LAB 10 REPORT

A PREPRINT

Daniele Cecca
Artificial Intelligence for Science and Technology
Milano Bicocca University
Supervised Learning

June 9, 2024

1 Introduction

This lab report details the training and analysis of a Transformer architecture.

In particular, we will analyze this network on a Sequence-to-Sequence task, using different sequence lengths.

2 Sequence to Sequence

A Sequence-to-Sequence task involves input and output sequences that are not necessarily of the same length. Popular tasks in this domain include machine translation and summarization. In this case, the task will be much simpler: Given a sequence of N numbers between 0 and M, the task is to reverse the input sequence. Although this task sounds very simple, RNNs can struggle with it because the task requires long-term dependencies. Transformers are designed to handle such dependencies, so we expect them to perform very well.

3 Dataset

To create our dataset, we divide our original data into sequences of fixed length.

We create two different datasets with sequences of different sizes:

- First Dataset: sequence length 16
- Second Dataset: sequence length 50

```
Input data: tensor([9, 6, 2, 0, 6, 2, 7, 9, 7, 3, 3, 4, 3, 7, 0, 9])
Labels:     tensor([9, 0, 7, 3, 4, 3, 3, 7, 9, 7, 2, 6, 0, 2, 6, 9])

Input data: tensor([6, 1, 0, 8, 4, 1, 5, 1, 0, 1, 1, 6, 1, 1, 3, 3, 9, 5, 5, 3, 0, 5, 6, 5,
    0, 3, 4, 2, 6, 9, 7, 7, 9, 3, 1, 8, 4, 4, 0, 4, 7, 5, 0, 1, 9, 9, 6, 5,
    8, 2])
Labels:     tensor([2, 8, 5, 6, 9, 9, 1, 0, 5, 7, 4, 0, 4, 4, 8, 1, 3, 9, 7, 7, 9, 6, 2, 4,
    3, 0, 5, 6, 5, 0, 3, 5, 5, 9, 3, 3, 1, 1, 6, 1, 1, 0, 1, 5, 1, 4, 8, 0,
    1, 6])
```

4 Training

We trained our model using the following hyperparameters:

- Epochs: 20

- Drop-out: 0
- Batch size: 128

We used Adam as the optimizer, and to achieve better results during training, we employed learning rate warm-up.

With this approach, we gradually increased the learning rate from 0 to our originally specified learning rate (0.001) in the first few iterations.

5 Evaluation

To evaluate the results of our architecture, we used accuracy, and we obtained 100% accuracy in both the test and validation sets:

```
Val accuracy:  100.00%
Test accuracy: 100.00%

Val accuracy:  100.00%
Test accuracy: 100.00%
```

5.1 Attention Map analysis

The attention maps provide visual insights into how the Transformer model attends to different parts o the input sequences.

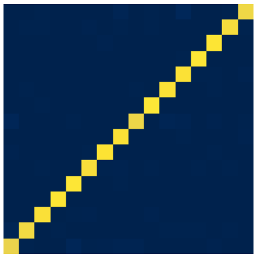

Figure 1: Attention Map 1

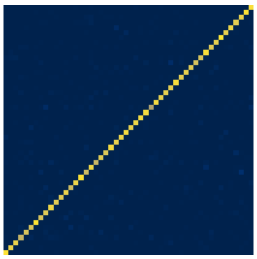

Figure 2: Attention Map 2

In both cases the attention maps shows clear diagonal patterns, indicating that the model attend strongly to corresponding positions,which is ideal for sequence reversal task

6 Conclusion

The Transformer architecture performed exceptionally well on the Sequence-to-Sequence task, achieving perfect accuracy. This confirms its effectiveness in handling tasks requiring long-term dependencies.