

Master note

Components needed

1. Communication layer
2. Linear algebra operation
3. Autodiff system
4. Optimizer
5. Neural network block
6. TUI

Neural network block

Attention

$$\mathbf{Q} = \mathbf{W}_q \mathbf{X} + b_q \mathbf{1}^T$$

$$\mathbf{K} = \mathbf{W}_k \mathbf{X} + b_k \mathbf{1}^T$$

$$\mathbf{V} = \mathbf{W}_v \mathbf{X} + b_v \mathbf{1}^T$$

$$\mathbf{S} = \mathbf{K}^T \mathbf{Q}$$

$$\tilde{\mathbf{V}} = \mathbf{V} \cdot \text{softmax}\left(\frac{\mathbf{S}}{\sqrt{d_{\text{attn}}}}\right)$$