

Quiz 3: Coursework

The Training Process

***Required**

1. Please enter your name: *

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 }

We consider the positive batch and the negative batch discussed in the previous quiz

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

Positive Batch

(6, 5)	→	1
(6, 3)	→	1
(6, 1)	→	1
(6, 2)	→	1

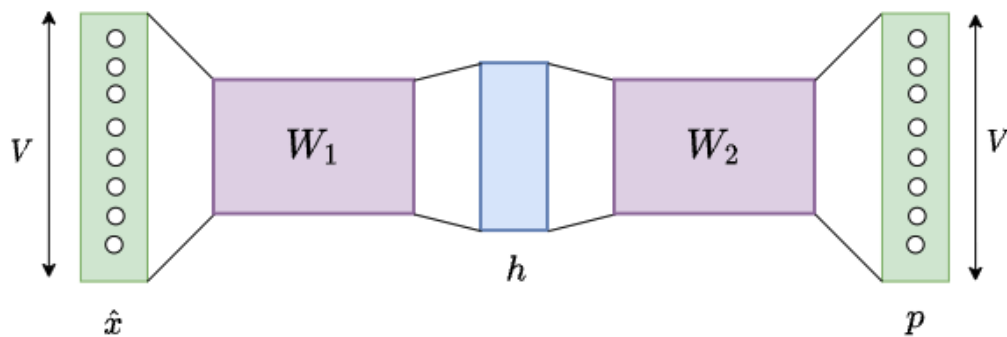
$\mathcal{D}_2 =$ LSTMs are Attention Neural Networks
 5 3 7 1 2

Negative Batch

(7, 5)	→	0
(7, 3)	→	0
(7, 1)	→	0
(7, 2)	→	0

The Forward Propagation

The following figure represents the Forward propagation. The objective is to predict the context words from the center word. We have the following hyperparameters: $V=8$, $D=3$



\hat{x} in the previous figure represents the one hot vector associated with an index x in $\{1, \dots, V\}$ representing a center word

The equations involved in the Forward propagation are summarized as follows:

A first linear transformation maps \hat{x} to the D -dimensional vector h as follows:

$$h = W_1^T \hat{x}$$

A second transformation maps the hidden vector h to the V -dimensional vector $p = (p_1, \dots, p_V)$ as follows:

$$p = \sigma(W_2^T h) \quad \text{where } \sigma \text{ is the sigmoid activation function.}$$

2. Which classification problem are we dealing with ?

1 point

Mark only one oval.

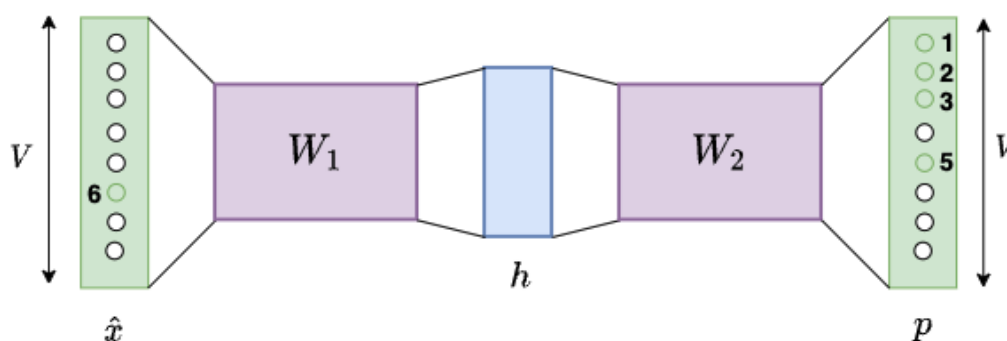
- ☐ A single binary classification problem
- ☐ A multiclass classification problem
- ☐ Several binary classification problems

3. Let us consider o in $\{1, \dots, V\}$. What is the interpretation of p_o (the o -th dimension of the output vector p) ? 1 point

Mark only one oval.

- ☐ The probability that the word of index o is in the context of the center word x
- ☐ The probability that the couple (x, o) is a fake couple.

Let us consider the positive batch. From the true center word 6 we compute p_1, p_2, p_3 and p_5 .



4. What is p_5 ? 1 point

Mark only one oval.

- ☐ The probability that the word "Neural" is in the context of "Sequential" ?
- ☐ The probability that the word "LSTMs" is in the context of "Sequential" ?
- ☐ The probability that the word "Sequential" is in the context of "Neural" ?

5. What is the loss function associated with a binary classification problem ? 1 point

Mark only one oval.

