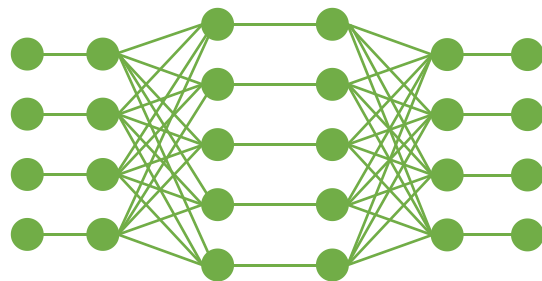
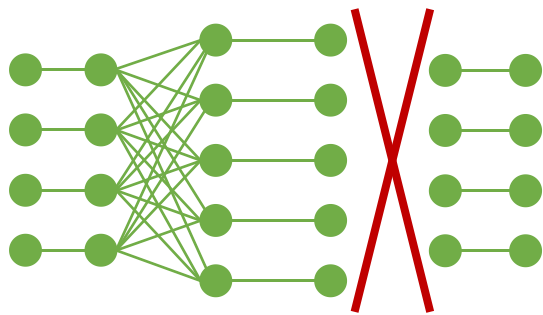


Attn-Attention
Act -Activation(Gelu)



Pruning
Linear 2

Pruning
Activation



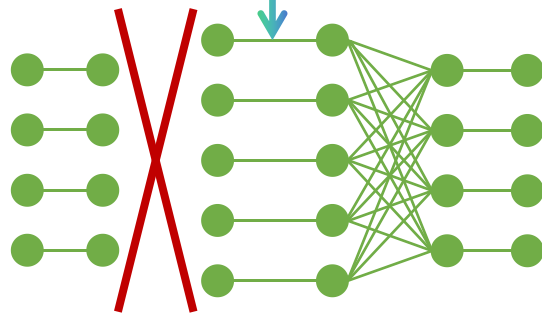
Attn Linear Act Linear Attn



dimensions mismatch

Attn Linear Act Linear Attn

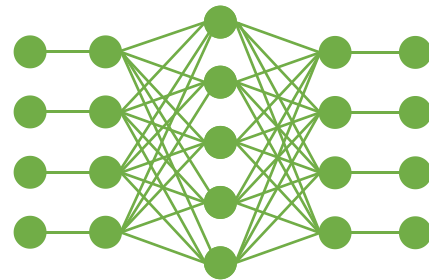
Pruning Linear 1



Attn Linear Act Linear Attn



dimensions mismatch



Attn Linear Linear Attn



dimensions match