NLP week 7

Question 1:

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

Multi-choice multi-correct: a) CBOW (contin. Bag of words), c) PCA (hand-digit recogn.dataset MNSIT dataset)

Question 2: Page no. 135

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

Option C) is correct

Question 3:

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

Option C) is correct

Question 4:

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

Option A) is correct.
Question 5:

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

KLD(p, q), KLD(q, p)

Option C)

Question 6:

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

W1 = [2, 8, 5] |W1| = sqrt(2^2 + 8^2 + 5^2)
W2 = [4, 9, 7] |W2| = sqrt(4^2 + 9^2 + 7^2)
Cosine sim(W1, W2) == numer / denominator
Numerator = (2 * 4) + (8 * 9) + (5 * 7)

Denominator = sqrt(2^2 + 8^2 + 5^2) * sqrt(4^2 + 9^2 + 7^2) = sqrt(4 + 64 + 25) * sqrt(16 + 81 + 49)

0.9869 == 0.987
Option C) is correct

Question 7:

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

Option a) is correct

Question 8:

Suppose you are computing the word vectors using Skip-gram architecture. You have 5 words in your vocabulary, {passed, through, relu, activation, function} in that order and suppose you have the window, '*through relu activation*' in your corpora. You use this window with `relu' as the center word and one word before and after the center word as your context.
Also, suppose that for each word, you have 2-dim in and out vectors, which have the same value at this point given by [-1,-1],[1,1],[2,-1],[1,-1],[1,0] for the 5 words, respectively. As per the Skip-gram architecture, what would be the total loss for this window?