

12 Mar 2025

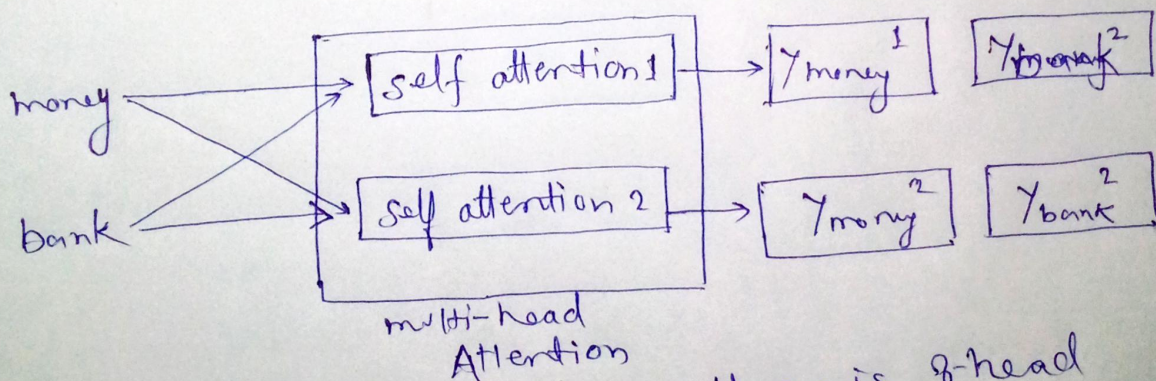
MULTI-HEAD ATTENTION

- The drawback of self attention is that it is only capture the one meaning/perspective of the given sentence
- Because there is may be possibility that given sentence have a multiple meaning example.

"The man saw the astronomer with a telescope"

- Hence the drawback of ~~multihed~~ self attention is solve with multihead attention (multiple self attention)

— let's take an example "money bank"



- in original research paper there is 8-head Attention is used in transformer.