Understanding Encoder-Decoder (Seq2Seq) Models

What Are Word Embeddings?

Embeddings are dense vector representations of words. They start as random vectors and are learned during training. Words with similar meanings end up with similar vectors in high-dimensional space. Example: 'I' -> [0.2, -0.1, 0.5, 0.0]

What Happens After Embeddings?

After embedding, the sequence of vectors is passed through an RNN (LSTM/GRU). Each word updates the hidden and cell states. Final hidden and cell states represent the encoded sequence.

Encoder Time Step Breakdown (LSTM)

Each LSTM time step does the following:

- Input Gate: i = sigmoid(Wi * [x; h_prev] + bi)
- Forget Gate: f = sigmoid(Wf * [x; h_prev] + bf)
- Output Gate: o = sigmoid(Wo * [x; h_prev] + bo)
- Candidate: g = tanh(Wg * [x; h_prev] + bg)
- Cell state: c = f * c_prev + i * g
- Hidden state: h = o * tanh(c)

Meaning of Hidden and Cell States

Hidden state (h): summarizes the sequence up to this point (short-term memory)

Cell state (c): carries long-term dependencies

Final (h, c) from encoder are used as initial state in decoder.

Encoder Summary

- Input sentence: 'I am fine'
- Embedding each word to vector

- Passing through LSTM gives final (h, c)

- These are fed into the decoder to start prediction.