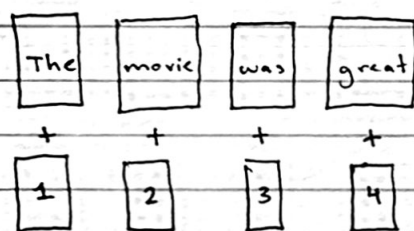


## Position Encodings

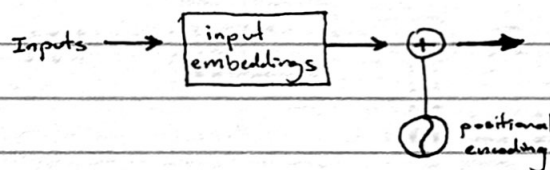
I visited New York and ate           

\* Visited and ate look the same from perspective of self attention, unless we provide position information

\* Positional Encoding: Refers to a family of schemes to provide this information to Transformer models



\* Encode each sequence position as integer and add it to word embedding vector



• Other types of encoding:

- Sines/cosines of diff. frequencies from Vaswani et. al.
- Relative position encoding: depends on distance between 2 tokens
- ALiBi: Adds linear bins to disprefer attending farther back
- No positional encoding (NoPE) works too!

↳ Transformers w/ self attention into past can learn to count tokens.