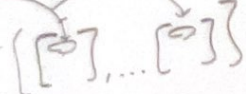


Deriving Embeddings Table (Position, Identity)

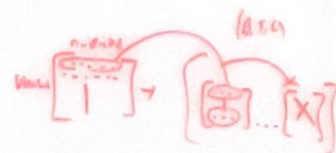
• Pos-Emb

- Affects all batches @ same pos.
- Sum across dim



• IP Embeddings

- Affects everywhere the index-element is present



For each embedding index in 'idx'

- Find correct gradient in E_{idx} , add to the ones in idx

$$\frac{\partial L}{\partial E_{idx}} = \sum_{\text{sources}} \frac{\partial L}{\partial \text{sources}}$$

where 'i' denotes index in idx