gradients forwards through time in some parts of the network and backwards through time in others.
- ☐ Backpropagation in a bi-directional RNN only propagates gradients backwards through time.
- ☒ Backpropagation is independently performed in the forward and backward components of a bidirectional block.
- ☐

The derivatives computed from the forward network are subsequently propagated onwards into the backward network.



IncorrectQuestion 5

0 / 1 pts

Select all statements that are true. You will have to refer to:

<http://deeplearning.cs.cmu.edu/F20/document/readings/Bidirectional%20Recurrent%20Neural%20>
[_ \(https://ieeexplore.ieee.org/document/650093\)](https://ieeexplore.ieee.org/document/650093)



Bi-directional RNNs are unsuited to problems where output predictions must be made online, as newer samples from the input sequence arrive.



Bi-directional RNNs capture distant long-distance (i.e. over time) dependence between the input and the output, whereas uni-directional RNNs only capture short-distance dependencies.



A uni-directional RNN can incrementally compute its outputs in a streaming manner as inputs are processed sequentially over time, whereas a bi-directional RNN must process the entire input sequence before computing its outputs.



You can get the same output as a bi-directional RNN, as it is defined in the paper, by training two uni-directional RNNs, one processing the input left-to-right, and the other right-to-left, and averaging their outputs.



PartialQuestion 6

0.5 / 1 pts

You are training an RNN using Relu activation for hidden units, which of the following are possible (Select all that apply)

Hint: Draws concepts from Lecture 14



The outputs of the RNN may blow up



The network is not guaranteed BIBO stable



Since the input-output relationship of the Relu is piecewise-linear, the network is BIBO stable



Saturation will occur



Question 7

1 / 1 pts

A recurrent neural network with a single hidden layer has 3 neurons in its recurrent hidden layer (i.e. the hidden representation is a 3-dimensional vector). The recurrent weights matrix of the hidden layer is given by

$$W_r = \begin{bmatrix} 0.75 & 0 & 0.25 \\ 0.75 & -0.2 & 0.0 \\ 0.5 & 0. & -0.35 \end{bmatrix}$$

The hidden unit activation is the identity function $f(x) = x$. All bias values are 0.

Let us define the “memory duration” of the network as the minimum number of time steps N such that for an isolated input at time t (with no previous or subsequent inputs) the length of the hidden activation vector at time $t+N$ certainly falls to less than $1/100$ th of its value at time t . This is effectively the amount of time in which the influence of the input at time t all but vanishes from the network.

What is the memory duration of the above network (choose the closest answer). Assume $h(t-1) = 0$ (the zero vector).

Hint: Lecture 14 Slides 28-30

Hint: Remember that for any eigenvector E , $W_r E = \lambda_e E$, where λ_e is the Eigenvalue corresponding to E . If the hidden response of the system is exactly E (with length 1) at $t = 0$, what is the response of the system after t time steps, and what is its length?

☒ 30

☐ 10

☐ 40

☐ 20



Question 8

1 / 1 pts

A recurrent neural network with a single hidden layer has 3 neurons in its recurrent hidden layer (i.e. the hidden representation is a 3-dimensional vector). The recurrent weights matrix of the hidden layer is given by:

$$W_r = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0.2 & -0.1 & 0.7 \\ -0.2 & 0.65 & 0.15 \end{bmatrix}$$

The hidden units have $\tanh()$ as their activation function. All bias values are 0.

Let us define the “memory duration” of the network as the minimum number of time steps N such that for an isolated input at time t (with no previous or subsequent inputs) the length of the hidden activation vector at time $t+N$ almost certainly falls to less than $1/100$ th of its value at time t . This is effectively the amount of time in which the influence of the input at time t all but vanishes from the network. The weight matrix for the inputs to the RNN (W_c in the slides) is an Identity matrix (3×3). Consider the input

vector at time step t , $x(t)$ be $\left[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right]^T$.

What is the memory duration of the above network (choose the closest answer). Assume $h(t-1) = 0$ (the zero vector). You may want to simulate the network to determine this value (an analytical solution is difficult to get).

Hint: Draws concepts from Lecture 13. For practical implementation, slides discussing the RNN architecture, equations governing the state updates, and examples of how inputs are processed through time would be most relevant. (Slide 73 and beyond)

- ☒ 20
- ☐ 35
- ☐ 25
- ☐ 30

The simulation models the behavior of the recurrent neural network over time. It starts by calculating the initial hidden state $h(t)$ using the given input and the identity matrix for input weights. Then, it iteratively applies the recurrent weight matrix W_r to the hidden state, passing the result through the \tanh activation function in each time step. This process continues until the magnitude (Euclidean norm) of the hidden state vector falls below 1/100th of its initial value. The number of iterations required to reach this threshold is counted, representing the memory duration of the network.



Question 9

1 / 1 pts

An RNN has a single recurrent hidden layer. The hidden state activations are all $\tanh()$. You will recall from class that the recurrence of the hidden layer in such a network has the form:

$$\mathbf{z}_t = W_{hh}\mathbf{h}_{t-1} + W_{xh}\mathbf{x}_t + \mathbf{b}$$

$$\mathbf{h}_t = \tanh(\mathbf{z}_t)$$

where \mathbf{h}_t is the recurrent hidden state (vector) at time t , \mathbf{x}_t is the input (vector) at time t , W_{hh} is the recurrent weight matrix, W_{xh} is the input weight matrix and \mathbf{b} is the bias.

During backpropagation, the length of the derivative vector $\nabla_{\mathbf{h}_t} Div$ for the hidden activation at time t is found to be exactly 1. This derivative is further backpropagated through the activation of the hidden layer to obtain $\nabla_{\mathbf{z}_t} Div$. What is the *maximum* length of $\nabla_{\mathbf{z}_t} Div$?

Hint: Lecture 13- slides 80 - 100

1



Question 10

1 / 1 pts

An RNN has a single recurrent hidden layer. You will recall from class that the recurrence of the hidden layer in such a network has the form:

$$\mathbf{z}_t = W_{hh}\mathbf{h}_{t-1} + W_{xh}\mathbf{x}_t + \mathbf{b}$$

$$\mathbf{h}_t = \text{activation}(\mathbf{z}_t)$$

where \mathbf{h}_t is the recurrent hidden state (vector) at time t , \mathbf{x}_t is the input (vector) at time t , W_{hh} is the recurrent weight matrix, W_{xh} is the input weight matrix and \mathbf{b} is the bias.

The hidden activation \mathbf{h}_t is also input to the output layer to produce the output \mathbf{y}_t at time t .

The hidden layer in the network has 2 neurons and the recurrent matrix W_{hh} is

$$W_{hh} = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}$$

During backpropagation, the length of the derivative vector $\nabla_{\mathbf{y}_t} \text{Div}$ for the output at time t is found to be 0, and $\nabla_{\mathbf{z}_{t+1}} \text{Div}$ for the recurrent affine value at time $t + 1$ is found to be exactly 1. What is the *maximum* length of the derivative for \mathbf{h}_t (i.e. for $\nabla_{\mathbf{h}_t} \text{Div}$)?

Hint: Recall that for any relation $\mathbf{u} = W\mathbf{v}$, the maximum value for the ratio of the lengths of \mathbf{u} and \mathbf{v} is the largest singular value of W , i.e. $\frac{|\mathbf{u}|}{|\mathbf{v}|} \leq \max(\text{singular value}(W))$

Quiz Score: 7 out of 10