Quiz 2:

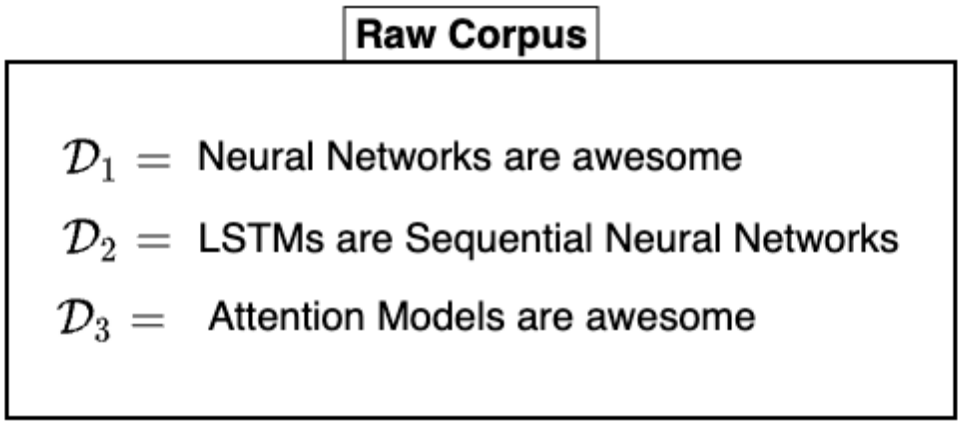
Creating the training dataset

* Indicates required question

1. Please enter your name: *

The dataset

Let us consider the following corpus



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

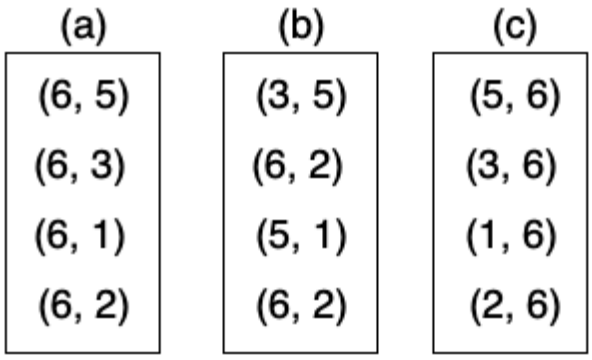
Let us consider the processed document 2 and the center word 6.

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

5 3 6 1 2

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

2. Which of these answers represent the positive batch ? 1 point



Mark only one oval.

- (a)
- (b)
- (c)

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

Mark only one oval.

- 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

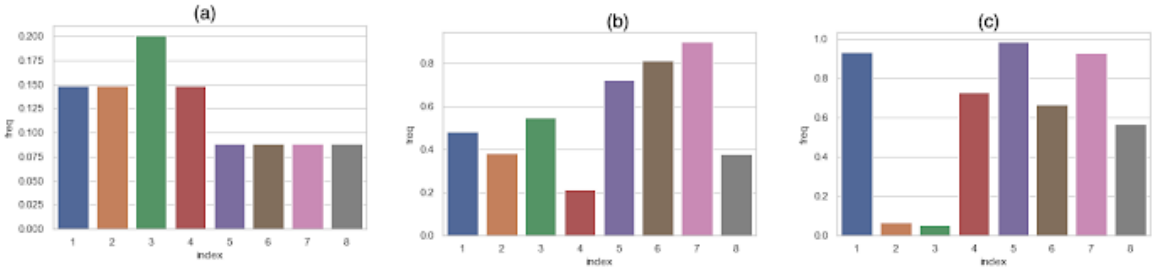
4. The "true couples" composing the positive batch are associated with the label 1 point

Mark only one oval.

