

Practical deep learning



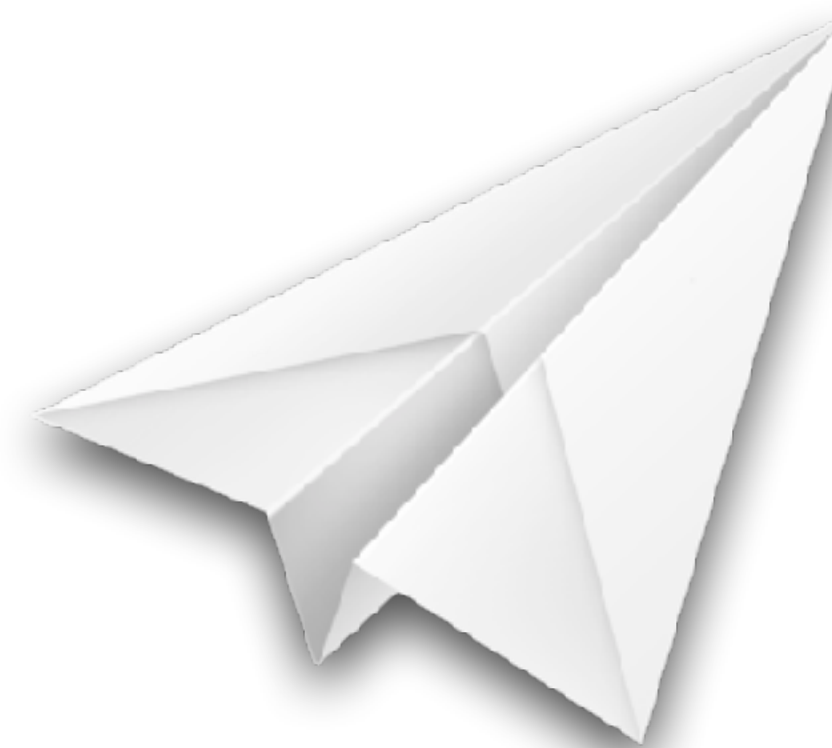
Transformers

$$\text{Atten}(\mathbf{h})(t) = \mathbf{h}(t) + \sum_{i=1}^h \mathbf{W}_o^i \sum_{s=1}^{\tau} \sigma[(\mathbf{W}_Q^i \mathbf{h}(t))^\top \mathbf{W}_K^i \mathbf{h}(s)] \mathbf{W}_V^i \mathbf{h}(s)$$
$$\text{FF}(\mathbf{h})(t) = f(\mathbf{h})(t),$$

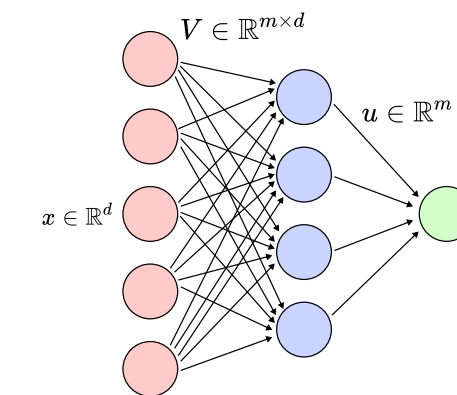
Intuition
through
experiments



Theoretical deep learning



Simplified network



Understanding
and
Improvements



Future of deep learning (??)



Architecture of the future

“Nothing is more practical than a good theory”