

$$[32, 2] \leftrightarrow [32, 3]$$

dim.

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

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

- Row 1 [0, 0, 0]

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

- Row 2 [1, 0, 0]

\vdots

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

- Row 32

[32, 0, 0]

$$\text{Hidden layer} = [6 \times 100]$$

weights

\downarrow randomly chosen
number of c/p's from the first layer
(3x2)

$$\text{Hidden layer bias} = [100]$$

Concatenation

$$\text{emb}[i, 0, :] + \text{emb}[i, 1, :] + \text{emb}[i, 2, :]$$

$$\Rightarrow 32 \times 6 [0, 0, 0, 0, 0, 0]$$

- row 1

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

- row 2

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

- row 32

$$[\text{batch-size}, \text{block-size}, \text{emb-dim}] \rightarrow [\text{batch-size}, (\text{block-size})^* \text{emb-dim}]$$

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

tanh

$\mapsto -1$ to make it generic

- hidden layer $h = \text{emb-view}(32, 6) @ w_1 + b_1$

- b_1 - same bias vector will be added to

each of the 32 rows of $h = \text{emb-view}(32, 6) @ w_1$