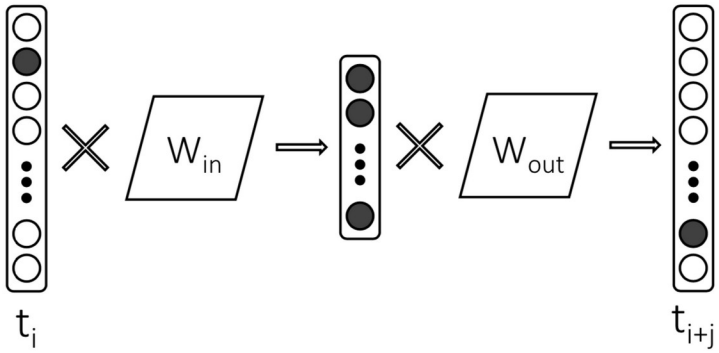
Name _____ Roll No _____ _____
Section

National University of Computer and Emerging Sciences, Lahore Campus

Course:	Information Retrieval	Course Code:	CS317
Program:	BS(Computer Science)	Semester:	Fall 2019
Duration:	25 Minutes	Total Marks:	10
Paper Date:	5-Nov-19	Weight	4%
Section:	A	Page(s):	2
Exam:	Quiz 3 solution	Roll No:	

Question 1:

Given a collection of 3500 documents with 5000 unique words, draw the architecture of neural network for WordToVec training. Suppose the number of dimension of word vector is 300. Clearly write size of input layer, hidden layers and output layer. [5 Marks]



Size of Input layer (t_i) = 1*5000

Size of Hidden layer = 1*300

Size of Output layer (t_{i+j}) = 1*5000

Size of Target Matrix (W_{in}) = 5000 * 300

Size of Context Matrix (W_{out}) = 300*5000

Name _____ Roll No _____ _____
Section

Question 2:

Represent the word “oranges” as vector using following corpus. Use TF.IDF weights. Assume the window size for word context is 3. [5 Marks]

Document 1: I like to ride cycle often.
Document 2: Ali and Hassan ate apple and oranges in the park.
Document 3: Ali ate apple not oranges in his house.
Document 4: Ali did not cross the street.

Solution:

Vector of Oranges:

Context words of oranges = ate, apple, and, in, the, park, not, his house

Dimensions	I	like	to	ride	cycle	often	Ali	and	Hassan	ate	apple	oranges	in	the
IF.IDF Weights	0	0	0	0	0	0	0	1*2	0	1.3*1	1.3*1	0	1.3*1	1*1

Dimensions	park	not	his	house	did	cross	street
IF.IDF Weights	1*2	1*1	1*2	1*2	0	0	0

Context Words	**IDF**
Ate	Log (4/2) = 1
Apple	Log (4/2) = 1
And	Log (4/1) = 2
In	Log (4/2) = 1
The	Log (4/2) = 1
Park	Log (4/1) = 2
Not	Log (4/2) = 1
His	Log (4/1) = 2
House	Log (4/1) = 2