|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **National University of Computer and Emerging Sciences, Lahore Campus** | | | | |
| C:\Users\saif\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.Word\final design.jpg | **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:** | **B** | **Page(s):** | **2** |
| **Exam:** | **Quiz 3 solution** | **Roll No:** |  |

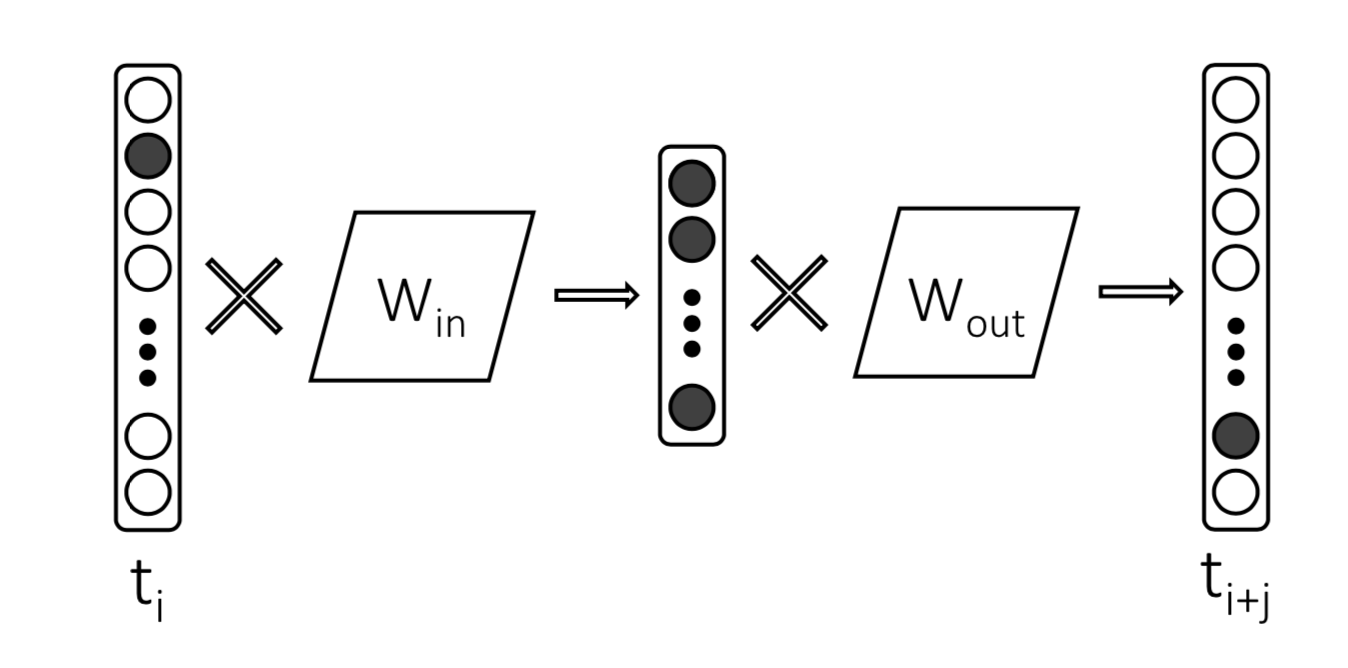
**Question 1:**

Briefly write down steps used by logistic regression for creating word embeddings as discussed in class. What are positive and negative examples and what objective is optimized by the classifier? [5 Marks]

Solution:

1. Treat the target word and a neighboring context word as positive examples.
2. Treat other words in corpus which do not occur in context as negative samples
3. Use logistic regression to train a classifier to distinguish those two cases
4. Use the weights of classifier (hidden layer) as the word embeddings

Objective of classifier is to maximize the similarity of target words with its context words. The similarity is measured by taking dot product of target word vector with all other words.



**Question 2:**

Represent the word “apple” as vector using following corpus. Use TF.IDF weights. Assume the window size for word context is 2. [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 apple:

Context words of apple = Hassan, ate, and, oranges, Ali, not

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dimensions | I | like | to | ride | cycle | often | Ali | and | Hassan | ate | apple | oranges | in | the |
| IF.IDF Weights | 0 | 0 | 0 | 0 | 0 | 0 | 1\*0.38 | 1\*2 | 1\*2 | 1.3\*1 | 0 | 1.3\*1 | 0 | 0 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Dimensions | park | not | his | house | did | cross | street |
| IF.IDF Weights | 0 | 1\*1 | 0 | 0 | 0 | 0 | 0 |

|  |  |
| --- | --- |
| **Context Words** | **IDF** |
| Ate | Log (4/2) = 1 |
| Hassan | Log (4/1) = 2 |
| And | Log (4/1) = 2 |
| Ali | Log (4/3) = 0.38 |
| Oranges | Log (4/2) = 1 |
| Not | Log (4/2) = 1 |