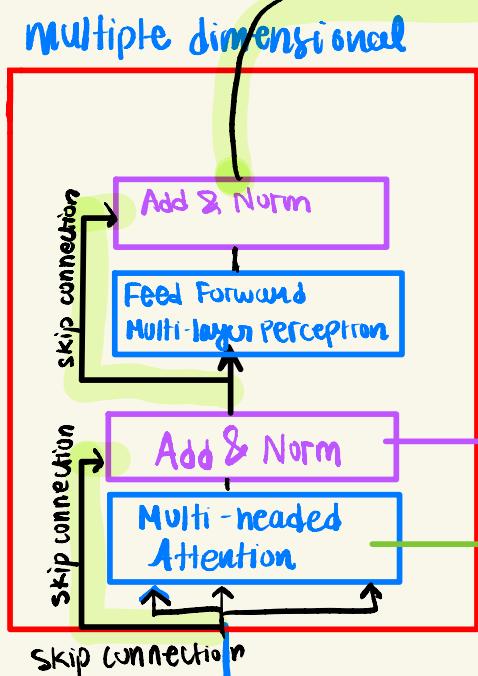


Encoder

goes to decoder



Skip Connection: $\text{Output} = \text{LayerNorm}(\text{input} + \text{Sublayer}(x))$

between 0 and 1.

Add: adds the original input delivered by the skip connection and the output of the layer.

- Input + Attention Output
- Normalization → adjusting values.

• Skip connections take the original input of a layer and add it to the output of that layer.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{D_K}}\right) \cdot V$$

Positional Encoding \oplus (Time encodings using Absolute Time Stamps) Recommended

- Start with 2 layers
- Final - 6 layers

Embedding Layer (linear)

Input (dataset → position, etc.)
 raw data → vector form

- Arouses
- to prevent losing data
- takes data of positional encoding and adds it to the output of the multi-headed attention.
- takes output of the

Decoder