- ☐ The categorical cross entropy
- ☐ The binary cross entropy

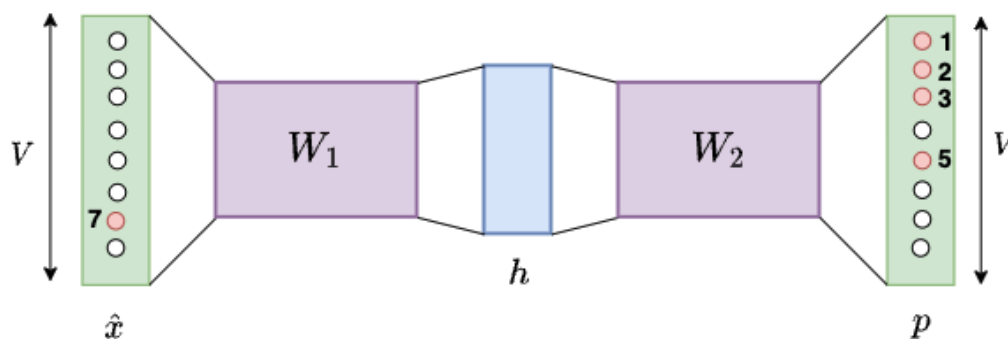
6. Here is the loss function associated with the positive batch. What are the elements of W_1 and W_2 which are involved in this expression ?

$$J_+ = -\frac{1}{4} \sum_{k \in \{1,2,3,5\}} \log(\sigma(W_1[6]^T W_2[k]))$$

Mark only one oval.

- ☐ The 6-th row of W_1 and the columns 1, 2, 3 and 5 of W_2
- ☐ The 6-th column of W_1 and the rows 1, 2, 3 and 5 of W_2
- ☐ All the rows and columns of W_1 and W_2

Let us consider the negative batch. From the fake center word 7 we compute p_1, p_2, p_3 and p_5 .



7. What is p_5 ?

1 point

Mark only one oval.

- ☐ The probability that the word "Neural" is in the context of "Attention" ?
- ☐ The probability that the word "LSTMs" is in the context of "Sequential" ?
- ☐ The probability that the word "LSTMs" is in the context of "Attention" ?

8. Here is the loss function associated with the negative batch. What are the elements of 1 point W_1 and W_2 which are involved in this expression ?

$$J_- = -\frac{1}{4} \sum_{k \in \{1,2,3,5\}} \log(1 - \sigma(W_1[7]^T W_2[k]))$$

Mark only one oval.

- ☐ The row 7 of the matrix W_1 and the columns 1, 2, 3, 5 of the matrix W_2
- ☐ The column 7 of the matrix W_1 and the rows 1, 2, 3, 5 of the matrix W_2
- ☐ All the rows and columns in W_1, W_2

The Backward Propagation For the positive batch

We have the following expressions of the gradients:

$$\begin{aligned} \nabla_{W_1[6]} (\log(\sigma(W_1[6]^T W_2[k]))) &= (1 - \sigma(W_1[6]^T W_2[k])) W_2[k] \\ \nabla_{W_2[k]} (\log(\sigma(W_1[6]^T W_2[k]))) &= (1 - \sigma(W_1[6]^T W_2[k])) W_1[6] \quad k \in \{1, 2, 3, 5\} \end{aligned}$$

9. Which expression of the gradient is correct ?

1 point

(a) $\nabla_{W_1[6]} J_+ = \frac{1}{4} \sum_{k \in \{1,2,3,5\}} (\sigma(W_1[6]^T W_2[k]) - 1) W_2[k]$

(b) $\nabla_{W_1[6]} J_+ = -\frac{1}{4} \sum_{k \in \{1,2,3,5\}} (\sigma(W_1[6]^T W_2[k]) - 1) W_2[k]$

(c) $\nabla_{W_1[6]} J_+ = \sum_{k \in \{1,2,3,5\}} (\sigma(W_1[6]^T W_2[k]) - 1) W_2[k]$

Mark only one oval.

☐ (a)

☐ (b)

☐ (c)

10. We have the following update equations associated with the positive batch. What is the number of parameters updated ? 1 point

$$W_1[6] \leftarrow W_1[6] - \eta \nabla_{W_1[6]} J_+$$

$$W_2[k] \leftarrow W_2[k] - \eta \nabla_{W_2[k]} J_+ \quad k \in \{1, 2, 3, 5\}$$

Mark only one oval.

☐ 5*D

☐ 4*D

☐ 2*V*D

The Backward Propagation For the negative batch

We have the following expressions of the gradients:

$$\nabla_{W_1[7]} (\log(1 - \sigma(W_1[7]^T W_2[k]))) = -\sigma(W_1[7]^T W_2[k]) W_2[k]$$

$$\nabla_{W_2[k]} (\log(1 - \sigma(W_1[7]^T W_2[k]))) = -\sigma(W_1[7]^T W_2[k]) W_1[7] \quad k \in \{1, 2, 3, 5\}$$

11. Which expression of the gradient is correct ?

1 point

(a) $\nabla_{W_2[k]} J_- = \frac{1}{4} \sum_{k \in \{1, 2, 3, 5\}} (\sigma(W_1[7]^T W_2[k]) - 1) W_1[7]$

(b) $\nabla_{W_2[k]} J_- = \frac{1}{4} (\sigma(W_1[7]^T W_2[k]) - 1) W_1[7]$

(c) $\nabla_{W_2[k]} J_- = \frac{1}{4} (\sigma(W_1[7]^T W_2[k])) W_1[7]$

Mark only one oval.

☐ (a)

☐ (b)

☐ (c)

12. Any question ?

This content is neither created nor endorsed by Google.

Google Forms