Attention is all you need

Shubham Gupta

January 26, 2020

1 Introduction

- This paper review is following the blog from Jay Alammar's blog on the **Illustrated Transformer**. The blog can be found here.

2 Paper Introduction

- New architecture based solely on attention mechanisms called **Transformer**. Gets rids of recurrent and convolution networks completely.
- Generally, RNN used to seq-to-seq tasks such as translation, language modelling, etc.
- Transformer allows for significant parallelization and relies only on attention.

3 Background

- *Self attention* Attention to different positions of a sequence in order to compute a representation of the sequence.

4 Model Architecture

- Transformer uses the following:
  - Encoder decode mechanism
  - Stacked self attention
  - Point wise fully connected layer for encoder and decoder

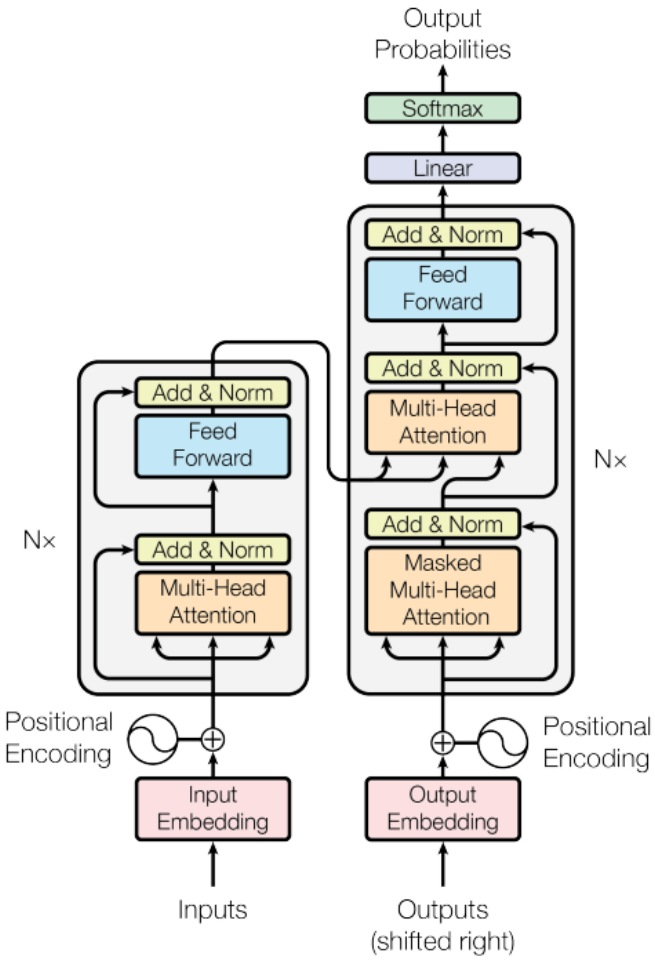

Figure 1: The Transformer - model architecture.

Figure 1: transformer