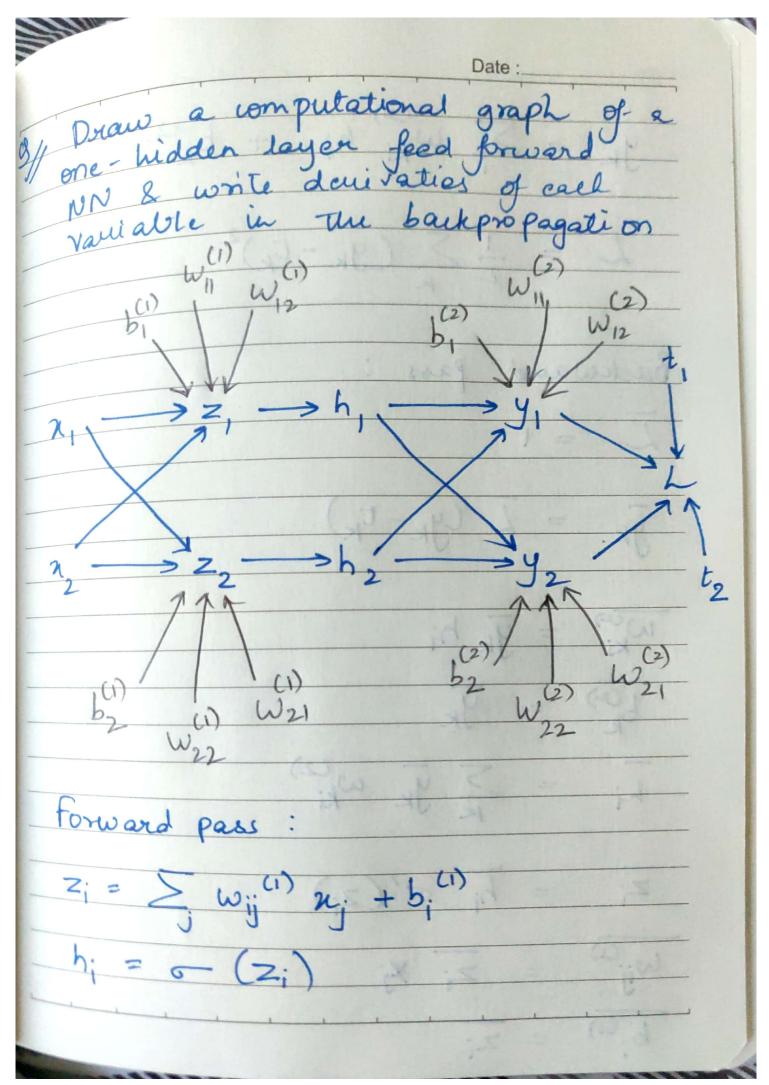
Date :
DL CS 583 A Fall 2021
Quiz 3
December 1, 2021
JH Maw
Instructor! Jia Xu
Student name: Shreeya Kokate
Student ID: 20005256
Student email addices: skokati@stevens.edu
Student email addiess: skokati@stevens.edu shreeya.csai@gnail.com

Date:
91
Ans lompaving attention models & RNNs we can say that, Advantages 1. Instead of RNN which have maximum path length access time propostional to the number of detirne steps, Attention based models have constant path length between the encoder input & decoder widden state. Because of this leaving becomes easieur.
2. Attention based models arlieve efficient content based soldiersing at the cost of recomputing content vectors at each time step.
3. #RNN architecture limits the parallelism potential for longer sequences whereas attention models supports this & are best at it.
4. Unlike RNN encodeus, the attention

niodr output do not depend on to redu of a puts. Disadvantage RNNs, detail the encoder & de so they have different lengths attention model:	moder outp	ut do not depend on t
RNNs, detach tur enroder & de so they have different lengths attention model:	eder of a	iputs.
RNNs, detach tur en vo der & de so they have different lengths attention model:	A KINDA	CONTROL BALBO PARSA
seep the energeter separation been decent and their been decent and their been decent and their been department to the content time of their content time of their content time of their content time of the content time of time of the content time of t	Disduarta	ge was an and
seep the energeter separation been decent and their been decent and their been decent and their been department to the content time of their content time of their content time of their content time of the content time of time of the content time of t	DAMAL CAN	and the day & day
seep the energeter separation been decent and their been decent and their been decent and their been department to the content time of their content time of their content time of their content time of the content time of time of the content time of t	KINIUS, act	ach mentoder se
seep the energeter separation been decent and their been decent and their been decent and their been department to the content time of their content time of their content time of their content time of the content time of time of the content time of t	so they he	ave different lengths
Secretary to the following the	attention	model.
Secretary to the following the	The same of the sa	The state and the sense
tention hand head addicasing which the possibility and the large sequences when the seaton and t		
teation based models subjecting when the time of the control of the control of the passibility on the control of the passibility on the control of the contr		
the cost of recomposition with the costs. The cost of recomposition with the costs. The costs time cheep.		0-13 10-10 10-11 100
the cost of recomposition with the costs. The cost of recomposition with the costs. The costs time cheep.	10 AD 18	teution hand rotates
The contract to the state of the contract of t		
the asterior limits the possibilities the second with the possibilities the second with the se		
addition of the large sequences when		at east line to able
ential for large sequences where		
ential for large sequences where		J. Dini) months on 40
courties properly support this is an		
	sussity and	

Date:
I Choose the connect one.
2) In practice, what is The most accurate description for activation function used in neural networks?
1. They must be differentiable 2. They can be non-differentiable, but only for a small number of points. 3. They can be any continuous function 4. They must be non-linear to be learnable.
Ans (B) or (2)
2) Given a neural network with N input nodes, no hidden layers, one output node, with entropy loss & sigmoid activation function, which algorithm can be used to find the global optimum.

D	ate:
	Bereit durching Manhald
	3. A bidirectional RNN because The
	prediction benefits from future lavel
	4. A one-directional amoder-devoder
	authitection can generate a sequen
	of future prices based on all
	historical input prices
4	Ans (4) A 1D encoder-devode
1	auchitecture.
-	a predict the next 30 daily price using
-	the previous 30 dally prices as in pute.
-	which model solection & englaration
	Janes Court and Court
	the fully converted deep feed forward
	returned because it considere all input
	prices in the hidden benjame to melice
	the best desistant with
	to the string to our disroch cored MAN become
	all considers the ander of the prival
	is the output hangth is the visions to
	the sight hough.
-	



Date: = 1 5 (yk-tk)2 Backward Pass yk hi S yk wki