Quiz 2: Coursework

Creating the training dataset

Please enter your name: *

HM

The dataset

Let us consider the following corpus

Raw Corpus

$\mathcal{D}_1 =$ Neural Networks are awesome

$\mathcal{D}_2 =$ LSTMs are Sequential Neural Networks

$\mathcal{D}_3 =$ Attention Models are awesome

The word2idx dictionary associated with the Raw Corpus is the following dictionary:

Word2idx = {   Neural   : 1,
                Networks   : 2,
                are   : 3,
                awesome   : 4,
                LSTMs   : 5,
                Sequential : 6,
                Attention   : 7,
                Models   : 8 }

The positive batch

Let us consider the processed document 2 and the center word 6.

Raw document $\mathcal{D}_2 = $ LSTMs are Sequential Neural Networks
                        5   3   6   1   2

Processed document $\mathcal{D}_2 = [\ 5\ ,\ 3\ ,\ 6\ ,\ 1\ ,\ 2\ ]$

Which of these answers represent the positive batch ?   1 point

(a)	(b)	(c)
(6, 5)	(3, 5)	(5, 6)
(6, 3)	(6, 2)	(3, 6)
(6, 1)	(5, 1)	(1, 6)
(6, 2)	(6, 2)	(2, 6)

◉ (a)

◯ (b)

◯ (c)

Why do we call the elements of the positive batch "the true couples" ?   1 point

◉ Because they correspond to couples (center_word, context_word) that exist in the corpus

◯ Because they correspond to couples (context_word, center_word) that exist in the corpus

◯ I have no idea

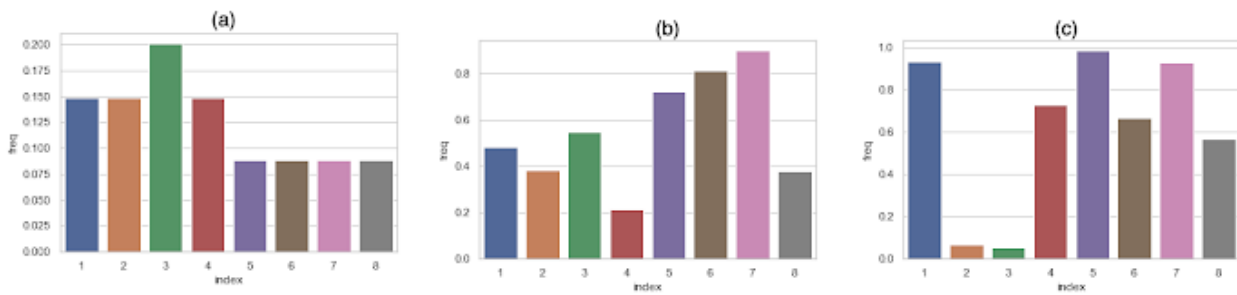The "true couples" composing the positive batch are associated with the label   1 point

◉ 1

◯ 0

The negative batch

Which figure represents the negative sampling distribution associated with the corpus (of 3   **2 points**
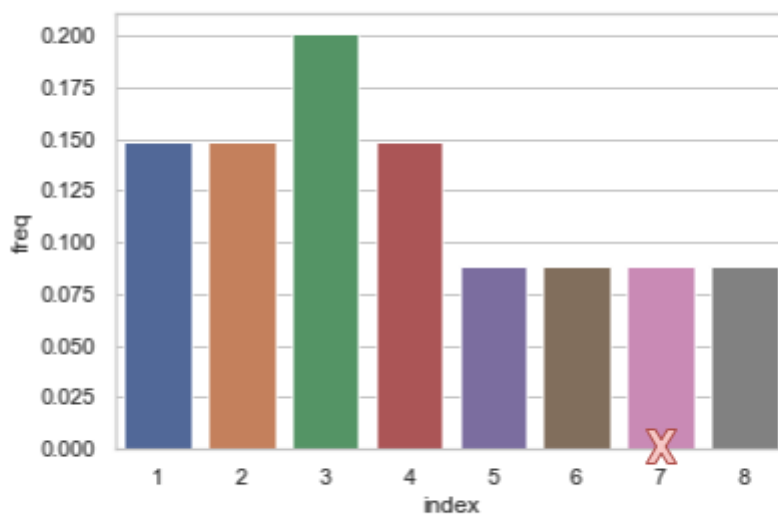documents) ?

⦿ (a)

◯ (b)

◯ (c)

By sampling a fake center word using the negative sampling distribution, we obtained the index 7.

What is the raw document associated with the fake center word that we have just sampled from the   1 point
negative sampling distribution?

(a) $\mathcal{D}_2 = $ LSTMs are Attention Neural Networks
                                     5   3   7   1   2

(b) $\mathcal{D}_2 = $ LSTMs are awesome Neural Networks
                                     5   3   4   1   2

(c) $\mathcal{D}_2 = $ LSTMs are Models Neural Networks
                                     5   3   8   1   2

- ⦿ (a)
- ◯ (b)
- ◯ (c)

The document created by replacing the true center word by the fake center word (sampled using   1 point
the negative sampling distribution) is very likely to be incoherent.

- ⦿ True
- ◯ False

Which of these answers represents the negative batch

1 point

(a)	(b)	(c)
(7, 5)	(3, 5)	(5, 7)
(7, 3)	(6, 2)	(3, 7)
(7, 1)	(5, 1)	(1, 7)
(7, 2)	(6, 2)	(2, 7)

◉ (a)

◯ (b)

◯ (c)

Why do we call the elements of the negative batch "the fake couples" ?

1 point

◉ Because they correspond to couples of (fake center word, context word) that don't exist in the corpus (with very high probability)

◯ Because they correspond to couples of (context word, fake center word) that don't exist in the corpus (with very high probability)

◯ I have no idea

The "fake couples" composing the negative batch are associated with the label

1 point

◉ 0

◯ 1

Any question ?

Google Forms