

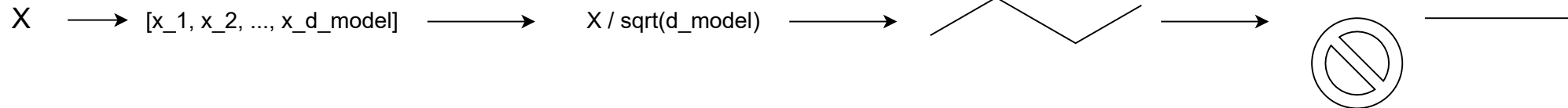
Decoder

Word Embedding

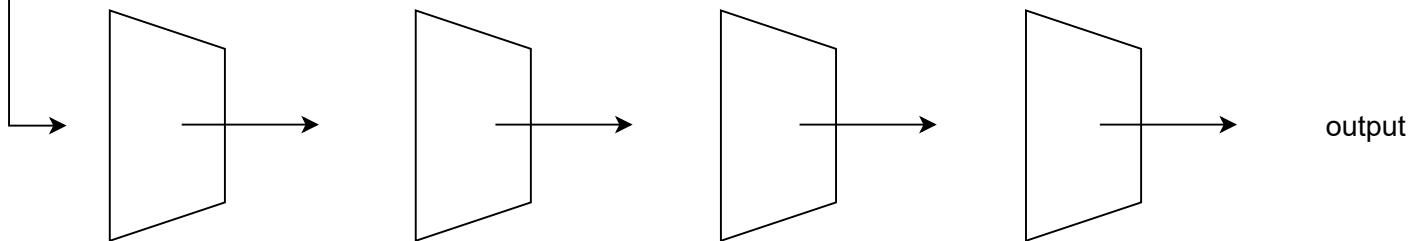
Scaling

Positional Encoding

dropout
0.1



Decoder Layers



$N = 4$