

MLP is All You Need ?



recent work.

MLP-Mixer

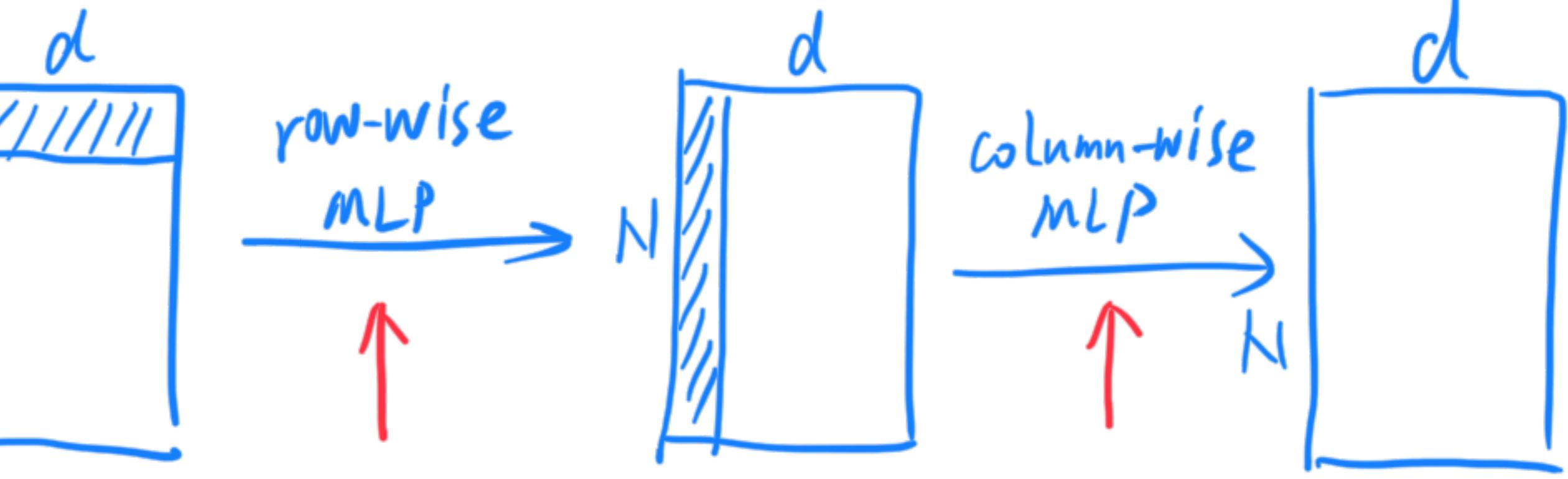
External Attention

Attention

essence

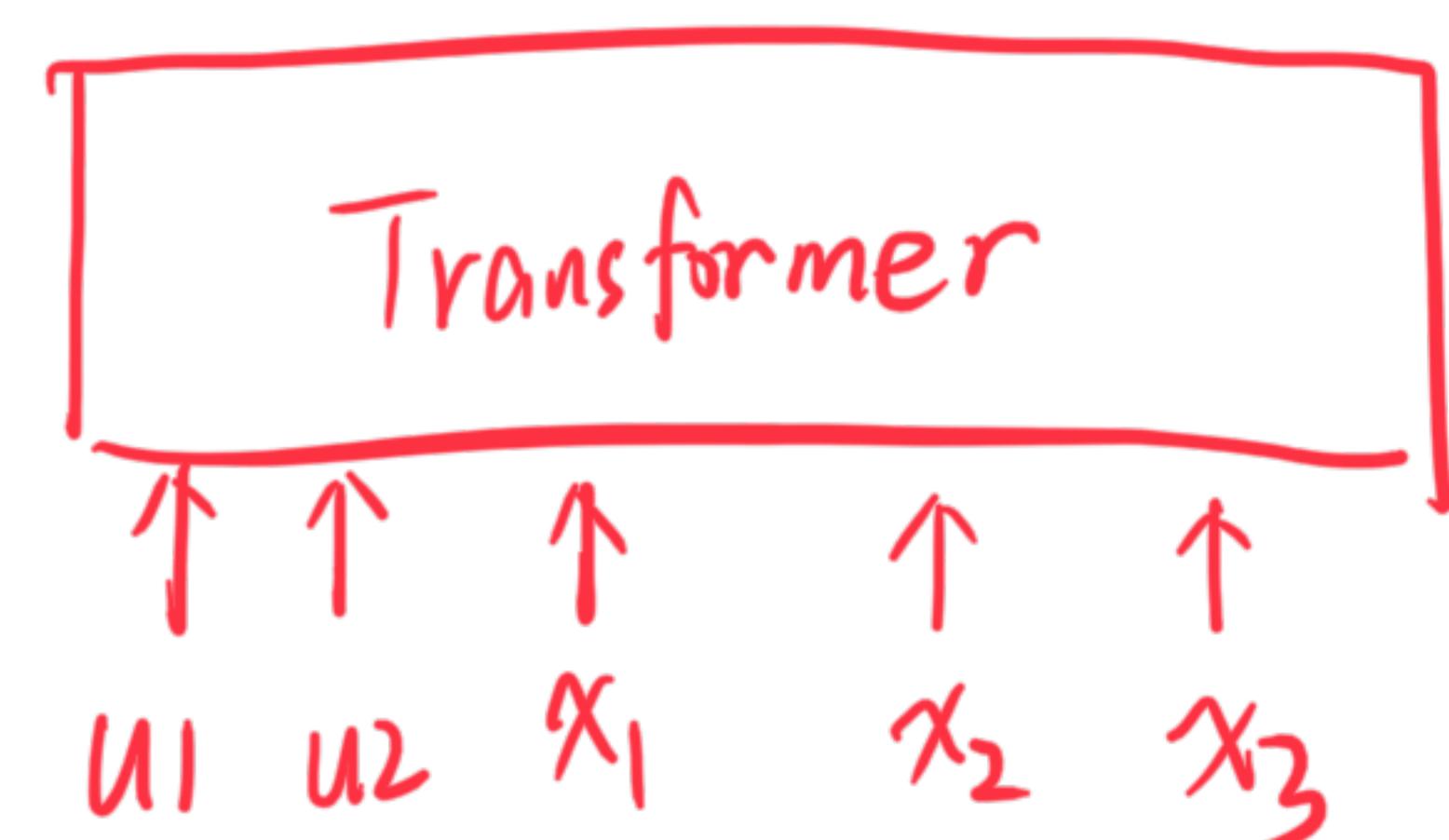
Implement with Transformer.

prefix-tuning
P-tuning



$$\begin{cases} A = \text{softmax}(QK^T) \\ X_{\text{out}} = AX_{\text{in}} \end{cases} \Rightarrow \begin{cases} A = \text{Norm}(XX^T) \\ X_{\text{out}} = AX \end{cases}$$

$$\begin{cases} A = \text{Norm}(XW_K^T) \\ X_{\text{out}} = AW_V \end{cases}$$



1	0	0	0	0
0	1	0	0	0
1	1	0	0	0
1	1	0	0	0
1	1	0	0	0

标题

- ▶ Prefix-Tuning: Optimizing Continuous Prompts for Generation 2021
- ▶ IS ATTENTION BETTER THAN MATRIX DECOMPOSITION? 2021
- ▶ ResMLP: Feedforward networks for image classification with data-efficient... 2021
- ▶ RepMLP: Re-parameterizing Convolutions into Fully-connected Layers for... 2021
- ▶ Do You Even Need Attention? A Stack of Feed-Forward Layers Does Surpri... 2021
- ▶ Beyond Self-attention: External Attention using Two Linear Layers for Vis... 2021
- ▶ MLP-Mixer: An all-MLP Architecture for Vision 2021
- ▶ GPT Understands, Too 2021