Advanced Sequence Models: Summary & Preview

What We Learned Today

1

Seq2Seq Architecture

- Encoder-Decoder
- Context vector
- Variable length I/O

2

RNN Variants

- LSTM cells
- GRU cells
- Gradient flow

3

Bidirectional RNN

- Forward + Backward
- Context from both
- Better encoding

4

Teacher Forcing

- Training strategy
- Exposure bias
- Scheduled sampling

5

Attention Mechanism

- Query, Key, Value
- Dynamic context
- Alignment weights

6

Implementation

- Batching & Padding
- Masking strategies
- Practical tips

Next Lecture: Transformers & Modern Architectures

- Self-Attention
- Transformer Architecture
- Multi-Head Attention
- BERT & GPT

Key Takeaway: Attention mechanism revolutionized sequence modeling by allowing models to selectively focus on relevant parts of input, solving the information bottleneck problem!