- 1
- 0

The negative batch

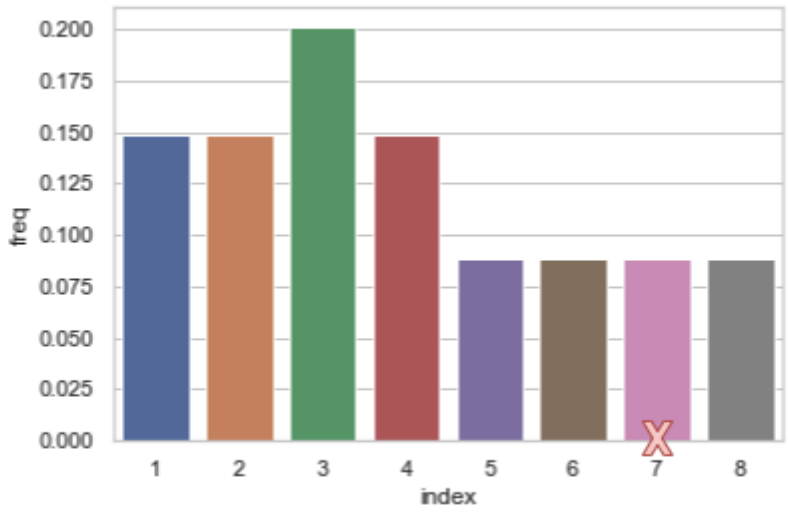
5. Which figure represents the negative sampling distribution associated with the corpus (of 3 documents) ? 2 points



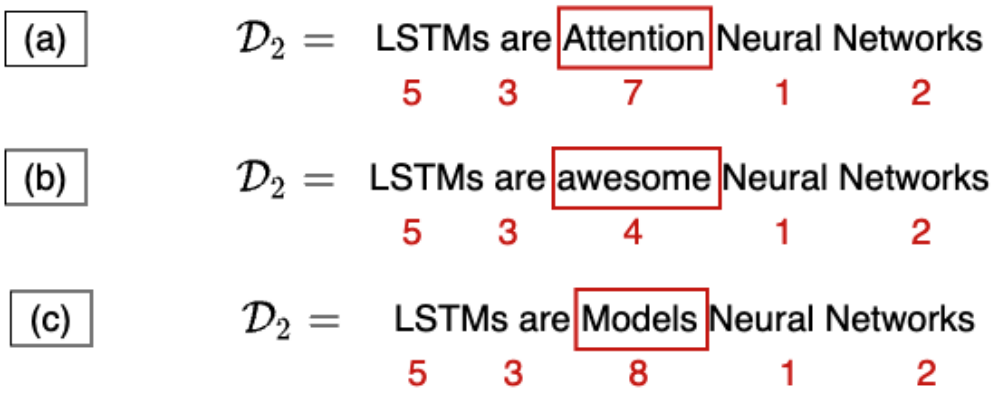
Mark only one oval.

- (a)
- (b)
- (c)

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



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



Mark only one oval.

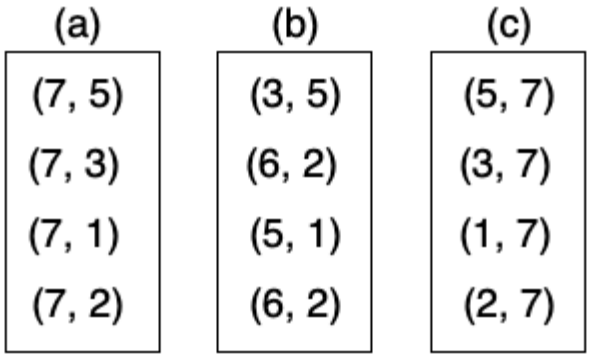
- (a)
- (b)
- (c)

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

Mark only one oval.

- True
- False

8. Which of these answers represents the negative batch 1 point



Mark only one oval.

- (a)
- (b)
- (c)

9. Why do we call the elements of the negative batch "the fake couples" ? 1 point

Mark only one oval.

- 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

10. The "fake couples" composing the negative batch are associated with the label 1 point

Mark only one oval.

- 0
- 1

11. Any question ?