

$[27, 2] \leftarrow [32, 3]$

dim 1

embedding = $C[X] \Rightarrow [32, 3, 2]$

$[0, 0], [0, 0], [0, 0]$

Row 1 $[0, 0, :]$

$[0, 0], [0, 0], [0, 0]$

Row 2 $[1, 0, :]$

$[0, 0], [0, 0], [0, 0]$

Row 32

$[32, 0, :]$

Hidden layer weights = $[6 \times 100]$

number of ips from the first layer
(3×2) randomly chosen

Hidden layer bias = $[100]$

concatenation

emb $[i, 0, :]$ + emb $[i, 1, :]$ + emb $[i, 2, :]$

32×6 $[0, 0, 0, 0, 0, 0]$

Row 1

$[0, 0, 0, 0, 0, 0]$

Row 2

$[0, 0, 0, 0, 0, 0]$

Row 32

$[batch_size, block_size, emb_dim] \rightarrow [batch_size, (block_size \times emb_dim)]$

- emb.view(32, 6) - with PyTorch

tanh

$\rightarrow -1$ to make it generic

- hidden layer $h = \text{emb.view}(32, 6) @ W1 + b1$

- $b1$ - same bias vector will be added to

each of the 32 rows of - emb.view(32, 6) @ W1