

## Language Model Head

takes  $h_N^L$  and outputs a distribution over vocabulary  $V$

