Understanding Logits and Targets in Language Models

1 Introduction

In language models, understanding the shapes of logits and targets is crucial for grasping how the model makes predictions and learns from data. This document explains these concepts using a simple example and appropriate tables.

2 Definitions

- **Logits**: The model's predictions for the next token in a sequence. They represent the likelihood of each possible token being the next one.
- **Targets**: The actual correct tokens that we want the model to predict.
- **Batch Size (B)**: The number of sequences being processed at the same time.
- **Sequence Length (T)**: The number of tokens the model considers for prediction at once.
- **Vocabulary Size (C)**: The total number of possible tokens in the model's vocabulary.

3 Logits and Targets Shapes

Given the shapes:

- Logits: (B, T, C)
- Targets: (B, T)

For our example, let's consider:

- $B = 4$ (number of sequences)
- $T = 1$ (predicting the next token)
- $C = 80$ (vocabulary size)

3.1 Logits Example

The logits shape can be represented as:

$$\text{Logits Shape} = (B, T, C) = (4, 1, 80)$$

When reshaped, it becomes:

$$\text{Reshaped Logits} = (B \times T, C) = (4 \times 1, 80) = (4, 80)$$

Table 1: Logits for Each Sequence

Sequence	**Logits (Next Word Guesses)**
1	[0.1, 0.2, 0.3, ..., 0.05, 0.02, 0.4]
2	[0.05, 0.2, 0.1, ..., 0.3, 0.1, 0.15]
3	[0.4, 0.1, 0.05, ..., 0.1, 0.2, 0.05]
4	[0.15, 0.3, 0.25, ..., 0.1, 0.05, 0.2]

3.2 Targets Example

The targets shape can be represented as:

$$\text{Targets Shape} = (B, T) = (4, 1)$$

Each sequence corresponds to one correct next token:

Table 2: Targets for Each Sequence

Sequence	**Target (Correct Next Word)**
1	1 (representing "on")
2	2 (representing "loudly")
3	3 (representing "with cheese")
4	4 (representing "every day")

4 Understanding the Shapes

- The logits are reshaped to $(B \times T, C)$ to facilitate direct comparison against the target values.
- The targets remain in (B, T) format since each sequence only has one correct answer to predict.

When calculating the loss, the model compares each row of logits (its predictions) with the corresponding target (the correct next token).

5 Conclusion

In summary, understanding the shapes of logits and targets is essential for evaluating how well a language model predicts the next token. The batch size B and sequence length T play a vital role in structuring these predictions, which ultimately guides the model's learning process.