Week 7 : Assignment 7

The due date for submitting this assignment has passed.

Due on 2024-03-13, 23:59 IST.

Assignment submitted on 2024-03-11, 20:55 IST

1) ***1 point***

Suppose you have a raw text corpus and you compute word co occurrence matrix from there. Which of the following algorithm(s) can you utilize to learn word representations? (Choose all that apply)

a. CBOW
b. SVM
c. PCA
d. Bagging

- [x] a.
- [ ] b.
- [x] c.
- [ ] d.

Yes, the answer is correct.
Score: 1
Accepted Answers:
a.
c.

2) ***1 point***

What is the method for solving word analogy questions like, given A, B and D, find C such that A:B::C:D, using word vectors?

a. $v_c = v_a + (v_b - v_d)$, then use cosine similarity to find the closest word of v_c.
b. $v_c = v_a + (v_d - v_b)$ then do dictionary lookup for v_c
c. $v_c = v_d + (v_a - v_b)$ then use cosine similarity to find the closest word of v_c.
d. $v_c = v_d + (v_a - v_b)$ then do dictionary lookup for v_c.
e. None of the above

- ○ a.
- ○ b.
- ◉ c.
- ○ d.
- ○ e.

Yes, the answer is correct.
Score: 1

Accepted Answers:
c.

3) ***1 point***

What is the value of $PMI(w_1, w_2)$ for $C(w_1) = 250$, $C(w_2) = 1000$, $C(w_1, w_2) = 160$, $N = 100000$? N: Total number of documents.
$C(w_i)$: Number of documents, w_i has appeared in.
$C(w_i, w_j)$: Number of documents where both the words have appeared in.
Note: Use base 2 in logarithm.

a. 4
b. 5
c. 6
d. 5.64

- ○ a.
- ○ b.
- ◉ c.
- ○ d.

Yes, the answer is correct.
Score: 1

Accepted Answers:
c.

4) ***2 points***

Solution (unit?unit=66&lesson=165)

Week 8 ()

Week 9 ()

Week 10 ()

Week 11 ()

Week 12 ()

Download videos ()

Text Transcripts ()

Books ()

Problem Solving Session - Jan 2024 ()

Given two binary word vectors w_1 and w_2 as follows:

w_1 = [1010101010]

w_2 = [0011111100]

Compute the Dice and Jaccard similarity between them.

a. 6/11, 3/8
b. 10/11, 5/6
c. 4/9, 2/7
d. 5/9, 5/8

- (●) a.
- ( ) b.
- ( ) c.
- ( ) d.

Yes, the answer is correct.
Score: 2

Accepted Answers:
a.

5) ***2 points***

Consider two probability distributions for two words be p and q. Compute their similarity scores with KL-divergence.

p = [0.20, 0.75, 0.50]

q = [0.90, 0.10, 0.25]

Note: Use base 2 in logarithm.

a. 4.704, 1.720
b. 1.692, 0.553
c. 2.246, 1.412
d. 3.213, 2.426

- ( ) a.
- ( ) b.
- (●) c.
- ( ) d.

Yes, the answer is correct.
Score: 2

Accepted Answers:
c.

6) ***2 points***

Consider the following word co-occurrence matrix given below. Compute the cosine similarity between
(i) w1 and w2, and (ii) w1 and w3.

	w4	w5	w6
w1	2	8	5
w2	4	9	7
w3	1	2	3

a. 0.773, 0.412
b. 0.881, 0.764
c. 0.987, 0.914
d. 0.897, 0.315

- a.
- b.
- c.
- d.

Yes, the answer is correct.
Score: 2

Accepted Answers:
c.

7) ***1 point***

Which of the following type of relations can be captured by word2vec (CBOW or Skipgram)?

a. Analogy (A:B::C:?)
b. Antonymy
c. Polysemy
d. All of the above

- a.
- b.
- c.
- d.

No, the answer is incorrect.
Score: 0

Accepted Answers:
a.