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

Attempt History

	Attempt	Time	Score
KEPT	Attempt 3	457 minutes	7 out of 10
LATEST	Attempt 3	457 minutes	7 out of 10
	Attempt 2	159 minutes	6 out of 10
	Attempt 1	139 minutes	7 out of 10

Correct answers are hidden.

Score for this attempt: 7 out of 10
Submitted Mar 2 at 11:51pm
This attempt took 457 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].

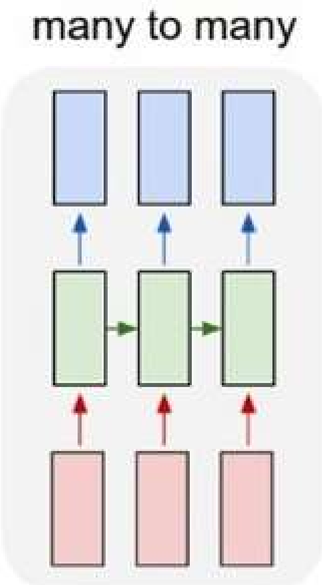
Hint: Lecture 13- slides 10 - 24

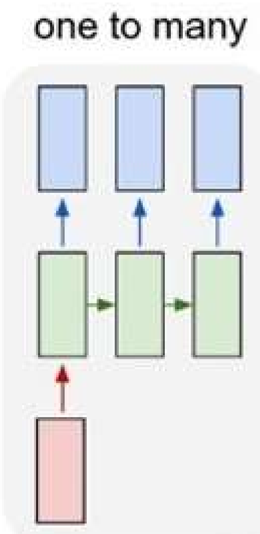
- [x] Weekly trends
- [ ] Bi-Weekly trends (once every two weeks)
- [ ] Annual Trends
- [x] Daily Trends

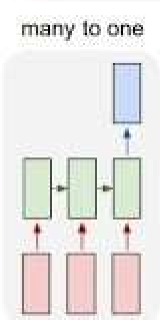
IncorrectQuestion 2
0 / 1 pts

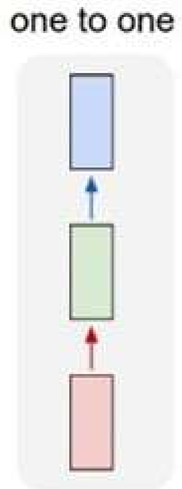
In a problem where an RNN is used to analyze a sequence of words to predict the next word, what kind of model would best fit the problem. Please refer to the definition of different types of recurrent networks mentioned in the lecture.

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

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

Question 4

1 / 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

- [ ] The derivatives computed from the forward network are subsequently propagated onwards into the backward network.
- [ ] Backpropagation in a bi-directional RNN only propagates gradients backwards through time.
- [x] Backpropagation propagates gradients forwards through time in some parts of the network and backwards through time in others.
- [x] Backpropagation is independently performed in the forward and backward components of a bidirectional block.

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)](https://ieeexplore.ieee.org/document/650093)

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

Question 6

1 / 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

- [ ] Saturation will occur
- [x] The outputs of the RNN may blow up
- [ ] Since the input-output relationship of the Relu is piecewise-linear, the network is BIBO stable
- [x] The network is not guaranteed BIBO stable

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

$$W_r = \begin{bmatrix} 0.05 & 0.7 & -0.1 \\ 0.7 & -0.1 & 0 \\ -0.1 & 0 & 0.7 \end{bmatrix}$$

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 1/100th 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?

- ○ 24
- ◉ 18
- ○ 11
- ○ 30

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 sigmoid() 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 T such that for any two isolated inputs x and y of the same length at time 0 (with no previous or subsequent inputs) the length of the difference in the length of the hidden activation vectors for the two inputs at time T almost certainly falls to less than 1/100th of its value at time 0.

In other words, representing the hidden state at time t, for the isolated input x at 0 as $h(t;x)$ (where we have introduced the “x” in the notation to explicitly show that this is the response to an isolated input of x at time 0), the memory duration as defined is the minimum T such that, for all x, y such that $|x| = |y|$, $|h(t;x) - h(t,y)| \ \leq \ 0.01|h(0;x) - h(0;y)|$ for all $t \geq T$.

This is effectively the amount of time in which the influence the of the direction of the input at time 0 all but vanishes from the network.

This is clearly difficult to compute analytically, but we can definitely attempt an estimate through specific examples and we will do that here.

The weight matrix for the inputs to the RNN (W_c in the pseudocode in the slides) is an Identity matrix (3×3). Consider the input vector $x = \left[\frac{1}{\sqrt(3)}, \frac{1}{\sqrt(3)}, \frac{1}{\sqrt(3)}\right]^\top$, and $y = -x$ (so our two exemplar vectors are diametrically opposite).

What is the memory duration of the above network, as estimated from the given x and y (choose the closest answer). Assume $h(-1) = 0$ (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)

- ○ 12
- ○ 8
- ◉ 4
- ○ 2

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

A deep neural network has 2 neurons in its $(l-1)$-th layer and 3 neurons in its l-th layer. The weight matrix for the l-th layer (i.e. the connections between the $(l-1)$-th layer and the l-th layer) is

$$\mathrm{W}_l = \begin{bmatrix} \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0.25 & -0.25 \\ -0.25 & 0.25 \end{bmatrix}$$

During backpropagation, the derivative vector $\nabla_{\mathbf{z}_l} Div$ for the affine input to the l-th layer, $\mathbf{z}_l$, is found to have length exactly equal to 1.0. What is the maximum length for the derivative $\nabla_{\mathbf{y}_{l-1}} Div$ for the output of the $(l-1)$-th layer, $\mathbf{y}_{l-1}$?

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. $\frac{|\mathbf{u}|}{|\mathbf{v}|} \leq \max(singular\ value(W))$

0.5

Quiz Score: 7 out of 10