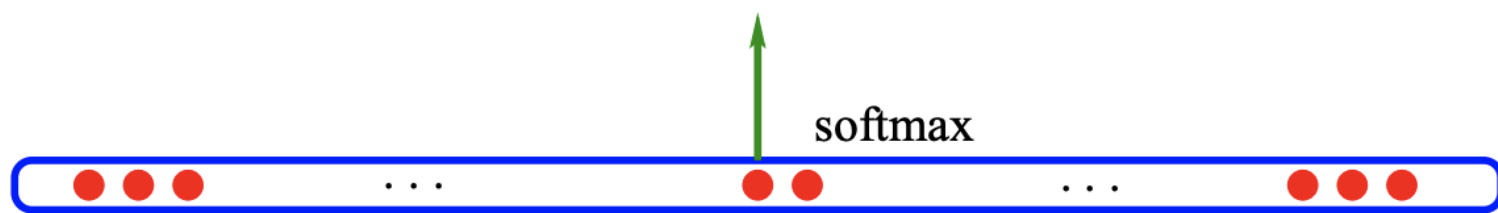


$$i\text{-th output} = P(w_t = i \mid \text{context})$$



softmax

most computation here

tanh



$C(w_{t-n+1})$

$C(w_{t-2})$

$C(w_{t-1})$

Table  
look-up  
in  $C$

Matrix  $C$

shared parameters  
across words

index for  $w_{t-n+1}$

index for  $w_{t-2}$

index for  $w_{t-1}$

