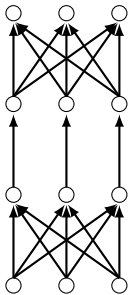


MLP

linear

relu
(neuron-wise)

linear



Transformer (vanilla)

self attn

MLP
(token-wise)

self attn

