

Started on	Thursday, 12 April 2018, 10:47 AM
State	Finished
Completed on	Thursday, 12 April 2018, 11:04 AM
Time taken	16 mins 25 secs
Grade	4.00 out of 8.00 (50%)

Question 1
 Incorrect
 Mark 0.00 out of 1.00
 Flag question

When constructing a word embedding, negative samples are:

Select one:

- ☐ a. All less frequent words that do not occur in the context of a given word
- ☐ b. Word - context word combinations that are not occurring in the document collection
- ☐ c. Context words that are not part of the vocabulary of the document collection
- ☒ d. Only words that never appear as context word ✖

The correct answer is: Word - context word combinations that are not occurring in the document collection

Question 2
 Correct
 Mark 1.00 out of 1.00
 Flag question

Which of the following statements on Latent Semantic Indexing (LSI) and Word Embeddings (WE) is **correct**?

Select one:

- ☐ a. LSI does take into account the frequency of words in the documents, whereas WE does not
- ☒ b. LSI does not take into account the order of words in the document, whereas WE does ✔
- ☐ c. The dimensions of LSI can be interpreted as concepts, whereas those of WE cannot
- ☐ d. LSI is deterministic (given the dimension), whereas WE is not

The correct answer is: LSI does not take into account the order of words in the document, whereas WE does

Question 3
 Incorrect
 Mark 0.00 out of 1.00
 Flag question

Given the following teleporting matrix (E) for nodes A, B and C:

$$\begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{2} & 1 \end{bmatrix}$$

and making no assumptions about the link matrix (R), which of the following is **correct**:

(Reminder: columns are the probabilities to leave the respective node.)

Select one:

- ☐ a. A random walker can always leave node B
- ☐ b. A random walker can never reach node A
- ☐ c. A random walker can always leave node C
- ☒ d. A random walker can never leave node A ✖

The correct answer is: A random walker can always leave node B

Question 4
 Incorrect
 Mark 0.00 out of 1.00
 Flag question

When representing the adjacency list of a Web page in a connectivity server by using a reference list from another Web page, the reference list is searched only in a neighboring window of the Web page's URL, because:

Select one:

- ☒ a. Typically many URLs in a web page are similar to each other ✖
- ☐ b. Often many URLs among two pages with similar URL are similar
- ☐ c. Subsequent URLs in an adjacency list have typically small differences
- ☐ d. Most extra nodes are found in the neighboring window

The correct answer is: Often many URLs among two pages with similar URL are similar

Question 5
 Correct
 Mark 1.00 out of 1.00
 Flag question

When computing PageRank iteratively, the computation ends when:

Select one:

- ☐ a. All nodes of the graph have been visited at least once
- ☒ b. The norm of the difference of rank vectors of two subsequent iterations falls below a predefined threshold ✔
- ☐ c. The probability of visiting an unseen node falls below a predefined threshold

- ☐ d. The difference among the eigenvalues of two subsequent iterations falls below a predefined threshold

The correct answer is: The norm of the difference of rank vectors of two subsequent iterations falls below a predefined threshold

Question 6

Incorrect

Mark 0.00 out of 1.00

Flag question

Given the 2-itemsets {1,2}, {1,5}, {2,5}, {1,4}, {1,3}, when generating the 3-itemsets we will:

Select one:

- ☐ a. Generate 4 3-itemsets after the join and 2 3-itemsets after the prune
- ☐ b. Generate 6 3-itemsets after the join and 1 3-itemsets after the prune
- ☒ c. Generate 5 3-itemsets after the join and 2 3-itemsets after the prune ✖
- ☐ d. Generate 4 3-itemsets after the join and 1 3-itemsets after the prune

The correct answer is: Generate 6 3-itemsets after the join and 1 3-itemsets after the prune

Question 7

Correct

Mark 1.00 out of 1.00

Flag question

For his awesome research, Tugrulcan is going to use the PageRank with teleportation and HITS algorithm, not on a network of webpages but on the retweet network of Twitter! The retweet network is a directed graph, where nodes are users and an edge going out from a user A and to a user B means that "User A retweeted User B". Which one is **false** about a Twitter bot that retweeted other users frequently but got never retweeted by other users or by itself?

Select one:

- ☐ a. Its authority value will be equal to the hub value of a user who never retweets other users
- ☐ b. It will have an authority value of zero
- ☒ c. It will have a PageRank of zero ✔
- ☐ d. It will have a non-zero hub value

The correct answer is: It will have a PageRank of zero

Question 8

Correct

Mark 1.00 out of 1.00

Flag question

Given the following list of transactions: {apple,milk}, {milk, bread}, {apple, bread, milk}, {bread}, which of the following is **correct**?

Select one:

- ☒ a. apple → milk has support ½ and confidence 1 ✔
- ☐ b. milk → bread has support ½ and confidence 1
- ☐ c. bread → milk has support ½ and confidence 1
- ☐ d. milk → apple has support ½ and confidence 1

The correct answer is: apple → milk has support ½ and confidence 1