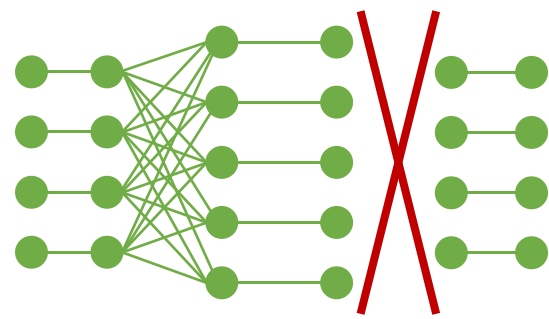


Pruning  
Linear 2



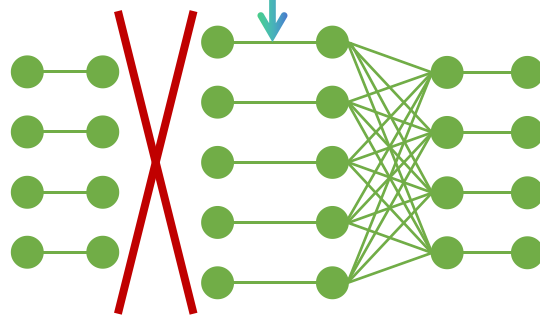
Attention   Linear   Gelu   Attention



dimensions mismatch



Pruning Linear 1



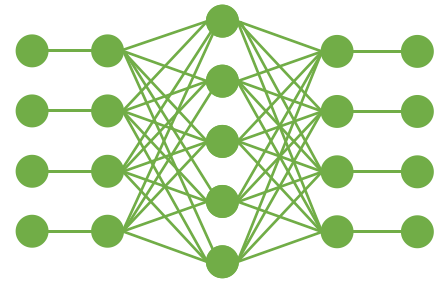
Attention   Gelu   Linear   Attention



dimensions mismatch



Pruning  
Activation



Attention   Linear   Linear   Attention



dimensions match