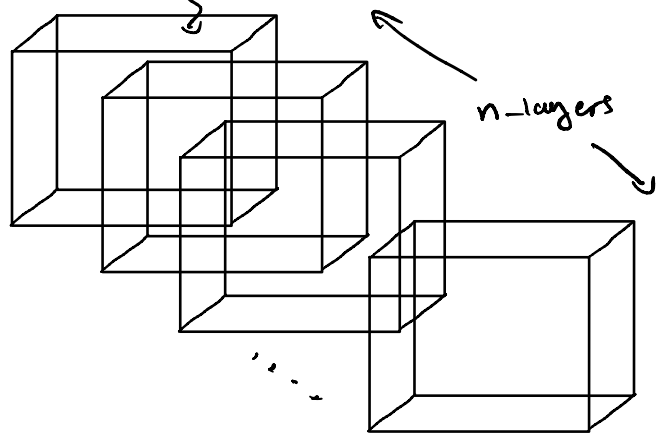
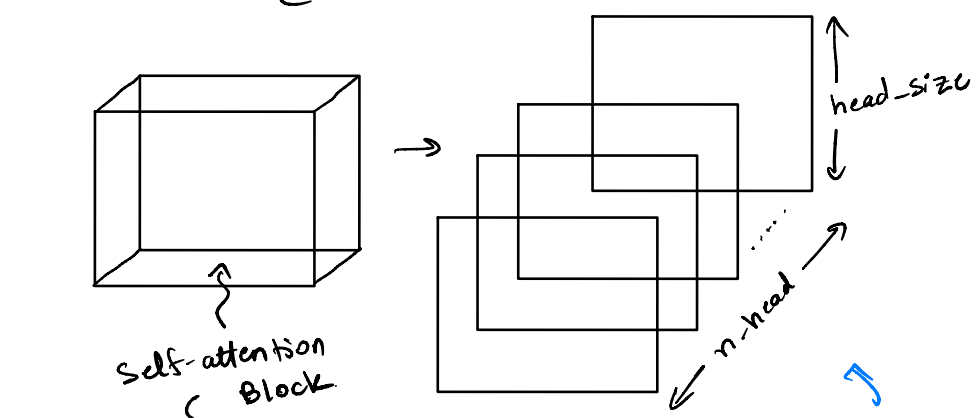
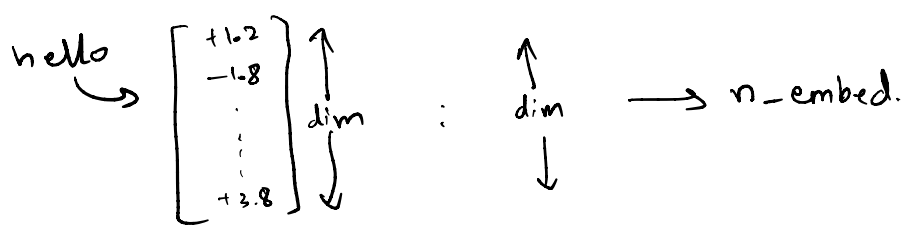
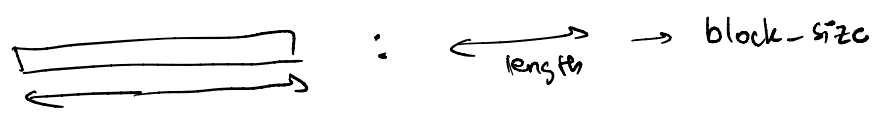
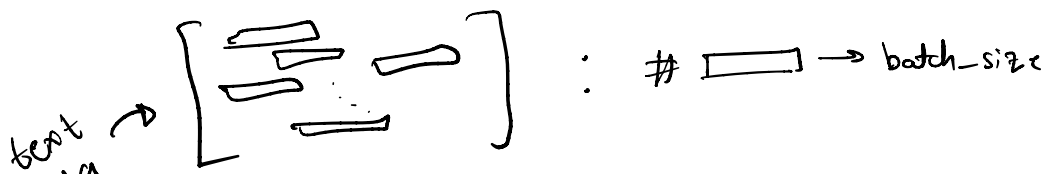
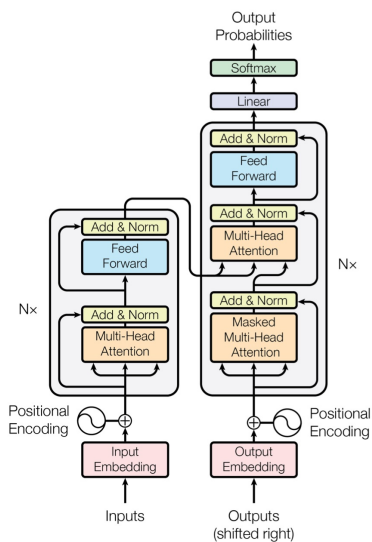


Understand the hyper parameters

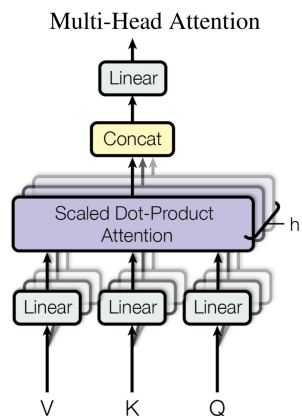


This has number of heads in each Block

Number of Blocks or Number of layers



Transformer architecture



multi-head attention.